

DOKTORI ÉRTEKEZÉS

SZABÓ MARTINA KATALIN

**A szentimentérték módosulásának a problémája a magyar
nyelvű szövegek szentimentelemzésében,
különös tekintettel az értékvesztésre és az értékváltásra**

Témavezetők:

Dr. Bibok Károly egyetemi docens

Dr. Vincze Veronika tudományos munkatárs

Készült:

Szegedi Tudományegyetem

Bölcsészettudományi Kar

Nyelvtudományi Doktori Iskola

Magyar Nyelvészet Doktori Program keretében

Szeged, 2018.

Köszönetnyilvánítás

Hálás köszönettel tartozom témavezetőimnek, Dr. Bibok Károlynak és Dr. Vincze Veronikának, akik egészen az egyetemi évektől töretlenül biztattak és támogattak a munkámban. Köszönöm Dr. Forgács Tamásnak, hogy felfigyelt az érdeklődésemre, és elindított a tudományos kutatás felé, valamint Dr. Kenesei István professzornak, hogy mindig nyitott volt a témámra és azt a kutatószemináriumaira befogadta, előadásaimat értékes kritikával illette.

Hálával tartozom Dr. Sass Bálintnak és Dr. Szécsényi Tibornak, hogy a dolgozat tématervéről írt bírálatukkal értékes hozzájárulást tettek a disszertációmhoz.

Köszönettel tartozom a Precognox Informatikai Kft. vezetőinek, Jóföldi Endrének és Kása Károlynak, hogy anyagi és szellemi háttérrel biztosítottak a munkámhoz, és munkatársaimnak, elsősorban Varjú Zoltánnak, Morvay Gergelynek, Balogh Kittinek, Ilyés Virágnak, Nyíri Zsófinak, Lázár Bernadettnek, Bors Bencének és Sipos Baláznak, hogy együtt dolgozhattunk számos, a disszertációhoz kapcsolódó részfeladatban.

Hálás vagyok Simkó Katalinnak és Varga Viktornak, hogy a Dívány-korpusz kézi feldolgozásában alapos, lelkiismeretes munkát végeztek.

Köszönöm az SZTE BTK Magyar Nyelvészeti Tanszéke minden munkatársának, hogy éveken át együtt lehettem és dolgozhattam velük, és a mai napig barátsággal fordulnak felém. Köszönöm az SZTE BTK Szláv Intézetének, valamint az MTA Társadalomtudományi Kutatóközpont "Lendület" RECENS Kutatócsoportjának, hogy hittek bennem, és támogatásukkal, biztatásukkal segítettek a munkában.

Végül, de egyáltalán nem utolsó sorban, hálás köszönet illeti drága szüleimet, Katalint és Martint, valamint páromat, Leventét, akik mindvégig mellettem álltak, és olyan családi légkört teremtettek, amelyben ezt az értekezést megírhattam.

A jelen kutatást Az Emberi Erőforrások Minisztériuma Új Nemzeti Kiválóság Programja, a Precognox Informatikai Kft., az MTA Társadalomtudományi Kutatóközpont „Lendület” RECENS Kutatócsoportja, a European Research Council (ERC) az Európai Unió „Horizon 2020” nevű, 648693-as számú kutatási és fejlesztési programja keretében, valamint az EFOP-3.6.1-16 kódjelű projekt támogatta.

Szabó Martina Katalin

2018. június 01.

„Unlike formal systems, humans are capable of (...) contextualizing parts of their entire belief system, thus making the truth of some propositions vary with the change of internal or external context.”

(Givón 1984: 321, idézi Klégr 2013: 18)

Előszó

Doktori értekezésemben a számítógépes értékeléselemzés, másképpen szentimentelemzés egy kardinálisnak tekinthető problémáját, a szentimentérték módosulásának a jelenségét tárgyalom részletesen. A szentimentelemzés a számítógépes nyelvészet, azon belül a tartalomelemzés egy részterülete. Arra irányul, hogy a digitalizált szövegek tartalmából kivonatolja és feldolgozza azokat az információkat, amelyek értékelést fejeznek ki.

A feladat számos komoly elméleti és alkalmazott nyelvészeti kihívást rejt magában, e problémákat azonban csupán csekély számú dolgozat igyekszik behatóbban, elméleti megfontolások mellett tárgyalni.

A számítógépes értékeléselemzés egyik kardinális, alkalmazásszintű problémája a szentimentérték módosulása, amely különösen a szentimentelemzés talán legelterjedtebb módszere, a szótáralapú megoldás szempontjából bír kimagasló jelentőséggel. Értékmódosulás esetén ugyanis egy adott elemnek a szótárban rögzített értéke eltér attól az értéktől, amelyet az elemzett megnyilatkozás összességében, ténylegesen hordoz (pl. *nem jó, alig olvasható, nagyon kedves* stb.).¹

A disszertációban az értékmódosulás problémakörén belül két különösen figyelemre méltó, mégis ritkán tárgyalt sajátossággal, az általam *értékvesztésnek* és *értékváltásnak* nevezett módosulási típusokkal foglalkozom a legbehatóbban. *Értékvesztő*ként utalok azokra az elemekre, amelyek negatív szemantikai tartalommal rendelkeznek ugyan, azonban fokozó funkcióban képesek azt részben vagy egészben elveszíteni (pl. *brutálisan jó, rohadt nagy*), *értékváltó*ként pedig azokra, amelyek negatív tartalmuk ellenére bizonyos esetekben pozitív értékelést is hordozhatnak (pl. *brutális verseny, durva buli*). Vizsgálataim fókuszát azért ezek a jelenségek képezik, mert kimerítő és helytálló elméleti nyelvészeti leírásuk még nem történt meg, hatékony nyelvtechnológiai kezelésük pedig egyelőre nem megoldott.

A legfontosabb célkitűzésem az volt, hogy a szentimentérték módosulásának a problémáját – különös tekintettel a fentebb említett típusokra – úgy tegyem alapos és sokrétű vizsgálat tárgyává, hogy az elemzés közben mind az elméleti, mind az alkalmazott nyelvészeti és nyelvtechnológiai vonatkozásokat egyaránt szem előtt tartom. Emellett a jelenségeket nagy mennyiségű valós adat alapján, különböző adatbázisokra építve

¹ Az elmondottaknak megfelelően a disszertációban a szentimentérték-módosulási jelenség szinkrón, és nem diakrón vonatkozásaival foglalkozom.

vizsgáljam, feltárva ezzel a jelenségek eddig ismeretlen kvantitatív és kvalitatív sajátosságait.

Az értekezésben részletesen tárgyalom az értékeléselemzés legalapvetőbbnek tekinthető dilemmáját, a nyelvi értékelés mibenlétének kérdését, elméleti és alkalmazott nyelvészeti szempontból egyaránt, és az irodalmi megállapításokat a nyelvi értékelés sajátágaival ütköztetem (l. 2). Elvégzem a nyelvi értékelés témaköréhez kapcsolódó fogalmak tisztázását, valamint a terminológiai hiányok pótlását (l. 3). Valós példák alapján elemzem az értékvesztést és az értékváltást nyelvtechnológiai, leíró nyelvészeti, szemantikai motivációs, valamint pszichológiai szempontból, és ez utóbbi kapcsán rendszerbe foglalom a vizsgált elemeket (l. 4.1-4.5). Vizsgálati eredményeim fontos adalékul szolgálnak az irodalom megállapításaihoz, valamint több helyütt cáfolják azokat.

Az értékvesztés jelenségét a lexikai pragmatika elméleti keretében, valamint a mentális lexikon hálózatos modelljének a segítségével is vizsgálom, kísérletet téve az elemek szótári reprezentációjának megadására (4.6). Megmutatom, hogy a két megközelítési mód összeegyeztethető, sőt bizonyos problémák vonatkozásában hatékonyan kiegészítheti egymást. Mindkét elméleti keret fontos potenciált jelent a szentimentelemzés számára. Az előbbi azért, mert bár a pragmatikai vonatkozásokat szinte meg sem említik a téma kutatói, amint azt a disszertációban is látjuk, bizonyos sajátosságok esetében a pragmatikai szempontok bevonása elengedhetetlen. Az utóbbi pedig többek között azért, mert a mentális lexikont hálózatként reprezentáló modellek hozzájárulhatnak a számítógépek számára értelmezhető és kezelhető fogalmi rendszerek kialakításához.

Az értékmódosulással kapcsolatos kutatásokat három vizsgálati eszköz alapján végzem el, amelyek a doktori kutatómunka keretében jöttek létre, és hiánypótló magyar nyelvű kutatási és fejlesztési eszközök (l. 5-6). Az adatbázisok vizsgálata számos, eddig ismeretlen összefüggést tárt fel, amelyeket az értekezésben részletesen ismertetek. Emellett az eszközök e kutatáson túl is számos, a magyar nyelvű szövegek szentimentelemzésével, illetve a szentimentérték módosulási jelenségével kapcsolatos további vizsgálati és fejlesztési tevékenységhez is fontos eszközként szolgálhatnak.

A korpuszvizsgálati eredményekre alapozva létrehoztam egy saját fejlesztésű szentimentelemző rendszert, amely egy saját fejlesztésű szentimentszótár és kombinációs szabályok² alapján rendel összesített szentimentértéket az egyes mondatokhoz (l. 7). A programhoz egy felhasználói felületet is készítettünk. A munka során egy kézzel,

² l. 7.1.

mondatszinten annotált szentimentkorpuszt is létrehoztunk azért, hogy az automatikus feldolgozás eredményességét mérni lehessen. A dolgozatban ismertetem az eszköz hatékonyságát a humán annotációval, valamint az egyszerű szótáralapú elemzési eredményekkel összehasonlítva. Rámutatok, hogy a kombinációs szabályok alkalmazása az egyszerű szótáralapú elemzési megoldásnál egy jelentősen jobb minőségű feldolgozást tesz lehetővé.

Tartalomjegyzék

1. Bevezetés.....	1
1.1. A témaválasztás indoklása, a kutatás célja	1
1.2. A kutatás módszere	6
1.3. A kutatás újszerűsége.....	6
1.4. A kutatás legfontosabb hozadékai	8
1.5. A dolgozat felépítése	9
 2. A nyelvi értékelés mibenlétének a kérdése	11
2.1. A probléma általános bemutatása	11
2.2. A nyelvi értékelés megközelítési módjai az elméleti nyelvészeti irodalomban	13
2.3. A bemutatott megközelítési módok értékelése	17
2.3.1. A nyelvi értékelés mint modális kategória	17
2.3.2. A nyelvi értékelés mint szemantikai komponens	19
2.3.2.1. A domén, a target, illetve kontextus szerepe a nyelvi értékelésben	20
2.3.2.2. A kompozíciós szemantika lehetőségei a szentimentelemzésben	23
2.3.3. A nyelvi értékelés mint a beszélői szubjektivitás kifejeződése	25
2.4. A tárgyalta összegzése	26
 3. Terminológiai kérdések.....	28
 4. Az értékvesztés és az értékváltás jelenségeinek a vizsgálata	37
4.1. A szerkezeti sajátosságok áttekintése	37
4.2. Az értékvesztés és az értékváltás megközelítési módjai a leíró nyelvészeti irodalomban	40
4.3. A nyelvtechnológiai irodalom megközelítési módja	45
4.4. Pszichológiai szempontú megközelítés	47
4.5. Az értékvesztés és az értékváltás vizsgálata a szemantikai motiváció szempontjából	55
4.5.1. Az értékvesztés vizsgálata.....	60
4.5.2. Az értékváltás vizsgálata.....	68
4.6. Az értékvesztő elemek szótári reprezentációja	70

4.6.1. Az értékvesztő elemek vizsgálata a lexikai pragmatika elméleti keretében	71
4.6.2. Az értékvesztő elemek vizsgálata az aktivációterjedéses hálózati elméleti keretben	81
5. A szentimentelemzési feladat részletes ismertetése	91
5.1. A szentimentelemzés tárgya és fontossága	91
5.2. A szentimentelemzéssel kapcsolatos kutatások áttekintése	95
5.2.1. A feladat nemzetközi vonatkozásai.....	95
5.2.1.1. A targetek kezelésének a feladata	95
5.2.1.2. A szentimentshifterek kezelésének a feladata	99
5.2.1.3. A legfontosabb elemzési módszerek és eszközök	101
5.2.2. A feladat hazai vonatkozásai.....	104
5.2.3. A szentimentérték módosulása, különös tekintettel az értékvesztési és értékkváltási típusokra	108
6. A szentimentérték változásának vizsgálata adatbázisok segítségével	111
6.1. A szövegtörzsek mint a kutatás és fejlesztés eszközei	111
6.2. A szóasszociációs adatbázisok mint a kutatás és fejlesztés eszközei	117
6.3. A kutatáshoz létrehozott adatbázisok és a vizsgálati eredmények bemutatása.....	122
6.3.1. A szentimenttörzs.....	123
6.3.1.1. A törzs létrehozásának az oka és célja.....	123
6.3.1.2. A szövegválasztás indoklása, a vizsgált szövegtípus jellemzése.....	126
6.3.1.3. A nyers törzs alapvető adatai	128
6.3.1.4. A törzs feldolgozásának a módszere és eszköze.....	128
6.3.1.5. A törzs annotálásának egyetértés-mérési eredményei	135
6.3.1.6. Az annotált törzs alapvető statisztikai adatai	135
6.3.1.7. A törzs vizsgálati eredményei.....	136
6.3.2. A NegEmotív törzs.....	145
6.3.2.1. A törzs létrehozásának oka és célja	145
6.3.2.2. A szövegválasztás indoklása, a vizsgált szövegtípus jellemzése	146
6.3.2.3. A kiinduló (nyers) törzs mérete és alapvető statisztikai adatai	148
6.3.2.4. A NegEmotív törzs létrehozásának a menete és eszköze	148
6.3.2.5. A törzs annotálásának egyetértés-mérési eredményei	151

6.3.2.6. Az annotált korpusz alapvető statisztikai adatai	153
6.3.2.7. A korpusz részletes vizsgálati eredményei	153
6.3.2.8. A korpusz vizsgálati eredményei a nemek közötti nyelvhasználati eltérésekre vonatkozóan.....	158
6.3.3. A negatív emotív elemek szóasszociációs adatbázisa.....	164
6.3.3.1. Az adatbázis létrehozásának az oka és célja.....	164
6.3.3.2. A tesztjellegű adatfelvétel anyaga, eszköze és módszere	165
6.3.3.3. Az új adatfelvétel anyaga, eszköze és módszere	165
6.3.3.4. A vizsgálat hipotézisei.....	168
6.3.3.5. Az adatbázis alapvető statisztikai adatai.....	170
6.3.3.6. Az adatbázis vizsgálati eredményei.....	173
7. Egy szótár- és szabályalapú szentimentelemző eszköz létrehozása	190
7.1. A kombinációs szabályok bemutatása	191
7.2. Az elemző rendszerben alkalmazott szótárak bemutatása	195
7.3. A teszteléshez használt korpusz bemutatása.....	196
7.3.1. A korpusz létrehozásának menete és módszere	197
7.3.2. Az annotátorok közötti egyetértésmérés eredményei	199
7.3.3. A kézzel annotált korpusz alapvető statisztikai adatai	200
7.4. A korpusz szótáralapú elemzése és az eredmények kiértékelése.....	202
7.5. A korpusz szabályalapú elemzése, a módszer kiértékelése és összevetése a szótáralapú módszerrel.....	207
8. Összegzés	216
IRODALOMJEGYZÉK	229
FORRÁSJEGYZÉK	253
FELHASZNÁLT ELEMZŐ ESZKÖZÖK	253
MELLÉKLETEK	254
JELMAGYARÁZAT	263

Táblázatok jegyzéke

1. Nyolc angol nyelvű határozószó szentimentértéke három szentimentadatbázisban (vö. Dragut–Fellbaum 2014: 39).....	46
2. Négy különböző folyadéknév és a hideg szó kapcsolatainak a gyakorisága az MNSZ2-ben, valamint közülük a fokozó elemmel módosított kapcsolatok száma.....	63
3. A korpuszannotáció alapvető statisztikai adatai.....	136
4. A negáló elemeknek a szentimentkifejezésekhez képesti pozíciója az ellentétes értékű fragmentumokban	141
5. A nyers Twitter-korpusz alapvető statisztikai adatai.....	147
6. Az annotálási rendszer rövid összefoglalása	151
7. A Cohen-féle Kappa-statisztika eredményeinek értelmezési tartományai (vö. Landis and Koch 1977)	152
8. Az annotált értékvesztési esetek részletes statisztikai adatai.....	154
9. Az öt leggyakoribb elem összes előfordulásának és kollokációinak a száma, valamint azok aránya	156
10. A hívószók listái, szemantikai tartalom szerinti bontásban.....	167
11. Az egyes hívószók asszociátumainak legnagyobb és legkisebb átfedései	182
12. Az asszociátumok szófaji megoszlásának összevetése a nők és a férfiak adatai között	186
13. Az asszociátumok szentiment- és emócióértékeinek összevetése a nők és a férfiak között	187
14. A kombinációs szabályok, példákkal	194
15. A pontosság, a fedés és az F-mérték a szótáras elemzés esetében, a kézi annotáció öt értéke szerinti bontásban és összesítve	204
16. A humán és a szótáralapú elemzés három alapvető értékének különböző variációjú együttes előfordulása, azok gyakorisági adataival együtt.....	206
17. A pontosság, a fedés és az F-mérték a szabályalapú elemzés esetében, a kézi annotáció öt értéke szerinti bontásban és összesítve	211
18. A humán és a szabályalapú elemzés három alapvető értékének különböző variációjú együttes előfordulása, azok gyakorisági adataival együtt.....	213

Ábrák jegyzéke

1. Az írott nyelvi szövegtörzsek osztályozásának egy lehetséges módja.....	114
2. Példa a szentimenttörzsek annotálási megoldására	134
3. A negatív és a pozitív szentimentkifejezések együttes előfordulásának gyakorisági megoszlása a pozitív és a negatív fragmentumokban	139
4. A negatív és a negatív szentimentkifejezések együttes előfordulásának gyakorisági megoszlása a pozitív és a negatív fragmentumokban	140
5. A negatív és a szentimentkifejezések egymáshoz képesti pozíciós sajátosságai az ellentétes értékű fragmentumokban	142
6. Részlet a törzsek annotációjából a Brat nevű programban.....	151
7. A <i>baromi</i> elem melléknévi kollokációi, a törzsbeli frekvenciájukkal együtt az MNSZ2-ben	157
8. Részlet a <i>baromi meglepő</i> kollokáció 115 előfordulásából az MNSZ2-ben	158
9. A deszemantizált fokozó elemek gyakorisága a két nemnél a különböző szemantikai tartalmú alaptörzsek mellett	160
10. Az értékvesztő elemek gyakorisági megoszlása a férfiak és a nők törzsében	162
11. A kitöltők száma az egyes korcsoportokban	171
12. A válaszadók asszociátumainak átlagos száma, korcsoportok szerinti bontásban	172
13. Az asszociátumok szófaji megoszlása az összes válaszszóra nézve	174
14. Az asszociátumok szófaji megoszlása a hívószók szemantikai tartalma szerinti bontásban	175
15. Az asszociátumok szófaji megoszlása a hívószók grammatikai sajátossága szerinti bontásban	176
16. Az első helyen adott asszociátumok szófaji megoszlása.....	177
17. Az első helyen adott asszociátumok szófaji megoszlása a hívószók grammatikai sajátossága szerinti bontásban	178
18. Az asszociátumok összesített szentimentelemzési eredményei	179
19. Az asszociátumok összesített szentimentelemzési eredményei, a hívószók hatos csoportosítás szerint.....	180
20. Az asszociátumok átfedési mátrixa	181

21. Az értékvesztésre – és esetlegesen értékváltásra is – képes asszociátumok gyakorisági megoszlása a negatív és a nem negatív hívószók szerinti bontásban	184
22. Az értékvesztésre – és esetlegesen értékváltásra is – képes asszociátumok gyakorisági megoszlása a hívószók hat kategóriája szerinti bontásban	185
23. Az annotált korpusz alapvető statisztikai adatai.....	200
24. A kézi annotálás eredménye a négy alkorpusz szerinti bontásban.....	201
25. A szótáras elemzési eredmények statisztikai adatai	203
26. A szótáras elemzés eredményei a humán annotáció öt értéke szerinti bontásban.....	204
27. A szótáras elemzés eredménye a négy alkorpusz szerinti bontásban.....	205
28. A szabályalapú elemzési eredmények statisztikai adatai	208
29. A szabályalapú elemzési eredmények a humán annotáció öt értéke szerinti bontásban	210
30. A szabályalapú elemzés eredménye a négy alkorpusz szerinti bontásban	212
31. A létrehozott szentimentelemző eszköz jelenlegi felhasználói felülete	214

1. Bevezetés

1.1 A témaválasztás indoklása, a kutatás célja

Doktori értekezésemben a számítógépes értékelelemzés, másképpen szentimentelemzés egy kardinálisnak tekinthető problémáját, a szentimentérték módosulásának a jelenségét tárgyalom részletesen.

A szentimentelemzés a számítógépes nyelvészet, azon belül a számítógépes tartalomelemzés egy részterülete. Arra irányul, hogy a digitalizált szövegek tartalmából kivonatolja és feldolgozza azokat az információkat, amelyek értékelést fejeznek ki. E kivonatolási és feldolgozási munka alatt a következő lépéseket értjük: az elemző eszköz az értékelő tartalmakat megtalálja a szövegekben, meghatározza ezeknek az értékeknek a típusát, vagyis pozitív vagy negatív voltát, esetlegesen az adott érték intenzitásának a mértékét, valamint, hogy megállapítsa, mi képezi az értékelés tárgyát (számítógépes nyelvészeti terminussal: *target*ét), tehát azt, hogy az értékelés mire irányul (vö. Szabó 2016: 153). Mindezt a természetesnyelv-feldolgozás (az angol nyelvű terminológiában: *Natural language processing*, NLP) kínálta automatikus eszközökkel igyekszik megvalósítani. A feladat az alábbi egyszerű példával szemléltethető:

(1) A szentimentelemzés nagyon érdekes feladat.

Az (1) alatti példában a *target* a *szentimentelemzés*, amelyet a beszélő nagyon érdekes feladatnak minősít. Az *érdekes* kifejezésre úgy tekintek, mint inherens, azaz lexikai szinten pozitív értékelő tartalommal rendelkező elemre, amely által a megnyilatkozásban a *target* pozitív értékelő minősítést kap, s ezt a minősítést a *nagyon* elem fokozza.

Ugyanakkor, egy nyelvi elem lexikai szintű szentimentértéke nem feltétlenül azonos vagy teljes mértékben azonos az őt magában foglaló megnyilatkozás szentimentértékével (vö. Bolinger 1972: 15–20, Feldman és mtsai 2011: 1643, Liu 2012: 36–38, Szabó 2015a: 630–633). A lexikai szintű szentimentérték számos okból kifolyólag eltérhet a bennfoglaló megnyilatkozás értékétől. Így például egy kifejezés negálható (2a), az értéke csökkenthető (2b) vagy növelhető (2c), sőt, például az irónia eszközével az ellentétére is fordítható (2d) (vö. Szabó 2015a: 631–632).

- (2) a. **Nem** hazudott.
b. **Valamennyire** tehetséges.
c. **Rendkívül** okos.
d. Ezt jól megcsináltad!

E problémakörhöz kapcsolódnak a dolgozatban különös részletességgel tárgyalt ún. *értékvesztés* és *értékváltás*. Mindkét nyelvi jelenség esetében a vizsgált elemek lexikai szinten negatív emotív szemantikai tartalommal rendelkeznek, azonban fokozó funkcióban képesek azt részben vagy egészben elveszíteni (3a), sőt bizonyos esetekben arra is, hogy negatív tartalmuk ellenére pozitív értékelést fejezzenek ki (3b) (vö. Szabó 2015a: 632, 2015c: 52–54). A példákban a vizsgált elemeket kövér szedéssel emelem ki.

- (3) a. **mocskos** nagy, **rohadt** jól
b. **brutális** volt a hétvégi koncert

Az előbbi esetben a vizsgált elemek más, értékelő kifejezések jelentését fokozzák (ezt a jelenséget nevezem *értékvesztés*nek), az utóbbi esetben pedig maguk töltik be értékelő kifejezés (szentimentkifejezés) funkcióját (ezt a jelenséget nevezem *értékváltás*nak) (vö. Szabó és mtsai 2017a: 251–252).

A szentimentelemzés tehát mindezeket az értékelő tartalmakat igyekszik automatikus módszerrel megtalálni és helyesen feldolgozni a szövegekben.

Az automatikus értékeléselemzés a nyelvtechnológiai kutatások palettáján még viszonylag új keletűnek tekinthető, napjainkban azonban a nemzetközi kutatásban és fejlesztésben egyre nagyobb figyelmet kap (vö. Pang–Lee 2008, Liu 2012). A szentimentelemzés népszerűvé válásának oka egyrészt a feladat elméleti nyelvészeti, valamint nyelvtechnológiai kihívásaiban, másrészt az eredmények – javarészt gazdasági – hasznosítási lehetőségeiben keresendő. A téma különböző aspektusaival számtalan nemzetközi publikáció foglalkozik (vö. pl. a gépi tanulás különböző lehetőségeivel pl. Pang és mtsai 2002, Turney 2002, szövegszintű szentimentosztályozással pl. Hu–Liu 2004, tagmondat-szintű elemzéssel pl. Wilson és mtsai 2005, a szótáralapú szentimentelemzés hatékonyságának kérdéseivel pl. Young–Soroka 2012, doménorientált szentimentszótárak létrehozásával pl. Kanayama–Nasukawa 2006, Qiu és mtsai 2009, Loughran–McDonald 2011, tőzsdeindexek mozgásának az előrejelzésével pl. Feldman és mtsai 2010, a teljes problémakör áttekintésével Liu 2012, a keretszemantikára építő elemzés lehetőségeivel pl.

Ruppenhofer–Rehbein 2012, szintaktikai elemzésre építő megoldásokkal pl. Socher és mtsai 2013), és megjelentek a hazai, kifejezetten a magyar nyelvű szövegek szentimentelemzésére fókuszáló dolgozatok is (vö. Berend–Farkas 2008, Miháltz 2010, 2013, Hangya és mtsai 2015, 2017). Fontos rámutatnunk azonban, hogy a publikációk túlnyomó többsége nyelvtechnológiai oldalról közelít a nyelvi értékeléselemzés problémáihoz; azok alkalmazott, és különösen elméleti vonatkozásait csekély számú dolgozat igyekszik behatóbban tárgyalni. Ezeknek a kérdéseknek az alapos, nyelvészeti szempontú vizsgálata azonban a fejlesztői oldal számára is fontos lenne, hiszen az elméleti és alkalmazott nyelvészeti kutatások eredményei hatékonyan képesek támogatni a nyelvtechnológiai fejlesztési feladatokat. A jelenség elméleti szintű leírása tehát kétségkívül nélkülözhetetlen egy helytálló alkalmazott nyelvészeti kutatás alapjainak a megteremtéséhez.

A számítógépes értékeléselemzés legalapvetőbb elméleti problémája a következőben határozható meg: a nyelvi értékelés mibenléte nem tisztázott, s ezzel összefüggésben az sem egyértelmű, hogy milyen, illetve melyik nyelvészeti kategóriaként képzelhető el. A jelen munka ezért részben ennek a hiánynak a pótlását is célozza.

A szentimentérték módosulását, illetve az esetleges módosító elemek különböző típusait illetően megállapítható, hogy a szerzők az egyes típusokat áttekintő jelleggel, behatóbb vizsgálat keretében a szentimentelemzés szempontjából nem tárgyalják. Nem találtam a nemzetközi irodalomban egyetlen olyan dolgozatot sem, amely a problémakört rendszerező igénnyel vizsgálná. A kapcsolódó kutatások alapvetően a jelenség egy-egy részproblémájára fókuszálnak, például a negáció kérdésére (vö. pl. Na és mtsai 2004, Kennedy–Inkpen 2006, Wilson és mtsai 2009).

A szentimentérték módosulásának két speciális, az általam *értékváltásnak* és *értékvesztésnek* nevezett típusait a disszertáció különös figyelemmel tárgyalja. A két jelenség mind elméleti, mind alkalmazott nyelvészeti, különösen nyelvtechnológiai szempontból figyelemre méltó. Kimerítő elméleti nyelvészeti leírásuk még nem történt meg, hatékony nyelvtechnológiai kezelésük pedig egyelőre nem megoldott (vö. pl. Andor 2011, Kugler 2014, Szabó 2015a: 52, Szabó és mtsai 2017a: 252, Drávucz és mtsai 2017: 236). Ezekre az elemekre ugyanis a szótáralapú automatikus értékeléselemzés – a lexikai szintű szentimentértékük miatt – téves elemzési eredményt ad.

Az értékváltás és az értékvesztés jelenségeivel csekély számú dolgozat foglalkozik (vö. Tolcsvai Nagy 1988, Partington 1993, Wierzbicka 2002, Jing-Schmidt: 2007, Laczkó 2007, Andor 2011, Kugler 2014, Dragut–Fellbaum 2014), és ezek többsége is csupán az

értékvesztésre fordít figyelmet. Az értékvesztéssel foglalkozó dolgozatok többsége a pszichológia, valamint az elmélet oldaláról teszi vizsgálatát tárgyává a problémát (vö. Tolcsvai Nagy 1988, Partington 1993, Wierzbicka 2002, Jing-Schmidt 2007, Laczkó 2007, Andor 2011, Kugler 2014). Egyetlen olyan dolgozatról (Dragut–Fellbaum 2014) van tudomásunk, amely az értékvesztés jelenségét nyelvtechnológiai szempontból közelíti meg.

A jelenlegi, szentimentelemzést célzó kutatásokra általánosságban jellemző, hogy a feladatot az úgynevezett *szenimentkifejezések*¹ (*sentiment words*) szótári formába rendezett listáival igyekeznek – legalább részben – megoldani (vö. Vázquez–Bel 2012, Liu 2012: 12, Szabó 2014, 2015b). Mindez a gyakorlatban azt jelenti, hogy az elemzők szólisták (másképpen: szentimentszótárak) formájába rendezik azokat a legkisebb nyelvi egységeket, amelyek a szótár készítőinek megítélése szerint önmagukban, tehát lexikai szinten pozitív vagy negatív értékelő tartalmat hordoznak. A szótár létrehozását követően az automatikus szentimentelemzést szótáralapon, egyszerű szóillesztéses módszerrel hajtják végre, a következőképpen: minden olyan kifejezéshez, amelyet a program a szentimentszótár alapján az elemzés tárgyát képező nyers szövegben azonosít, a szótár szerinti pozitív vagy negatív értéket rendel.

E szótárak a legtöbb szentimentelemzési megoldásban kulcsszerepet töltenek be. Ugyanakkor a szentimentérték módosulásának a jelensége a szótáralapú szentimentelemzés hatékonyságát jelentősen csökkenti, hiszen emiatt a szótárban rögzített szentimentértékek összegzése nem feltétlenül azt a szentimentértéket adja, amellyel az adott megnyilatkozás ténylegesen rendelkezik (vö. Hu–Liu 2004).²

Az angol nyelvű nyelvtechnológiai irodalom *szenimentshiftereknek* (*sentiment shifters*) nevezi azokat a nyelvi elemeket, amelyek mintegy értékbeli eltolódást okoznak a teljes megnyilatkozás szentimentértékében a benne szereplő szenimentkifejezések inherens értékelő tartalmának összegéhez képest (vö. Hu–Liu 2004, Liu 2012: 60). A shifterek problémáját a szerzők a legtöbb esetben úgy igyekeznek megoldani, hogy a szentimentszótárakkal együtt különböző kombinációs szabályokat alkalmaznak (vö. pl. Feldman és mtsai 2011: 1643, Ruppenhofer–Rehbein 2012). A kombinációs szabályok a szentimentszótárakban foglalt pozitív és negatív szenimentkifejezések, valamint bizonyos egyéb nyelvi elemek egymáshoz képesti pozícióit kódolják. Azt határozzák meg, hogy

¹ Abból a megfontolásból használom a *szenimentkifejezés* terminust (eltérően az angol *sentiment word* terminus szó szerinti fordításától), hogy szenimentkifejezésként nem csupán egy-, de többszavas frazeológiai egységek is szolgálhatnak; például *értéktelen ~ egy fillért nem ér* (vö. Szabó 2014: 14).

² Az egyes szenimentkifejezések értékének a súlyozása fontos szereppel bír az automatikus értékeléselemzésben, erről a problémáról a dolgozatban később részletesen lesz szó, l. 5.2.1.1, 6.3.1.7).

amennyiben egy adott elemzési egységen belül egy bizonyos elem egy szentimentkifejezéshez képest valamilyen pozíciót foglal el, úgy az elemző rendszer az adott szentimentkifejezéshez a szótár alapján rendelt értéket a szabályban foglaltak szerint módosítja. Ezek a megoldások arra törekszenek, hogy a szentimentszótárak elemeinek lexikai szintű szentimentértékéből, a shifterek hatását figyelembe véve mintegy "kiszámítsák" a megnyilatkozás értékét. Ezzel összefüggésben, bizonyos elemzések (vö. pl. Loughran–McDonald 2011) már nem csupán a pozitív és a negatív szentimentkifejezések szótárát alkalmazza, hanem a különböző shiftereket is listákba rendezik. Az ún. Loughran–McDonald pénzügyi szentimentlexikon (vö. Loughran–McDonald 2011) például a hagyományos szentimentszótárak mellett egyéb specifikus listákat is tartalmaz.

A kombinációs szabályok alkalmazása a szentimentelemzésben több problémát is felvet, ráadásul nem fedhető le velük minden jelenség. Egyrészt, amint arra a (2a-c) alatti példa is rámutat (l. fentebb), amíg bizonyos megnyilatkozások esetében van olyan nyelvi elem, amely azonosítható az értékmódosulási folyamatban (pl. egy tagadó operátor), addig más esetekben ilyen elemet nem, vagy nem feltétlenül tudunk azonosítani (pl. a (2d) alatti példa esetében).

Másrészt, az értékelést kifejező megnyilatkozások gyakorta nem teljesen kompozicionálisak (részletesebben l. 2.3.2.2). Ahogyan arra több szerző (vö. Israel 2004: 711, Feldman és mtsai 2010: 4) is rámutat, a szentimentérték nem kalkulálható ki minden esetben maradéktalanul kombinációs szabályok segítségével.³ A kombinációs szabályokat alkalmazó megoldás ugyanis szintén szótáralapú, és ebből kifolyólag – akárcsak a pusztán szótáralapú megoldás – a lexémák inherens értékelő tartalmával dolgozik. Ez például a disszertáció fókuszában álló értékvesztés és értékváltás jelenségeinek a szempontjából is problematikus lehet.

A disszertációban az említetteken túl arra is vállalkozom, hogy megvizsgálom a szentimentszótárak és a kombinációs szabályok működési hatékonyságát. A vizsgálathoz szükséges korpuszt, szótárakat és szabályokat ebben az esetben is magam hoztam létre.

³ A dolgozat bírálója, Sass Bálint felveti, hogy nem jelentene-e megoldást a problémára, ha a nem kompozicionális kifejezéseket a targetekkel együtt vennénk fel a szótárba. Ez a módszer egyszerre orvosolhatná a „nemkompozicionalitás”, valamint a targetfüggés problémáját. Véleményem szerint ez két okból problémás: Egyrészt, a targetek meglátásom szerint kimeríthetetlen halmazt alkotnak, egyszerűen lehetetlen minden lehetséges targetet listába foglalni. Másrészt úgy gondolom, hogy bizonyos kifejezések értékelő tartalma nem is a target, hanem a tágabb kontextustól, illetve az aktuális doméntól függ. Ezekben az esetekben tehát értelmetlen volna a kifejezéseket targetekkel együtt szótárazni.

1.2 A kutatás módszere

A disszertációban bemutatott kutatómunka több, különböző típusú munkafázisból tevődik össze.

A nyelvi értékelés mibenlétének tárgyalásához áttekintem a nyelvi értékelés legfontosabb megközelítési módjait az elméleti nyelvészeti irodalomban.

Az értékvesztést és az értékváltást illetően bemutatom a szakirodalmi, köztük az elméleti nyelvészeti, a nyelvtechnológiai, valamint a pszichológiai megállapításokat, majd a jelenségeket a szemantikai motiváció szempontjából, korpuszadatok⁴ (korpuszbeli (MNSZ2, vö. Oravecz és mtsai 2014), valamint további valós forrásokból származó példák) segítségével vizsgálom. Ezt követően, a kutatási tapasztalatokra építve elemzem a jelenséget a lexikai pragmatika, valamint a hálózatos jelentésreprezentáció elméleti kereteiben, és kísérletet teszek a két keret összehangolására.

A szentimentérték módosulásának nyelvtechnológiai szempontú elemzéséhez áttekintem a hazai és a nemzetközi irodalom legfontosabb megállapításait a szentimentelemzésre vonatkozóan, valamint a különböző adatbázisok nyelvtechnológiai felhasználási lehetőségeit illetően. Az értékmódosulási jelenségek kvantitatív és kvalitatív sajátosságait különböző, saját szerzőségű adatbázisok (korpuszok, valamint egy szóasszociációs adatbázis) alapján tárom fel. A megismert kvantitatív és kvalitatív sajátságokat az irodalom megállapításaival, valamint saját, a disszertációban ismertetett elemzési eredményeimmel is összevetem.

Végül, a kutatási eredményekre építve létrehozok egy szabályalapú automatikus szentimentelemző rendszert, amellyel a szentimentszótárak és a kombinációs szabályok létjogosultságát kívánom felmérni.

1.3 A kutatás újszerűsége

A disszertációban bemutatott kutatás fontos újdonsága, hogy behatóan tárgyalja a szentimentérték módosulásának jelenségét mind elméleti, mind alkalmazott nyelvészeti szempontból, alapvetően a magyar nyelvű szövegek automatikus értékeléselemzésére fókuszálva. A szentimentérték módosulási jelenségét az elméleti irodalom nem tárgyalja

⁴ A dolgozat a *korpuszadat* megnevezést az *introspektív adat* terminussal szembeállítva használja (vö. Kertész-Rákosi 2008: 43).

kimerítően, illetve a megközelítési módok nem nyújtanak megfelelő elméleti alapot az alkalmazott kutatásokhoz (l. 1.2). Emellett, a magyar nyelvű szövegek értékelemzésének problémájával behatóbban, nyelvészeti szempontokat is figyelembe vevő vizsgálat keretében rendkívül csekély számú dolgozat foglalkozik (vö. Berend–Farkas 2008, Miháltz 2010, 2013, Hangya és mtsai 2015). Ugyanakkor, ezek a dolgozatok nem nyújtanak kellő információt azokról a kardinális nyelvészeti problémákról, valamint az e problémák vonatkozásában alkalmazott elméleti és gyakorlati szintű megoldásokról, amelyek meghatározzák az automatikus értékelemzés eredményességét.

A szentimentérték módosulását, illetve a szentimentshifterek különböző típusait illetően megállapítható, hogy a szerzők az egyes típusokat áttekintő jelleggel, behatóbb vizsgálat keretében a szentimentelemzés szempontjából nem tárgyalják. Nem találtam a nemzetközi irodalomban egyetlen olyan dolgozatot sem, amely a problémakört rendszerező igénnyel vizsgálná. A kapcsolódó kutatások alapvetően a jelenség egy-egy részproblémájára fókuszálnak, például a negáció kérdésére (vö. pl. Na és mtsai 2004, Kennedy–Inkpen 2006, Wilson és mtsai 2009).

Ami a magyar nyelvre vonatkozó, szentimentelemzéshez kapcsolódó kutatásokat illeti, az említett problémákat saját dolgozataimon (Szabó 2015a, 2015b, 2015c, 2017, Szabó–Vincze 2015, 2016, Szabó és mtsai 2016a, 2016b, 2017a, 2017b, 2017d) kívül más nem tárgyalja. A jelen kutatás tehát ebből a szempontból is hiánypótlónak tekinthető.

A dolgozat fontos eredménye, hogy az értékvesztés és az értékváltás jelenségeit különös alaposággal tárgyalja. Mindkét említett jelenség figyelemre méltó, elméleti és alkalmazott nyelvészeti szempontból egyaránt, és az ezeket tárgyaló dolgozatok mennyisége, illetve az ezeket a jelenségeket illető megállapítások száma kifejezetten kevés. A hazai nyelvtechnológiai irodalomban az említett problémákat – egyetlen rövid megjegyzéstől eltekintve (vö. Hangya és mtsai 2015: 212) – saját munkáimon (Szabó 2015a, 2015c, 2016, 2017 és Szabó és mtsai 2017a, 2017b, 2017d) kívül más nem vizsgálja.

Az értékvesztés és az értékváltás jelenségeit a szemantikai motiváció szempontjából részletesen tárgyalom, valamint a lexikai pragmatika, valamint a hálózatos jelentésreprezentáció elméleti kereteiben is elemzem, különös tekintettel az értékvesztési típusra.

A dolgozatban bemutatott kutatás további fontos újdonsága, hogy a jelenségeket nagy mennyiségű nyelvi adat alapján, különböző adatbázisokra építve vizsgálom, feltárva ezzel a jelenségek eddig ismeretlen kvantitatív és kvalitatív sajátosságait. Mindemellett, a

kutatáshoz felhasznált vizsgálati eszközök (két, kézzel annotált korpusz és egy szóasszociációs adatbázis) e kutatómunka keretében jöttek létre, és rendre hiánypótlóak.

A kutatási eredményekre alapozva egy szabályalapú automatikus szentimentelemző rendszert is létrehozok, amelyhez tesztkorpuszt és szentimentszótárakat is fejleszték.

1.4 A kutatás legfontosabb hozadékai

A kutatás három legfontosabb elméleti szintű hozadéka a következő: Egyrészt, behatóan tárgyalom a szentimentelemzés legalapvetőbb dilemmáját, a nyelvi értékelés mibenlétének kérdését, valamint a szentimentérték módosulási lehetőségeit (l. 2). Másrészt, az értékváltás és az értékvesztés jelenségeit illetően feltárom a szemantikai motivációs sajátosságokat (l. 4.5). Harmadrészt, a lexikai pragmatika, valamint a hálózatos jelentésrepresentáció elméleti kereteit használva szótári reprezentációt adok a vizsgált elemekhez (l. 4.6).

A szentimentérték módosulási jelenségét különböző adatbázisokon alapuló saját vizsgálati eredményekre támaszkodva tárgyalom (l. 6). A feltárt sajátosságok eddig ismeretlenek voltak, és remélhetőleg fontos adalékul szolgálhatnak a jelenség elméleti és alkalmazott nyelvészeti kutatásához, valamint hozzájárulhatnak az automatikus értékeléselemzés eredményességének a javulásához is.

A munka további jelentős hozadéka, hogy három vizsgálati eszköz, köztük két kézzel annotált szövegkorpusz és egy szóasszociációs adatbázis, valamint egy a fejlesztő munkához szükséges, ugyancsak manuálisan feldolgozott korpusz a doktori kutatás keretében jöttek létre (l. 6). Mindezek hiánypótló magyar nyelvű kutatási és fejlesztési eszközök, amelyek számos, a magyar nyelvű szövegek szentimentelemzésével, illetve a szentimentérték módosulási jelenségével kapcsolatos további kutatási és fejlesztési tevékenységhez is hathatós eszközként szolgálhatnak.

A vizsgálati tapasztalatok alapján létrehoztam egy automatikus szentimentelemző eszközt, amely hatékonyabb feldolgozási megoldást nyújt a szótáras szentimentelemzési módszerhez képest (l. 7).

A kutatás keretében különböző, a szentimentelemzésben használható szótárak is létrejöttek, amelyek közül a két szentimentlexikon, amely a pozitív és a negatív értékelő tartalommal rendelkező kifejezéseket tartalmazza, szabadon hozzáférhető kutatási célokra

(opendata.hu) (l. 7). A tervezetteknek megfelelően a többi eszköz is elérhetővé válik a közeljövőben.

1.5 A dolgozat felépítése

Az értekezés öt nagyobb részből épül fel.

Az első egységben (2. fejezet) a nyelvi értékelés mibenlétének kérdését tárgyalom részletesen, elméleti és alkalmazott nyelvészeti szempontból egyaránt. Áttekintem az elméleti nyelvészeti irodalom vonatkozó megállapításait, és azokat a nyelvi értékelés sajátágaival ütköztetem.

A második részben az értékvtáltási és az értékvesztési jelenségeket tárgyalom behatóan. Elvégzem a nyelvi értékelés témaköréhez kapcsolódó fogalmak tisztázását és a terminológiai hiányok pótlását (3. fejezet). Részletesen ismertetem azt a rendszerezést, amelyet a szentimentérték módosulásának egyes típusaira alkalmazok. Ezt követően (4. fejezet) áttekintem a kapcsolódó elméleti nyelvészeti, nyelvtechnológiai, valamint pszichológiai vizsgálati eredményeket. Korpuszadatok alapján feltárom a jelenség nyelvi sajátságait a szemantikai motiváció és a produktivitás szempontjából. Végül a problémát lexikai pragmatikai, valamint a hálózatos jelentésreprezentáció elméleti keretében tárgyalom.

A disszertáció harmadik nagyobb egységében (5. fejezet) a számítógépes értékeléselemzéssel foglalkozom. A szentimentelemzéssel kapcsolatos nemzetközi és hazai kutatások áttekintését követően a szentimentérték módosulását tárgyalom, kifejezetten a nyelvtechnológia szempontjából, különös figyelmet fordítva az automatikus elemzés számára komoly kihívást jelentő értékvesztésre és értékvtáltásra.

A dolgozat negyedik nagy egységében (6. fejezet) azokat az értékvesztéssel és értékvtáltással kapcsolatos vizsgálatokat mutatom be, amelyeket az erre a célra létrehozott három adatbázis alapján végeztem el. Áttekintem a korpuszok és a szóasszociációs adatbázisok felhasználási lehetőségeit és típusait, valamint legfontosabb nemzetközi és hazai példáit. A korpuszok kapcsán foglalkozom az azokban feldolgozott szövegtípusok jellemzőivel. Ezt követően közlöm az adatbázisok létrehozásának elméleti alapvetéseit, pontos menetét, eszközeit, forrásait, valamint statisztikai adatait.

Az értekezés utolsó részében (7. fejezet) ismertetek egy olyan automatikus szentimentelemzési megoldást, amelyet a bemutatott vizsgálati eredményekre alapozva

hoztam létre, és amellyel a szentimentérték módosulásának problémája részben kezelhetővé válik a magyar nyelvű szövegek automatikus értékeléselemzésének a feladatában. Részletesen tárgyalom az implementáció előkészítő lépéseit, a megvalósítás menetét, tesztelem a módszer hatékonyságát, és a mérési eredményeket részleteiben is ismertetem.

2. A nyelvi értékelés mibenlétének a kérdése

2.1 A probléma általános bemutatása

„A számítógépes nyelvészeti rendszerek nagy része valamilyen elméleti nyelvészeti keretrendszeren alapul, azaz az elméleti nyelvészeti kutatások eredményeit felhasználja a nyelvtechnológia” (Gábor és mtsai 2008: 855). Rendkívül fontos tehát, hogy a nyelvtechnológiai kutatási és fejlesztési feladatok stabil elméleti nyelvészeti eredményekre támaszkodhassanak.

A nyelvi értékeléselemzésének a legalapvetőbb dilemmája – akár az elmélet, akár az alkalmazott nyelvészet, így az automatikus megoldások oldaláról közelítjük azt meg – a nyelvi értékelő jelentés mibenléte. A probléma megoldása tehát kiemelkedően fontos a szentimentelemzés szempontjából is, hiszen az automatikus elemzés során mindenekelőtt pontosan meg kell tudni határozni az elemzés tárgyát. A jelen fejezetben ezért a nyelvi értékelés mibenlétét tárgyalom részletesen.

Mindenekelőtt annak céljából, hogy a problémát behatóbb vizsgálat tárgyává tehessek, röviden összefoglalom a nyelvi értékelés alapvető megközelítési módjait az automatikus értékeléselemzés feladatkörében.

Amint arról a dolgozat bevezető fejezetében (l. 1.1) már említést tettem, a szentimentelemzést célzó kutatások legtöbbször a feladatot az úgynevezett *szentimentkifejezések* (*sentiment words*) szótári formába rendezett listáival igyekeznek – legalább részben – megoldani. Ezt az eljárásmodot nevezzük szótáralapú elemzésnek (*Lexicon-based method*) (vö. Vázquez-Bel 2012, Liu 2012: 12–13). Az elemzéshez a szakértők szólisták formájába rendezik azokat a legkisebb nyelvi egységeket, amelyek megítélésük szerint önmagukban, lexikai szinten pozitív vagy negatív értékelő tartalmat hordoznak. Ezt követően az automatikus értékeléselemzést szótáralapon, egyszerű szóillesztéses módszerrel hajtják végre, a következőképpen: minden olyan kifejezéshez, amelyet a program a két szentimentszótár alapján a nyers szövegben azonosít, a szótár szerinti pozitív vagy negatív értéket rendeli. Ezt követően a szoftver a fejlesztők által meghatározott méretű elemzési egységekhez (a teljes dokumentum, mondat vagy a mondatnál kisebb szövegrész) azt az értéket rendeli, amelyet a kapott értékek összegéből kiszámít (vö. Hu–Liu 2004, Liu 2012: 49–50).

A szentimentelemzés terén már-már vezérelvnek tekinthető az a megállapítás, hogy az értékelő szemantikai tartalmak legfontosabb indikátorai e pusztán szentimentkifejezések, s ezzel összefüggésben a szótárak a legtöbb szentimentelemzési megoldásban kulcsszerepet töltenek be (vö. Liu 2012: 12).

A szentimentelemzésben a szótárhasználatra építő megközelítési mód azon a feltételezésen alapul, hogy bizonyos nyelvi elemek pozitív vagy negatív értékelő jelentéskomponenssel rendelkeznek a kontextuális sajátságoktól függetlenül; másképpen, egy meghatározott szentimentértéket hordoznak mintegy szemantikai szerkezetük részeként (vö. Baroni–Vegnaduzzo 2004: 18, Kamps és mtsai 2004, Vegnaduzzo 2004, Moghaddam–Popowich 2010: 3). Ugyanakkor, amint azt korábban már tárgyaltam (l. 1.1) a szótáralapú szentimentelemzés hatékonyságát jelentősen csökkenti – többek között – a domén-, illetve targetfüggés jelensége: számos nyelvi elem egy adott szövegdoménben, illetve egy adott targetre (vagy célpontra) vonatkozóan eltérő szentimentértékkel rendelkezik valamely más doménbeli sajátságához képest (vö. Ahn és mtsai 2012, Vázquez–Bel 2012: 3557, Vechtomova és mtsai 2014: 553–554). Ennek okán egyre elterjedtebb az úgynevezett doménfüggő szentimentszótárak alkalmazása, amelyek lexikai anyagukban, valamint a bennfoglalt kifejezések szentimentértékét tekintve igyekeznek az aktuális szövegdomén sajátságaihoz igazodni (vö. pl. Hangya és mtsai 2015).

E speciális, az elemezni kívánt szövegek tartalmi sajátságait is figyelembe vevő szentimentszótárak ugyanakkor a szentimentérték módosulásának a jelenségét egyáltalán nem képesek kezelni (vö. pl. Feldman és mtsai 2011: 1643). Ahogyan azt már a dolgozat bevezetőjében röviden ismertettem (l. 1.1), a számítógépes értékeléselemzés egy kardinális, alkalmazásszintű problémája az ún. szentimentérték-módosulás jelensége, amelynek esetében egy nyelvi elem lexikai szintű szentimentértéke nem azonos, vagy nem teljes mértékben azonos az őt magában foglaló teljes megnyilatkozás szentimentértékével (vö. Bolinger 1972: 15–20, Feldman és mtsai 2011: 1643, Liu 2012: 36–38, Szabó 2015a: 630–633). A korábban (2) alatt közölt példákat itt (4) alatt megismétlem.

- (4) a. **Nem** hazudott.
- b. **Valamennyire** tehetséges.
- c. **Rendkívül** okos.
- d. Ezt jól megcsináltad!

A problémát a szerzők a legtöbb esetben úgy igyekeznek megoldani, hogy az említett szentimentszótárak mellett különböző szemantikai kombinációs szabályok megalkotását és alkalmazását javasolják. Olyan szentimentelemző rendszer létrehozását célozzák tehát, amely a szentimentszótárral, valamint az egyéb megoldásokkal azonosított elemek különböző, előre programozott kombinációiból képes mintegy kiszámítania teljes elemzési egység szentimentértékét (vö. Feldman és mtsai 2010: 3). A megoldás két, talán legfontosabb sajátága a következő: Egyrészt, alapvetően azokra az esetekre fókuszál, amikor az értékmódosulásnak van a módosított elem környezetében valamilyen detektálható okozója. Ennek értelmében például a (4) megnyilatkozások közül a (4a-c) alattiak azok, amelyeknek a kezelését célozza. Másrészt, a kombinációs szabályokat alkalmazó megoldás szótáralapú, és ebből kifolyólag – akárcsak a pusztán szótáralapú megoldás – a lexémák inherens értékelő tartalmával dolgozik.

Érdemes kiemelni, hogy a szentimentelemzést a szakirodalom a nyelvtechnológiai kutatások között alapvetően szemantikai vonatkozású problémaként tartja számon. Csupán elvétve találni olyan, a témában született dolgozatot, amely egyáltalán említést tenne a feladat pragmatikai vonatkozásairól (vö. pl. Honkela és mtsai 2012).

2.2 A nyelvi értékelés megközelítési módjai az elméleti nyelvészeti irodalomban

A nyelvi értékelés mibenlétének tárgyalásához tekintsük át annak megközelítési módjait a hazai elméleti nyelvészeti irodalomban!

Kiefer (2008: 507–515) a melléknevek jelentése kapcsán tárgyalja a nyelvi értékelést. Annak okán, hogy a szerző vizsgálatának fontos szempontja a melléknevek kontradiktórikus tulajdonságának megléte, illetve hiánya, ez a szempont a szerző osztályozási rendszerét alapjaiban határozza meg. Megkülönbözteti az abszolút és a relatív mellékneveket, amelyek közül az utóbbiak jelentése „az összehasonlítás alapjától függ” (Kiefer 2008: 511). Így például, a *nagy elefánt* és a *nagy egér* esetében nem beszélhetünk abszolút, csupán a jellemzett entitások csoportjához viszonyított tulajdonságról.

Kiefer (2008: 511) alapján a mértéket jelentő (ún. dimenzionális, *kicsi–nagy*) és az értékelő (ún. evaluatív, pl. *rossz–jó*) melléknevek a relatív melléknevek két legfontosabb alosztályát képezik. Kiefer (2008: 507–508) ez utóbbiakat illetően is hangsúlyozza, hogy némely esetben „az alaptag előre nem jelezhető módon határozza meg a melléknév

jelentését” (Kiefer 2008: 507). A *jó étel* és a *jó fűrés* példák esetében „a *jó* melléknév jelentése meghatározó mértékben az alaptag jelentésétől függ” (Kiefer 2008: 508).

Más – a nyelvi értékelés elemzése szempontjából ugyancsak releváns – melléknevek, például Kiefer (2008: 510) rendszerében „a tulajdonság magas fokát jelentő melléknevek (pl. *gyönyörű, kiváló*)”, „a valamely tulajdonság hiányát jelölő, ún. privatív melléknevek (pl. *süket, sánta*)”, „a valamely emberi tulajdonságot jelölő” melléknevek (pl. *féltékeny, fukar*) az abszolút melléknevek között kapnak helyet.

Azt látjuk tehát, hogy a kieferi megközelítés (2008) az értékelő tulajdonságot a melléknevek egy szűk csoportjához köti, és nem a nyelvi értékelés leírására és rendszerbe foglalására törekszik.

Laczkó (2008: 326) „minősítés”-nek nevezi azt az eljárásmodot, amelyben egy élő személyt, egy élettelen tárgyat, egy eseményt vagy egy folyamatot minősége, jellege alapján valamilyennek nyilvánítunk. A szerző szerint a magyar nyelvben a minősítés kifejezésére a leggyakrabban a minősítő és fajtajelölő mellékneveket használjuk, ugyanakkor az ezekben kifejezett szemantikai tartalom „szerkezetekkel vagy az irodalmi nyelvben gyakori formával, körülírással is” átadható. A „szerkezeteket” illetően Laczkó (2008: 326) megemlíti, hogy bizonyos esetekben a kifejezés tagjait „logikai, pragmatikai jelentés vagy „csupán” a nyelvszokás tartja össze”. Kiemeli továbbá, hogy a minősítés nyelvi kifejezése alá- és mellérendelő szintagmákkal egyaránt lehetséges, ily módon minősíthetünk alany–állítmány-viszonnyal (*Péter tanuló*), tárgyas szintagmával (*Péter szeret olvasni*), határozós szintagmával (*uszodába jár*), minősítő jelzős szerkezettel (*magas diák*), azonosító értelmező szerkezettel (*Péter, a barátom*), valamint a mellérendelő szintagmák különböző altípusaival (*érdekes, sőt izgalmas; lassan, megfontoltan dolgozott*). Mindemellett említést tesz a minősítés kifejezésére alkalmas szerkezetek sajátos típusáról, az állandósult szókapcsolatokról is (*letargiába esik; szegény, mint a templom egere*).

Szilágyi (1996: 11) „értékjelentés”-nek nevezi a szójelentésnek azt az összetevőjét, amelyben „kifejeződik a beszélőközösségnek a jel tárgyához való viszonya”. Megállapítja, hogy az értékjelentés pozitív és negatív attitűdöt egyaránt tükrözhet, illetve léteznek az értékjelentés szempontjából semleges kifejezések is (például *nátrium-klorid, fonéma* stb.). A szerző (1996: 11–12) szerint értékjelentésről csupán akkor beszélhetünk, ha az a szótári jelentés részét képezi; tehát „az értékjelentés vizsgálata [...] a szójelentésre irányul”.

Berényi (1988: 75) Szilágyi (1996)-hoz hasonlóan ugyancsak szemantikai komponensként tételezi a nyelvi értékelést, amikor a főnevek és a melléknevek sajátóságaként említi az esetlegesen megjelenő, úgynevezett „értékelő jegy”-et. Véleménye

szerint a főnév „önmagában lebecsülésre, vagy éppen ellenkezőleg, felértékelésre utaló szemantikai jeggyel rendelkezik”. Mindennek fényében ellentmondásosnak tűnhet több megállapítás is, amelyeket Berényi (1988) dolgozatában más helyütt tesz. Egyrészt arról beszél, hogy a denotatív lexikai elemeket a hozzájuk társuló melléknév affektív jegye is felruházhatja „a beszélői becslésre utaló szubjektív értékkel” (Berényi 1988: 77). Másrészt felhívja a figyelmet arra a jelenségre is, amikor a jelzős összetétel elemei közül önmagában egyik sem telített a szubjektív megítélés jegyével, de az összetétel már rendelkezik ezzel a sajátsággal (vö. Berényi 1988: 77). Ez utóbbi szemantikai változás véleménye szerint mind a kapcsolódás pusztán tényéből, mind pedig a tágabb szövegösszefüggés értelmezésre gyakorolt hatásából fakadhat. Berényi (1988: 77) a jelenségre *a külgazdasági politika egyéni arcú* és *a gazdasági tömörülés vezető ereje* kifejezéseket hozza példaként.

Fentebbi észrevételein túl Berényi (1988) a nyelvi modalitás jelenségét is tárgyalja az értékelés kapcsán. Az igéket illetően megállapítja, hogy a propozicionális tartalomra vonatkozó beszélői becslés „a bizonyosságtól a bizonytalanságig, a fenntartásig vagy az elutasításig terjed”, és mindezek kifejezése történhet igemódokkal és módbeli segédigékkel egyaránt (Berényi 1988: 78). Emellett olyan úgynevezett „modalizáló elemeket” is feltételez, amelyek mondathatározóként explicit vagy implicit módon utalnak a beszélői állásfoglalásra az állított tartalommal kapcsolatban; például *természetesen*, *kétségtelen*, *köztudott* stb. (vö. Berényi 1988: 76, 77–78).

Berényihez hasonlóan Péter Mihály (2008: 2) is a modalitás fogalmával hozza összefüggésbe a nyelvi értékelést, amikor azt a modalitás egyik fajtájaként határozza meg, s szemantikai kategóriának tekinti. Ugyanakkor megjegyzi, hogy „a nyilatkozat modális kereteként a pragmatika szférájába tartozik” (Péter 2008: 2). Berényi (1988)-cal és Szilágyi (1996)-tal megegyezően ő is amellett érvel, hogy az értékelés szerves részét képezi a nyelvi elemek jelentésszerkezetének (vö. Péter 1991b: 47; 2008: 2). Ami a nyelvi értékelés típusait illeti, Péter (1991b, 2008) két különböző szempont alapján a következő kategorizálási megoldásokat tartja lehetségesnek: egyrészt megkülönböztethetjük az értékelés „általános” és „speciális” típusát (vö. Péter 2008: 2), s az elsőre a *jó* és a *rossz* mellékneveket, a másodikra a *szép*, a *hasznos*, a *rút* és a *haszontalan* fogalmakat hozza példaként. Másrészt megkülönböztethetjük az „emocionális” és a „racionális” értékelés típusát (vö. Péter 1991b: 46). A szerző az előbbire *a főnököm remek ember*, az utóbbira *a habbeton rossz hővezető* mondatokat hozza példaként. Az emocionális típushoz kapcsolódik a szerző 2008-as dolgozatában tett megállapítása, mely szerint az értékelés rendszerint társadalmilag és történelmileg meghatározott (például egy jó autó ismérvei

napjainkban nem azonosak a múlt század harmincas éveinek arra vonatkozó ideáival), s „a társadalmi érvényű értékelések mindenkor bizonyos normákra, a valóság idealizált modelljének megfelelő elképzelésekre, követelményekre vonatkoznak” (Péter 2008: 2). Ugyanakkor megjegyzi, hogy az értékelés lehet nem e társadalmi norma által meghatározott, hanem egyéni is.

A fentebb tárgyalt szerzők tehát Laczkó (2008) és Kiefer (2008) kivételével rendre a modalitás fogalmához kapcsolják a nyelvi értékelést. Vegyük hát górcső alá a nyelvi modalitás szakirodalmának a nyelvi értékelés jelenségéhez kapcsolható tételeit!

Kiefer (1986: 24–27) a nyelvi modalitást illetően a *faktív* és a *nem faktív attitűd* fogalmaival operál. Közülük a faktív attitűdöt kifejező szerkezetek a szerző rendszerében olyan faktív attitudinális operátorokat tartalmaznak, amelyek „a beszélőnek az adott tényálláshoz való érzelmi-értelmi viszonyulását” fejezik ki (Kiefer 1986: 25), például:

- (5) a. Pisti **sajnos** beteg.
- b. Pisti **szerencsére** már egészséges.

Murvai (2001: 59) és Pete (2002: 189–190) megkülönböztet egy úgynevezett *értékelő modalitás*-típust, azonban e fogalmat a két szerző egymástól teljesen eltérően értelmezi. Murvai (2001: 59) az értékelő modalitást – a dolgozatában közölt példák alapján (*jó–rossz, kíváncsi–sajnálatos*) – a kieferi (1986: 24–27) faktív attitűdöt kifejező attitudinális operátorokhoz hasonlóan képzei el, lévén, hogy ezek az elemek is a beszélőnek a proposíciós tartalomhoz való érzelmi-értelmi viszonyulását fejezik ki. Pete (2002: 189–190) véleménye szerint ugyanakkor az értékelő modalitás „a beszélő ismeretei bizonyosságának a fokát [...] jelöli” (Pete 2002: 189). Pete (2002: 189) rendszerében tehát az értékelő modalitás a kieferi nem faktív attitűd fogalmával rokonítható, ahol a szerkezet olyan attitudinális operátort tartalmaz, amely a beszélői bizonyosság fokára utal, például:

- (6) Pisti **valószínűleg** beteg.

Pete (2002: 189–190) alapján tehát például a következő, (7) alatti megnyilatkozások „nem tartoznak az értékelő modalitás körébe” (Pete 2002: 189):

- (7) a. Éva **rendkívül** okos lány.
- b. Éva **nem** buta lány.

Összegzőképpen a következőket mondhatjuk el: az áttekintett, nyelvi értékeléssel kapcsolatos irodalmi tételek megállapításai alapvetően három sajátosság köré csoportosulnak. Egyrészt, az értékelést szemantikai komponensnek, a szótári jelentés részének tekintik. Másrészt, a legtöbb szerző a nyelvi modalitáshoz tartozó fogalomként képzelel el. Harmadrészt, az értékelés kapcsán rendre a beszélői szubjektivitással hozzák összefüggésbe a nyelvi értékelést, s egyetlen kivételtől (Péter 1991b) eltekintve nem szólnak az objektív nyelvi értékelés lehetőségéről.

2.3 A bemutatott megközelítési módok értékelése

A jelen fejezetben az értékelés nyelvi viselkedése alapján górcső alá veszem, hogy a (2.2) részben bemutatott megközelítési módok megfelelő elméleti alapot kínálnak-e a nyelvi értékelés elemzéséhez. Amellett fogok érvelni, hogy az elméleti nyelvészeti irodalom megállapításai nem kielégítőek sem az elméleti oldal, sem a nyelvtechnológia számára. A nyelvi értékelés néhány sajátosságának körüljárása alapján továbbá meg kívánom mutatni, hogy a nyelvi értékelés jelensége – legalább részben – túlmutat a lexikai és a kompozíciós szemantika szféráján, és a pragmatikai sajátosságok jelentősebb figyelembevételével nélkülözhetetlen egy pontosabb eredményre képes elemző rendszer létrehozásához.

2.3.1 A nyelvi értékelés mint modális kategória

Amint azt a megelőző fejezetben megállapítottam (l. 2.2), a legtöbb szerző a nyelvi értékelést szemantikai kategóriának tekinti, és a modalitás fogalmával hozza kapcsolatba (vö. Kiefer 1986, Berényi 1988, Murvai 2001, Pete 2002, Péter 2008). Ugyanakkor, a modalitás elméleti keretében az értékelés fogalma több megközelítés szerint (vö. Kiefer 1986, Berényi 1988, Pete 2002) a faktivitás fogalmával fonódik össze, tehát az értékelést a beszélő a kifejezett tartalom bizonyossági fokáról fogalmazza meg. Nem magát a tartalmat értékeli tehát annak minősége szempontjából, hanem azt, hogy mennyire biztos annak tényszerű voltában. Emellett a szerzők kiemelik, hogy értékelő modalitás esetében a mondatban valamilyen operátornak (pl. attitudinális) is meg kell jelennie (vö. Kiefer 1986,

Berényi 1988, Murvai 2001, Pete 2002). Mindez azonban nem szerencsés a számítógépes értékeléselemzés szempontjából. Lássuk a problémát részleteiben is!

Először is, az értékeléselemzés nem korlátozódhat azokra a megnyilatkozásokra, amelyek faktív attitudinális operátort (vö. Kiefer 1986, Murvai 2001, Pete 2002) vagy egyéb „modalizáló elemet” (vö. Berényi 1988) tartalmaznak, hiszen a feladat szempontjából érdemi információt hordozhatnak az ilyen operátort nem szerepeltető adatok is, például:

- (8) a. Pisti beteg.
- b. Pisti már egészséges.

Emellett a szentimentelemzés során elsődlegesen nem a beszélői bizonyosság fokát, hanem a kifejeződő értékítéletet keressük a szövegben. Kiefer (1986), Berényi (1988) és Pete (2002) megközelítési módja alapján a (8) alatt közölt példák e szempontból sem tartoznak az értékelés kategóriájába.

A modalitás elméleti kerete ugyanakkor más szempontból is problematikus az értékeléselemzés számára. Az indirekt formájú értékelést például ez a rendszer nem teszi kezelhetővé. Kiefer (1986: 32) alapján a kérdés egy olyan beszédaktus, amely „nem fejez ki kognitív vagy emotív viszonyulást” a benne megfogalmazott tényállással szemben, a kérdő formájú adatokat tehát ki kellene zárni a vizsgálat köréből. Ugyanakkor tudvalevő, hogy a kérdő mondatfajta számos funkciója között informáló szerepet is betölthet (vö. Perrault–Allen 1980: 170). Nyelvi produktumaink számtalan olyan kérdő formájú megnyilatkozást tartalmaznak, amelyeknek nem információkérési, hanem indirekt formájú információközlő szándékuk van, például:

- (9) a. Hallottad, milyen rossz minőségű fék van rajta?
- b. El tudsz képzelni ennél borzalmasabb telefont?

A fentebb tárgyaltak alapján tehát a következőket állapíthatjuk meg: Mindenekelőtt, a modalitáselméletben a nyelvi értékelés fogalma a különböző attitudinális operátorokkal, valamint a beszélői bizonyosság fokával kapcsolódik össze, ami a jelen feladathoz nem ad megfelelő elméleti háttérrel. Egyrészt, amint fentebb láttuk, az értékeléselemzés szempontjából érdemi információt hordozhatnak az attitudinális operátort nem szerepeltető nyelvi adatok is. Másrészt, a nyelvi értékelés és a beszélői bizonyosság két önálló

dimenzióként rajzolódik ki, és nem szerencsés ezeket a fogalmakat egy problémaként kezelni. Érdemes felhívni a figyelmet arra, hogy a beszélői bizonyosság fontos sajátja az automatikus értékeléselemzés számára is. A bizonytalanságot jelölő kifejezések automatikus azonosítása napjaink nyelvtechnológiai kutatásainak egyik fontos problémaköre (vö. Vincze és mtsai 2014a, 2014b, Szabó–Vincze 2015: 221–222, Drávucz–Szabó 2017). A szövegek automatikus tartalmi elemzése során ugyanis csupán akkor tudunk megfelelő elemzési eredményt produkálni, ha az elemző rendszerünk képes a tényszerűen közölt és a bizonytalan nyelvi tartalmakat elkülöníteni egymástól. A szentimentelemzés feladatában ez azt jelenti, hogy elsősorban azokra az információkra tudunk teljes értékű adatokként támaszkodni, amelyeket a megnyilatkozó tényként közöl, tényként fogad el. Az automatikus értékeléselemzésben tehát egy fontos részfeladatként jelentkezik a faktív és nem faktív olvasatú adatok elkülönítése egymástól, és ezzel összefüggésben hiba lenne a modalitás elméleti irodalma alapján az értékeléselemzést a nyelvi bizonytalanság kezelésének a feladatával azonosítani.

Az elmondottakon túl azt is megállapítottam, hogy a modalitás elméleti kerete szűk a nyelvi értékeléselemzés számára, hiszen nem teszi lehetővé az indirekt beszédaktusokban kifejezett értékelő nyelvi tartalmak kezelését sem.

2.3.2 A nyelvi értékelés mint szemantikai komponens

Amint azt a nyelvi értékelés megközelítési módjait áttekintő részben láttuk (l. fentebb, 2.2), Berényi (1988), Péter (1991b) és Szilágyi (1996) egyaránt a szó szemantikai komponenseként tételezi a nyelvi értékelést. Szilágyi (1996: 11–12) az értékjelentést a szótári jelentés részeként kezeli, ennek értelmében az értékjelentés vizsgálatát a szó szemantikai sajátságának vizsgálataként képzei el. Hasonlóan, Péter (1991b: 47, 2008: 2) is a nyelvi elemek jelentésszerkezetének szerves részeként tekint az értékelésre, Berényi (1988: 75) pedig egy úgynevezett „értékelő jegy”-ről beszél, amely a főnevek és a melléknevek sajátságaként hordozza a lebecsülés vagy felértékelés szemantikai tartalmát.

Az áttekintett irodalmak közül csupán Laczkó (2008) és Kiefer (2006) utal az értékelés szempontjából a kontextus szerepére. Laczkó (2008: 216) említést tesz a pragmatikai jelentés értékelésbeli szerepéről. Kiefer (2006: 507–508) az értékelő mellékneveket a relatív melléknevek osztályába sorolja, lévén, hogy megállapítása szerint jelentésük gyakorta az általuk módosított elem jelentésétől függ.

Annak céljából, hogy megállapítsam, vajon tekinthető-e az értékelés a szó jelentéskomponensének, a következő sajátságokat, illetve kérdéseket vizsgálom meg részletesebben: a domén- és targetfüggés problémáját, a kontextus szerepének a kérdését, valamint a kombinációs szabályok alkalmazhatóságának a lehetőségét.⁵ Az értékvesztés és az értékváltás jelenségeit, amelyek ugyancsak fontosak e probléma szempontjából, a dolgozat önálló fejezetében tárgyalom részletesen (l. 4).

2.3.2.1 A domén, a target, illetve a kontextus szerepe a nyelvi értékelésben

Amint arról a dolgozatban már több helyütt is szóltam (l. 1.1 és 2.1), számos kifejezés egy adott szövegdoménben, illetve egy adott targetre vonatkozóan eltérő értékelő szemantikai tartalommal rendelkezik valamely más doménbeli vagy targetfüggő sajátságához képest (vö. Ahn és mtsai 2012, Vázquez–Bel 2012: 3557, Vechtomova és mtsai 2014: 553–554). Például, amennyiben egy telefontöltő kicsi, azt általában pozitív tulajdonságnak tekintjük, míg ugyanez a sajátság egy számítógép-memória vonatkozásában alapvetően negatív. Emellett az is lehetséges, hogy egy adott kifejezés az aktuális kontextusban nyer valamilyen értékelő tartalmat, és egy másik kontextusban esetlegesen nem ugyanazt az értékelő tartalmat hordozza vagy semmilyen értékelő tartalmat nem hordoz.

Amint a (2.2)-ben láttuk, Kiefer a mellékneveknek csupán egy szűk csoportját vonja az ún. „értékelő melléknevek” körébe. Ugyanakkor a szerző más csoportjaiban helyet kapó melléknevek, például a „mértéket jelölő (*kicsi–nagy*)”, „a tulajdonság magas fokát jelentő melléknevek (pl. *gyönyörű, kiváló*)”, „a valamely tulajdonság hiányát jelölő, ún. privatív melléknevek (pl. *süket, sánta*)”, „a valamely emberi tulajdonságot jelölő” melléknevek (pl. *féltékeny, fukar*) éppúgy releváns információk lehetnek a nyelvi értékelés elemzése során, hiszen mindezek a melléknevek – akár az adott kontextusban, domén vagy target függvényében – jelölhetnek értékelést. A kieferi rendszer (2008), mivel nem a nyelvi értékelés leírására és rendszerbe foglalására törekszik, hanem más szempont szerint szerveződik, nem is tudja a nyelvi értékelés elemzésének feladatát az elmélet oldaláról hatékonyan támogatni.

A jelen fejezetben a domén- és targetfüggés jelenségét, valamint a kontextus szerepét teszem alaposabb vizsgálat tárgyává, abból a szempontból, hogy a jelenleg a

⁵ Az ironia jelenségével e disszertáció keretei között nem foglalkozom.

szenzimentelemzés gyakorlatában leggyakrabban alkalmazott megoldások mennyire tekinthetők helytállónak, és – ezzel összefüggésben – alkalmazhatónak.

A jelenségek okán Ahn és munkatársai (2012: 5–6) két „polaritástípus”-t tételeznek fel, s rendszerüket a koreai melléknévek példáján keresztül a következőképpen mutatják be: Amennyiben egy melléknév úgynevezett *abszolút polaritást* (*Absolute Polarity*) hordoz, úgy értékelő tartalmának minősége (pozitív vagy negatív volta) nem függ az adott doméntól, amelyben szerepel, illetve targettől, amelyet minősít. Amennyiben azonban a melléknév *relatív polaritással* (*Relative Polarity*) rendelkezik, úgy értékelő tartalmának milyensége az aktuális doméntól, illetve targettől.⁶ Tekintsük az alábbi példákat a szerzők rendszeréből!

- (10) Abszolút pozitív polaritás: *cohta* 'jó'; *chinhata* 'barátságos'
Abszolút negatív polaritás: *telepta* 'piszkos'; *sikkulepta* 'zajos'
Relatív polaritás: *mukepta* 'nehéz'; *kanghata* 'erős'

A szerzők (Ahn és mtsai 2012: 6) összesen 2135 melléknévi predikátumot megvizsgálva arra a megállapításra jutnak, hogy azok 58,6%-a rendelkezik relatív, s csupán 41,4%-a abszolút polaritással.

Magam amellett érvelek, hogy az abszolút és a relatív polaritás feltételezése nem helytálló megoldás a doménfüggés jelenségének a kezelésére. Egy elem lexikai szinten vagy hordoz értékelő szemantikai tartalmat, és ezzel összefüggésben ez az értékelő tartalom eleve, a szótár szintjén meghatározottan pozitív vagy negatív, vagy nem hordoz ilyen tartalmat ezen a szinten. Ez természetesen nem jelenti azt, hogy a kontextusban nem nyerhetne értékelő tartalmat egy a szótár szintjén ilyen tartalommal nem rendelkező elem is. Elfogadom tehát, hogy bizonyos lexémák (l. pl. *undorító*, *selejtes*, *alantas* stb.) inherens értékjelentéssel rendelkeznek (vö. Kiefer 2008: 508). Relatív értékjelentést azonban nem tételezek a lexika szintjén kódolt információként, hiszen megközelítésem szerint a lexikai szintű értékelő tartalom esetében a pozitív vagy a negatív sajátságot is kódolni kell, nem csupán a tartalom meglétét.

⁶ A bemutatott kategóriák nem azonosíthatóak Péter (1991: 46) korábban tárgyalt *emocionális* és *racionális értékelés* fogalmaival (l. 2.2). Amíg ugyanis a szerző rendszere az érvényességi kör alapján osztályozza a nyelvi értékelést, vagyis aszerint, hogy azok egyéni megállapításon vagy általános igazságon alapulnak-e, addig az utóbbi felosztás a „puszta” target–szentiment-viszonnnyal operál. Ezzel összefüggésben tehát Péter (1991: 46) mindkét korábban idézett példája (l. 2.2) targettől függetlenül rendre ugyanazzal a szentimentértékkel rendelkező elemet (*remek* és *rossz*) tartalmaz.

Függetlenül attól, hogy a fentebbi oppozíciót nem fogadjuk el, azok az eredmények, miszerint Ahn és szerzőtársai (2012: 6) alapján a vizsgált melléknévi predikátumoknak mindössze 41,4%-a mutatott inherens értékjelentést, ugyancsak arra mutat, hogy az értékjelentést nem határozhatja meg pusztán a lexikai szemantika szintje, és ezzel összefüggésben annak vizsgálata nem merülhet ki a pusztán lexikai jelentés vizsgálatában.

A relatív inherens értékelő szemantikai tartalom feltételezésénél nem kevésbé problematikusak azok a megközelítési módok, amelyekben a szerzők (vö. pl. Moghaddam–Popowich 2010) nem csupán a pozitív és negatív polaritásúnak ítélt kifejezéseket rendezik lexikonba, hanem bizonyos, a polaritás szempontjából semlegesnek tekintett kifejezéseket is szótárba foglalnak. Ennek megfelelően például Moghaddam és Popowich (2010: 3) a (11a) alatti példákat pozitív, a (11b) alattiakat negatív szemantikai tartalmúaknak ítéli, és velük szemben a (11c) alatt közöltek a szentimentérték szempontjából semleges elemeknek tartja (a szerzők dolgozatában bemutatott listából most csupán az első 5-5 elemet közlöm):

- (11) a. *good* 'jó', *nice* 'jó, kedves', *awesome* 'tök jó', *excellent* 'kiváló', *great* 'nagyszerű'
b. *bad* 'rossz', *awful* 'borzasztó', *defective* 'hibás, tökéletlen', *faulty* 'hibás', *poor* 'szegény'
c. *mediocre* 'közepes, középserű', *average* 'átlagos', *enough* 'elég, elégséges', *fair* 'korrekt, méltányos', *okay* 'rendben'

Úgy vélem ugyanakkor, hogy mindazok a nyelvi elemek, amelyek valamilyen sajátságot jelölnek, kivétel nélkül képesek valamilyen értékelés kifejezésére. A (11c) alatti példák semlegessége így szótári szinten nem tételezhető.

A kontextus szerepe kapcsán érdemes említést tenni az úgynevezett polaritási skáláról. A szentimentelemzés nemzetközi gyakorlatában (vö. pl. Ruppenhofer és mtsai 2014) nem ritka megoldás, hogy a különböző nyelvi elemeket elhelyezik egy matematikai értelemben vett diszkrét skálán az alapján, hogy a szerzők megítélése szerint az adott elemek egymás viszonylatában milyen fokú pozitív vagy negatív értékelést fejeznek ki (vö. Ruppenhofer és mtsai 2014: 117). Az így létrehozott skála tehát arra ad lehetőséget, hogy a

pozitív és a negatív értéktartományon belül további konszekutív értékeket tegyenek fel az elemzők.⁷

Annak céljából, hogy rámutassunk a megoldás kérdéses voltára, tekintsünk két elemet Ruppenhofer és munkatársai (2014: 117) kérdőíven alapuló skaláris rendszeréből! A szerzők értékskáláján a *second-rate* ('másodosztályú') kifejezéshez negatívabb értékelő tartalmat rendelnek a szerzők a *so-so* ('tűrhetően') kifejezés értékelő tartalmánál. Belátható azonban, hogy a fentebbi elemek értékelő tartalma, illetve annak foka erősen kontextusfüggő, nem helytálló tehát minden előfordulásukhoz ugyanazokat a szentimentértékeket rendelni az értékhierarchiára támaszkodva.

A jelen alfejezetben tett megállapításaim elsősorban a szentimentszótár összeállításakor hasznosítottam (l. 7.2).

2.3.2.2 A kompozíciós szemantika lehetőségei a szentimentelemzésben

Amint azt a (2.1)-ben tárgyaltam, a legtöbb szentimentelemzést célzó munka a szentimentszótárak alkalmazása mellett a szemantikai kombinációs szabályszerűségekre fókuszál, és azokat igyekszik a szótárak mellett alkalmazni (pl. Feldman és mtsai 2011: 1643, Ruppenhofer–Rehbein 2012). Ezek a dolgozatok olyan szabályalapú szentimentelemző rendszer megalkotását célozzák, amely képes az értékmódosulás jelenségét hatékonyan kezelni (vö. Feldman és mtsai 2010: 3, Liu 2012: 60).

A (2.1)-ben azt láttuk, hogy amíg bizonyos megnyilatkozások esetében van olyan nyelvi elem, amely azonosítható az értékmódosulási folyamatban (pl. egy tagadó operátor), addig más esetekben ilyen elemet nem, vagy nem feltétlenül tudunk azonosítani.

Az angol nyelvű nyelvtechnológiai irodalom *szentimentshiftereknek* (*sentiment shifters*) nevezi azokat a nyelvi elemeket, amelyek értékbeli eltolódást okoznak a teljes megnyilatkozás szentimentértékében a benne szereplő szentimentkifejezések inherens értékelő tartalmának pusztá összegéhez képest (vö. Ding és mtsai 2008, Liu 2012: 60).

A kombinációs szabályok alkalmazásával ezeknek a shiftereknek a „hatását” igyekeznek megfelelően kezelni (vö. Feldman és mtsai 2011: 1643). Az eljárás mód alapvető dilemmája az, hogy az értékelést kifejező megnyilatkozások gyakorta nem teljesen kompozicionálisak. Ahogyan arra több szerző (vö. Israel 2004: 711, Feldman és

⁷ A megoldás alapjának tekinthető többek között Osgood és mtsai (1957).

mtsai 2010: 4) is rámutat, a szentimentérték nem kalkulálható ki minden esetben maradéktalanul szemantikai kompozíciós szabályok segítségével. Tekintsük az alábbi példákat!

- (12) a. Mari jó tanuló.
b. Mari **nem** jó tanuló. → Mari rossz tanuló

- (13) a. Mari buta.
b. Mari **nem** buta. → Mari okos

A (13b) alatti két példa között nem lehet kontradiktórikus viszony, hiszen az egyik tagadásából nem következik a másik állítása (vö. Mogyoródi 2012:17). Israel (2004: 711) alapján, amíg a pozitív tartalmú elem tagadása esetén negatív értékeléssel kell számolnunk, addig ugyanez a művelet a negatív érték esetében nem feltétlenül vezet hasonló eredményre. Hasonlóan érvel Kiefer (2008: 513), aki szerint a *széles–keskeny* és a *szép–csúnyatípusú* melléknevek esetében „az *x széles* állításból ugyan következik, hogy *x nem keskeny*, de az *x nem keskeny* állításból nem következik, hogy *x széles*”.⁸

A kombinációs szabályokat alkalmazó megoldás a nyelvi jel egyfajta kódmodell-szerű felfogását tükrözi. A humán kommunikáció e pragmatikai modellje a kommunikációt úgy tekinti, mint „explicit módon közölt üzenetek kódolása és dekódolása” (Németh T. 2011: 45). Eszerint konvencionális kapcsolat áll fenn a jelzésként felfogott szavak és az üzenetként értékelt szójelentések között, azaz a jelentések bizonyos értelemben „áttetszenek a mondatokon” (Reboul–Moeschler 2000: 23). A kombinációs szabályokat alkalmazó megoldás tehát inherensen korlátozott, hiszen a részekből, kombinációk útján igyekszik kiszámítani az egész kifejezés jelentését.

Az elméleti problémák megfontolása mellett ugyanakkor nélkülözhetetlen meggondolás tárgyává tenni, hogy mi az, ami a jelenlegi nyelvtechnológiai eszközökkel, az automatikus értékeléselemzés számára megvalósítható, és – nem mellékesen – költséghatékony.

A disszertációban a későbbiekben bemutatok egy korpuszvizsgálatot (l. 6.3.1.7), amelynek egyik fő célja, hogy megismerjem a kombinációs szabályok alkalmazási

⁸ A sajátságot vessük össze a kontrárius viszonnyal, amelyben a két állítás nem lehet egyidejűleg igaz, de ha az egyik igaz, a másik szükségszerűen hamis. Ugyanakkor az is lehetséges, hogy egyidejűleg hamisak (vö. Brugger 2005: 114).

lehetőségeit. Ezt követően, az implementációt tárgyaló részben (l. 7), a létrehozott adatbázisokon végzett elemzések eredményeire alapozva azt is megvizsgálom, bizonyos kombinációs szabályok alkalmazása mennyiben hozhat javulást az automatikus értékeléselemzés eredményességében, az egyszerű szótáralapú feldolgozási módhoz képest.

2.3.3 A nyelvi értékelés mint a beszélői szubjektivitás kifejeződése

Amint ismertettem (l. 2.2), a hazai elméleti nyelvészeti irodalomban, az értékelés kapcsán a szerzők rendre a beszélői szubjektivitással hozzák összefüggésbe a nyelvi értékelést (vö. Kiefer 1986, Berényi 1988, Szilágyi 1996), s egyetlen kivételtől (Péter 1991b) eltekintve nem szólnak az objektív nyelvi értékelés lehetőségéről. A nyelvi értékelés elemzése ugyanakkor nem korlátozódhat kizárólagosan a Péter (1991b: 46) által „emocionális”-nak nevezett, azaz a beszélő szubjektív ítéletét megfogalmazó értékeléstípus vizsgálatára; a „racionális” értékelés éppúgy releváns információ (vö. Péter 1991b: 46).

A probléma kapcsán érdemes a következő sajátságokat is megemlítenünk: A nemzetközi gyakorlatban a szentimentelemzés feladatát gyakran a szubjektivitáselemzés fogalmával jelölik, e két terminust tulajdonképpen egymás ekvivalenseként használják (vö. pl. Wilson és mtsai 2005, Liu 2010: 627). Ahogyan Liu (2010: 627) érvel, a szöveges információk két alapvető kategóriára oszthatók: ténytípusokra és véleményekre. Ez utóbbiakat nevezi a szerző szubjektív kifejezéseknek, amelyek a megnyilatkozó értékelését vagy érzését tartalmazzák valamely entitás, esemény, vagy azok valamely sajátsága vonatkozásában. Mindezek alapján a szerző nem tekinti a szentimentelemzés feladatához tartozónak a ténytípus közlések kivonatolását. Liuhoz (2010) hasonlóan, Wilson és munkatársainak (2005) szentimentelemző rendszere is a szubjektív mondatok azonosítását célozza, és dolgozatukban a feladatot szubjektivitáselemzésként említik.

Magam a szentimentelemzésben nem választom szét e két információtípust (vö. Szabó 2016, Drávucz–Szabó 2017). A szentimentelemzés eredményének felhasználása szempontjából ugyanis nincs jelentősége annak, hogy az adott értékelés szubjektív megítélés eredménye-e. Az automatikus értékeléselemzés során mindkét információtípus kinyerésére szükség van (vö. Szabó 2016: 160).

Amint arról korábban már említést tettem (l. 2.3.1), a valódi jelentőséggel bíró sajátságot, és egyben problémát a szentimentelemzésben az 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 2014: 99). A szövegek automatikus tartalmi elemzése során elsősorban azokat az értékelő tartalmakat kívánjuk kinyerni, amelyeket a megnyilatkozó tényként közöl.

Az elmondottakat illetően fontos tehát hangsúlyozni, hogy a beszélői szubjektivitás, valamint a beszélői bizonytalanság fogalmai – bár részben összefüggenek – nem azonosíthatóak egymással.

A bizonytalanság nyelvi megnyilvánulási formái a Dívány-korpusz feldolgozásakor külön annotációt kaptak abból a célból, hogy a korpusz az e jelenséget érintő kutatási és fejlesztési feladatokat is támogatni tudja (l. 6.3.1.4).

2.4 A tárgyaltak összegzése

A jelen fejezetben az értékeléselemzés legalapvetőbbnek tekinthető dilemmáját, a nyelvi értékelő jelentés mibenlétét vizsgáltam részletesen.

Annak céljából, hogy a problémát behatóbb vizsgálat tárgyává tehessem, röviden összefoglaltam a nyelvi értékelés alapvető megközelítési módjait az automatikus értékeléselemzés feladatkörében. Ezt követően áttekintettem azokat a megközelítési módokat, amelyekkel a nyelvi értékelés kapcsán az elméleti irodalomban találkozhatunk, majd görcső alá vettem a bemutatott megközelítési módok helytállóságát.

Az irodalom áttekintése alapján megállapítottam, hogy a legtöbb szerző a szó szemantikai komponenseként tételezi a nyelvi értékelést és a modalitás fogalmához kapcsolja. Emellett a szerzők a beszélői szubjektivitással hozzák összefüggésbe a jelenséget, s egyetlen kivételtől (Péter 1991b) eltekintve nem szólnak az objektív nyelvi értékelés lehetőségéről.

Annak céljából, hogy megállapítsam, vajon megfelelő elméleti keretet nyújthatnak-e a bemutatott megközelítési módok az automatikus értékeléselemzés számára, a kezelési megoldásokat több kapcsolódó jelenség, valamint a szentimentelemzésben alkalmazott megoldás szempontjából vizsgáltam meg.

A vizsgálat során a következő legfontosabb megállapításokat tettem: Egyrészt, a modalitás elméleti kerete az értékeléshez úgy közelít, mint a kifejezett tartalom bizonyossági fokáról tett beszélői ítélethez. Emellett különböző operátorok meglétét is

feltételezi az értékelő megnyilatkozásban. Mindez azonban nem szerencsés a számítógépes értékeléselemzés szempontjából. A modalitás elméleti kerete ugyanakkor más szempontból is problematikus az értékeléselemzés számára. Az indirekt módon megfogalmazott értékelést például ez a rendszer nem teszi kezelhetővé. Másrészt, több olyan nyelvi sajátásra is rámutattam, amely alapján a nyelvi értékelést – legalább a nyelvi elemek egy csoportja esetében – bizonyosan nem tekinthetjük szemantikai komponensnek.⁹ Azt is megmutattam, hogy a nyelvtechnológiai irodalom – bár felfigyel ezekre a jelenségekre – nem kezeli azokat kellő körültekintéssel, illetve helytállóan. Harmadrészt, a nyelvi értékelés elemzése nem korlátozódhat kizárólagosan a beszélő szubjektív ítéletét megfogalmazó értékeléstípus vizsgálatára.

A fentebb elmondottak alapján amellet érveltem, hogy az elméleti nyelvészeti irodalom megállapításai nem nyújthatnak megfelelő elméleti alapot a nyelvi értékelés jelenségének sem az elméleti, sem a nyelvtechnológiai szintű kezeléséhez, és a pragmatikai sajátságok jelentősebb figyelembevételére volna szükség egy pontosabb elemző rendszer létrehozásához.

⁹ Ezzel összefüggésben, amint azt a terminológiai kérdések kapcsán ki fogom fejteni (l. 3), lehetségesnek tartom a lexikai szintű értékelő tartalmat kontextustól függetlenül, azaz feltételezem, hogy bizonyos nyelvi elemek értékelő tartalommal rendelkeznek már a szótár szintjén. Ugyanakkor azt is feltételezem, hogy az értékelő tartalom létrejöhet a kontextus függvényében is: egy, a lexika szintjén értékelő tartalmat nem hordozó elemekből álló megnyilatkozás is kifejezhet értékelést, valamint a lexikai szintű értékelő tartalom is megváltozhat az aktuális kontextusban.

3. Terminológiai kérdések

A jelen dolgozatrészben tisztázom mindazokat a fogalmakat, amelyek a disszertáció alapvető terminológiáját képezik. A rendszerezés ismertetése során górcső alá veszem az elméleti és a nyelvtechnológiai irodalom kapcsolódó fogalmait, és részletesen tárgyalom a problémákat és hiányosságokat, alátámasztva ezzel a disszertációban alkalmazott terminológiát.

Szentimentkifejezésnek (*sentiment words*) vagy *nyelvi értékelő kifejezésnek* nevezem a nyelvi értékelés elemzésének azokat a minimális, egy vagy több szóból álló nyelvi egységeit, amelyek önmagukban, lexikai vagy szótári szinten egyetlen meghatározott szentimentértékkel (pozitív vagy negatív) rendelkeznek (pl. *jó, rossz, szép, csúnya*) (vö. Szabó 2014: 7). A lexikai szintű szentimentértéknek a feltételezése teszi lehetővé, hogy az automatikus értékelelemzést a szentimentkifejezések szótári formába rendezett listáival próbáljuk – legalább részben – megoldani (vö. Vázquez–Bel 2012, Liu 2012: 12, Szabó 2014, 2015b, Drávucz és mtsai 2017, Drávucz–Szabó 2017). A szentimentértékkel rendelkező lexémák megnevezésére a nemzetközi nyelvtechnológiai irodalom az angol *sentiment word* terminust használja (vö. pl. Liu 2012). Magam abból a megfontolásból használom a *szentimentkifejezés* terminust, hogy szentimentkifejezésként nem csupán egy-, de többszavas egységek is szolgálhatnak; például *értéktelen ~ egy fillért nem ér* (vö. Szabó 2014: 14).

A nemzetközi számítógépes nyelvészeti terminológia alapján magam is *targetnek* (másképpen: *célpont*) nevezem azokat a nyelvi elemeket, amelyekre az értékelés irányul (vö. Szabó–Vincze 2015). A nyelvi értékelés számtalan típusú targetre irányulhat, például személyre (14a), eseményre (14b), helyre (14c), tárgyra (pl. egy termékre) (14d), de akár egy hosszabban kifejtett tartalomra (14e) is (A példákban a targeteket dőlt szedéssel emelem ki):

- (14) a. Ha a környezetben olyan *ember* van, aki negatív energiát sugároz, vagy számodra egyszerűen nem tűnik szeretetre méltónak (...) [24.hu]
b. Nem tartja magát felelősnek a törtétekért a botrányos *szilveszteri buli*t szervező Off Production Kft. [sci-fi.mandiner.hu]
c. Imádtam *Marokkót, Törökországot* [cotcot.hu/cikk/cimke/stoppolás]

d. Az elromlott *laptop töltő* nem világít, esetleg villog vagy szikrázik [laptophardware.hu/blog]

e. *Főzéssel kezdték, majd a közönséget szórakoztatták történeteikkel, feladványaikkal.*[...] Edina beszédstílusa rém idegesítő, vontatottan beszélt, biztos unta az egészet. [velvet.hu/celeb]

A target szerepét egy entitás (15a), vagy annak valamely aspektusa (15b) is betöltheti, például

(15) a. egy olyan *mobilt* akart a vállalat piacra dobni, ami szép és ellenálló egyaránt. [tech2.hu]

b. a szélek felé ívelő, rendkívül finoman megmunkált *kijelző* [tech2.hu]

Az értekezésben *szentimentfragmentum*nak nevezek minden olyan szövegrészt, amely egyetlen meghatározott szentimentértékű (pozitív vagy negatív) értékelést fejez ki valamely target vonatkozásában (vö. Szabó és mtsai 2016a: 176).¹⁰ Az alábbi példában az elmondottak értelmében egy szentimentfragmentumot találunk (határait szögletes zárójellel jelöltem), és a fragmentum egy szentimentkifejezést és egy targetet tartalmaz (a szentimentkifejezést félkövérrel, a targett dőlttel emeltem ki).¹¹

(16) [_{targ} A szentimentelemzés] [_{szent} [_{int} nagyon] bonyolult] feladat.

A példában a *szentimentelemzés* a target, azaz az értékelés tárgya. Ezt minősíti a beszélő nagyon bonyolult feladatnak. A minősítés értéke ebben az esetben negatív, ugyanis a nyelvtechnológiai doménben, amennyiben egy számítógépes nyelvészeti feladat bonyolultnak bizonyul, azt alapvetően nem tekintjük kívánatosnak. A „bonyolultság” mint sajátság ugyanakkor nem feltétlenül negatív minden doménben (vö. pl. a *mentális lexikon egy bonyolult hálózat*), azonban a számítógépes értékeléselemzésben fontos célunk többek között az is, hogy a szentimentértékeket az adott doménnek megfelelően, helyesen tudjuk

¹⁰ A szentimentfragmentum tehát nem feltétlenül tartalmaz szentimentkifejezést, azaz olyan elemet, amely szótári szinten értékelő tartalmat hordoz. Ez a megközelítési módomban teremt lehetőség azokra a vizsgálataimra is (l. 4.5.1, 6.3.2), amelyekben szentimentértékkel nem rendelkező alaptagokat is górcső alá veszek.

¹¹ Amint azt a dolgozatban látni fogjuk, a szentimentfragmentumnak nem kritériuma, hogy egyetlen tagmondatban tartalmazza az értékelést és a targetet, l. pl. a (48) alatti megnyilatkozásokat a 4.5.2-ben.

megállapítani. A jelen esetben tehát a megnyilatkozásnak negatív szentimentértéke van, amelyet a *nagyon* elem fokoz.

A disszertációban a nyelvi értékelő tartalmat a *szentimentérték* fogalmával jelölöm, amelyet definícióm szerint egyaránt hordozhat egy lexéma és bármely annál nagyobb elemzési egység (pl. egy megnyilatkozás vagy egy egész dokumentum is). Szentimentértéke van tehát annak a nyelvi elemnek, amely lexikai szinten hordozza azt, és annak a megnyilatkozásnak is, amelynek az értékét a bennfoglalt szentimentértékek együtt adják.

Amint arra a (16) alatti példa is rámutat, a szentimentfragmentumban levő szentimentkifejezés lexikai szintű szentimentértéke nem feltétlenül azonos a teljes, bennfoglaló szentimentfragmentum értékével (vö. Bolinger 1972: 15–20, Feldman és mtsai 2011: 1643, Liu 2012: 36–38, Szabó 2015a: 630–633). A módosulás lehet fokozati, de akár eredményezheti azt is, hogy a két érték teljesen ellentétes lesz egymással. Ezt a jelenséget tekinti a nyelvtechnológiai irodalom (vö. pl. Hogenboom et al. 2011, Muhammad és mtsai 2013) a *szentimentérték módosulásának* (angolul: *sentiment modification* vagy *sentiment score modification*).

A szakirodalom az értékmódosulás kapcsán alapvetően a módosító elemekre fókuszál, azaz rendszerint feltételez egy olyan elemet a kontextusban, amelynek hatására egy másik elem értéke annak lexikai szintű értékéhez képest módosul. Magam ugyanakkor azt az esetet is értékmódosulásnak tekintem, amikor egy adott nyelvi elem (amely nem feltétlenül a fragmentum szentimentkifejezése) lexikai szintű értékelő tartalma és annak kontextuális értékelő tartalma között eltérés van. Ennek megfelelően értékmódosulásként kezelem például az értékváltást, amikor egy szentimentkifejezés, annak negatív lexikai szintű szemantikai tartalma ellenére pozitív értékelést fejezhet ki (pl. *durva koncert*, *brutális telefon* stb.) valamint az értékvesztést is, amikor a lexikai szinten negatív értékelő tartalmat hordozó elem fokozó szerepben, tehát nem szentimentkifejezés funkciójában áll a fragmentumban (pl. *brutálisan jó*, *durván nagy* stb.). Rá kell mutatnunk tehát a lexikai szintű és a kontextuális szentimentérték eltéréseinek lehetőségére. Fontos hangsúlyozni továbbá, hogy az értékmódosulási jelenség a megközelítésem szerint nem korlátozódik azokra az elemekre, amelyeket a fragmentum szentimentkifejezéseiként azonosítunk, hiszen az értékmódosulás lehetséges bármely, lexikai szinten értékelő szemantikai tartalommal rendelkező nyelvi elem esetén.

Ez a megközelítési mód azért is szerencsésebb az általános gyakorlatnál, mert lehetőséget ad az olyan módosulások figyelembe vételére és vizsgálatára is, ahol a

szenimentkifejezés és a bennfoglaló megnyilatkozás értéke a módosulás ellenére is azonos. Ezt tapasztalhatjuk például az értékvesztés jelenségének bizonyos típusainál (részletesebben l. lentebb). Például, a *brutálisan rossz* szókapcsolatban a módosító elem fokozó szerepben áll, és nem a lexikai szintű szemantikai tartalma szerint interpretáljuk azt. A teljes kifejezés ugyanakkor e módosulás ellenére is nagyon negatív szenimentértékkel rendelkezik, akárcsak a vizsgált elem lexikai szintű szemantikai tartalma. Amennyiben a módosulás jelenségét a nyelvtechnológiai gyakorlat szerint tekintjük, úgy az ehhez hasonló jelenségeket nem vehetjük figyelembe. A definíció kiegészítése emellett azért is szerencsés, mert lehetéssé teszi az olyan módosulások figyelembe vételét is, amelyekben további módosító elemeknek, azaz shiftereknek (l. lentebb) nincs szerepük. Az említett esetekben egyazon lexéma hordoz ugyanis eltérő értékelő jelentést a szótár és a kontextus szintjén, amellyel az értékmódosulással foglalkozó nyelvtechnológiai irodalom nem számol. A megközelítési módom tehát lehetővé teszi a probléma behatóbb és szélesebb körű áttekintését, kezelését. A nyelvtechnológia irodalma mind a szenimentkifejezések, mind a nagyobb elemzési egységek (pl. egy megnyilatkozás vagy egy egész dokumentum) összesített értékére a *polaritás* fogalommal utal. A *polaritás* fogalma ugyanakkor több szempontból is problematikus, és bővebb tárgyalásra érdemes.

Mindenekelőtt a nyelvtechnológia *polaritás* fogalma nem azonos azzal a *polaritás* fogalommal, amelyet az elméleti nyelvészet használ. Amíg ugyanis az elméleti nyelvészet azt kétpólusúként tételezi (vö. Kiefer 2008: 523), addig a nyelvtechnológia skaláris jelenséggként tekint rá. Ezzel összefüggésben, a nyelvtechnológiában gyakran találkozunk olyan megközelítési módokkal, amelyek a polaritás valamilyen mértékű módosulásáról beszélnek (pl. egy szenimentkifejezés lexikai szintű értékének a fokozásáról különböző intenzifikáló elemek segítségével, vö. pl. *jó* és *nagyon jó*).

A nemzetközi nyelvtechnológiai irodalomban a polaritás jelenségéhez a skaláris sajátságon túl legtöbbször azt a jellemzőt kapcsolják, hogy az egyes szenimentkifejezések nem csupán egy típusú, hanem két polaritási sajátsággal is rendelkeznek: egyrészt a lexikai szintű, másrészt a kontextustól függő, aktuális polaritással. E két polaritástípus megnevezésére a nyelvtechnológia a *prior polarity* és a *contextual polarity* terminusokat használja (vö. Liu 2012:67, Wilson és mtsai 2005, 2009). Az előbbi terminus alatt a szerzők tehát azt az értéket értik, amelyet egy nyelvi elem kontextustól függetlenül, lexikai szinten hordoz. Az utóbbi fogalommal pedig arra a szemantikai tartalomra utalnak, amellyel a szenimentkifejezés az aktuális kontextusban rendelkezik. Ahogyan azt Wilson

és szerzőtársai (2009: 3) magyarázzák, a két polaritástípus elkülönítése azért fontos, mert, bár a kifejezések gyakorta azonos prior és kontextuális polaritással rendelkeznek, számos esetben e két érték különbözik egymástól.

Az elmondottakon túl– amint azt már tárgyaltam (l. 2.3.2.1) – a nyelvtechnológia irodalmában találkozni az *abszolút* (*Absolute Polarity*) és a *relatív polaritás* (*Relative Polarity*) fogalmaival is (vö. Ahn és mtsai 2012: 5–6).¹²A szerzők magyarázata alapján, amennyiben egy melléknév abszolút polaritást hordoz, úgy értékelő tartalmának minősége (pozitív vagy negatív volta) nem függ az adott doméntól, amelyben szerepel, illetve targettől, amelyet minősít. Amennyiben azonban a melléknév relatív polaritású, úgy értékelő tartalma az aktuális doméntól, illetve targettől függ. Ezzel a megoldással igyekeznek tehát azt a jelenséget kezelni, hogy bizonyos szentimentkifejezések lexikai szintű, mintegy inherens értékelő tartalommal rendelkeznek, míg más nyelvi elemek ugyancsak hordozhatnak értékelő tartalmat, annak ellenére, hogy az előzőekhez hasonló inherens értékelő tartalmat nem hordoznak.

Magam a *polaritás* fogalmát nem használom, és a döntést a következő érvekkel támasztom alá. Egyrészt, tekintettel arra, hogy a disszertációban a probléma körüljárásához az elméleti nyelvészeti irodalom megállapításait is vizsgálom, illetve annak eredményeire alapozom az alkalmazott nyelvészeti kutatómunkát, nem volna szerencsés egy olyan fogalmat alkalmazni az értekezésben, amely a két területen más-más értelemben használatos. Másrészt, az úgynevezett prior és kontextuális polaritás kettőssége nem fedi le a teljes problémakört, amely a szentimentérték módosulásának különböző típusait magában foglalja. Ez ugyanis csupán arra a jelenségre utal, amelyben egy adott szentimentkifejezés lexikai szintű értéke az aktuális kontextusban módosul, például ironikus használat miatt (17a-b), és nem utalhat arra a jelenségre, amikor a szentimentkifejezés és az azt magában foglaló szentimentfragmentum szentimentértéke valamely további, a fragmentumban szereplő elem hatására tér el egymástól, például negáció esetén (17c-d).

¹² A nyelvtechnológiában e fogalmak nem azokat a jelenségeket igyekeznek lefedni, mint amelyeket az elméleti nyelvészetben ugyanezek a terminusok. Kiefer (2008: 509) rendszerezésében például az abszolút melléknévek a következő implikációnak tesznek eleget: A *Péter egy fiú* és a *Péter okos* állításokból következnie kell a *Péter egy okos fiú* állításnak. Ezzel szemben a relatív melléknévek esetében azok jelentése „az összehasonlítás alapjától függ” (Kiefer 2008: 511). Nem beszélhetünk tehát abszolút, csupán a jellemzett entitások csoportjához viszonyított tulajdonságról (l. fentebb, 2.2).

- (17) a. Ez szép volt!
 b. Pisti megint okosat mondott!
 c. Ez nem volt szép.
 d. Pisti megint nem viselkedett megfelelően.

Harmadrészt, nem fogadom el azt a fogalmi oppozíciót, mely szerint bizonyos nyelvi elemek relatív értékelő tartalommal rendelkezhetnek, azaz olyan lexikai szintű értékelő tartalmat feltételezünk nekik, amely se nem pozitív, se nem negatív. Magam úgy tekintek a szentimentértékre, amely bizonyos elemek esetében inherens, lexikai szinten meglevő tulajdonság, és amennyiben ezzel a tulajdonsággal egy adott elem rendelkezik, úgy az érték minőségét (azaz pozitív vagy negatív voltát) is magában hordozza (pl. *jó, rossz, szép, csúnya*).¹³ A további elemekre (pl. *lassú, gyors, kemény, puha*) úgy tekintek, hogy azok ilyen inherens, lexikai szintű szentimentértékkel nem rendelkeznek. Ezek az elemek egy adott megnyilatkozásban jelölhetnek értékelést, azonban ez nem jelenti azt, hogy az értékelő jelentés a lexikai szintű sajátosságuk lenne. Ebből fakad az is, hogy a kontextusban esetlegesen létrejövő értékelő tartalmuk lehet pozitív és negatív egyaránt.

A nyelvtechnológiai irodalomban gyakran találkozni a *polarity shifter* vagy *contextual valence shifter* (*shift*: 'változás, eltolódás, elmozdulás') terminusokkal is mint egymás szinonimáiként használt fogalmakkal (vö. pl. Polanyi–Zaenen 2004, Ikeda és mtsai 2008, Li és mtsai 2010, 2013, Boubel és mtsai 2013). A szerzők ezekkel a megnevezésekkel utalnak minden olyan elemre, amely egy adott szentimentfragmentumban szerepelve eltérést okoz a szövegben szereplő szentimentkifejezés(ek) lexikai szintű értéke, valamint a fragmentum értéke között (vö. Li és mtsai 2010: 635). A korábban (5a-c) alatt közölt példákat itt (18) alatt megismétlem (a példákban a shiftereket kövér szedéssel emelem ki):

- (18) a. **Nem** hazudott.
 b. **Valamennyire** tehetséges.
 c. **Rendkívül** okos.

A (18a) alattiakhoz hasonlókat – az elméleti irodalommal összhangban – negálóknak (angolul: *negator*), a (18b) és (18c) alattiak típusait csökkentő és növelő intenzifikáló

¹³ Ezeknek az elemeknek az ironikus használati lehetőségeit e dolgozat keretei között nem tárgyalom.

elemeknek (angolul: *intensifier*) nevezi a nyelvtechnológiai irodalom. E fogalmakat az értekezésben magam is alkalmazom.

A fenti értékmódosulási típusokat tárgyalja a problémakörrel foglalkozó legtöbb irodalmi tétel (vö. pl. Na és mtsai 2004, Polanyi–Zaenen 2004, Kennedy–Inkpen 2006, Ikeda és mtsai 2008, Wilson és mtsai 2009, Li és mtsai 2010, 2013, Boubel és mtsai 2013), és ezzel a jelenséggel igyekeznek elszámolni a már tárgyalt, szemantikai kombinációs szabályok alkalmazásával is (vö. pl. Feldman és mtsai 2010, Ruppenhofer–Rehbein 2012) (l. 2.1).

A szentimentérték módosulását, illetve a szentimentshifterek különböző típusait illetően megállapítható, hogy a szerzők az egyes típusokat áttekintő jelleggel, behatóbb vizsgálat keretében a szentimentelemzés szempontjából nem tárgyalják. Nem találtam a nemzetközi irodalomban egyetlen olyan dolgozatot sem, amely a problémakört rendszerező igénnyel vizsgálná. A kapcsolódó kutatások alapvetően a jelenség egy-egy részproblémájára fókuszálnak, például a negáció kérdésére (vö. pl. Na és mtsai 2004, Kennedy–Inkpen 2006, Wilson és mtsai 2009). Ami a hazai, szentimentelemzéshez kapcsolódó kutatásokat illeti, az említett problémákat saját dolgozataimon (Szabó 2015a, 2015b, 2015c, 2017, Szabó–Vincze 2015, 2016, Szabó és mtsai 2016a, 2016b, 2017a, 2017b) kívül más nem tárgyalja.

Magam az angol nyelvű terminus alapján ugyancsak shifterekként utalok azokra a nyelvi elemekre, amelyek egy adott szentimentfragmentumban szerepelve eltérést okoznak a szövegben szereplő szentimentkifejezések lexikai szintű értéke, valamint a teljes fragmentum értéke között (vö. Li és mtsai 2010: 635). Ugyanakkor hangsúlyozom, hogy a szentimentérték módosulását nem minden esetben shifter okozza, hiszen bizonyos esetekben az okot nem tudjuk egyetlen elemhez kapcsolni. Ezt a jelenséget példázza az általam *értékváltásként* nevezett típus, amelyet a fejezet további részében részletesen tárgyalok.

A szentimentérték-módosulási jelenségein belül megkülönböztetem az úgynevezett *értékvesztést*, valamint az *értékváltást*. Az értékeléselemzés szempontjából (legyen az akár elméleti, akár alkalmazott nyelvészeti megközelítés) különösen figyelemre méltó sajátsággal bírnak azok az elemek, amelyek lexikai szinten negatív szemantikai tartalommal rendelkeznek ugyan, azonban fokozó funkcióban képesek azt részben vagy egészben elveszíteni (19a-b), sőt bizonyos esetekben arra is, hogy negatív tartalmuk ellenére pozitív értékelést fejezzenek ki (20a-b) (vö. Szabó 2015a: 632, 2015c: 52, 2016: 165–167). A vizsgált elemeket a példákban kövérrel emelem ki.

- (19) a. Szerintetek is **borzasztó** nagy a fiam lába? [gyakorikerdesek.hu]
 b. Nagyon egyszerű leszek: **rohad**t jól vannak kitalálva a szereplők személyiségei [szeretlekmagyarorszag.hu]
- (20) a. azt is említettük, hogy hamarosan hazánkba is ellátogat a banda egy borítékolhatóan **brutális** koncert keretében [rockstation.blog.hu]
 b. **durva** volt, amit a hegedűs művelt a hangversenyen [Nádasdy 2003]¹⁴

A (19) alatti példákban a vizsgált elemek más, értékelő kifejezések jelentését fokozzák, és ilyenkor a lexikai szintű negatív tartalmukat részben vagy egészben elveszítik (erről a problémáról a disszertáció későbbi, a szemantikai motivációt vizsgáló részében részletesen lesz szó, l. 4.5). A (20) alatti esetekben a vizsgált elemek maguk töltik be az értékelő kifejezések funkcióját, amikor is az aktuális szentimentértékük ellentétes lesz a lexikai szintű szentimentértékükkel. Az előbbi jelenséget *értékvesztés*nek, az utóbbit *értékváltás*nak nevezem (vö. Szabó és mtsai 2017a: 251–252).

Az értékvesztés kapcsán érdemes felhívni a figyelmet arra, hogy ezek az elemek, miután szentimentértéküket részben vagy egészben elveszítették, fokozó értelmű shifterek funkcióját töltik be. Vessük össze a (21a) és a (21b) alatti példát (a példákban a vizsgált szókapcsolatokat dőlt szedéssel emelem ki)!

- (21) a. az Európai Unió *nagyon jó volt* Írországnak [linguee.de]
 b. *őrült jó volt* már csak nézelődni is a megszokott pályákon [gamestar.hu]

Azt látjuk, hogy mindkét esetben fokozó értelmű módosító elem szerepel a megnyilatkozásban, ugyanakkor, amíg a (21b) alatti példa esetében ez a fokozó szerepű elem értékvesztést mutat (hiszen lexikai szinten negatív szemantikai tartalmat hordoz), addig a (21a) alatti példában szereplő elem nem.

Összegezve, értékvesztéskor tehát a vizsgált negatív emotív elem shifter funkcióját tölti be, és ebben a szerepben egy szentimentkifejezés értékét fokozza. Az értékváltás

¹⁴ Nádasdy Ádám (2003) példája. Bár a megnyilatkozásban a vizsgált elem értelmezhető annak lexikai szintű szemantikai tartalmának megfelelően negatív értékelésként is, a szerző alapján, valamint a dolgozat céljainak megfelelően itt az értékváltás esetét vizsgálom.

ugyanakkor maga a szentimentkifejezés esetében jelentkezik, tehát egy adott negatív emotív elem szentimentkifejezés funkciójában aktuálisan nem negatív értékelés fejez ki.

Értékvesztés esetében lehetséges az az eset is, hogy a lexikai szinten pozitív elem kerül fokozó pozícióba (pl. *jó csúnya*, *tökéletes hülye*) (vö. Nemesi 1998: 26). Ebben az értekezésben azonban csak a negatív szemantikai tartalmú elemekkel foglalkozom.¹⁵

¹⁵ A dolgozat keretei nem tették lehetővé, hogy a pozitív elemcsoportra is kiterjesszem a vizsgálatokat.

4. Az értékvesztés és az értékváltás jelenségeinek a vizsgálata^{16,17}

A jelen fejezetben a szentimentérték módosulásának a típusai közül az értékváltás és az értékvesztés jelenségeit tárgyalom részletesen, áttekintve a probléma elméleti nyelvészeti, pszichológiai, valamint alkalmazott megközelítési módjainak a legfontosabb eredményeit.

4.1 A szerkezeti sajátságok áttekintése

Ebben az alfejezetben röviden áttekintem, milyen szófajúak lehetnek, valamint milyen szófajú szavakhoz kapcsolódhatnak azok a nyelvi elemek, amelyek értékvesztést vagy értékváltást mutatnak. A rendszerezéshez valós szövegekből származó nyelvi adatokat hívok segítségül.

Ami az értékvesztő elemeket illeti, azok melléknévi, határozatlan számnévi vagy határozószói alaptaghoz, valamint igéhez szoktak kapcsolódni. (A példákban a vizsgált elemeket kövérrel, a szerkezeti alaptagokat dőlt szedéssel emelem ki.)

- (22) a. a korabeli vadászgépek, a Nieuport 17-ek, Fokker DR1-ek és a légi harcok **mocskosul jól** néznek ki a filmvásznon [port.hu]
b. Szerintetek is **rohadtul sok** lesz már a call of dutykból? [gyakorikerdesek.hu]
c. rögtön dupla vagy tripla adagot kell készíteni, mert **eszméletlenül gyorsan** fogynak [trendalelke.hu]
d. **Halálosan megszerette** De Rênálnét.¹⁸ [Stendhal: *Vörös és fekete*, fordította: Illés Endre 2013]

¹⁶ A jelenségek pontos definícióját l. 3.

¹⁷ A vizsgált elemek listáját l. az 1. számú mellékletben.

¹⁸ A dolgozat bírálója, Sass Bálint megjegyzi, hogy az igés szerkezeteket esetlegesen szerencsésebb lett volna nem bevonni a vizsgálat körébe, mert a targetek megállapítását illetően problémásak. Egyetértek a bírálóval, ezekben az esetekben előfordulhat, hogy az adott elemzési egységben több elem is tekinthető targetnek, azonban ezt az elemzés célja egyértelműsíteni tudja. Így például a (22d) alatti példa esetében az, hogy a rejtett alanyt vagy *De Rênálnét* tekintjük a targetnek, attól függ, hogy az aktuális tartalomelemzési feladat szempontjából mely szubjektumra nézve fontos az információ, azaz melyikről szeretnénk értékelést kinyerni a szövegből.

Az első három esetben a vizsgált elem – Cseh (1998: 34) terminusa alapján – ragtalan melléknévi határozó alakjában is előfordulhat (23a-c).¹⁹ Ugyanez azonban igei alaptag mellett nem lehetséges (23d).

- (23) a. Megjegyzem, **mocskos jól** tolja ezt is, de nyilván őt nem ez a karrier érdekelte elsősorban. [totaldamage.hu]
b. Mit fogunk csinálni ezzel a **rohadt sok** hulladék üveggel Győrben? [boldoguljgyorben.blog.hu]
c. Elsősők lettek....**eszméletlen gyorsan** telik az idő. [facebook.hu]
d. ***Halálos** megszerette De Rênalnét.

A vizsgált elem, annak lexikai szintű szemantikai tartalmát tekintve nem csak negatív, hanem pozitív szentimentértékkel is rendelkezhet (vö. Nemesi 1998: 26), például

- (24) a. (...) a megfelelőnek messziről sem nevezhető terep **szépen** kicsinálta a bokaízületeket [edzesonline.hu/blog]
b. Beyonce rajongó vagy? Biztos **jó** hülye lehetsz! [ize.hu]

Ezt az esetet azonban a jelen disszertációban részletesen nem tárgyalom.

Említsük meg, hogy a ragtalan melléknévi határozó esetében egybe írt változattal is találkozni, például

- (25) Harcsa Veronika jazz-énekesnő mesélt mindenféle izgi dolgot, és egy **marhajó** playlistet hozott nekünk! [petofilive.hu]

Az értékváltás jelenségét illetően az alábbi három típust kell számba vennünk: Az egyik lehetőség szerint a vizsgált elem önmagában áll, és a vele kifejezett értékelés a mellékmondatban kifejezett tartalomra mint targetre vonatkozik (vö. Andor 2011: 37). Tekintsük ismét a korábban (20b) alatt közölt példát, amelyet itt (26) alatt megismétlek!

- (26) **durva** volt, amit a hegedűs művelt a hangversenyen.

¹⁹ Kugler (2014: 130) – Cseh (1998)-tól eltérően – határozói és jelzői szerepű elemekről beszél, magam inkább Cseh (1998)-at követem a *Mai magyar nyelv rendszere* (II: 163–164) és a *Mai magyar nyelv* (309) alapján.

A másik esetben az elem egy target funkciójában álló főnévhez kapcsolódik (vö. Andor 2011: 37). Tekintsük ugyancsak Andor (2011: 37) példáját!

(27) Nagyon **durva** sportkupét mutatott be a Volvo.

Amint azt a fentebbi példa is mutatja, lehetséges, hogy a vizsgált elemhez egy további határozószó is kapcsolódik, és az is lehetséges, hogy az ugyancsak egy értékvesztést mutató elem (vö. Partington 1993: 189–190), például

(28) hát **mocskos durva** buli volt tegnap, fél 8-11ig voltunk kb. a ride club-ban, innen nem messze (...) Ja és tényleg 50 cent volt a koktél.... 11 csúszott le (...) :) [pellagriasdiary.blogspot.hu]²⁰

A harmadik lehetőség az, hogy az értékváltó elem egy igéhez kapcsolódik, például:

(29) amcsi énekesnők közül egyedül Aguilera, aki tényleg **durván** énekel, de egyrészt nem Idolos, másrészt ezt Ő is elég jól titkolta a legutóbbi lemezéig [sorozatjunkie.hu]

Ez utóbbi szerkezetben, bár hasonlít az értékvesztési eseteknél tárgyalt igés szerkezetre (l. fentebb, (22d) alatti példa), a vizsgált elemet nem azonos módon interpretáljuk: amíg a (22d) alatti példában a *halálosan* elem az alaptag jelentését intenzívebbé teszi (vö. nagyon, rendkívüli módon megszerette), addig az itt vizsgált esetben a *durván* szó az éneklés rendkívüliségét, kiemelkedő minőségét jelzi. Ez a sajátság véleményem szerint döntő érv amellet, hogy a két típust a terminológia szintjén is külön kezeljük egymástól.

Vizsgálati tapasztalataim alapján az értékváltás jelensége a szentimentértéket tekintve csupán egy irányban lehetséges: az elem lexikai szinten negatív, az aktuális kontextusban pozitív értékelő tartalommal rendelkezik. Ezzel kapcsolatosan figyelemre méltó Schmelev (2012: 842) megjegyzése, amely szerint a francia nyelvben találni az ellentétes irányú módosulásra is adatot: a *sacré* 'szent' pozitív szentimentértékkel rendelkezik, ennek ellenére képes negatív értékelő funkciót is betölteni. Ilyenkor a szerző

²⁰ A példa értékváltást is tartalmaz, de jelen esetben nem erre, hanem a *durva* fokozó elem továbbfokozási lehetőségére (*mocskos*) szerettem volna felhívni a figyelmet.

(Schmelev 2012: 842) szerint az angol *damned* 'átkozott, rohadt' elemhez hasonló aktuális szemantikai tartalmat hordoz, pl. *sacré menteur* 'átkozott hazug'.

Magam a kutatómunka és a vizsgálatok során a fenti példához hasonló jelenséget a magyar nyelvben nem találtam.²¹

4.2 Az értékvesztés és az értékváltás megközelítési módjai a leíró nyelvészeti irodalomban

Az alábbiakban röviden áttekintem a vonatkozó nyelvészeti kutatások legfontosabb eredményeit.

Az értékváltás jelenségével az elméleti oldalon részben az *enantioszémia* (néhány helyen: *antagonímia* és *kontronímia*) terminusa alatt foglalkoznak. Enantioszémia esetén egy adott lexémának egyszerre, azaz egyazon nyelvéllapotban két ellentétes jelentése van (vö. Kicsi 2010, Schmelev 2012). Klégr (2013: 18) úgy tekint az enantioszémiaira, mint a poliszémia egy speciális esetére, amely ötvözi az antonímia sajátosságát is.

A jelenség többféleképpen is megnyilvánulhat. Schmelev (2012: 837) például tíz, Klégr (2013) hét altípust különböztet meg. Voznesenskaja (Вознесенская 2011: 759) alapján az ún. emocionális-értékelő enantioszemiát a jelenség egy külön altípusaként tartják számon a téma kutatói. Ebben az esetben az adott lexéma egyszerre két ellentétes értékkel (értékelő jelentéssel) bír, pl. az angol nyelvben a *terrific* ('iszonyú, szörnyű' vagy óriási, csodálatos'), vagy az oroszban a *бесценный* ('felbecsülhetetlen értékű' vagy 'értéktelen') (vö. Schmelev 2012: 842). Schmelev (2012: 841–842) alapján azok a szavak, amelyek erős emocionális értékelést fejeznek ki, gyakran nyernek ezzel ellentétes szemantikai tartalmat. A szerző (Schmelev 2012: 842) megemlíti továbbá, hogy a jelenség az értékek elmozdulását tekintve az ellentétes irányba is lehetséges, azaz egy pozitív értékelést jelölő elem válik negatív értékelő tartalmat kifejező elemmé. Schmelev (2012: 842) a jelenségre a francia *sacré* 'szent, szentséges' szót hozza példaként, amely gyakorta az 'átkozott, rettentő' jelentésben szerepel, például *sacré menteur* 'átkozott hazudozó'.²²

Ami a jelenség létrejöttének módját illeti, Rahmati (2015: 81–82) és Schmelev (2012: 837) amellett érvelnek, hogy az az ún. szemantikai eltolódás (*semantic shift*) egyik

²¹ Hasonló, de nem azonos jelenség az, ha pozitív értékű kifejezéseket szitokszóként használunk, pl. *A hétszentségit!*, *szent szar* stb.

²² vö. a magyarban a *hétszentség* kifejezéssel.

lehetséges következménye. Ahogyan Rahmati (2015: 81–82) fogalmaz, a szó „eltolódik” eredeti jelentésétől egy másik jelentéshez.

Annak ellenére, hogy a többjelentésűség problémájával számos dolgozat foglalkozik (vö. pl. Pustejovsky–Boguraev 1996, Nerlich és mtsai 2003, Kiefer 2007, Rakova és mtsai 2007), az enantioszémia problémáját kevesek tárgyalják. Ahogyan arra Umar (1982) alapján Rahmati (2015: 84) is felhívja a figyelmet, a modern nyelvészeti irodalomban a jelenség csekély figyelmet kap, annak ellenére, hogy nyelvfüggetlen sajátosság. Kicsi (2010: 197), valamint Voznesenskaja (Вознесенская 2011: 759) az orosz nyelv vonatkozásában szintén hangsúlyozza, hogy az a kevésbé tárgyalt jelenségek közé tartozik.

Az enantioszémia és az értékváltás kapcsolatát illetően fontosnak tartok kitérni egy részproblémára, amely a rendszerezést érinti. Kicsi (2010: 197) alapján nem tartoznak „a szűkebb értelemben vett enantioszémiahoz” az ironikus nyelvhasználat eszközei, bár hozzáteszi, hogy a két jelenség szoros kapcsolatban áll egymással. A szerző (Kicsi 2010: 197) Szerdahelyi (1995: 282) irónia-definíciójával indokolja érvelését, amely szerint az ironikus nyelvhasználatban „a szöveg igazi jelentése a szó szerinti értelem ellenkezője”. Kicsi (2010: 197)-tel szemben Voznesenskaja (Вознесенская 2011: 759) az orosz frazeologizmusokban jelentkező enantioszémia kapcsán az iróniát a jelenség egy lehetséges altípusaként említi.

A jelen dolgozatnak nem célja, hogy arra a kérdésre választ adjon, miszerint tekinthető-e az enantioszémia alesetének az irónia. Azt azonban fontosnak tartom, hogy az általam vizsgált értékváltási jelenséget az iróniától az enantioszémiaán belül is biztonságosan el tudjuk különíteni egymástól. Úgy vélem, hogy az értékváltás esetében két szempontból sem ugyanaz a folyamat játszódik le, mint ami az irónia esetében történik. Mindenekelőtt, az ironikus beszédmódban, amint azt Szerdahelyi (1995: 282) alapján is láttuk, az aktuális jelentés „a szó szerinti értelem ellenkezője”, például:

- (30) Ez az ember csupa kedvesség (amelyben a *kedvesség* aktuális jelentése: 'undokság')

Meglátásom szerint a fenti példában az ironikus jelentés úgy jön létre, hogy a *kedvességszó* lexikai szemantikai tartalmához annak antonim párját kapcsoljuk aktuális jelentésként. Ugyanakkor, értékváltás esetében nem ilyen jellegű antonimikus viszonyról van szó. Az aktuális és a lexikai szintű jelentés ugyanis pusztán az általuk hordozott értékelő tartalmat

tekintve ellentétesek egymással, egyéb szemantikai tartalmuk nem. Vessük össze a korábban (20b) alatt közölt példát (amelyet itt (31a) alatt megismétlek) a (31b) alatti példával!

- (31) a. **durva** volt, amit a hegedűs művelt a hangversenyen [Nádasdy 2003]
(amelyben a *durva* aktuális jelentése nem a 'puha, gyengéd')
- b. egy igazán **brutális** és **kegyetlen** buli vár a szegedi Underground fanokra [facebook.com] (amelyben a *brutális* és a *kegyetlen* aktuális jelentése nem az 'emberséges, együttérző')

A fentebbi eltérés mellett az értékváltást az is elkülöníti az irónia jelenségétől, hogy értékváltáskor a lexikai szinten negatív elem válik az aktuális kontextusban pozitív értékelést kifejező elemmé. Ez utóbbi kritériumot illetően az értékváltás ahhoz a jelenséghez áll közelebb, amelyet Klégr (2013: 16) az *anti-irony* terminussal jelöl. Ebben az esetben a megnyilatkozást tevő egy negatív értékelést kifejező elemet használ pozitív értékelés kifejezésére, például

- (32) Látom, sokat lustálkodszt mostanában (miután a megnyilatkozást tevő értesült a partner szorgalmas munkájáról, aktív ténykedéséről)

Amíg tehát irónia esetében a megnyilatkozást tevő egy látszólagosan barátságos értékeléssel valójában támadást hajt végre, addig anti-ironia esetében az ellenkező történik: egy látszólagos támadással pozitív értékelést fejez ki (l. még *ugratás elve*, angolul *Banter Principle*, vö. Leech 1983: 142–145). Ugyanakkor ettől a nyelvhasználati formától is világosan elkülöníti az értékváltási jelenséget a korábban, első kritériumként tárgyalt sajátság, tehát, hogy anti-ironia esetében mindenképpen a lexikai és a kontextuális szemantikai tartalom antonim viszonyban áll egymással, értékvesztés esetében pedig az antonimikusság csupán az értékekre vonatkozik. (Az enantioszémia jelenségére a lexikai-szemantikai és a mentális reprezentáció kapcsán még visszatérek, l. 4.6.)

Ami az értékvesztés jelenségét illeti, az enantioszémia fogalmi keretében ezt a problémát nem lehetséges – legalábbis teljes egészében – tárgyalni. Amint azt a dolgozat későbbi, szemantikai motivációt tárgyaló részében meg fogom mutatni (l. 4.5), az értékvesztés alapeseteiben a fokozó elem vagy teljesen elveszíti negatív szemantikai

tartalmát, vagy azt részben megőrzi. Bár előfordulhat, hogy a lexikai szinten negatív fokozó elemmel a beszélő egy pozitív effektus kifejezését célozza, ez kimondottan ritka.

Az alábbiakban azokról a dolgozatokról szólok részletesebben, amelyek az enantioszémia fogalmától függetlenül, kifejezetten az értékvesztés vagy az értékváltás jelenségével foglalkoznak.

Andor (2011: 33) alapján „a lexikális egységek, kifejezések polaritásváltása [...] a jelentésváltozások folyamataiban meglehetősen gyakori jelenség”. Magyarázata szerint annak „leggyakoribb eseteiben a negatív jelentéstartalmú és használatú lexikális egységek pozitív irányú jelentésváltozását vagy jelentésbővülését, jelentésük kiterjesztését figyelhetjük meg” (Andor 2011: 33). Andor véleménye (2011: 35) szerint a jelenség különösen „az értékítéletet, fokozást kifejező ún. intenzifikáló szavak körében jellemző”.

Andor (2003) bizonyos magyar és angol nyelvű szavak (pl. *awfully* (*good/clever*), *terribly* (*intelligent*), *borzasztó(an)/elképesztően* (*jó*), *hihetetlenül* (*okos*) stb.) korpuszalapú elemzése alapján megállapítja, hogy a vizsgált kifejezések a két nyelvben nagy számban fordulnak elő, és gyakorta szinonimacsoportokba rendeződnek. Emellett gyakran kollokálódnak negatív értéket hordozó melléknevekkel amellet, hogy egyre nagyobb arányban fordulnak elő pozitív tartalmú kifejezésekben, konstrukciókban (vö. Andor 2003: 44–45). Andor megállapításaival összhangban áll Székely (2007: 51) megjegyzése, mely szerint a negatív fokozó kifejezések gyakran kapcsolódnak pozitív értékjelentésűekhez (pl. *pokolian ügyes, félelmetesen/borzasztóan jó* stb.).

Andor (2011: 39–41) a magyar és az angol nyelvű adatok mellett román, orosz, japán és pandzsábi nyelvű példákat is megvizsgálva arra a következtetésre jut, hogy a jelenség a vizsgált nyelvekben is megfigyelhető. Andor (2011) megállapításai mellett Nemes (1998: 26), Jing-Schmidt (2007) és Székely (2007) vizsgálati eredményei is arra mutatnak, hogy a jelenség számos nyelvben megtalálható, ezért valószínűleg nyelvfüggetlen sajátosság.

Nemes (1998: 26), Székely (2007: 51), Andor (2011: 33) és Kugler (2014: 133–134) egyaránt említést tesznek arról, hogy a jelenség az intenzifikáló elemek kapcsán az ellentétes érték irányába is lehetséges, azaz pozitívból negatívba (pl. *tökéletesen buta, jól elhibázta*).²³ Ugyanakkor mindhárman megjegyzik, hogy ez utóbbi eset sokkal ritkább jelenség az előbbinél; ahogyan Andor (2011: 33) fogalmaz „az ellenkező irányú, pozitívból

²³ A jelenségre Kiefer (2008: 514) is hoz példát (*tökéletes bolond*), de egészen más szempontból elemzi azt (az implikáció érvényességi feltételeit tárgyalja).

negatív tartalmúvá alakulás nyelvhasználatának előfordulása jóval ritkább eset a nyelvek sokféleségében”.

Andor (2011: 35) véleménye szerint a jelenség főleg „az értékítéletet, fokozást kifejező ún. intenzifikáló szavak körében jellemző”. A szerző utóbbi megállapítása alapján arra következtethetünk, hogy lehetségesnek tart hasonló jelentésváltozást nem csupán a fokozó elemek körében, azonban azt, hogy milyen elemekre gondol, nem fejti ki.

Kugler Nóra (2014) felhívja a figyelmet arra, hogy „a lexikális fokozás leggyakoribb kifejezései között sok a negatív polaritású, negatív érzelmekkel összekapcsolódó szó” (Kugler 2014: 133). Elsősorban a *félelmetesen* fokozó értelmű elemet, valamint a kifejezés fokozó funkciójú használatának lehetséges pszichológiai hátterét teszi vizsgálata tárgyává (Kugler (2014) megközelítését a későbbiekben részletesebben is tárgyalom, l. 4.4).

Balogh (2009) a magyar nyelvben a fokozó elemek csoportján belül megkülönbözteti azokat, amelyek elsődleges funkciójuk szerint fokozó szerepet töltenek be (pl. *nagyon, módfelett, felettébb, szerfelett* stb.), valamint azokat, „amelyeknek van elsődleges jelentése, és csak másodlagosan szolgálnak fokozásra” (pl. *határtalanul, mélyen, mértéktelenül, kimondhatatlanul, alaposan, egészen* stb.).

Jing-Schmidt (2007: 433–435) emotív intenzifikáló elemeknek nevezi azokat, amelyekre magam a dolgozatomban értékvesztésként utalok. A szerző (Jing-Schmidt 2007: 433–435) az emotív intenzifikáló elemek interpretációjának a problémájával a kognitív metaforaelmélet keretében számol el (vö. többek között Johnson 1987, Lakoff–Johnson 1980; Lakoff 1987, 1990, Lakoff–Turner 1989, Sweetser 1990). Amellett érvel, hogy az interpretáció során metaforikus leképezés történik két konceptuális tartomány, az emóciók és a nyelvi tartomány között, ugyanakkor ebben a folyamatban a metonímiának is kiemelt szerep jut. A folyamatot a német nyelvű *schrecklich* ’szörnyű’ elem példáján keresztül magyarázza, a következőképpen (vö. Jing-Schmidt 2007: 434): A forrástartományban a félelem érzelm metonimikusan a félelemmel kapcsolatos magas emotív intenzitásra redukálódik a céltartományban. A két tartomány közös jegye tehát a MAGAS INTENZITÁS, és a forrástartomány emotív intenzitása képeződik le a céltartományban (Jing-Schmidt megközelítéséről l. még 4.5).

4.3 A nyelvtechnológiai irodalom megközelítési módja

Amint arról a dolgozat 1.1. bevezető fejezetében szóltam, egyetlen olyan dolgozatról (Dragut–Fellbaum 2014) van tudomásunk, amely legalább az egyik jelenséget, az értékvesztés problémáját nyelvtechnológiai szempontból közelíti meg. A szerzők (Dragut–Fellbaum 2014: 41) maguk is megjegyzik, hogy saját munkájukon kívül nem ismernek olyan dolgozatot, amely az általuk vizsgált fokozó határozószók (köztük értékvesztő elemek) nyelvtechnológiai szerepét tárgyalná. Ami a magyar nyelvű szövegek automatikus értékeléselemzésének a feladatkörét illeti, a jelenségeket – egyetlen rövid említéstől (vö. Hangya és mtsai 2015: 212) eltekintve – saját dolgozataimon (Szabó 2015a, 2015c, 2016, 2017 és Szabó és mtsai 2017a, 2017b, 2017d) kívül más nem vizsgálja. A két probléma tárgyalása ugyanakkor a nyelvtechnológia, különösen a szentimentelemzés szempontjából fontos volna, hiszen azt a jelenleg legelterjedtebbnek tekinthető, szótáralapú automatikus elemzési megoldással nem lehet kezelni. Az értékvesztésre és az értékváltásra képes nyelvi elemek ugyanis rendszerint a lexikai szintű szentimentértéküknek megfelelően szerepelnek a szentimentlexikonban, ennek következtében automatikus, szótáralapú elemzésük téves eredményt ad.

A jelen dolgozatrészben rövid összefoglalását adom annak a megközelítési módnak, amelyet Dragut és Fellbaum (2014) alkalmaz az értékvesztés vonatkozásában.

A szerzők azt vizsgálják, hogy a fokozó értelmű határozószók milyen szemantikai szerepet töltenek be az értékelést kifejező megnyilatkozásokban. Azt kívánják feltárni, hogy vajon ezek az elemek maguk hordoznak-e értékelő tartalmat az aktuális kontextusokban, vagy csupán a szentimentkifejezések szemantikai tartalmát erősítik. A *polarity intensifier* terminussal utalnak azokra az elemekre, amelyek e szemantikai tartalom erősítésére alkalmasak (vö. Dragut–Fellbaum 2014: 38). Ezen az elemcsoporton belül megkülönböztetik azokat, amelyek határozói szerepűek, és *adverbial intensifier* elnevezéssel utalnak rájuk, például:

(33) I was absolutely delighted by the simple story and amazing animation.
'Teljesen el voltam ragadtatva az egyszerű történetről és a lenyűgöző animációról'

Hét angol nyelvű határozószót (*absolutely* 'teljesen', *awfully* 'borzasztóan', *enormously* 'roppantul', *extremely* 'rendkívül', *horribly* 'szörnyen', *incredibly* 'hihetetlenül', *pretty* 'meglehetősen', *seriously* 'komolyan, súlyosan') vizsgálva többek

között a következő kérdésekre keresik a választ: A vizsgált elemek megléte vagy hiánya okozhat-e módosulást a teljes megnyilatkozás szentimentértékében oly módon, hogy azt az ellenkezőjére változtatja, illetve ezzel a problémával összefüggésben, rendelkeznek-e ezek az elemek szentimentértékkel?

A kutatás első lépéseként a szerzők áttekintik, milyen értékkel rendelkeznek a vizsgált elemek bizonyos szentimentlexikonokban (Opinion Finder (OF): Wilson és mtsai 2005), Appraisal Lexicon (AL): Taboada–Grieve 2004), valamint a SentiWordNetben (SWN, Esuli–Sebastiani 2006). Az áttekintés eredményét az alábbi táblázatnak megfelelően ismertetik:

Adverbs	OF	AN	SWN
<i>absolutelly</i>	Neu	-	Neu
<i>awfully</i>	Neg	Neg	Neu
<i>enormously</i>	Neg	-	Neu
<i>extremely</i>	Neg	-	Pos
<i>horribly</i>	Neg	Neg	Neu
<i>incredibly</i>	Pos	Pos	Neu
<i>pretty</i>	Pos	Pos	Neu
<i>seriously</i>	Neg	-	Neu

1. táblázat: Nyolc angol nyelvű határozószó szentimentértéke három szentimentadatbázisban (vö. Dragut–Fellbaum 2014: 39)²⁴

A vizsgálatban a szerzők humán megítélésre alapoztak. Filmes témájú véleményesszövegekből kiválasztottak összesen 160 olyan mondatot, amelyben a vizsgált határozószók valamelyike szerepelt, majd elkészítették ezeknek a pármondatait is, amelyek csupán abban különböztek az eredetiektől, hogy törölték belőlük a vizsgált elemet. Figyeltek arra, hogy a vizsgálati anyagban a határozószók azonos arányban szerepeljenek, és arra is, hogy ugyancsak azonos gyakorisággal szerepeljenek pozitív és negatív szentimentértékű mondatokban. Ezt követően az adatközlőket arra kérték, hogy rendeljenek szentimentértéket a kapott mondatokhoz egy ötfokozatú skála alapján, a következő lehetőségek közül választva: nagyon pozitív, pozitív, semleges, negatív és nagyon negatív.

A vizsgálat többek között a következő sajátságokat tárta fel: Arra vonatkozóan, hogy ellentétére fordíthatja-e a határozószó a megnyilatkozás szentimentértékét, azt

²⁴ A jelen disszertációban az itt feltüntetett értékek megítélésével nem foglalkozom, a saját vizsgálataimat a magyar nyelvre korlátozom.

állapították meg, hogy ez az összes eset 8%-ában következett be. A határozószók alapvetően intenzitásbeli elmozdulást okoztak a mondatokban, tehát azokat a mondatokat, amelyeket ezek nélkül az elemek nélkül is negatívnak vagy pozitívnak értékelték az adatközlők, a határozószóval bővített változatukban is azonos szentimentértékűnek találták, csupán magasabb intenzitással. A kutatók szerint a vizsgálati eredmény arra mutat, hogy a szentimentlexikonokban alkalmazott megoldás, miszerint különböző inherens szentimentértékeket tulajdonítanak a vizsgált határozószóknak (l. fentebb, 1. táblázat), nem helytálló. A fentebbieknek megfelelően ugyanis a vizsgált elemek alapvetően fokozó szereppel bírnak, s ezt a sajátsgot kellene tükrözniük a szentimentelemzésben alkalmazott különböző lexikai erőforrásoknak is.

4.4 Pszichológiai szempontú megközelítés

Az értékvesztéssel foglalkozó dolgozatok közül több a pszichológia oldaláról vizsgálja a problémát (vö. Partington 1993, Jing-Schmidt 2007, Andor 2011, Kugler 2014). Az enantioszemiát tárgyaló dolgozatokban (vö. pl. Rahmati 2015: 84) is találni utalást arra, hogy az egymással ellentétes tartalmak létrejöttében az egyik fő motivációnak az emóciók felnagyítása, túlzása tekinthető. Nemesi (1998: 26) szerint a vizsgált elemek használatának a magyarázatát „a mindenkor beszélők pszichikumában érdemes keresni”. Ahogyan a szerző érvel, „az irodalomban, a filmezésben és a mindennapi érintkezésben egyaránt rendkívüli hatást lehet elérni szorongást, félelmet okozó témák fölvetésével, boncolgatásával, és e hatás legfeltűnőbb megnyilvánulása a befogadók részéről az odafigyelés” (uő 27).

Emellett Schmelev (2012: 841–842) alapján azok a szavak, amelyek erős emocionális értékelést fejeznek ki, gyakran nyernek ezzel ellentétes szemantikai tartalmat. Azt tehát, hogy az értékvesztő és értékváltó elemeket a pszichológiai motiváció szempontjából is tárgyaljuk, a szakirodalom megállapításai is indokolják.

Az ezekben a dolgozatokban tett megállapítások közül néhány, bár alapvetően az értékvesztési jelenségre vonatkozik, az értékváltás vizsgálatában is használható, lévén, hogy az értékváltás és az értékvesztés jelenségei hasonló működési mechanizmussal bírnak.

A jelenséget a pszichológiai folyamatok oldaláról elemző dolgozatok alapvetően arra a kérdésre keresik a választ, hogy mi az oka, illetve célja annak, hogy bizonyos,

lexikai szinten negatív tartalommal rendelkező elemeket pozitív elemek fokozóiként, sőt bizonyos esetekben pozitív elemekként használunk, továbbá, hogy mi és hogyan történik az interpretáció során a befogadói oldalon. A fejezetben ezeknek a vizsgálatoknak a legfontosabb megállapításait összegzem.

A szakirodalom alapján a negatív emotív elemek használatát alapvetően három jelenséggel hozhatjuk összefüggésbe: a *negatív torzítással*, a *pozitív torzítással* vagy *Pollyanna-jelenséggel*, valamint a *kongruenciával* (vö. Szabó 2017).

Jing-Schmidt dolgozatában (2007) az értékvesztő elemeket teszi vizsgálat tárgyává a pszichológiai folyamatok szempontjából. A szerző (Jing-Schmidt 2007: 417) ezeknek az elemeknek a megnevezésére – saját terminológiámtól eltérően – az *emotív intenzifikáló* (*emotive intensifiers*) terminust használja. A szerző a dolgozatában amellet érvel, hogy a jelenség az ún. *negatív torzítás* (*negativity bias*) kognitív-affektív modelljében hatékonyan magyarázható. Tekintsük a *negatív torzítás* elméletét közelebbről is!

A *negatív torzítás* jelensége azt takarja, hogy a kellemetlen, fenyegető információknak tudattalanul is szignifikánsan nagyobb figyelmet tulajdonítunk, mint a kellemes, pozitív tartalmú információknak (vö. Jing-Schmidt 2007: 418). Ahogyan Baumeister és mtsai (2001) fogalmaznak: a rossz erősebb a jónál.²⁵ A jelenséget mára számos empirikus vizsgálat eredménye alátámasztotta (vö. pl. Peeters–Czapinski 1989, Pratto–John 1991, Cacioppo–Berntson 1994). A negatív torzítás kialakulásának okát több szerző evolúciós okokra vezeti vissza (vö. Pratto–John (1991) és Baumeister és mtsai (2001), idézi Jing-Schmidt (2007: 418–419)). Úgy vélik ugyanis, hogy a negatív hatásokra, jelenségekre történő fokozott odafigyelés elengedhetetlen a faji alkalmazkodóképességhez, ami pedig a faj fennmaradásának a kulcsa.

A különböző vizsgálatok eredményei alapján (vö. pl. Rozin–Fallon 1987, Hansen–Hansen 1988: 922–923, Cosmides–Tooby 2000: 93, Rozin és mtsai 2000: 637–638) a negatív jelenségek közül azok váltanak ki fokozott figyelmet, amelyek az egyén fenyegetettségének indikátorai lehetnek (vö. Öhman és mtsai 2001: 392–394). Ennek értelmében nagyobb figyelmet keltenek a félelmet vagy undort kiváltó jelenségek, emellett a düh különböző megnyilvánulási formái (pl. haragos arckifejezés) is. (Ez a sajátság valószínűleg összefüggésben áll az általam vizsgált elemek lexikai szintű szemantikai tartalmával, l. lentebb.) Az undor – a félelem és a düh mellett – azért játszik fontos

²⁵ „bad is stronger than good.”

szerepet a negatív torzításban, mert önvédő mechanizmusként funkcionál a fertőzésekkel, mérgezésekkel szemben (vö. Rozin és mtsai 2000: 637–638).

Kugler Nóra (2014: 131–132) dolgozatában, amelyben a *félelmetesen* fokozó értelmű elem nyelvhasználati okait kutatja, a félelem emóció figyelemre gyakorolt hatását vizsgálja. Kugler (2014: 131) alapján, a limbikus, azaz a magatartás szabályozásáért felelős neurális rendszerben találhatók a félelem negatív emóciójának átélésekor aktiválódó idegsejthálózatok. E területeket tehát a félelem központjainak is tekinthetjük. Ugyanakkor, ezek az idegsejtek nem csupán akkor aktiválódnak, ha maga az alany félelmet él át. Az aktiváció olyankor is jellemző, ha az alany valaki mást félelmetes helyzetben lát, sőt, ha egy másik ember fenyegetettségéről, félelméről csupán verbális információt kap (vö. Bauer (2010: 41), idézi Kugler (2014: 131)). Aktivációt figyeltek meg ugyanis különféle képalkotó eljárásokkal olyan esetekben is, amikor a vizsgálati személy pusztán hallgatta, amint egy másik személy beszélt a félelméről, vagy egy általa félelmetesnek tartott helyzetről.

Jing-Schmidt (2007: 425) amellett érvel, hogy az emotív fokozó elemek használata szorosan összefügg a negatív torzítás jelenségével. Véleménye szerint ugyanis a használatukat az a cél motiválhatja, hogy a beszélő intenzívebb figyelmet váltson ki a hallgatóból. Mivel a hallgató ösztönösen nagyobb figyelmet fordít a potenciális veszély jeleire, a beszélő az emotív fokozó elemekkel kiaknázza a negatív torzításban „rejlő” kommunikációs lehetőségeket.

Kugler (2014) Jing-Schmidt (2007)-hez hasonló eredményre jut. Véleménye szerint „érzelmi működésünk neurológiája képezi a biológiai hátterét annak, hogy erős hatást tudunk tenni a beszédpartnerre olyan témaválasztással és olyan nyelvi kifejezések alkalmazásával is, amelyekhez erős érzelmek kapcsolódnak” (Kugler 2014: 132).

Úgy vélem, bár a fentebbi, az értékvesztés esetét vizsgáló dolgozatok az értékváltást nem vizsgálják, valószínűsíthetjük, hogy a pozitív torzítás az utóbbit is hasonlóképpen motiválhatja.

A negatív torzítás jelenségével szoros összefüggést mutat az a jelenség, amelyet magyarul talán az *izgalom téves tulajdonításának* fordíthatnánk (angolul: *misattribution of arousal*).²⁶ A fogalom azt a pszichológiai folyamatot jelöli, amelynek során az egyén tévesen más emócióval azonosítja az átélt izgatottság okozóját, mint ami azt valójában kiváltotta (vö. Schachter and Singer 1962). Például, az alany a félelem emócióra adott

²⁶ Köszönöm Balázs Lászlónak, hogy az elméletre, és annak az általam vizsgált problémával való lehetséges összefüggésére felhívta a figyelmemet.

fiziológiai választ a vonzalom érzelmének tulajdonítja. A tévedés oka valószínűleg az, hogy az összetévesztett érzelmekre adott válaszoknak azonos fiziológiai tünetei vannak, mint például az említett két emóció esetében a vérnyomás megemelkedése vagy a szaporább légzés.

A jelenséget először Schachter és Singer azonosította 1962-es dolgozatában (vö. Foster és Mtsai 1998: 86). Amint rámutatnak, az emóciók azonosítása a megtapasztalt fiziológiai tünetek alapján történik. Ez a sajátság ad lehetőséget arra, hogy az egyén az átélt tüneteket tévesen ne annak az emóciónak tulajdonítsa, amely azokat valójában kiváltotta. Schachter és Singer (1962) elméletét azóta számos vizsgálati eredmény igazolta. Így például, Dutton és Aron (1974) vizsgálati tapasztalatai alapján, félelmet keltő helyzetben (egy magasan levő és keskeny hídon áthaladva) a férfi vizsgálati személyek szignifikánsan gyakrabban számolnak be arról, hogy a látott nőt vonzónak találták, mint a kontrollcsoport tagjai. A vizsgálati személyek tehát a félelem által kiváltott fiziológiai tüneteket tévesen a vonzalom emóciónak tulajdonítják.

Véleményem szerint az értékvesztő és értékváltó elemek használatának pszichológiai motivációit illetően a fentebbi jelenséget is figyelembe kellene vennünk. Lehetségesnek tartom ugyanis, hogy a negatív emotív elemek – legalább részben – éppen e pszichológiai folyamat okán váltanak ki nagyobb figyelmet. A jelen dolgozatnak nem célja, hogy e hipotézist bizonyítsa, ugyanakkor fontosnak tartom kiemelni, hogy nincs tudomásom olyan dolgozatról, amely a pszichológiai jelenség, valamint az értékvesztő és az értékváltó elemek használatának összefüggéseit vizsgálná.

A következő tényező, amelyet az értékvesztési jelenség kapcsán fontolóra kell vennünk, az a *pozitív torzítás*. Ez a tendencia, amelyet *Pollyanna-jelenségnek* is neveznek, valószínűsíthetően tovább erősíti az értékvesztő elemek affektív hatását (vö. Jing-Schmidt 2007: 422–425). Az ún. *Pollyanna-hipotézis*, amely Boucher és Osgood (1969) nevéhez fűződik, nyelvi univerzáléként tételezi a pozitív töltetű kifejezések magasabb használati arányát a negatív töltetű nyelvi elemekkel szemben.²⁷

Amint arra Jing-Schmidt (2007: 423) felhívja a figyelmet, a negatív és a pozitív torzítás jelenségei csupán látszólag mondanak ellent egymásnak. Amíg ugyanis a negatív

²⁷ A Pollyanna-jelenség lehetséges okaival e dolgozat keretei között részletesen nem foglalkozom, csupán annyit említek meg, hogy Jing-Schmidt (2007: 423) alapján a szociális elvárások hatására alakulhat ki. Így például az ún. nyelvi tabu, azaz a közlésben tiltott nyelvi elemek kerülése (vö. Crystal 2006: 127–133, Bakró-Nagy 2008: 78–79) vagy az eufemizmusra való általános törekvés (*euphemism treadmill*, vö. Pinker 1994) mind szociális elvárásokon alapszanak.

torzítás a neurológiai folyamatok szintjét érinti, addig a pozitív torzítás a nyelvhasználat síkjára jellemző. Hatásukat tehát nem azonos folyamatokban tapasztaljuk meg.

A Pollyanna-jelenség kapcsán Jing-Schmidt (2007: 423) úgy véli, hogy a pozitív torzítás miatt a megnyilatkozásokban megjelenő negatív szemantikai tartalmú kifejezések intenzívebb hatást váltanak ki a befogadói oldalon. Tekintettel arra, hogy a kommunikációban a pozitív kifejezések használata képezi a normát, a negatív elemek magasabb illokúciós erővel rendelkeznek, ezáltal maximalizálhatják a kommunikációs hatást (vö. Jing-Schmidt 2007: 425).

A Pollyanna-jelenség és a negatív tartalmú kifejezések használata közötti összefüggésre hívja fel a figyelmet Kugler (2014: 132) is, amikor megállapítja, hogy „miközben a nyelvi tevékenységre alapvetően a pozitív beállítódás a jellemző (...), a hatáskeltésben (tudatosan vagy tudattalanul) kiaknázzhatjuk a fenti tapasztalatot.”

A pozitív torzításnak az általam vizsgált jelenségekre gyakorolt hatását illetően úgy vélem, hogy e tendenciának is – akárcsak a negatív torzítás jelenségének – jelentős szerepe lehet az értékvesztő és az értékváltó elemek használati motivációjában egyaránt.

A harmadik tényező, amely megfontolásra érdemes az értékvesztés és az értékváltás jelenségeinek vizsgálatakor, az a *kongruencia* tendenciája (vö. Kugler 2014: 133). Ahogyan Kugler (2014: 133) fogalmaz, a *kongruencia*, másképpen értékbeli egyez(t)és okán „a legtermészetesebb és ezért a legkönnyebben feldolgozható szerkezetekben azonos polaritású kifejezések kapcsolódnak össze; vagy másként: a kifejezések összekapcsolásának ebben az esetben feltétele az, hogy értékjelentésükben egyezzenek”. A szerző Péter (2008), valamint Imrényi (2012, 2013) alapján a *könnyen / jól megoldotta* kapcsolatokat hozza példaként a jelenség hatásának az igazolására. Amennyiben a kifejezés tagjai között kongruencia van, úgy a szerkezet feldolgozása kisebb mentális erőfeszítést igényel. Azonban, ha nincs, a szerkezet nagyobb mentális erőfeszítéssel dolgozható fel. Ezt látszik alátámasztani az is, hogy a kongruencia hiánya miatt fellépő szemantikai komplexitás jelöltebb formát is eredményezhet, mint például a magyar nyelvben az úgynevezett korlátozó kifejezések esetében; pl. *nehezen / rosszul oldotta meg (*megoldotta)* (vö. Péter (2008) és Imrényi (2012, 2013), idézi Kugler (2014: 133)).

Kugler (2014: 134) az értékvesztés jelenségéről a kongruencia tendenciájával összefüggésben a következő megállapításokat teszi: A beszélő ebben az esetben ellentétes értékeket rendel egymás mellé, és e nehezen összeegyeztethető értékek integrálásához az

interpretáció oldalán nagyobb mentális erőfeszítés szükséges. A nagyobb erőfeszítés pedig tovább erősítheti a szándékolt hatást (Baum Krisztina közlése alapján Kugler (2014: 134)).

Partington (1993: 188–189) – Kuglerhez (2014: 134) hasonlóan – érvel. Véleménye szerint az ellentétes értékekkel rendelkező elemek kapcsolatai –jelen esetben az értékvesztő elemek és alaptaguk kapcsolatai – intenzívebb hatást képesek kiváltani a befogadói oldalon a már megszokott, „kiszámítható” kapcsolatoknál. A szerző arról is említést tesz, hogy az ezekhez hasonló „szokatlan” kollokációkkal a retorikai hatás fokozása céljából gyakran élnek mind az irodalmi szövegekben, mind a nyelvi humorban.

A kongruencia valószínűsíthetően az értékváltás jelenségét nem, csupán az értékvesztést motiválhatja, lévén, hogy elsősorban azokban a szerkezetekben kollokálódnak az ellentétes értékeket hordozó elemek.

Az értékvesztéssel kapcsolatban érdemes felhívni a figyelmet az ún. *szemantikai prozódia* jelenségére, amelyet ritkábban *diskurzus-* vagy *pragmatikai prozódia*ként is említene a szerzők (vö. Andor 2011: 34). Andor szerint Sinclair (1991: 112) észrevételezte először, hogy bizonyos kifejezések a szövegekben dominánsan negatív vagy pozitív szemantikai tartalmú nyelvi elemekkel kollokálódnak. Ilyen kifejezés például az angolban az *utterly* ('teljesen, teljes mértékben') határozószó vagy a *provide* ('nyújt, ad') ige. Közülük az előbbi jellemzően negatív vagy ironikus értelmű megnyilatkozásokban szerepel (vö. Louw (1993: 49, 163–164), idézi Andor (2011: 34–35)), míg az utóbbi legtöbbször pozitív szemantikai tartalmú főnevekkel kollokálódik (vö. Stubbs (2002: 65–66), idézi Andor (2011: 35)). A példák annak ellenére, hogy azok lexikai szinten nem rendelkeznek értékelő szemantikai tartalommal, kollokálódási sajátosságuk következtében „felveszik” az adott értékelő tartalmat. Ahogyan Partington (1998: 68) fogalmaz, a „konnotációs színezet” mintegy „áttérjed” az értékelő kifejezésről a vele kollokálódó elemre is. Ezt a jelenséget nevezzük Louw (1993) alapján szemantikai prozodiának (vagy *diskurzus-*, illetve *pragmatikai prozodiának*, vö. Stubbs 2002: 65–66, Andor 2011: 34–35).

A szemantikai prozódia az értékvesztés szempontjából azért figyelemreméltó, mert valószínűsíthetően folyamatosan csökkenti az olyan kollokációknak az affektív értékét, amelyben egymás ellentétes tartalmak kapcsolódnak össze. Úgy vélem, a tendencia hatással lehet arra, hogy bizonyos értékvesztő elemek egyre több elemmel kollokálódhassanak módosítóként, és emellett talán arra is, hogy bizonyos elemek egyre gyakrabban forduljanak elő pozitív elemek módosítóiként negatív elemek helyett. A csökkenő affektív értékre hívja fel a figyelmet Kugler (2014: 137), is, amikor Nemesi (1998) alapján a következőt említi: „bár intenzifikáló szerepüket hatásosságuknak

köszönhetik, minél kiterjedtebbé válik az alkalmazásuk (...), annál ártalmatlanabbak is lesznek az érzelmi működésünk szempontjából”. A dolgozat egy későbbi, a jelenségeket korpuszok és egy szóasszociációs adatbázis segítségével, kvantitatív szempontok szerint is tárgyaló fejezetében (l. 6) ezt a problémakört is igyekszem körüljárni: korpuszalapon megvizsgálom az értékvesztő elemek kollokálódási sajátságait, valamint a szóasszociációs adatbázis segítségével megmutatom a különböző negatív emotív elemekre adott hasonló, illetve eltérő asszociációs mintázatokat.

Végezetül, az értékváltást és értékvesztést mutató elemeket aszerint tárgyalom, hogy azok lexikai szintű szemantikai tartalmuk szerint az alapemóciók közül mely érzelmetípushoz kapcsolhatóak.

Ekman és Friesen (1969), valamint más pszichológusok, például LeDoux (2000: 130–132) bizonyos emóciókról azt tételezik, hogy azok az evolúció folyamatában fejlődtek ki, ezzel összefüggésben az azokhoz kapcsolódó arckifejezések és gesztusok kultúrafüggetlenül produkálhatóak és interpretálhatóak. Másképpen, “intuitíve érthetőek” (Bánczerowski 2005: 203). Ezeket az érzelmetípusokat nevezik a pszichológiában alapemócióknak, amelyek közé az öröm, a düh, a bánat, a félelem, az undor, valamint a meglepődés tartoznak, és a nyelvtechnológiai tartalomelemző megoldások is gyakorta erre a rendszerezésre építenek (vö. Szabó–Morvay 2015, Szabó és mtsai 2016c: 284).

Jing-Schmidt (2007: 427–433) a lexikai szintű emocionális szemantikai tartalom szerint három csoportra osztja az általa emotív fokozóknak nevezett elemeket: a félelem, az undor, valamint a düh érzelmehez kapcsolható elemekre. Megvizsgáltam, hogyan illeszkednek Jing-Schmidt (2007) rendszeréhez a magyar nyelvű értékvesztést és értékváltást mutató elemek. A szerző felosztását némi változtatással alkalmaztam. A félelem által motivált fokozó elemek háromféleképpen kapcsolódhatnak az emócióhoz: a lexikai szintű szemantikai tartalom takarhatja magát az emóciót, utalhat a félelem okára, valamint jelölheti a félelem által kiváltott valamely testi tünetet is. Az undor érzelmehez kapcsoltam azokat a kifejezéseket, amelyek vagy magát az emóciót, vagy annak kiváltó okát jelölik. Végül, a düh által motiváltak tekintetem azokat, amelyek szitkozódásként, szitokszavakként is felfoghatóak.

A kialakított rendszerezést az alábbiakban, példaanyaggal együtt közlöm.

1. a félelem emóció által motivált elemek:

- a) a fogalom magát az emóciót takarja: *félelmetes(en)*; *borzalmas(an)*; *borzasztó(an)*; *rettenetes(en)*; *rettentő(en)*

- b) a félelem okára utal: *állat(i); baromi(ra/an); brutális(an); bődületes(en); durva/durván;éktelen(ül); embertelen(ül); irdatlan(ul); irgalmatlan(ul); istentelen(ül); kegyetlen(ül); könyörtelen(ül); marha/marhára; pokoli(an); rém(es/en); szörnyen;szörnyű; szörnyűséges(en)*
- c) a félelem által kiváltott valamely testi tünetet jelöli: *eszméletlen(ül); észvesztő(en); halál(i/os)(an); örületes(en); örült; pusztító(an); szédítő(en); szédületes(en); veszett(ül)*
2. Az undor emóció által motivált elemek:
- büdös(ül); disznó; dög; irtó(ra); irtózatós(an); iszonyatos(an); iszonyú(an); mocskos(ul/an);mocskos; piszkos(ul); piszok(ul), rohad(ul)*
3. A düh emóció által motivált elemek:
- átkozott(ul); (ki)baszott(ul); kurva/kurvára*

A bánat érzelmét által motivált, értékvesztésre és/vagy értékváltásra képes elemet a magyar nyelvben – akárcsak Jing-Schmidt (2007) az angolban – nem találtam. Ennek az oka összefügghet a negatív torzítás működési mechanizmusával, mely szerint a negatív jelenségek közül azok váltanak ki fokozott figyelmet, amelyek az egyén fenyegetettségének indikátorai lehetnek (vö. Öhman és mtsai 2001: 392–394) (l. fentebb). A bánat pedig nem tartozik ezek közé.

Érdemes felhívni a figyelmet arra, hogy azok a magyar nyelvre vonatkozó dolgozatok, amelyek a negatív fokozó elemek emotív szemantikai tartalmával is foglalkoznak, az érzelmek közül kizárólag a félelmet említik (vö. Nemesi 1998, Kugler 2014). Nemesi (1998: 26) szerint például ezek a kifejezések „az ijesztés, ijedés képzetkörébe tartoznak”, és „ijesztő túlzásoknak” nevezi őket. Ugyanakkor, amint látjuk, mind az emotív tartalmak vizsgálati tanulságai, mind a jelen alfejezetben feltárt pszichológiai motivációs tényezők azt mutatják, hogy ezek az elemek a félelem mellett a düh és az undor érzelmekhez is kapcsolódhatnak.

A magyar nyelvű vizsgálatok során találtam még néhány olyan fokozó elemet, amely véleményem szerint az érzelmek rendszerezését alapul véve külön csoportot, a meglepődés érzellel kapcsolatos fokozó elemek csoportját képezik:

4. A meglepődés emóció által motivált elemek:
- elképesztő(en); döbbenetes(en); megdöbbentő(en); meghökkentő(en)*

Ezeket az elemeket azonban nem tekinthetjük a negatív szemantikai tartalmú elemek közé tartozóknak, lévén, hogy a meglepődés érzelm pozitív és negatív egyaránt lehet (vö. Drávucz és mtsai 2017: 232). Ugyanakkor emocionális tartalmuk miatt a kutatómunkához kapcsolódó szóasszociációs adatfelvételnél ilyen elemek is szerepelnek ingeranyagként (l. 6.3.3).

Az esetlegesen az öröm érzelmhez kapcsolható elemek a jelen kutatási kérdés szempontjából nem relevánsak. Csupán megjegyzem, hogy a fokozó elemek közül a *fantasztikus(an)* az, amelyet egyedülként ehhez az érzelmhez kapcsolhatunk. (A szóasszociációs adatfelvételben ez az elem is szerepel ingeranyagként kontrasztív vizsgálati céllal (l. 6.3.3).)A jelen alfejezetben tett megállapításaim szóasszociációs adatbázis ingeranyagának az összeállításakor hasznosítottam (l. 6.3.3.3).

4.5 Az értékvesztés és az értékváltás vizsgálata a szemantikai motiváció szempontjából²⁸

Ebben a fejezetben az értékvesztést és az értékváltást mutató nyelvi elemeket a szemantikai motiváció szempontjából vizsgálom.

Áttekintem a szakirodalom legfontosabb, e problémakörhöz kapcsolódó megállapításait, és rámutatok, hogy azok a vizsgált elemek szemantikai motivációját illetően nem egységesek, sőt gyakran az egyes szerzők véleménye is ellentmondásos. Ezt követően néhány, korpuszból (MNSZ2, Oravecz és mtsai 2014), valamint egyéb, valós szövegekből származó példaanyag alapján górcső alá veszem azokat a szemantikai motiváció szempontjából. Többek között amellet fogok érvelni, hogy az értékvesztőknek nevezett elemek viselkedése e tekintetben nem egységes (vö. Szabó 2017).

Az értékvesztő elemeket emellet megvizsgálom a skalaritás szempontjából is, és több olyan sajátosságukra is rámutatok, amely a negatív emotív fokozó elemeket a nem ebbe a típusba tartozó skaláris elemekkel rokonítja.

Amint arról már korábban említést tettem (l. 4.2), az értékvesztés és az értékváltás jelenségeivel csekély számú dolgozat foglalkozik, és ezek többsége is csupán az értékvesztésre fordít figyelmet. Az értékváltás jelenségét külön egyáltalán nem tárgyalják a szerzők, és csupán néhány olyan megállapítást találni, amelyet valószínűsíthetően mindkét

²⁸ A vizsgált elemek listáját l. az 1. számú mellékletben.

típusra vonatkoztathatunk. A jelen fejezet elsődleges célja az, hogy a szakirodalmi megállapítások helytállóságát valós példákon felmérje, a jelenségeket a szemantikai motiváció szempontjából leírja, ezáltal az irodalmi eredményekhez adalékul szolgáljon.

Ugyanakkor, a fejezet végén az értékváltás eseteit is górcső alá veszem, és ugyancsak valós példák vizsgálatán keresztül rámutatok, hogy az értékvesztésre képes elemeknek csupán egy része használható értékváltóként. Arra is kísérletet teszek, hogy a sajátságot a vizsgált elemek szemantikai tartalmával összefüggésbe hozzam.

Andor (2011: 33) – mint fentebb tárgyaltam (l. 4.1) – az értékvesztési jelenséget a „jelentésváltozás”, illetve a „jelentésbővülés” esetei közé sorolja. Nem szól azonban arról, hogy e módosulás milyen mértékű a vizsgált kifejezések lexikai szintű szemantikai tartalmához képest. Tekinthetjük-e tehát teljes mértékűnek a deszemantizálódási folyamatot, vagy csupán – esetleg bizonyos esetekben – részleges módosulásról van szó.

Tolcsvai Nagy (1988: 402) a jelenséget kifejezetten a fokozó elemek funkciójában tárgyalja. Amellett érvel, hogy amíg az általa „hagyományosnak” nevezett jelzői csoport tagjai, így például a *kiváló*, a *nagyszerű* vagy a *csodálatos* elemek magukban hordozzák „a szóval jelölt érték valóságát”, addig az *állati* vagy a *baromi* típusú jelzők „a jelzett értékek relativitására utalnak”. Ennek következtében véleménye szerint azok jelentésében „ironizáló magatartás” mutatkozik meg. Megállapítható tehát, hogy Tolcsvai Nagy (1988) – például Andorral (2011) ellentétben – nem tartja lehetségesnek, hogy a vizsgált elemek bizonyos esetekben, lexikai szintű negatív értéküket elveszítve pusztán fokozó funkciót tölthetnének be. Úgy véli, hogy ezeknek a fokozó elemeknek a használata szükségszerűen az általuk módosított kifejezés pozitív értékének a redukcióját eredményezi.

Tolcsvai Nagy (1988) érvelésével ellentétben Székely (2007: 30, 45) úgy véli, hogy a vizsgált fokozó elemek szemantikailag „gyakorta motiválatlanok”, s ezt a motiválatlanságot a szerző szerint (Székely 2007: 43–45) két sajátság is alátámasztani látszik: Egyrészt, ezekben a szerkezetekben olyan elemek is kollokálódhatnak, amelyekben a fokozó elem önmagában a módosított elem szemantikai tartalmával ellentétes tartalmat hordoz (pl. *pokoli hideg*). Másrészt, bizonyos szókapcsolatok esetében az elemek között semmiféle szemantikai–logikai kapcsolat nincs (pl. *borzalmasan szeret*). Tolcsvai Nagy (1988) ugyanakkor azt, hogy mit ért „gyakorta” alatt, és mitől függ a szemantikai motiválatlanság, nem tárgyalja.

Kugler (2014) állásfoglalása az értékvesztést mutató fokozó elemek szemantikai motivációjára vonatkozóan ugyancsak nem egyértelmű. A szerző arra a következtetésre jut, hogy „a kifejezés fogalmi szerkezetében a negatív tartomány aktiválódhat ugyan, a

fokozati skála felől nézve azonban a nagy érzelmi hatással analóg módon, de a skála pozitív pólusán, 'nagy intenzitás' jelentést képvisel a kifejezés mindkét polaritású alaptag esetén" (Kugler 2014: 134). A lexikai szinten negatív kifejezések fokozó funkciójában tehát pusztán a „nagy intenzitás” szemantikai tartalmat hordozzák mind pozitív, mind negatív kifejezések módosítójaként. Ugyanakkor, a szerző megjegyzése, mely szerint „a negatív tartomány aktiválódhat”, arra utal, hogy a negatív tartalom elvesztése nem minden esetben következik be teljes mértékben vagy szükségszerűen. Ehhez illeszkedik az a megállapítás is, amelyet a dolgozat egy későbbi részében tesz a szerző (Kugler 2014: 137): „a negatív polaritású alapszóból létrejött fokozó kifejezések részlegesen neutralizálódnak, ezáltal egyre több kifejezéssel kapcsolhatók össze” (Kugler 2014: 137). E helyütt tehát részleges deszemantizálódásról beszél. Mindemellett – ahogy azt már korábban is közöltem (l. 4.4) – Nemesi (1998) alapján azt is megemlíti, hogy „bár intenzifikáló szerepüket hatásosságuknak köszönhetik, minél kiterjedtebbé válik az alkalmazásuk [...], annál ártalmatlanabbak is lesznek az érzelmi működésünk szempontjából” (Kugler 2014: 137).

Hasonló megállapítást tesz Péter (1991a: 141, 1991b: 66) alapján Nemesi (1998: 26) is, aki szerint a jelenség „a szójelentés ábrázoló elemének elhomályosodásához, és ezzel összefüggésben [...] a kommunikációban betöltött szerepének a növekedéséhez vezet”.

Balogh (2009) a vizsgált jelenséget tekintve a teljes deszemantizáció mellett foglal állást. Amint azt korábban ismertettem (l. 4.1), a magyar nyelvű elemek csoportján belül megkülönbözteti azokat, amelyek elsődleges funkciója a fokozás, valamint azokat, amelyek másodlagos funkciójukban szerepelhetnek fokozó értelmű elemekként. Ez utóbbi típust illetően megállapítja, hogy amennyiben „az ilyen, másodlagos fokozó elemeket egy-egy megfelelő kulcsszóhoz kapcsoljuk, elveszítik elsődleges, azaz lexikális jelentésüket és átveszik a fokozó értelmű „nagyon” adverbium jelentését”.

Balogh (2009)-hez hasonlóan gondolkodik Cseh (1998: 36) is: a negatív fokozó elemeket „a *nagyon* határozószó szinonimáinak, pontosabban [...] a határozószóként 'nagyon, nagy fokban, erősen' jelentésben használatos szavaknak” tekinti. Ugyanakkor az ő magyarázata alapján sem viselkedik minden vizsgált elem szemantikailag egységesen. Amíg például az *iszonyúról* megállapítja, hogy „alapjelentésétől mind jobban eltávolodva” alkalmassá vált arra, hogy „pozitív jelentéstartalmat is kifejezzon”, addig véleménye szerint a *marha* vagy a *tetű* „durvítják a beszédet, argotikus jelleget adnak a nyelvi megnyilatkozásnak” (Cseh 1998: 37–38).

Jing-Schmidt (2007: 425, 434) amellezt érvel, hogy a lexikai szinten negatív emotív elemek fokozó funkcióban elveszítik literális jelentésüket, és elsődleges szemantikai tartalmuk a pusztá intenzitásra redukálódik. Jing-Schmidt (2007: 425–426) alapján tehát fokozó szerepben az emotív intenzitás aktiválódik, a literális jelentés pedig látens marad. Így például, a német *schrecklich gut* ('borzasztó jó') szókapcsolatban a *schrecklich* esetében a literális jelentés látens marad, míg az ezzel a szemantikai tartalommal összekapcsolódó magas emotív intenzitás aktiválódik. Az elmondottakkal összefüggésben Jing-Schmidt (2007: 425) azt is hangsúlyozza, hogy az emotív fokozókat az eltérő nyelvi viselkedés elkülöníti az általános fokozó értelmű kifejezésektől. A fentebb idézett *schrecklich gut* 'borzasztó jó' példával ellentétben ugyanis a szintén német nyelvű *sehr gut* 'nagyon jó' kifejezés fokozó elemével nem kapcsolódik össze magas emotív intenzitás, és ezzel összefüggésben a *sehr* – és a többi, általános értelmű – fokozó elemet a beszélő nem a hallgatói figyelem elnyerésének érdekében használja, hanem „pusztán” azért, hogy az értékelés egy bizonyos, „többé-kevésbé” meghatározható fokozatát fejezze ki (vö. Bolinger (1972: 17), idézi Jing-Schmidt (2007: 426)).²⁹ Az emotív fokozók használatával tehát a beszélő a magasabb fokot – eltérően az általános fokozó kifejezésektől – annak emotív hatásával szeretné átadni a hallgatónak.

Ugyanakkor, Jing-Schmidt (2007) – akárcsak Balogh (2009) (l. fentebb) – a tartalmat illetően nem minden elem esetében vélekedik egyformán. Az ugyancsak német nyelvű *stink-* ('bűdös') elemről például úgy véli, hogy az ellenszenvesség, taszítás negatív konnotációja az esetében a fokozó funkcióban is megmarad (Jing-Schmidt 2007: 430).³⁰

Partington (1993) szerint a fokozó elemek csoportjának bővülése két tendencia mentén történik: az ún. *relexikalizáció* és *delexikalizáció* folyamataiban. A két fogalmat a szerző Sinclair (1992) terminológiájából kölcsönzi, és némiképp módosított formában alkalmazza a folyamat leírására. Tekintsük ezeket közelebbről is!

Partington (1993: 180–183) *relexikalizáció*nak nevezi többek között azt az általa *modal-to-intensifier shift*-ként említett jelenséget, amelynek során bizonyos modális jelentésű kifejezések fokozó funkciójú elemekké válnak. Véleménye szerint számos, ma

²⁹ „[...] a common degree word such as *sehr* 'very' [...] communicates a positive evaluation that is more or less accountable. Like the German degree word *sehr*, English *pretty*, *fairly*, *quite*, *very* and many others and Chinese *xiangdang* 'quite', *hen* 'very', *shifen* 'completely', etc. do not work to enact the speaker's emotion but provide accountable subjective evaluations of measurable degrees and thus are not considered as emotive intensifiers.” (Jing-Schmidt 2007: 426)

³⁰ „While the negative connotation of offense and repulsiveness remains in the semantics of the emotive intensifier *stink-*, the prefixoid *sau-* has completely risen above its lexical origin *Sau* 'sow, pig', a widely held prototype of filthy, disgusting creatures, to signal high emotive intensity without any negative aftertaste.” (Jing-Schmidt 2007: 430)

fokozó szerepet betöltő elem a nyelv egy korábbi állapotában modális szemantikai tartalmat hordozott, amelynek segítségével a beszélő kifejezhette a tartalom igazságára vonatkozó értékelését, vagy pedig megerősíthette szavai őszinteségét. A relexikalizáció következtében ezek az elemek eredeti modális tartalmukat elveszítve új nyelvi funkciót, fokozó szerepet kapnak. A szerző a folyamat példaként hozza az angol *very* 'nagyon' elemet, amely egykor az 'őszintén, valóban' jelentést hordozta.

A *delexikalizáció* fogalmát Partington (1993: 183) úgy definiálja, mint olyan redukciót, amely a nyelv egyes szavainak vagy szókapcsolatainak lexikai szintű szemantikai tartalmát illetően következik be. A jelenség következtében az adott nyelvi egység alkalmassá válik egy új funkció betöltésére, és ebben a funkcióban nincs egyéb szemantikai tartalom, amellyel hozzájárulna a bennfoglaló kifejezés jelentéséhez. Partington (1993: 183) a jelenség egy tipikus példaként – akárcsak a relexikalizáció esetében (l. fentebb) – a *very* 'nagyon' intenzifikáló elemet hozza, amely megállapítása szerint „csekély önálló jelentést” ad a kontextusához.

Partingtonnak (1993) az angol *very* 'nagyon' elem intenzifikálónak válására vonatkozó gondolataival érdemes összevetni Jing-Schmidt-nek (2007: 437) a német *sehr* 'nagyon' nyelvtörténeti fejlődéséhez kapcsolódó érvelésével. Jing-Schmidt alapján, a német *sehr* fokozó elem az ófelnémet *sēr* főnévből fejlődött ki, amely 'fájdalom, baj, bánat' jelentést hordozott. Az elmondottak véleményem szerint arra mutatnak, hogy akár a *sehr* fokozó elem fejlődésének története is megegyezhet azzal a folyamattal, amelyet az általunk vizsgált értékvesztő elemek esetében látunk.

Partington (1993: 184) a delexikalizációt olyan kontinuumként képzei el, amelyen a különböző fokozó elemeket találhatjuk attól függően, hogy a folyamatnak éppen mely fokozatát mutatják. A végpontban elhelyezkedő elemek kizárólagosan a fokozó funkciót töltik be, ettől a ponttól távolodva pedig olyan elemeket találunk, amelyek a fokozó szerepükben is őrzik valamelyest elsődleges szemantikai tartalmukat. Az angol *awfully* 'szörnyen, borzasztóan' kifejezést illetően például a szerző (Partington 1993: 183–184) úgy látja, hogy az majdnem teljes delexikalizáltságot mutat. Ezt támasztják alá véleménye szerint azok a korpuszvizsgálati eredmények, amelyek szerint a vizsgált elem negatív és pozitív polaritású kifejezésekkel megközelítőleg egyforma gyakorisággal kollokálódik, a kollokálódási lehetőségei tehát „relatív szabadok”. Az *awfully* elemtől eltérően ugyanakkor – ugyancsak korpuszvizsgálati eredmények alapján – a *terribly* 'szörnyen, borzasztóan' gyakrabban kollokálódik negatív tartalmú kifejezésekkel, ami a szerző

(Partington 1993: 184) szerint arra mutat, hogy ez utóbbi egy kevésbé delexikalizálódott elem.

A továbbiakban korpuszból (MNSZ2, Oravecz és mtsai 2014), valamint egyéb, valós szövegekből származó példák alapján görcső alá veszem azok az értékvesztést, valamint az értékváltást mutató elemek nyelvi viselkedését a szemantikai motiváció szempontjából.

4.5.1 Az értékvesztés vizsgálata

Az értékvesztést illetően amellet fogok érvelni, hogy a nyelvi viselkedésük a szemantikai motiváció szempontjából nem egységes, és alapvető összefüggésben van az általuk módosított elem szentimentértékével.

Véleményem szerint azokban az esetekben, ahol a vizsgált fokozó elemek lexikai szinten pozitív vagy negatív szentimentértékkel rendelkező alaptagokat módosítanak, azok deszemantizálódnak, és csupán az emotív tartalmukból fakadó intenzitási jegyük őrződik meg. Ezt a megállapítást érvényesnek tekintem mind a pozitív (34a,c), mind a negatív (34b,d) alaptagok melletti fokozó elemekre (a módosított szentimentkifejezést dőlt, a fokozó elemet kövér szedéssel jelöltem).

- (34) a. No magáról a versenyről nem mondok semmit, csak annyit tudok mondani, hogy **rohad** jó lesz, remélem! Hajrá Sumi! [MNSZ2, magyarországi, személyes, fórum]
- b. Nem értem a fanyalgókat, igen vannak **rohad** rossz magyar filmek, de ez kifejezetten jó! [MNSZ2, magyarországi, személyes, fórum]
- c. Bár azt nem gondoltam volna, hogy 40 fölötti, hiszen még így is **borzasztóan szép**! [starity.hu]
- d. Ugyan ne beszélj – mondták sógornői –, a te férjed **borzasztóan csúnya**, olyan, mint egy farönk, ha nappal látnád, megijednél tőle. [*Farönk királyfi / Tibeti legendák és mesék*, fordította: Róna Tas András 1958]

A fenti példák vizsgálata alapján úgy vélem, hogy a szentimentértékkel rendelkező alaptagok mellett a fokozó elemekkel a beszélő hatáskeltési céllal él. Ezekben az esetekben a negatív emotív szemantikai tartalom a megnyilatkozás jelentésében nem játszik szerepet,

csupán fokozza a megnyilatkozás erejét. Úgy is fogalmazhatnánk, hogy ezekben az esetekben a vizsgált elemeknek a jelentése kimerül az intenzitásban mint szemantikai tartalomban, ehhez semmilyen egyéb jelentéskomponens nem adódik hozzá.

A fentebb leírttól valamelyest eltérőnek tekintem azt az esetet, amikor a pozitív alaptag mellett a fokozó elem emotív szemantikai tartalmán interperszonális funkció alapul. Ilyenkor a beszélő a vizsgált elemmel nem csupán az intenzitást kívánja fokozni, hanem egyben a közölt tartalom rá gyakorolt hatását is közvetíteni szeretné. Tekintsük az alábbi példákat!

- (35) a. Bírom, amikor valaki tudja magáról, hogy **rohadt szép**, de mégis rondának mondja magát, hogy kapjon pár kedves szót. [ilajkyou.hu]
b. Nem vagyok vak, soha nem is voltam, így tudom, hogy Zalán még mindig **kegyetlenül vonzó** [Frányó Kriszta: *Legény(álom)lak*]

A fentebbi példákban a fokozó elemek arra is rámutatnak, hogy a közölt tartalom milyen emotív hatást tesz a beszélőre.³¹ Ilyenkor véleményem szerint az elem affektív jegye fontos szerepet kell, hogy betöltsön az interpretáció során.

Most vizsgáljuk meg a szemantikai motiváció kérdését a szentimentértékkel nem rendelkező alaptagok módosítói esetében!

Véleményem szerint, amennyiben a szerkezeti alaptag nem rendelkezik lexikai szintű szentimentértékkel, a vizsgált elemek mellettük a szemantikai motiváció szempontjából nem viselkednek egységesen. Tekintsük az alábbi példákat!

- (36) a. A gála idén **borzasztóan hosszú**ra nyúlt a rengeteg beiktatott zenés betét miatt (...) [MNSZ2, magyarországi, sajtó, article]
b. 166 centi 46 kiló az milyen egy 14 éves lánynak? Nem vagyok **rohadt magas**? [gyakorikerdesek.hu]

Amennyiben az alaptag nem rendelkezik lexikai szintű értékkel, három alapvető esetet tartok lehetségesnek. Az egyik lehetőség az, hogy a beszélő negatív viszonyulásának a kifejezéséhez a fokozó elem lexikai szintű negatív szemantikai tartalma is hozzájárul, sőt bizonyos esetekben a negatív értékelést maga a fokozó elem fejezi ki. Ezt az esetet

³¹Köszönöm Kenesei Istvánnak, hogy felhívta a jelenségre a figyelmemet.

példázzák a fentebbi (36) alatt közölt adatok. Úgy vélem, hogy a (36a) alatti megnyilatkozásban a negatív beszélői értékítélethez a *borzasztóan* fokozó elem lexikai szintű negatív tartalma is hozzájárul. Ennek következtében a teljes szemantikai kiüresedésről, tehát a negatív tartalom teljes mértékű elvesztéséről nem beszélhetünk. A (36b) alatti példával kapcsolatosan azt állítom, hogy abban a beszélő a fokozó elem lexikai szintű negatív szemantikai tartalmára építve fejezi ki arra vonatkozó kétségét, hogy megfelelő-e a magassága. A vizsgált elem használata nélkül a negatív viszonyulás nem lenne egyértelmű. Ezzel összefüggésben tehát itt sem lehet teljes szemantikai motivátlanság, megőrződik a negatív tartalom.

Ugyanakkor, a példákat illetően az is megállapítható, hogy a fokozó elem az intenzitáson túl pusztán a negatív komponenssel járul hozzá a teljes kifejezés információtartalmához, egyéb szemantikai összetevővel nem. Ez alatt azt értem, hogy ha a (36b) alatti megnyilatkozást tekintjük, a *rohadt* melléknevet nem úgy interpretáljuk, mint például a *rohadt alma* szókapcsolat esetében tennénk. (A megközelítést vessük össze Székely (2007: 43–45) megállapításával, mely szerint a vizsgált kifejezésekben a módosító és a módosított elem között nincs szemantikai-logikai kapcsolat (l. fentebb).)

Úgy vélem, hogy a fentebbi érvelésem, miszerint a szemantikai kiüresedés nem feltétlenül teljes mértékű, alátámasztja a következő sajátság is: a semleges alaptagok mellett, pozitív értékelésben a negatív fokozó elemek használata korlátozott akkor, ha az adott sajátság az adott targetre vonatkozóan pozitív és negatív egyaránt lehet.³² A (37) alatti valós példát alább a szögletes zárójelbe tett elemekkel magam egészítettem ki a jelenségbemutatása céljából.

- (37) Nincs is jobb annál, mikor a nagy melegben magunkhoz veszünk egy esernyős, [**szörnyű/*borzasztóan/*dög*] *hideg* koktélt és elszürcsölgetjük a medence mellett [az eredeti példa forrása: szilveszter.hu]

A *hideg* elem a szentimentérték szempontjából semleges, és folyadékokkal kapcsolatban egyaránt jelölhet negatív és pozitív sajátságot. Ezzel összefüggésben, a (37) alatti példához hasonló esetekben, amikor is megnyilatkozással a beszélő a pozitív viszonyulását szeretné kifejezni, a negatív fokozó elemek használata nem megengedett. Másképpen közelítve a problémához azt is mondhatjuk, hogy amennyiben ilyen esetben a beszélő negatív fokozó

³² Köszönöm Sass Bálintnak, hogy a dolgozat tématervéről írt bírálatában felhívta a figyelmemet a sajátság korlátozott érvényére.

elemet használ, az értékelés szükségszerűen negatív lesz, tehát nem lehet azt pozitívként értelmezni.

A kérdés körüljárásához az MNSZ2 korpuszt használtam, de ahhoz, hogy a jelenséget nagyobb adatmennyiségen (és esetleg a korpuszban nem reprezentált szövegtípusokon) is ellenőrizhessem, az elemzett szókapcsolatokat a Google-ben is lekérdeztem. Az MNSZ2 segítségével a következőképpen jártam el: először négy választott folyadéknévnek (*víz*, *koktél*, *üdítő* és *sör*) a *hideg* szóval módosított kapcsolatára kerestem rá, majd megvizsgáltam, hogy melyfokozó elemekkel használhatóak együtt. Az elemzés során csak azokat a kollokációkat vettem figyelembe, amelyek legalább kétszer szerepeltek a korpuszban. A lekérdezések alapvető statisztikai adatait az alábbi táblázat foglalja össze:

A vizsgált kollokációk	összes előfordulás	ebből azok száma, amelyeket fokozó elem módosít
<i>hideg víz</i>	3229	71
<i>hideg koktél</i>	10	7
<i>hideg üdítő</i>	167	134
<i>hideg sör</i>	471	106

2. táblázat: *Négy különböző folyadéknév és a hideg szókapcsolatainak a gyakorisága az MNSZ2-ben, valamint közülük a fokozó elemmel módosított kapcsolatok száma*

Megvizsgálva a fokozó elemmel módosított szókapcsolatokat megállapítottam, hogy a vizsgált folyadéknévek *hideg* tulajdonságát a beszélő pozitív viszonyulása esetén nem módosíthatja negatív fokozó elem. Úgy is fogalmazhatnánk, hogy azokban az esetekben, ahol a beszélő pozitív értékelés szeretne kifejezni, megnyilatkozásában a semleges alaptag a negatív fokozó elemet „nem tűri meg” maga mellett. Amennyiben a megnyilatkozással a beszélő pozitív viszonyulást szeretne kifejezni, úgy az esetlegesen megjelenő fokozó elem semleges (38a) vagy pozitív tartalmú (38b-c).

- (38) a. **Nagyon** *hideg* sört akarok és azzal szeretnék lenni akit Nagyon-Nagyon szeretek [MNSZ2, magyarországi, személyes, Facebook]
- b. Készítsd elő a zubbonyomat, lakkcipőmet és legyen **jó** *hideg* víz a mosakodáshoz. [Vass: *Élvemaradottak útja / Az utolsó hadgyakorlat a Bácskában*]

c. A kerti munka után jól esik egy üdítő zuhany, majd jöhet egy **jó hideg** sör és az édes pihenés! :) [MNSZ2, magyarországi, személyes, Facebook]

Ha a vizsgált kifejezésekhez mégis negatív fokozó elem kapcsolódik, úgy a megnyilatkozás rendre negatív értékelést fejez ki. Ezt mutatják az alábbi, Google-kereséssel kapott példák is:

- (39) a. A pancsi jól sikerült, a **dög hideg** víz ellenére [babaszoba.hu]
b. Nem valószínű, hogy partot értek a **borzasztóan hideg** víz és a sziget körüli erős áramlatok miatt. [the6thousandmileblog.wordpress.com/]

A jelenséggel kapcsolatban a magyarázat elején azt a kitételez tettem, hogy a semleges alaptagok mellett a negatív fokozó elemek használata csupán akkor korlátozott, ha az alaptaggal jelölt sajátság az adott targetre nézve pozitív és negatív is lehet. Vizsgáljuk meg most azt az esetet, ahol a semleges alaptag az adott targetre vonatkozóan mindig pozitív tulajdonságot jelöl. Tekintsük az alábbi példákat! (A példákban a targeteket aláhúzással jelöltem.)

- (40) a. **Marha gyors** lesz az internet 2016 végére az iskolákban [minuszos.hu]
b. Az elején még **brutálisan gyors** számítógépünk idővel idegesítően lassú lesz.[MNSZ2, magyarországi, sajtó, article]
c. A kínálat szűkössége miatt az 5 perc lehetőség elképesztően felértékelődik, mély fókuszra eredményez, amire az **őrült gyors** navigáció és a játékoság rá is tesz egy lapáttal. [highlightsofhungary.hu]
d. Ez 565 ppi pixelsűrűséget eredményez, ami **brutálisan éles** képet jelent. [napidroid.hu]

A példák alapján azokban az esetekben, ahol az adott target vonatkozásában az adott sajátság mindig pozitív, negatív fokozó elemek is előfordulhatnak a semleges alaptagok mellett a pozitívan értékelő megnyilatkozásokban. A (40) alatti példákban leírt sajátságok az adott targetek viszonylatában nem lehetnek negatívak. Ez a pragmatikai sajátság ad lehetőséget arra, hogy a vizsgált kontextusokban negatív fokozó elemeket is használhassunk. (Nem így a folyadékok vonatkozásában a *hideg* sajátság, amely – mint fentebb láttuk – negatív is lehet).

Az előzőekben, a (36) alatti példák kapcsán azt az esetet tárgyaltam, amikor a beszélő negatív viszonyulásának a kifejezéséhez a fokozó elem lexikai szintű negatív szemantikai tartalma is hozzájárul, sőt bizonyos esetekben a negatív értékelést maga a fokozó elem fejezi ki. A semleges alaptagok melletti másik lehetőség az, amikor a fokozó elem itt is deszemantizált, és vele a beszélő pusztá hatáskeltési cél. Tekintsük az alábbi példát!

- (41) **Brutálisan** nagy halat találtak a bűvárok, nem fogsz hinni a szemednek³³
[hirek.ma/hirek]

A fenti példában a *brutálisan* negatív tartalma részlegesen sem vesz részt az interpretációban, az elemnek kizárólagosan intenzitási értéke van (vö. (34) alatti példák).

A fentebb bemutatott jelenségeken túl véleményem szerint megkülönböztethetünk még egy harmadik esetet is, amely a semleges alaptagok melletti fokozó elemek szemantikai tartalmát jellemzi. Azt, amikor a fokozó elem negatív emotív tartalmán interperszonális funkció alapul, és ennek nagy szerep jut a megnyilatkozás pozitív értékelő tartalmának a létrehozásában. Tekintsük az alábbi példát!

- (42) [...] szilaj barnaság és **kegyetlen** kék szemek kellettek még ahhoz, hogy az ötvenes évekbeni női és férfiszívek felmelegedjenek a bandavezér iránt [Krúdy Gyula: *A kandúr*]

A fenti példában a vizsgált elem nem pusztá fokozó szerepet tölt be, hanem lexikai szintű szemantikai tartalma általa jellemzett entitás rendkívüliségére, a beszélőre gyakorolt hatására utal. A pozitív értékelő szemantikai tartalom létrehozásában tehát aktív szerepet játszik a vizsgált fokozó elem elsődleges tartalma. Ez az eset hasonló ahhoz, amelyben a fokozó elemnek pozitív szentimentértékkal rendelkező elem mellett jut interperszonális szerep (l. fentebb a (35) alatti példákat), azonban itt az alaptagnak inherens értéke nincs.

A negatív fokozó elemeket illetően figyelemreméltó probléma az, hogy amíg bizonyos negatív elemek, fokozó szerepben állhatnak pozitívan értékelő megnyilatkozásokban, addig más hasonló elemek nem. Nemesi (1998: 26) példaként az *ijesztő*, a *nyomasztó*, *riasztó*, valamint a *fenyegető* elemeket említi (pl. **riasztóan kedves*

³³ Véleményem szerint negatív ugyanazon okból kifolyólag nem lehet a mondat, amiért a (40) alatti példák.

emberek, *fenyegetően sikeres fellépés). A jelenségnek az okáról Nemesi (1998: 26)-tal egyetértésben úgy vélem, hogy ezekben az esetekben a szemantikai motiváció annyira erős, hogy az – legalábbis a jelen nyelvéllapotban – nem engedi meg ezeknek az elemeknek a pozitív értékelő megnyilatkozásokban való használatát.³⁴

A negatív fokozó elemek kapcsán fontosnak tartom, hogy a vizsgált elemeket a skalaritás szempontjából is elemzés tárgyává tegyük.

Általánosságban a fokozó elemek között két nagy csoportot különböztethetünk meg: a skaláris (*scalar modifiers*), valamint a totalitást kifejező módosító elemeket (*totality modifiers*) (vö. Paradis 2001: 49–50). Paradis (2001: 49–50) alapján, a skaláris módosító elemek, mint például az angol nyelvben a *very* ('nagyon') vagy a *fairly* ('elégg, meglehetősen') kijelölnek egy tartományt annak a tulajdonságnak a skáláján, amelyet az általuk módosított elem szemantikai tartalma meghatároz. Ugyanakkor, a totalitást kifejező fokozók, mint például az *absolutely* ('teljesen, tökéletesen') vagy az *almost* ('majdnem') nem egy tartományra, hanem egy konkrét értékre mutatnak az adott skálán.

A vizsgálattal azt szeretném feltárni, hogy a negatív fokozó elemek a skalaritás szempontjából az általános fokozó értelmű kifejezésekkel azonosan viselkednek-e.

A megvizsgált korpuszadatok (MNSZ2, vö. Oravecz és mtsai 2014), valamint egyéb, valós szövegekből származó példák arról tanúskodnak, hogy a vizsgált fokozó elemek a skaláris módosító elemekkel mutatnak viselkedésbeli hasonlóságot, ugyanis mindkét csoport tagjai tovább fokozhatóak a *nagyon* elem segítségével (a példában a vizsgált szókapcsolatot kövér szedéssel emelem ki):

(43) míg este **nagyon brutálisan** jó nézettséget hozott az X-Faktor, 18-49-ben a TV nézők közel 2/3 a tehetségkutató műsort nézte. [sorozatwiki.hu]

A (43) alattihoz hasonló kifejezések tehát éppúgy adatolhatóak, mint a más skaláris fokozók és a *nagyon* elemkapcsolatai, például

(44) Nagyon örülök, és hálás vagyok, hogy mindig **nagyon fantasztikusan** jó videókat, cikkeket teszel fel [esettanulmányok.hu]

³⁴ Hasonlóan érdekesek azok a negatív fokozók, amelyek egy korábbi nyelvéllapotban gyakori előfordulásúak voltak, ma már azonban – a sztenderd nyelvváltozatban legalábbis – nem használatosak, például *égedelem/égödelöm jószívű, kietlen szép* (vö. Prohászka 1953: 199, 202, idézi Nemesi 1998: 26). A jelen dolgozatnak azonban történeti és dialektológiai vizsgálatok nem képezik a tárgyát.

A fentebbiekkel szemben a totalitást jelölő fokozók nem fokozhatóak tovább, pl. **nagyon teljesen jó, *nagyon maradéktalanul jó* stb.

Figyelemreméltó továbbá, hogy a vizsgált negatív fokozó elemek nem csupán a *nagyon* elemmel, de egymással is fokozhatóak (l. 4.3). A korábban (28) alatt közölt példát itt (45a) alatt megismétlem.

- (45) a. hát **mocskos durva** buli volt tegnap, fél 8-11ig voltunk kb. a ride club-ban, innen nem messze (...) Ja és tényleg 50 cent volt a koktél.... 11 csúszott le (...) :) [pellagriasdiary.blogspot.hu]
- b. Vajon mekkora haszon várható ebből a **szédületesen baromi** nagy üzletből. [MNSZ2, személyes, magyarországi, fórum]

Végezetül említsük meg, hogy a negatív fokozó elemek egyes beszélők számára egymáshoz képest eltérő fokot is képesek jelölni, például

- (46) Mert ez nem csak **embertelen** jó, de **félelmetesen** jó! [moly.hu]

Amint látjuk tehát, a beszélő a vizsgált elemeket egy fokozati skálán elhelyezkedőként használja együtt.

Összegezve a vizsgálat tanulságait a következő legfontosabb megállapításokat emelem ki: A lexikai szinten negatív emotív tartalommal rendelkező fokozó elemek a szemantikai motivátlanság tekintetében csupán inherens (pozitív vagy negatív) szentimentértékű alaptagok mellett viselkednek egységesen. Ilyenkor vagy csupán az intenzitási jegyük őrződik meg, vagy – interperszonális funkció esetében – az emotív komponens is. Azokban az esetekben, amikor a vizsgált elemek inherens értékkel nem rendelkező alaptagot módosítanak, a vizsgált fokozó elemek nem azonos mértékben veszítik el lexikai szintű szemantikai tartalmukat. Három alapvető esetet tartottam lehetségesnek: pusztán az intenzitási jegy érvényesül, a negatív tartalom is szerephez jut, és végül, – interperszonális funkció esetében – ha az emotív komponens érvényesül. E megállapítások a vizsgált elemek nyelvi viselkedéséről egy jelentősen árnyaltabb képet adnak a szakirodalom korábbi megállapításainál, illetve bizonyos mértékben cáfolják is azokat.

A jelen alfejezetben tett megállapításaim a szótári reprezentációval kapcsolatos vizsgálatokban (l. 4.6), a NegEmotív korpusz annotációjának a kidolgozásában (l. 6.3.2.4), valamint az automatikus elemző eszköz létrehozásakor hasznosítottam (l. 7.1).

4.5.2 Az értékváltás vizsgálata

A fejezet további részében az értékváltás jelenségét veszem górcső alá a szemantikai motiváció szempontjából.

Az MNSZ2 (vö. Oravecz és mtsai 2014), valamint egyéb, valós szövegekből származó adatok alapján az értékváltást illetően a következő megállapításokat tehetjük: Az értékváltó elemekkel a beszélő kétféleképpen információt közölhet. Az egyik eset az, amikor pozitív értékelést szeretne kifejezésre juttatni vele (a vizsgált elemeket itt is kövér szedéssel emelem ki), például

- (47) a. Azt a **durva** bulit! Hogy én eddig nem voltam itt!!! :DDD [MNSZ2, magyarországi, személyes, Facebook]
b. hogy ott milyen **brutális** net sebesség tud lenni, még a törzsasztalon is [MNSZ2, magyarországi, személyes, fórum]

A másik eset, amikor nem vagy nem elsődlegesen a pozitív értékelését szeretné közölni, hanem a megnyilatkozásban foglalt információtartalom figyelemreméltó, meglepő voltát hangsúlyozza, például

- (48) a. Nem az a **durva**, hogy ma betelefonáltam a humbárfalvi rádió kívánságműsorába, hogy kérjek egy fröccsöt, hanem hogy kaptam [Facebook]
b. Azt a **durva**, milyen ütős ez a cucc (...) [Lackfi János 2014: *Egészségedre, Sándor!*]

Vizsgálataim arra is rámutattak, hogy amíg bizonyos, értékvesztésre képes elemek betölthetik az értékváltó funkciót (pl. a *durva* és a *brutális* elemek, ahogyan a (47) és a (48) alatti példákban láttuk), addig más elemek nem. Tekintsük az alábbi példákat!

- (49) a. És a csuda ér rá ezen töprengeni, amikor itt van egy *átkozott* buli a nyakunkon. [Népújság, 1977. január 1, 1. szám, 28. évfolyam]
b. Amikor ugyan az a *rohad*t Avaya telefon van a képernyőn a csaj asztalán, amilyenén tárcsáztam kedves ügyfeleimet [...] [MNSZ2, magyarországi, személyes, Facebook]

Az elemzési eredményeim alapján például az *átkozott* és a *rohad*t, ahogyan a (49) alatti megnyilatkozásokban is, szentimentkifejezés funkciójában nem lehetnek értékvtók, azaz csupán negatív értékelést hordozhatnak, holott korábban azt láttuk (l. 4.5.1), hogy ezek az elemek éppúgy állhatnak értékvesztő funkcióban, mint például a *durva* vagy a *brutális*. A korábban (34a) alatt közölt példát itt (50) alatt megismétlem.

- (50) No magáról a versenyről nem mondok semmit, csak annyit tudok mondani, hogy **rohad**t jó lesz, remélem! Hajrá Sumi! [MNSZ2, magyarországi, személyes, fórum]

A tapasztalataim arra mutatnak, hogy az értékvtó funkció a korlátozottabb: minden általam vizsgált elem állhat értékvesztő funkcióban, és jóval kevesebb értékvtóként. Olyan elemet azonban, amely értékvtó lehet, de értékvesztő nem, egyet sem találtam. A jelenséget azzal magyarázom, hogy amíg értékvesztéskor már a szűkebb kontextus (azaz a kifejezés alaptagja) egyértelműsíti az adott elem aktuális szentimentértékét. Ugyanakkor az értékvtó funkcióban ilyen elem nincs, ezért ez a funkció kevésbé „nyitott”.³⁵

Úgy vélem, hogy a jelenség hasonló ahhoz, mint amit az értékvesztés kapcsán tapasztaltunk a semleges alaptagok esetében. Tudniillik, hogy semleges alaptagok mellett a vizsgált fokozó elemek használata ugyancsak korlátozottabb, mint pozitív vagy negatív szentimentértékű elemek mellett (kivéve, ha egyéb kontextuális és ko-textuális körülmények egyértelműsítik a használatot).

Megvizsgálva az elemeket abból a szempontból, hogy azok lexikai szintű szemantikai tartalma mely emócióval áll kapcsolatban (l. fentebb, 4.4), úgy tűnik, hogy azok az elemek, amelyek az undor (pl. *rohad*t, *iszonyú*) vagy a düh érzelemhez (pl.

³⁵ A dolgozat bírálója, Sass Bálint felveti, hogy esetlegesen azért van ez így, mert az értékvtók az értékvesztőkből alakulnak ki, mégpedig úgy, hogy elmarad a (szemantikailag nem hangsúlyos) szentimentkifejezés komponens. Ennek megfelelően az értékvtó elemek mögé odaérthető egy default szentimentkifejezés. Magam is átgondoltam ezt a lehetőséget, azonban e feltételezés esetén nem tudom megmagyarázni azokat az eseteket, amikor az értékvtó elem pusztán figyelemfelkeltő szerepű, tehát a tartalomnak nem a pozitív, hanem a meglepő voltát hangsúlyozza (l. a (48) alatti példák).

átkozott, kurva) kapcsolódnak, nem szerepelhetnek értékváltó, csupán értékvesztő funkcióban.³⁶ Ezek az elemek – lexikai szintű emotív tartalmuk okán – nem tudnak szentimentkifejezés funkciójában pozitív értékelést kifejezni. Velük szemben, mint a (47) és a (48) alatti példák alapján is láttuk, bizonyos elemek, mint például a félelem emócióhoz köthető *félelmetes, brutális, durva, szédületes* vagy *kegyetlen*, az értékváltó funkciót is képesek betölteni.

A vizsgálati tanulságokat a következőképpen összegezhetjük: Az értékváltó elemekkel a beszélő két típusú információt közölhet. Az egyik eset az, amikor pozitív értékelést szeretne kifejezésre juttatni vele, a másik, amikor a közölt tartalom figyelemreméltó, meglepő voltát hangsúlyozza.

Azt is megmutattam, hogy minden általam vizsgált elem állhat értékvesztő funkcióban, azonban jóval kevesebb értékváltóként. Olyan elemet, amely éppen ellenkezőleg, tehát értékváltó lehet, de értékvesztő nem, nem találtam.³⁷ A jelenséget az elemek emotív szemantikai tartalmával hoztam összefüggésbe.

A jelen alfejezetben tett megállapításaim a NegEmotív korpusz annotációjának a kidolgozásában (l. 6.3.2.4), valamint az automatikus elemző eszköz létrehozásakor hasznosítottam (l. 7.1).

4.6. Az értékvesztő elemek szótári reprezentációja

A dolgozat jelen fejezetében azt a problémát járom körül, hogyan lehetne az értékvesztést mutató elemeknek a megelőző fejezetekben feltárt sajátásaival helytállóan, ugyanakkor a számítógépes nyelvészeti alkalmazások szempontjából is hasznosítható módon elszámolni az elméleti oldalon. Arra a kérdésre keresem a választ, hogyan képzelhető el az értékvesztést mutató elemek szótári reprezentációja, illetve az itt alkalmazott elméleti keretek, a lexikai pragmatika, valamint a mentális lexikon hálózatos modellje erre milyen lehetőségeket kínálnak.

A szótári reprezentáció kérdése nem csupán az elméleti nyelvészeti megfontolások, de a nyelvtechnológia szempontjából is kiemelkedően fontos. A lexikon ugyanis, amely a nyelvi elemek reprezentációját tartalmazza, „nélkülözhetetlen szinte minden számítógépes

³⁶ A többi emóció ebből a szempontból – a már leírtak miatt (l. 4.4) – nem releváns, ezért velük itt nem foglalkozom.

³⁷ Az értékváltásra is képes elemeket az 1. számú mellékletben félkövér szedéssel emeltem ki.

nyelvészeti feladat megoldásához” (Gábor és mtsai 2008: 854). Ily módon a nyelvtechnológiai alkalmazások sikere javarészt azon múlik, hogy a lexikont, azaz a nyelvi tudás bázisát sikerül-e hatékonyan működtethető rendszerként létrehozni (vö. Gábor és mtsai 2008: 854). A lexikon hatékonysága pedig nyilvánvalóan csupán akkor képzelhető el, ha annak kialakítása mögött helytálló elméleti nyelvészeti keretrendszer van.

Az értékvesztő elemek szótári reprezentációját illetően úgy gondolom, hogy azokkal a lexikai pragmatika elméleti keretében, valamint a mentális reprezentáció aktivációterjedéses hálózati modelljében lehetséges hatékonyan elszámolni. E két megközelítési módot a jelen fejezetben részletesen tárgyalom.

4.6.1 Az értékvesztő elemek vizsgálata a lexikai pragmatika elméleti keretében

A számítógépes lexikonok (*computational lexicons*) szerveződését illetően Véronis (2003) két alapvető típust különböztet meg: az enumeratív, valamint a generatív megoldást (vö. Gábor és mtsai 2008: 863). Az előbbiben a jelentések explicit formában szerepelnek, míg az utóbbiban az egyes szavakhoz tartozó szemantikai információk alulspecifikáltak, így a pontos jelentés mindig az aktuális generáló szabályok segítségével jön létre (vö. Gábor és mtsai 2008: 863). Véronis (2003) alapján például a számítógépes nyelvészeti alkalmazások egy kardinális részfeladata, a jelentésegértelműsítés (*Word Sense Disambiguation, WSD*) csupán olyan lexikonok segítségével végezhető el hatékonyan, amelyek „a szavak jelentésének a leírása helyett a szavak használatának a leírását tartják szem előtt, így a lexikon szervező elvének a megfigyelhető jelenségeknek kell lenniük, például szintaktikai vagy kollokációs információknak” (Véronis (2003), idézik Gábor és mtsai (2008: 864)). A generatív lexikonok építési elvein alapszik többek között Pustejovsky és munkatársai (1995) ún. CoreLex szótára vagy Levin (1993) ún. *Verb classes and alternations* adatbázisa (vö. Gábor és mtsai 2008: 864–865).

A lexikai pragmatika (vö. Blutner 1998) úgy igyekszik szisztematikusan elszámolni a pragmatikai jelenségekkel, hogy abban a nyelvi elemek alulspecifikált szemantikai reprezentációjára épít. E megközelítési módja alapján, a lexéma önmagában pusztán olyan jelentésmaggal rendelkezik, amely mindig az aktuális kontextusban, a kontextuális és enciklopédikus információk által nyeri el a teljes jelentését (vö. Blutner 1998: 1). Ez az elv lehetőséget kíván teremteni arra, hogy olyan jelenségeket is kezelni lehessen, amelyekkel más elméleti keretekben nehezen tudnak elszámolni (mint például a lexémák motiválatlan

többértelműsége). A probléma körüljárásához röviden, példák segítségével tárgyalom a lexikai jelentés kérdését, kiindulva a kompozicionalitás elvéből.

A kompozicionalitás elve szerint egy adott lexéma rendre pontosan ugyanazzal a szemantikai tartalommal járul hozzá minden egyes kifejezés jelentéséhez, amelyben előfordul (vö. Fodor–Pylyshyn 1988). Amint azonban arra számos szerző rámutat (pl. Quine 1960, Lahav 1989, 1993), többek között a melléknevek nyelvi viselkedése gyakorta ezzel az elvvel nem áll összhangban. Quine (1960) alapján a *red apple* 'piros alma' és a *pink grapefruit* 'rózsaszín grépfrút' angol nyelvű szókapcsolatokhoz eltérő képzetek társulnak abban a tekintetben, hogy a gyümölcs színét annak mely részéhez kapcsolják a beszélők. Quine (1960) a *red hair* 'vörös haj' és a fentebbi *red apple* kifejezés összevetése kapcsán megállapítja, hogy a *red* lexémához eltérő képzeteket társítunk attól függően, hogy mely entitást minősítik (vö. Blutner 1998: 3, 2004: 496). Ugyanígy, a *fast* 'gyors', a *tall* 'magas' vagy az *intelligent* 'intelligens' interpretációját is alapvetően meghatározza az, hogy mely entitásokra vonatkoztatjuk őket (vö. Pustejovsky–Bogurajev 1993). Kiefer (2008: 531) a *kicsi* és a *nagy* mellékneveket vizsgálva úgy utal a jelenségre, mint „a kontextus dimenziócsökkentő hatása”-ra.

A problémát számos elméleti megoldással igyekeztek a téma kutatói orvosolni. Így például, a melléknevek esetében az ún. adnominális funktor (vö. Montague (1970), Keenan (1974) és Kamp (1975), idézi Blutner 1998: 3) vagy az ún. szabad változó feltételezésével (vö. Blutner 1998: 4, 2004: 497–498). (Ezeket az elméleteket többek között Blutner (1998: 2–9, 2004: 496–498) tárgyalja, e dolgozat keretei között nem részletezem).

A problémával az igék kapcsán ugyancsak számos dolgozat foglalkozik (vö. pl. Katz–Fodor 1963, Searle 1980, Bibok 2014). Katz és Fodor (1963) az alábbi példákkal mutatja be a problémát (vö. Blutner 2004: 489):

- (51) a. Should we *take* the lion *back* to the zoo?
'Vissza kell vinnünk az oroszlánt az állatkertbe?'
- b. Should we *take* the bus *back* to the zoo?
'Busszal kell visszamennünk az állatkertbe?'

Blutner (2004: 489) úgy fogalmaz, hogy a fentebbiekhez hasonló esetekben a két megnyilatkozás interpretálása során eltérő konceptuális illesztés (*setting*) történik, és ez az eltérés nem lehet a lexikonban rögzített szemantikai sajátosság.

Amint arra Searle (1980) felhívja a figyelmet, például a *cut* 'vág' szemantikai tartalma eltérő módon járul hozzá a teljes kifejezés tartalmához az aktuális megnyilatkozástól függően, például

- (52) a. cut the grass 'vágja a fűvet'
b. cut the cake 'vágja a tortát'

Searle (1980) alapján, mivel ezekben a megnyilatkozásokban az ige nem többértelmű, már a literális jelentés megfejtésekor figyelembe kell vennünk a kontextus szerepét (vö. Searle (1980), idézi Wilensky (1989)).

A jelenséget a főnevek szempontjából vizsgálja többek között Barsalou (1987, 1992) és Gluckberg (2001), akik Wilson (2003: 282) alapján arra a megállapításra jutnak, hogy a valamely kategóriával vagy konceptummal kapcsolatos enciklopédikus információk az interpretáció során képesek lehetnek az adott kategória vagy konceptum denotációjának a szűkítésére vagy a tágítására is. Például, a *bird* 'madár' főnév a lexéma denotációjának valamely specifikus részét is előhívhatja, vagy épp ellenkezőleg, azt a tágabb kategóriát, amelybe a denotátum az enciklopédikus információk alapján tartozik (így pl. a repülő entitások csoportját). Wilson (2003) e sajátosságokat a relevanciaelmélettel összefüggésben magyarázza.

Blutner (1998: 11–19) a jelenséget az ún. korlátozási probléma szempontjából tárgyalja részletesen. Felhívja a figyelmet arra, hogy Bierwisch (1983) alapján az angol nyelvben az olyan lexémák, mint a *school* 'iskola', a *university* 'egyetem', a *government* 'kormány' és a *parliament* 'parlament' nem azonos módon viselkednek abban a tekintetben, hogy hányféle jelöltjük lehet (épület, folyamat vagy intézmény) (vö. Blutner 1998: 11–14). A korlátozás kérdése szempontjából hasonlóan figyelemre érdemesek az angol nyelvben a *pork* 'sertéshús', a *beef* 'marhahús' és a *wood* 'fa' (anyagnévi értelemben), valamint a *pig* 'sertés', a *cow* 'marha' és a *tree* 'fa' (növény értelemben) lexémák is (vö. Blutner 2004: 503).

A fentebbi jelenségek elméleti leírásában alapvetően két megközelítési mód létezik. Az egyik megközelítés úgy tekint a jelenségre, mint tisztán nyelvészeti problémára, és a példákhoz hasonló esetekben az egyes lexémákat poliszém elemekként képzei el, amelyek lehetséges jelentéseit már a lexikon szintjén egyenként számba kell vennünk (vö. Cruse (2004), idézi Klégr (2013: 13)). Ez a megközelítés azzal az alaptézissel operál, hogy a lexikai szinten determinált elsődleges jelentésekből egy későbbi szinten, megszorítások

útján választódnak le azok a szemantikai tartalmak, amelyek az adott kontextusban értelmezésként nem lehetségesek. Ugyanakkor, ahogyan arra Blutner (1998: 12) is rámutat, e megoldást számos szerző kritizálta (pl. Meyer 1994, Taylor 1994, Blutner 1995). A másik elv szerint az adott lexémának inherens tartalomként csupán annyi különböző jelentését tételezzük, amely feltétlenül szükséges; az összes többi jelentésvariánst, amely más jelentés által motiváltnak tekinthető, nem kezeljük önálló jelentésként (vö. Klégr 2013: 8). Erre épít a Blutner (2004: 515) által alulspecifikáltnak nevezett elképzelés, amely a megszorításokkal operáló megközelítéssel ellentétben nem számos, már meglevő jelentés megszorításával, hanem egyetlen alulspecifikált jelentés gazdagításával operál (részletesebben l. lentebb).

A fentebb tárgyalt, az intézmények jelölőit érintő problémához hasonló jelenséget a melléknevek esetében is találunk, ráadásul az is előfordul, hogy több nyelv ekvivalens melléknevei esetében azok a különböző nyelvekben eltérő korlátozás alá esnek (vö. Kiefer 2008: 531). Kiefer a következő német és angol nyelvű példákat veti össze egymással:

- (53) a. Peter ist groß – Peter is tall 'Péter magas'
- b. großes Zimmer – large room 'nagy szoba'
- c. großes Auto – big car 'nagy autó'

Amint látjuk, a 'nagy' jelentésű *groß* a német nyelvben számos olyan kontextusban használható, amelyekben az angol nyelv esetében rendre különböző mellékneveket kell alkalmazni. Kiefer (2008: 531) megjegyzi, „az említett problémák miatt a *nagy-kicsi* mellékneveknek már a szemantikus ábrázolása is számos nehézséget támaszt”.

A poliszémia kérdésére visszatérve, Klégr (2013: 12–13) is tárgyalja azt a jelen dolgozat szempontjából különösen fontos enantioszémia jelenségének vonatkozásában (az enantioszemiáról részletesen volt szó a dolgozatban, l. 4.2). A kérdés az, hogy vajon enantioszémia esetén egyetlen poliszém lexémáról beszélhetünk, vagy pedig a különböző, egymással ellentétes jelentések önálló lexikai egységeket képeznek (vö. Klégr 2013: 13).

Blutner (1998: 12, 18, 2004: 491) számos, a fentebbiekhez hasonló jelenséget megvizsgálva úgy véli, hogy a konceptuális tudás jelentősebb figyelembevételre, és ezzel együtt az alulspecifikált szemantikai reprezentáció feltételezése vezethet csupán ezeknek a jelenségeknek a helytálló leírásához. Ennek megfelelően a tárgyalt, ún. intézmény-típusú fogalmakat részletesen vizsgálva megmutatja, hogy a korlátozási probléma nem szemantikai, hanem pragmatikai vonatkozású, és ezzel összefüggésben azt pragmatikai

szempontokat érvényesítve, alulspecifikált szemantikai reprezentáció feltételezésével lehetséges kezelni(vö. Blutner 1998: 41–46). A lexikai egységek reprezentációját szemantikailag alulspecifikált formában adjuk meg, és az aktuális jelentés ebből pragmatikai mechanizmusok útján jön létre, amely kontextuális sajátságokon és enciklopédikus tudáson alapul(vö. Blutner 1998: 18, 45, 2004: 491). Ezt a megközelítési módot támogatja többek között az a sajátság is, amelyre a korábban tárgyalt *pork*, *beef* és *wood* típusú anyagnevek használati korlátozottsága kapcsán hívja fel a figyelmet a szerző (Blutner 2004: 503). Amennyiben ugyanis a megnyilatkozásban megfelelő kontextuális feltételeket teremtünk, úgy a korlátozás akár fel is oldódhat, s eközben új korlátozások léphetnek fel, például

(54) Hindus are forbidden to eat cow / ?beef
'Az indiaiak nem ehetnek marhát'

Ez a tapasztalat is arra mutat, hogy a tárgyalt korlátozások alapvetően pragmatikai meghatározottságúak.

Blutner (1998) érvei számos szerző megközelítési módjával összhangban állnak. Így például, Wilson és Sperber (2004) alapján a jelentést a hallgatónak kell kidolgoznia az aluldeterminált jelentésből a kontextus alapján. Croft és Cruise (2004) véleménye szerint az egyes szavak szemantikai inputja a kontextusban képződik meg, azáltal, hogy az ún. jelentéspotenciált (*meaning potential*), illetve „általános értelmet vagy jelentésmagot” (*general purport*) a kontextus aktiválja (vö. Klégr 2013: 9). Wilensky (1989) a megnyilatkozás jelentését illetően szintén úgy gondolja, hogy a pragmatikai tudás már a reprezentációs folyamat legelső fokán be kell, hogy kapcsolódjon. Wilson (2003: 284) alapján pedig a „szószerintiség” egyenesen fel sem tételezhető sajátság, a nyelvileg kódolt jelentés ugyanis nem több mint a beszélői jelentés nyoma.

A továbbiakban az értékvesztő elemek problémáját veszem górcső alá a lexikai pragmatika elméleti keretét alkalmazva. A diszkusszióban azokra a valós példákra támaszkodom, amelyeket a szemantikai motiváció problémája kapcsán elemeztem (l. fentebb, 4.5). Az értékvesztési jelenség leírásában az alulspecifikált szemantikai reprezentáció elvét követem, és alapvetően Bibok (2014, 2016) érvelésére és reprezentációs megoldására támaszkodom.

Bibok (2016: 411) lexikai-szemantikai reprezentációjának egyik alaptézise az, mely szerint a szójelentés elidegeníthetetlen részét, a magjelentést (*core meaning*) el kell

különíteni az ún. prototipikus vagy sztereotipikus világtudástól, amely ugyancsak az adott szó jelentésére vonatkozik. Ezzel összefüggésben, a lexikai-szemantikai reprezentáció tartalmazhat olyan szemantikai predikátumokat, amelyek a sztereotipikus struktúrából adódnak ugyan, azonban az aktuális kontextusban, a megnyilatkozás interpretálása során mégsem épülnek rá a szó magjelentésére. A szerző (Bibok 2016) ezeket a szemantikai predikátumokat a reprezentációban kerek zárójelekkel különbözteti meg a nem opcionális predikátumoktól, és azt a megállapítást teszi, hogy mindezek az opcionális jelentéselemek akár külön-külön, akár együtt is ráépülhetnek a magjelentésre.³⁸

Bibok (2016: 410) egy másik fontos elméleti alapvetése, hogy a kódolt enciklopédikus információnak többféle típusát is megkülönbözteti. Úgy véli, bizonyos információkat a lexikai-szemantikai reprezentáció tárol, másokat a kontextus lexikai elemei hívnak elő, egy harmadik típusú információ pedig kontextusfüggetlenül jelenhet meg az általános világtudásunk alapján a megértés során. Annak okán, hogy háromféle enciklopédikus információt is megkülönböztet, lehetőséget ad az interpretációs mechanizmus következő leírására (vö. Bibok 2014: 227–228): Amennyiben az aktuális kontextus összhangban áll a lexikai-szemantikai reprezentáció valamely prototipikus vagy sztereotipikus struktúrájával, úgy a lexéma az aktuális kontextusban problémamentesen interpretálható. Amennyiben azonban az aktuális kontextus nem áll összhangban a lexéma lexikai-szemantikai reprezentációjának egyik prototipikus vagy sztereotipikus struktúrájával sem, „távolabbi” enciklopédikus információk elérésére van szükség.³⁹

A fentebb vázolt két tételt fontos alapelvnek tekintem az értékvesztő elemek leírásában. Ezek teszik lehetővé véleményem szerint ugyanis az értékvesztő elemek lexikai-szemantikai reprezentációjának helytálló megadását. A reprezentációban foglalt lexikai tétel a megközelítési mód szerint csupán akkor érvényes, ha a vizsgált elemek fokozó szerepet töltenek be, azaz fokozóként melléknévi vagy határozószói értékelő elemmel kapcsolódnak.

Amint arról a szemantikai motiváció problémája kapcsán részletesen szoltam (l. 4.5), a vizsgált elemek értékvesztés esetén, szentimentértékkel rendelkező alaptagok

³⁸ A vázolt megközelítési mód és a relevanciaelmélet (vö. Wilson 2003) összefüggéseivel, viszonyával a disszertációban nem foglalkozom, részletesebben l. Bibok (2016: 411).

³⁹ Bibok (2016) alapvetően igék jelentésszerkezetének a leírására törekszik, az orosz *rezat* 'vág' ige jelentését például a következőképpen adja meg (angol nyelvű tanulmányában):

(55) [[[x USE z] : [z PRESS y]] CAUSE [[BECOME [not WHOLE y]] (: [BECOME [EXIST w]])]],
where 1. w = parts of y, i.e. [w PARTS_OF y],
2. the round brackets express optionality.

mellett csupán intenzitást fejeznek ki⁴⁰, más esetekben az emotív tartalmukból fakadó affektív jegyük is részt vesz a megnyilatkozás interpretálásában, amikor is e szemantikai komponensen interperszonális funkció alapul. Az előbbi esetre hoztam a (34a-b) alatti, az utóbbira a (35a-b) alatti példákat, amelyeket itt (56) alatt ebben a sorrendben megismétlek:

- (56) a. No magáról a versenyről nem mondok semmit, csak annyit tudok mondani, hogy **rohadt jó** lesz, remélem! Hajrá Sumi! [MNSZ2, magyarországi, személyes, fórum]
- b. Nem értem a fanyalgókat, igen vannak **rohadt rossz** magyar filmek, de ez kifejezetten jó! [MNSZ2, magyarországi, személyes, fórum]
- c. Bírom, amikor valaki tudja magáról, hogy **rohadt szép**, de mégis rondának mondja magát, hogy kapjon pár kedves szót. [ilajkyou.hu]
- d. Nem vagyok vak, soha nem is voltam, így tudom, hogy Zalán még mindig **kegyetlenül vonzó** [Frányó Kriszta 2016: *Legény(álom)lak*]

Azt is megállapítottam, hogy a szentimentértékkel nem rendelkező alaptagok mellett három eset lehetséges a fokozó elemek szemantikai motivációját illetően: a fokozó negatív értékelő tartalommal rendelkezik, pusztán intenzitást jelöl vagy pedig – akárcsak az (56c-d) alatti esetekben – az emotív tartalmukból fakadó affektív jegyük is részt vesz a megnyilatkozás interpretálásában, és azon interperszonális funkció alapul. Az ezeket az eseteket demonstráló megnyilatkozásokat ((36b), (41), (42)) itt (57) alatt ebben a sorrendben megismétlem:

- (57) a. 166 centi 46 kiló az milyen egy 14 éves lánynak? Nem vagyok **rohadt magas**? [gyakorikerdesek.hu]
- b. **Brutálisan nagy** halat találtak a bűvárok, nem fogsz hinni a szemednek [hirek.ma/hirek]
- c. [...] szilaj barnaság és **kegyetlen kék** szemek kellettek még ahhoz, hogy az ötvenes évekbeni női és férfiszívek felmelegedjenek a bandavezér iránt [Krúdy Gyula: *A kandúr*]

⁴⁰ A megállapítást az automatikus elemző eszköz létrehozásakor hasznosítottam (l. 7.1).

Azt is megmutattam, hogy a semleges, azaz szentimentértékkel nem rendelkező alaptagok mellett, pozitív értékelésben a vizsgált fokozó elemek használata korlátozás alá esik, amennyiben az alaptaggal jelölt tulajdonság az adott targetre nézve negatív sajátság is lehet. Így például a *rohadt hideg víz* vagy a *borzasztóan hideg üdítő* szókapcsolatok rendre negatívak, pozitívként nem értelmezhetőek, és pozitívan értékelő megnyilatkozásokban vizsgálataim alapján nem is adathozhatóak. A fokozó tehát itt is negatív értéket hordoz a kifejezésben. Ugyanakkor azt is láttuk, hogy a korlátozás nem érvényes, ha az alaptaggal jelölt tulajdonság az adott targetre nézve csak pozitív sajátság lehet. A korábbi (40a-b) példákat itt (58) alatt megismétlem:

- (58) a. **Marha gyors** lesz az internet 2016 végére az iskolákban [minuszos.hu]
b. Az elején még **brutálisan gyors** számítógépünk idővel idegesítően lassú lesz. [MNSZ2, magyarországi, sajtó, article]

E megállapítások alapján a vizsgált fokozó elemeknek a következő lexikai-szemantikai reprezentációt tulajdonítom:

- (59) [intenzív] ([affektív]) ([negatív]),
ahol a kerek zárójelek opcionális jelölnek.

Az alkalmazott elméleti keretben (vö. Bibok 2014, 2016) a reprezentációban feltüntetett opcionális jegyek a sztereotipikus struktúrából erednek, és Bibok (2016) alapján azokat kerek zárójelbe tettem. A reprezentációt a következőképpen értelmezhetjük: A vizsgált elemek intenzitási jegye minden megnyilatkozásban jelen kell, hogy legyen, így ezt tekinthetjük a magjelentésnek. Azt is megállapíthatjuk, hogy bizonyos esetekben kizárólag ez a komponens vesz részt a megnyilatkozás interpretálásában (vö. Szabó 2016: 168). Ezt láttuk az (56a-b) és az (57b) alatti példák esetében. Ugyanakkor az, hogy az emotív tartalomtól fakadó affektivitás vagy a negatív értékelő tartalom ugyancsak részt vesz-e az interpretációban, az az aktuális használat függvénye. Amint azt a (56c-d) és az (57c) alatti példáknál láttuk, az affektív komponens aktív, és azon interperszonális funkció alapul. Más esetben, így az (57a) alatti, valamint a *rohadt hideg víz* kifejezésekhez hasonló példák rendre negatív értékelést fejeznek ki a fokozó elem negatív szemantikai komponense miatt. Ezekben az esetekben tehát valamely (esetlegesen több) fakultatív komponensnek rá kell

épülnie a magjelentésre, részt kell vennie a megnyilatkozás jelentésének az interpretálásában.

A reprezentációban (félelem/düh/undor) szemantikai komponenst a fokozó elemek lexikai tételében nem jeleníték meg. Ez meglátásom szerint kizárólag ott tételezhető, ahol a vizsgált negatív elemek nem értékvesztő, és nem is értékvtó szerepűek. Tekintsük az alábbi példákat!

- (60) a. **Brutálisan** bánt a gyerekekkel egy veszprémi óvónő - Felfüggesztett börtönt kapott. [szuloklapja.hu]
- b. **mocskos** volt a ruhája olyan mocskos volt a ruhája, hogy a szénbányászok 12 órás műszak után sem néznek ki ennyire koszosan. [mobilarena.hu]

A fentebbihez hasonló esetekben nem ugyanazt a lexikai tételt kell megadnunk, mint amelyet ugyanezeknél az elemeknél fokozó szerepben. Az értékvesztő elemek esetében tehát ezt a komponenst a reprezentációban opcionálisként sem jelenítem meg.

Ezzel a megközelítési móddal lehetőségünk van azzal is elszámolni, hogy például a *rohadt alma* szókapcsolatban a *rohadt* elemnek a szótári reprezentációjában miért nem jelenik meg az intenzitási jegy. Azért nem, mert ebben a használatban a vizsgált elemnek más lexikai tételt tulajdonítunk. Amíg például – a fejezetben tárgyaltaknak megfelelően – a *rohadt jó* kapcsolatban levő elemhez a fokozók esetében felírt tételt adjuk meg, addig e használat esetében nem.

Továbbgondolásra érdemes kérdés az, hogy az alulspecifikált jelentés feltételezésében rejlő lehetőségeket ki lehetne-e úgy aknázni, hogy a vizsgált elemek fentebbi különböző használati esetében is csupán egyetlen lexikai tételt tennénk fel, és abból hoznánk létre az egyes funkciókat a magjelentésre épülő opcionális jelentésjegyek segítségével. Ez azt jelenti, hogy például a *rohadt jó*, a *rohadt alma* (az alma rohadtsága), valamint a *rohadt alma* (szitokszó) kifejezések esetében, mindhárom használatához azonos lexikai tételt adnánk meg. A problémára a disszertáció zárásáig a lexikai pragmatika elméleti keretében kielégítő megoldást nem találtam, azonban ebből a szempontból is ígéretesnek tartom az aktivációterjedés hálózati modell alkalmazását. A kérdést a fejezet későbbi részében (l. 4.6.2) e modell segítségével is tárgyalom.

Fontos rámutatnunk, hogy a vizsgált elemek fentebb vázolt interpretációs mechanizmusát további, a tágabb kontextus sajátosságai is befolyásolhatják. Tekintsük az

alábbi valós példát! (A példában a vizsgált szókapcsolatot dőlt, a kontextus elemezni kívánt sajátságait kövér szedéssel emelem ki.)

- (61) Ti meg igyekezzetek, mert *végre dög meleg* lesz és **hozza a buli a formáját**. [juniorrezervatum.hu]

Korábbi érvelésemnek megfelelően a példa kiemelt szókapcsolatát negatív értékelésként kellene értelmeznünk (l. újra az (57) alatti példát és annak elemzését). Ugyanakkor, a kapcsolat tágabb kontextusa (pl. a megnyilatkozás *végre* eleme) ezt az értelmezést lehetetlenné teszi. Ilyenkor tehát a negatív szemantikai komponens nem vehet részt az interpretációban. Arra, hogy ezzel a jelenséggel is elszámolhassunk, a különböző típusú enciklopédikus információk feltételezése ad lehetőséget (l. fentebb) (vö. Bibok 2016: 410). E megközelítési mód szerint a következőképpen írhatjuk le a fenti példa megértési mechanizmusát: A vizsgált szókapcsolat olyan kontextusba kerül, amely ellehetetleníti, hogy a fokozó elem negatív komponense a magjelentésre épüljön, annak ellenére, hogy azt az alaptag szemantikai tartalma nem akadályozná. Ilyenkor tehát más, így a kontextus lexikai elemeivel összefüggő, vagy akár kontextusfüggetlen, az általános világtudásunkon alapuló enciklopédikus információk juthatnak fontos szerephez.

A lexikai pragmatika keretét illetően, ahogyan azt Wilensky (1987: 20–22) részleteiben tárgyalja, felmerül a kérdés, hogyan számolhatunk el a feldolgozási időre vonatkozó sajátságokkal. A legutóbb tárgyalt esetben például a megértés során a „távolabbi” információk keresése csupán azt követően kezdődik meg, hogy a hallgató a szemantikai reprezentációban foglalt enciklopédikus információt egyeztette a kontextussal. Ez a mechanizmus szükségszerűen valamennyi feldolgozási időt kellene, hogy igényeljen. Hasonlóan, akár a relevancia- (vö. Wilson 2003: 284), akár az optimalitáselmélet (vö. Blutner 2004: 509–512) oldaláról közelítve a problémához, mind a releváns interpretáció eléréséhez, mint az ún. optimális értelem kidolgozásához feldolgozási időre volna szükség. Ugyanakkor, ahogyan arra több szerző (vö. pl. Wilensky 1987: 20, Gibbs 1994, Nemesi 2007, Forgács és mtsai 2012) is felhívja a figyelmet, a metaforák vagy az indirekt beszédaktusok megértése bizonyítottan nem igényel hosszabb feldolgozási folyamatot a nem metaforikus vagy a direkt jelentések megértésénél.

E megértési folyamatok pontos kidolgozása még nem történt meg (vö. Wilensky 1987: 21). Ugyanakkor magam – Wilson (2003: 283–287)-tel egyetértésben – úgy vélem, hogy a mentális lexikon aktivációterjedéses működésének a feltételezése lehetőséget adhat

a lexikai pragmatika elméleti alapvetéseinek a továbbfejlesztéséhez, így módon a vizsgált jelenségek adekvátabb lexikai pragmatikai leírásához.

A következő dolgozatrészben a mentális lexikon aktivációterjedéses hálózati reprezentációjának a lehetőségét tárgyalom részletesen, és az értékvesztés problémáját ebben a megközelítésben is körüljáróm. Célom, hogy megmutassam, a mentális lexikon aktivációterjedéses működési mechanizmusának a feltételezése fontos adalékul szolgálhat a lexikai pragmatikai megközelítéshez, ezáltal a vizsgált fokozó elemek nyelvi működésének a pontosabb leírásához.

4.6.2 Az értékvesztést mutató elemek vizsgálata az aktivációterjedéses hálózati elméleti keretben

Westbury és Hollis (2007: 28), valamint De Boot (2012) alapján, a közeljövő kutatásaiban szintézisre van szükség ahhoz, hogy a nyelv (és a mentális lexikon) működését jobban megérthessük. Ez azt jelenti, hogy egyrészt az eddig egymástól függetlenül elemzett jelenségeket is szintetizálni kell, másrészt a beszéddel és gondolkodással kapcsolatosan nélkülözhetetlen olyan komplex modell(ek) megalkotása, amely(ek) figyelembe veszi(k) az aktuális kutatások eredményeit.

Az, hogy a mentális lexikont hogyan modellezzük, szorosan összefügg a pszicholingvisztika és a kognitív tudomány legfontosabb aktuális vitakérdéseivel (vö. Pléh és mtsai 2008: 814). A kognitív funkciók szerveződésére vonatkozóan három alapvető megközelítési módot különböztethetünk meg: a moduláris, az interakciós, valamint a konnekcionista felfogást (vö. Pléh és mtsai 2008: 814). A moduláris megközelítés úgy tekint a kognitív rendszerre, mint részben izolált komponensekből (másképpen: modulokból) álló szerveződésre, amelyben a modulok autonómak és információsan zártak (vö. Fodor (1983), idézik Pléh és mtsai (2008: 814)). A modell felépítéséből következik, hogy – mivel az bottom-up elrendezésű – „a meglevő tudás” csupán egy későbbi szakaszban léphet be a feldolgozási folyamatba. Az interakciós modellben ezzel szemben az egyes szintek befolyásolhatják egymás működését, és „sorrendjük is variálódhat a feldolgozandó anyagtól függően” (Pléh és mtsai 2008: 814). Az interakciós modellnél a konnekcionista megközelítés még radikálisabb álláspontot képvisel. Amíg ugyanis az előbbi – bár megengedi az egyes szintek közötti interakciót – nem veti el a hierarchikus szintek létét, az utóbbi „az egyes információforrások folyamatos és egyidejű

felhasználásaként képzele el a nyelvi működéseket, amelyek egy, az idegrendszerhez hasonló hálózatban mennek végbe” (Pléh és mtsai 2008: 814). A konnekcionista megközelítés úgy tekint a mentális rendszerre, mint egymáshoz különböző erősséggel kapcsolódó egységek hálózatára (vö. Rumelhart és mtsai 1986). A kognitív aktivitást ebben a modellben – szekvenciális feldolgozás helyett – párhuzamosan terjedő aktivációkként képzeljük el.

A mentális lexikon hálózatos modellezését támogatják azok a kutatási eredmények, amelyek a legkülönbözőbb hálózatos szerveződésekben szabályszerű mintázatok létrejöttére mutatnak rá. Barabási a következőképpen fogalmaz: „Hálózatok mindenhol vannak. Az agy axonok által összekötött idegsejtek hálózata, maguk a sejtek pedig biokémiai reakciók által összekötött molekulák hálózatai. A társadalmak szintén hálózatok [...]. A hálózatok átjárják a technológiát is: az internet, az elektromos hálózatok, valamint a szállítási rendszerek csupán néhány példa erre. Még a nyelv is, amit gondolataink közvetítésre használunk, önmagában véve nem más, mint szintaktikai kapcsolatokkal összekötött szavak hálózata” (Barabási 2006: 1298). A 1990-es évek végén jelennek meg az első olyan tudományos megállapítások, amelyek az eddig egymástól függetlennek tűnő hálózatos rendszereknek (pl. úthálózat, világháló, emberi kapcsolatrendszerek stb.) közös tulajdonságaira irányítják a figyelmet, továbbá arra, hogy ezek a tulajdonságok matematikailag leírhatók és elemezhetők (Barrat és mtsai 2008, Watts 1999, 2004, idézi Kovács (2013: 61)). A különböző hálózattípusok a következő legfontosabb közös sajátságokat mutatják: 1) skálafüggetlenek, 2) az előzővel összefüggésben robusztusak, 3) kisvilág-karakterük van, 4) bennük a kapcsolatok irányítottak és nem azonos súlyúak (vö. Fóris 2010: 124–126). A skálafüggetlenség azt jelenti, hogy a hálózatban az elemek fokszámeloszlása a hatványfüggvényt követi (vö. Fóris 2010: 124–125). Ez azt jelenti, hogy van egy, a hálózat méretéhez képest kis számú elemhalmaz, amelynek nagyon sok kapcsolata van, majd a kapcsolatok számának csökkenésével növekszik az adott kapcsolatmennyiséggel rendelkező elemek száma (vö. Fóris 2010: 125).⁴¹ A skálafüggetlen tulajdonságból következik a robusztusság. Egy hálózat akkor robusztus, ha a környezet változásaira kis érzékenységgel reagál (vö. Kovács 2013: 62). Mivel a skálafüggetlen hálózatban csupán kevés elem rendelkezik kiemelkedően sok kapcsolattal, a hálózatot érő random támadások viszonylagosan „kis veszélyt” jelentenek a teljes hálózatra nézve. E

⁴¹ Hatványfüggvény írja le a világ számtalan, egymással össze nem függő jelenségét, így többek között a szógyakoriságot, a tudományos cikkek idézettségének a számát, a könyvek eladási listáját, a telefonhívások eloszlását az USA-ban, a földrengések nagyságát, a Hold krátereinek átmérőjét, a háborúk intenzitását és a családnevek gyakoriságát is (vö. Newman (2005), idézi Kovács (2013: 66)).

sajátságok mellett a valós hálózatok kisvilág-karaktert mutatnak, azaz viszonylag kicsi az átlagos távolság az egyes elemek között (vö. Watts és Strogatz (1998), idézi Kovács (2013: 71–72). Másképpen, átlagosan kevesebb lépésben juthatunk el egy hálózati elemtől egy másik elemig, mint a mesterségesen generált, véletlen hálózatokban.⁴² Végezetül, a valós hálózatokban a kapcsolatok gyakorta irányítottak és nem azonos súlyúak (vö. Kovács 2013: 62). Irányítottak, hiszen számos kapcsolat csupán az egyik hálózati pontból mutat a másikra, fordítva nem. Nem egyenrangúak pedig azért, mert vannak erősebb és gyengébb kapcsolatok.

Több vizsgálat (vö. Fóris 2010: 126–127, Gangl 2012: 27–28, Imrényi 2012, Kovács 2013: 13, 97) eredményei arra mutatnak, hogy a mentális lexikon hasonló hálózatos elképzelése helytálló alternatívát kínálhat, és olyan problémákra is megoldást nyújthat, amelyekkel eddig nem sikerült maradéktalanul elszámolni (l. 4.6.1). Jelenleg nincs egységes elképzelés arra vonatkozóan, hogy a konceptuális szint külön kezelendő-e, létezik-e szemantikai szint, amennyiben igen, az nyelvfüggetlen vagy nyelvfüggő-e, illetve hogyan kapcsoljuk ezekhez az elképzelésekhez a prototipikalitás jelenségét (vö. Kovács 2013: 197). Amennyiben a mentális lexikont mint komplex hálózatot tekintjük, az alhálózatok közötti átfedésekkel, valamint azok különböző szintű aktiválásával, akkor „lehetséges, hogy a fent említett kérdéseket és megközelítéseket teljesen más szemszögből kell vizsgálni” (Kovács 2013: 197). Így például, amint azt a (4.6.1)-ben részletesebben tárgyaltam, a legtöbb melléknévhez eltérő képzetek társulnak attól függően, hogy mely entitást minősítik. Az igéket érintő hasonló problémákkal kapcsolatban ugyancsak számos dolgotra utaltam.

Úgy vélem, hogy a mentális lexikon aktivációterjedéses hálózatként való tételezése lehetőséget adhat arra, hogy ezeket a kontextuális sajátságokat a modellbe építhessük, és a kontextuális és lexikai szemantikai sajátságok együttes működésével elszámolhassunk. Azt gondolom továbbá, hogy ez a megközelítési mód összeegyeztethető azzal a lexikai pragmatikai kerettel is, amelyben a lexémák aktuális jelentése azok alulspecifikált szemantikai reprezentációjából, az enciklopédikus tudás hozzájárulásával jön létre, és amellyel a disszertációban vizsgált fokozó elemek interpretációs mechanizmusát is modelleztem (l. 4.6.1) (vö. Blutner 1998, 2004, Bibok 2014, 2016). E kontextuális sajátságok és az enciklopédikus tudás az aktivációterjedéses hálózatos modellben ígéretes

⁴² Az emberek közötti távolság és a kapcsolatok kisvilágkaraktere ma már részben mérhető a közösségi oldalak adatait használva. Backstrom és mtsai (2011) mérései alapján például a Facebook-felhasználók átlagos távolsága egymástól 4,74 lépés (vö. Kovács 2013: 72).

módon kapcsolhatóak be az interpretációs mechanizmusba, így lehetségessé válhat annak pontosabb és részletesebb kidolgozása.

A mentális lexikon hálózatos reprezentációjának lehetősége a jelen disszertáció kutatási témájával összefüggésben két szempontból is különösen figyelemreméltó. Egyrészt, ahogyan arra például Kovács (2013: 201) felhívja a figyelmet, e modellek hozzájárulhatnak a számítógépek számára értelmezhető és kezelhető fogalmi rendszerek kialakításához. Így például az intelligens (szemantikai alapú) keresőszoftverek fejlesztésében, az automatikus fordítói rendszerek fejlesztésében, szövegbányászati módszerek kialakításában és alkalmazásában, valamint szövegtörzsek elemzésének hatékonyabbá tételében is jelentős segítséget nyújthatnak (mindezekről részletesebben l. Kovács 2013: 14–15, 198–203). Mindezzel összefüggésben lehetővé válhat a mentális lexikon működésének informatikai eszközökkel történő pontosabb modellezése (vö. Kovács 2013: 15). Másrészt, Lengyel (2012: 16) alapján a szóasszociációs technika – amely a mentális lexikon hálózatos szerveződésének legfontosabb vizsgálati módszere (részletesen l. lentebb a jelen alfejezetben, valamint 6.3.3) – alkalmazásának kísérletei újabban megjelentek a piackutatás, így a termékek, terméktulajdonságok leírása és a márkaimázs területén. Tekintettel arra, hogy a szentimentelemzés eredményeinek egyik legfontosabb felhasználási területe a termékvélemények feltárása, úgy vélem, ez az összefüggés is a választott elméleti keret alkalmazását támogatja.

Az aktivációterjedés modellek jelenleg a legelfogadottabb hálózatos reprezentációs megoldások közé tartoznak (vö. Kovács 2013: 32–33). E modellek közös jellemzője, hogy a mentális lexikont csomópontok és az azokat összekötő kapcsolatok hálózataként írják le. A hálózat elemei a nyelvi produkció és interpretáció során különböző aktivációs szintekre kerülnek. Produkció során például a *doktor* szó hallásakor eltérő mértékben aktiválódnak a *fogorvos*, a *kórház*, a *láz*, valamint az *ágy* és a *bölcső* szavak (vö. Pléh 2008: 828). A sajátságát a produkcióra vonatkozóan ún. szóasszociációs adatfelvételek eredményei (pl. az ún. Agykapocs-kutatás eredményei, vö. Kovács 2013), az interpretációra vonatkozóan pedig a megértési időt vizsgáló (például az előfeszítés szerepére irányuló) vizsgálatok eredményei támasztják alá (e vizsgálatokról és azok eredményeiről részletesebben l. Pléh és mtsai 2008: 803–828).

Az aktivációterjedés modell egy további fontos alapvetése, hogy egyes aktiválási folyamatok erősebbek, míg mások gyengébbek, így nem aktiváljuk egyszerre az összes lehetséges kapcsolatot (vö. Lengyel 2012: 15, Huszár 2005: 63–64). Emellett az aktiválás

közben blokkolási folyamatok is tételezhetők, amelyek akadályozzák az aktiváció terjedését (vö. Kiss (1975) és Gósy (2005), idézi Kovács (2013: 33)).

A mentális lexikon hálózatos modellezésével kapcsolatosan megállapítható, hogy a modellek keveset szólnak a hálózati csomópontok belső felépítéséről. Arról, hogy hogyan, milyen komponensekből épülhet fel a csomópont, valamint azok hogyan és milyen szerepet játszanak a terjedő aktivációban. Levelt (1989: 182) szerint a mentális lexikon egységei a jelentést, a szintaktikai, a morfológiai, valamint a fonológiai információkat tartalmazzák. Ugyanakkor megjegyzi (1989: 183), hogy az egyes elemek valószínűleg további információkat is tárolnak, úgymint például pragmatikai, stiláris és affektív jegyeket. Ezek viszonyát azonban az említett négy komponenssel nem dolgozza ki. Levelt észrevételéhez illeszkedik az a megközelítés is, hogy a mentális lexikonban a világról alkotott enciklopédikus tudásunk is tárolódik (vö. Pléh és mtsai 2008: 79, Kovács 2013: 26).

Bár a nyelvtudományban a mentális lexikon különböző típusú sajátosságait, illetve hálózatként felfogható jelenségeit legtöbbször egymástól függetlenül vizsgálják, egyre több dolgozat mutat rá azok összefüggéseire (vö. pl. Aitchison 2003: 32–33, Kovács 2013: 186–188). A hazai vizsgálatok közül Kovács Fóris (2009), Balaskó (2010), valamint Benczes (2011) kutatási eredményeit emeli ki. Közülük Balaskó (2010: 103–106) a szójelentés meghatározásában látja a hálózatok szerepét: a *helyzet* szót korpuszokban vizsgálva arra a megállapításra jut, hogy a szavak jelentését „egymásba ágyazott hálózatoknak a dinamikus kapcsolata határozza meg” (uő 107). Ennek megfelelően tehát a nyelvi produkció során (például egy szóasszociációs adatfelvételkor) a lexikai, a grammatikai, a szemantikai és a pragmatikai háló együttes, dinamikus működése, interakciója feltételezhető (vö. Balaskó 2010: 106–107). Ennek az egyik látványos következménye lehet az a jelenség, hogy a fáradtság, illetve a csökkenő koncentráció hatására az asszociációs vizsgálatokban a fonetikai asszociációk kerülnek túlsúlyba (vö. Spitzer (1996), idézi Kovács (2013: 193)).

A hálózati pontok felépítését illetően, Baayen (2007)-re támaszkodva Kovács (2013: 33) azokat is mikrohálózatokként, másképpen a különböző információk szervezett egységeként képzei el.⁴³ Ily módon az egyes hálózati csomópontokat egységen kívüli és belüli kapcsolatok egyaránt jellemzik. Egységen belüli kapcsolat van például a szó különböző ragozott alakjai között, míg más egységek között például szemantikai (*hajó* –

⁴³ „Each node is itself a network that learns causes from its child nodes and forms beliefs that it propagates to its parent nodes” (Baayen (2007: 97), idézi Kovács (2013: 33)).

csónak) és fonetikai (*pár – szár*) kapcsolatok jöhetnek létre (vö. Kovács 2013: 26). Mindez azt jelenti tehát, hogy nem csupán az egyes lexémákat reprezentáló hálózati pontok között léteznek a kapcsolatok, hanem azt is, hogy a kapcsolat bármely mikro- és makrohálózati komponens között létrejöhet. Ily módon épülhet fel a többdimenziós mentális lexikon (vö. Lengyel 2012: 15, Kovács 2013: 188).⁴⁴

A fejezet további részében a negatív emotív fokozó elemek nyelvi viselkedését vizsgálom az aktivációterjedés hálózati modellben.

Levelt (1989: 182), Baayen (2007: 97) és Kovács (2013: 26, 194) megközelítését elfogadva, a mentális lexikon egyes pontjai, amelyek a szavakat reprezentálják, „mikrohálózatként” képzelhetők el. E hálózat részeit különböző típusú információk, köztük az adott szó szemantikai predikátumai teszik ki. A lexikai pragmatikai kerettel (l. 4.6.1) összegegyeztethető módon, a negatív fokozó elemek esetében az [intenzív], az [affektív], valamint a [negatív] a jelentéskomponenseknek kell itt csomópontok formájában megjeleníteniük.⁴⁵ Mindez azt jelenti, hogy az adott lexéma önmagában is egy hálózati pontot alkot, ugyanakkor a fentebbi komponenseivel maga is mikrohálózatot hoz létre. A makrohálózat további pontjaihoz a lexéma egy egységként, valamint annak bármely alhálózati pontja, tehát az egyes komponensek is kapcsolódhatnak. E feltevés összhangban áll Levelt (1989: 183) már ismertetett megközelítési módjával, amely alapján a hálózat egyes elemei valószínűleg az alapvető komponenseken túl további, pragmatikai és stiláris információkat, valamint affektív jegyeket is tárolnak.

A mentális lexikont Lengyel (2012: 15) és Kovács (2013: 188) alapján komplex, többdimenziós hálózatként képzelem el, amelyben a mikro- és makrohálózatok pontjai számos hasonló alhálózati ponthoz kapcsolódnak. Ennek megfelelően tehát a vizsgált fokozó elemek bekapcsolódnak – többek között – a kollokációs hálózatba.

Az interpretációs mechanizmus az aktivációterjedés hálózati elméleti keretben a következőképpen írható le: Az aktivációterjedés modell alapján a hálózati kapcsolatok alapvetően meghatározzák, hogy az aktiváció mely hálózati komponensekre terjedhet ki. Ily módon, amikor a vizsgált negatív elemek pozitív vagy negatív szentimentértékkel rendelkező alaptagot módosítanak, e kollokációs sajátság miatt a [negatív] szemantikai

⁴⁴ Hasonlóan érvel Klégr (2013: 19) az enantioszémia szemantikai kapcsolataira vonatkozóan. A szerző alapján, enantioszémia esetén az adott lexémának ún. belső és külső szemantikai kapcsolatai vannak. A belső kapcsolat az, amely az adott lexéma két, egymással ellentétes jelentése között áll fenn, a külső kapcsolatok pedig azok, amelyek a lexémát a lexikon további elemeinek a szemantikai tartalmához fűzik (pl. hiponima-, meronima- és szinonimarelációk).

⁴⁵ A jelen probléma szempontjából nem releváns esetleges komponenseket (pl. fonológiai, morfológiai stb.) itt nem tárgyalom.

hálózati komponens nem aktiválódhat. Ez a tartalom ugyanis nem egyeztethető össze az adott kontextusban (kollokációban) betöltött szerepével, az aktiváció tehát erre a csomópontokra nem terjedhet ki (vö. az (56) alatti példákat és azok elemzését fentebb, 4.6.1).

Amennyiben a fokozó elem a szentimentérték szempontjából semleges alaptagot módosít, megfelelő kontextuális feltételek mellett (l. 4.6.1) a fokozó elem által a szókapcsolat nyerhet negatív értékelő jelentést. Ebben az esetben a lexéma negatív minőségi jegye is aktiválódhat, mivel azt a vizsgált elemmel kollokálódó alaptag semleges értéke nem akadályozza meg (l. (57a) alatti példát és elemzését fentebb, 4.6.1).

Mindez összeegyeztethető a korábban feltárt korlátozási sajátsággal (l. 4.6.1), mely szerint, amennyiben az alaptaggal jelölt tulajdonság az adott targetre nézve negatív sajátság is lehet, a negatív fokozókkal együtt a megnyilatkozás nem tud pozitív értékelést kifejezni, kizárólag negatívot (vö. *rohadt hideg víz*). Ugyanakkor a korlátozás nem érvényes, ha az alaptaggal jelölt tulajdonság az adott targetre nézve csak pozitív sajátság lehet (vö. *rohadt gyors internet*). Az aktivációterjedéses hálózati modellt alkalmazva azt mondhatjuk, hogy amíg az utóbbi esetben a kontextuális sajátságok nem engedik meg a [negatív] komponens aktivációját, addig az előbbi esetben ilyen akadály nincs, ezért az szükségszerűen be is következik.

Amint azt a (61) alatti példa kapcsán tárgyaltam (l. 4.6.1), az itt vázolt interpretációs mechanizmust további kontextuális sajátság felülírhatja. Ebben az esetben a tágabb kontextuális sajátság megakadályozza a negatív minőségi jegy aktivációját a hálózati ponton belül, és így csupán az [intenzív] szemantikai komponens kerül magasabb aktivációs szintre.⁴⁶

Amint azt korábban tárgyaltam (l. 4.6.1), a lexikai pragmatikai elméleti keretben a negatív emotív elemek fokozó szerepű használatával úgy tudtam elszámolni, hogy azoknál önálló lexikai tételt adtam meg. Ugyanakkor felvettem annak a lehetőségét, hogy a vizsgált elemek fentebbi különböző használatai esetében is csupán egyetlen lexikai tételt tennénk fel, és egyetlen magjelentésből hoznánk létre az egyes funkciókat. Az alkalmazott lexikai pragmatikai megközelítésben ezidáig nem tudtam kielégítő megoldást adni, azonban ígéretesnek tartom az aktivációterjedéses hálózati modell alkalmazását. A modell lehetőséget ad ugyanis arra, hogy például a már említett *rohadt jó*, *rohadt alma* (az alma rohadt tulajdonsága), és *rohadt alma* (szitokszó) kifejezések esetében, mindhárom

⁴⁶ A bemutatott megközelítési mód Paradis (2001: 48) eredményeivel is összeegyeztethető, aki ugyancsak amellett érvel, hogy a nyelvi kifejezések jelentése mindig a kognitív rendszer konceptuális egységeinek adott kontextusbeli aktivációján keresztül jön létre.

módosítói használatnál ugyanazokat a hálózati komponenseket tegyük fel a lexémában, tehát az [intenzív], az [affektív], a [negatív] és a [félelem/düh/undor] komponenseket. Azt, hogy ezek közül mely komponensek jutnak szerephez az adott interpretációs folyamatban, mindig a terjedő aktiváció fogja meghatározni az aktuális kontextus függvényében.

Bár a fejezet célja az, hogy az értékvesztő elemek interpretálási mechanizmusát vizsgáljam, szeretnék röviden kitérni az aktivációterjedéses hálózati modellt illetően az értékvtó elemek nyelvi működésére is.

Amint a (61) alatti példa elemzésénél is láttuk, a negatív fokozó elemek interpretálásakor jelentős szerepe lehet a tágabb kontextusnak is. Az értékvtást illetően, ehhez hasonló működési mechanizmussal az aktuális doménnek is kiemelkedő szerep jut. Bizonyos domének ugyanis tipikusan megengedik az értékvtást, míg mások nem.⁴⁷ Éppen ezért azt, hogy valamely vizsgált elemet szentimentkifejezésként értékvtóként interpretáljuk-e, alapvetően meghatározza az, hogy az adott megnyilatkozás milyen szövegdoménben szerepel. Tekintsük az alábbi példát!

(62) A *brutális* énekes háza előtt ezek gyűltek össze.

Véleményem szerint ez a megnyilatkozás, amennyiben például reklámszövegben szerepel, megengedi a vizsgált elem interpretálását értékvtó funkcióban is. Ugyanakkor, amennyiben például bűnügyi hírszövegben jelenik meg, ez az olvasat tipikusan nem lehetséges. A probléma megközelítését illetően tehát egyetértek Kovács (2013: 189) véleményével, aki szerint az aktivációterjedés során az ún. doménhálónak kulcsszerepet kell kapnia.

Az érvelésem összeegyeztethető Klégr (2013: 15) megállapításával is, amely szerint az enantioszémia kapcsán meg kell különböztetnünk egy a kollokációs sajátságokkal összefüggésben jelentkező típust (a szerző elnevezése alapján: *collocational enantioseme*). A szerző alapján, az enantioszémia egyik típusánál az, hogy az aktuális megnyilatkozásban mely jelentés érvényesül, attól függ, hogy az adott elem milyen szemantikai tartalmú kollokátorhoz kapcsolódik, valamint, hogy mi szerepel a tágabb kontextusban. A szelekciós mechanizmusban ez alapján a megfelelő jelentés aktiválódik, amely

⁴⁷ Egy korábbi dolgozatomban (Szabó 2015c) a *brutális* szó előfordulási gyakoriságát vizsgáltam negatívan értékelő elemként és értékvtóként, nyolc különböző domén szövegeiben. Az eredményeket nem szerepeltetem a jelen disszertációban, azonban kiemelném, hogy amíg például az általam elemzett gazdasági szövegekben az előfordulások egyharmadában a *brutális* értékvtó funkcióban állt, addig politikai tárgyú szövegekben ez a használat gyakorlatilag kizárt.

megközelítés ugyancsak az aktivációterjedéses hálózati megközelítést támogatja, ebben az esetben az enantioszémia problémaköre oldaláról. Mindez megmagyarázhatja azt is, miért nem vezet interpretálási problémákhoz az egy adott lexémához tartozó, egymással ellentétes jelentések léte (vö. Schmelev 2012: 838).

A jelen fejezetben tárgyalt aktivációterjedéses hálózati modell alkalmazását – amint arra már a fejezet elején utaltam – több fontos vizsgálati eredmény is alátámasztani látszik. A modell előnyei közül az alábbiakban egyet, amely a jelen vizsgálat szempontjából kiemelkedően fontos, részletesebben is tárgyalok.

Amint arra számos szerző felhívja a figyelmet (vö. pl. Wilensky 1987: 20, Gibbs 1994, Nemesi 2007, Forgács és mtsai 2012), a nem literális jelentések megértése bizonyítottan nem igényel hosszabb feldolgozási folyamatot a literális jelentések megértésénél. Forgács és munkatársai (2012) például a főnév–főnév-kapcsolatok interpretációján keresztül azt vizsgálták, hogy mely típusú kifejezések megértéséhez van szükség nagyobb feldolgozási időre: a konvencionális literális, a konvencionális metaforikus, a nem konvencionális literális vagy a nem konvencionális metaforikus kapcsolatnak. A vizsgálat eredményei alapján a metaforák és a literális jelentések megértése közel azonos feldolgozási időt követelt a vizsgálati személyektől. Valódi eltérést csupán a kifejezések újszerűsége, a konvencionálistól eltérő volta eredményezett. A szerzők (Forgács és mtsai 2012: 1437) fMRI vizsgálatokkal azt is bizonyították, hogy a konvencionális kifejezések (mind a metaforikusak, mind a literálisak) nem ugyanolyan BOLD-aktivációt (*BOLD: blood-oxygen-level-dependent*) eredményeznek a két agyféltekében, mint a nem konvencionális kifejezések (a metaforikusak és a literálisak egyaránt).

Meglátásom szerint, e sajátságot hatékonyan magyarázhatja a mentális lexikon aktivációterjedéssel működő hálózatos felépítése. A modell megengedi ugyanis, hogy azt feltételezzük, metaforikus nyelvhasználat esetében az aktiváció csupán a megengedett irányba terjed tovább (vö. Caramazza (1999), idézi Kovács (2013: 187), Huszár (2005: 63–64)), vagy legalábbis egyszerre több jelentés aktivációja is bekövetkezik, azok között nincs feltétlenül sorrendi hierarchia.

Érdemes megjegyezni, hogy a nem literális jelentések megértésének e módja összhangban áll például az ún. vezérjelentés elméletével (vö. *graded salience hypothesis*, Giora 1997, 2003), amelyben – ellentétben például a grice-i elmélettel (1975), amely a metaforikus kifejezések megértését két lépésben tételezi – nem a figurativitás, hanem egyéb sajátságok (prototipikalitás, familiaritás, konvencionalitás és gyakoriság) határozzák

meg alapvetően a megértési mechanizmust (a problémáról részletesebben l. többek között Wilensky 1987: 20–22, Wilson 2013: 283–287).

A jelen fejezetben azt tárgyaltam részletesen, hogyan alkalmazható a mentális lexikon aktivációterjedéses hálózati elméleti kerete a negatív fokozó elemek nyelvi viselkedésének a modellezésére. Amint azt a vizsgálat során igyekeztem megmutatni, ez a megközelítési mód összeegyeztethető a megelőző alfejezetben tárgyalt lexikai pragmatikai elméleti kerettel, sőt bizonyos problémák vonatkozásában a két modell hatékonyan kiegészítheti egymást. Az aktivációterjedéses hálózati megközelítés további fontos előnyeként említettem az, hogy lehetőséget ad a vizsgált nyelvi jelenségnek a számítógépes értékeléselemzés számára is értelmezhető és kezelhető modellezésére.

A jelen alfejezetben áttekintett szakirodalmi eredmények, valamint az itt tett megállapításaim fontos alapot szolgáltattak a szóasszociációs vizsgálatomhoz (l. 6.3.3).

5. A szentimentelemzési feladat részletes áttekintése

Az értekezés jelen fejezetében a nyelvtechnológia tárgykörébe tartozó számítógépes értékeléselemzést, másképpen a szentimentelemzést tárgyalom részletesen.

Először tisztázom a feladat tárgyát és fontosságát, majd áttekintem a szentimentelemzéssel kapcsolatos nemzetközi és hazai kutatásokat. Figyelembe véve a hazai kutatások csekély számát, a szentimentelemzés legfontosabb problémáit, módszereit és eszközeit a nemzetközi kutatásokkal kapcsolatban ismertetem. A hazai kutatásokat kimerítő jelleggel tárgyalom. A disszertáció fókuszában álló szentimentérték-módosulási jelenséget a szentimentelemzés feladatköre szempontjából külön alfejezetben veszem górcső alá.

Az áttekintés az adatbázis-építési feladatok (l. 6.3), valamint az elemző eszköz elkészítésének (l. 7) az alapját képezi.

5.1 A szentimentelemzés tárgya és fontossága

Amint azt az értekezés bevezető fejezetében már tárgyaltam, a szentimentelemzés a számítógépes nyelvészet, azon belül a számítógépes tartalomelemzés egy részterülete. Arra irányul, hogy a digitalizált szövegek szemantikai tartalmából kivonatolja és feldolgozza azokat az információkat, amelyek értékelést fejeznek ki, meghatározza ezeknek az értékeknek a típusát és esetlegesen az intenzitásuk mértékét, valamint, hogy megállapítsa az értékelés tárgyát (számítógépes nyelvészeti terminussal: *targetét*) (vö. Szabó–Vincze 2015). A feladatot korábban egy példával szemléltettem (16) alatt, azt (63) alatt megismétlem.

(63) [targ A szentimentelemzés] [szent [int nagyon] érdekes] feladat.

Amint azt a példa sematikus elemzése is mutatja, a jelen esetben a *szentimentelemzés* a target, amelyről egy pozitív értékelés fogalmazódik meg. Ezt a minősítést a *nagyon* elem fokozza.

Bár a magyar nyelvű szakirodalomban az irodalmi tételek túlnyomó többsége a *szentimentelemzés* terminust alkalmazza a feladat megnevezésére, ritkább esetben a

véleménybányászat, a *véleménykinyerés* és a *véleménykivonatolás* kifejezésekkel is találkozhatunk.⁴⁸ E fogalmak ugyanakkor rendre a nyelvtechnológia szakterminusainak tekinthetők, s a feladat tárgyát az elméleti kutatásokban a *nyelvi értékelés* kifejezéssel fedhetjük le.

A szentimentelemzés a nyelvtechnológiai kutatások palettáján még viszonylag új keletűnek tekinthető, ugyanakkor napjainkban a nemzetközi kutatásban és fejlesztésben is egyre nagyobb figyelmet kap (vö. Pang–Lee 2008, Liu 2012). A témával számtalan nemzetközi publikáció foglalkozik (pl. Pang és mtsai 2002, Turney 2002, Hu–Liu 2004, Wilson és mtsai 2005, Kanayama–Nasukawa 2006, Qiu és mtsai 2009, Feldman és mtsai 2010, Loughran–McDonald 2011, Liu 2012, Ruppenhofer–Rehbein 2012, Young–Soroka 2012, Socher és mtsai 2013), és megjelentek a hazai, kifejezetten a magyar nyelvű szövegek szentimentelemzésére fókuszáló dolgozatok is (vö. Berend–Farkas 2008, Miháltz 2010, 2013, Hangya és mtsai 2015).

A szentimentelemzés népszerűvé válásának oka egyrészt a feladat tudománystechnológiai kihívásaiban, másrészt az eredmények – javarészt gazdasági – hasznosítási lehetőségeiben keresendő. Mindenekelőtt azonban fontos ráirányítani a figyelmet arra, hogy az online kommunikációnak köszönhetően ma olyan hatalmas mennyiségű, digitális formájú – így számítógépes programokkal elemezhető – szövegadat áll a rendelkezésünkre, amely korábban sem a kutatási, sem a gazdasági oldal számára nem volt elérhető (vö. Витехновский 2013: 34). Például, számos hivatalos dokumentum szabadon hozzáférhető, a legtöbb sajtótermék rendelkezik honlappal, és a kulturális termékek jelentős része is elérhető digitalizált formátumban (pl. Magyar Elektronikus Könyvtár). A webkettes technológiák megjelenésével a személyes tartalmak is elárasztották az internetet: megjelentek a blogok, valamint az olyan közösségi oldalak (pl. a Twitter és a Facebook), amelyeken a felhasználók rendszeresen megosztják tapasztalataikat, véleményüket (vö. Прохоров–Керимов 2012).

Ami a tudományos és technológiai kihívásokat illeti, a digitális formájú szöveges tartalmak ilyen mértékű megnövekedése számtalan új lehetőséget nyit a számítógépes nyelvfeldolgozásban, ami a természetesnyelv-feldolgozással kapcsolatos kutatási és fejlesztési tevékenység kibontakozását segíti. A gazdasági hasznosíthatósággal kapcsolatban elmondható, hogy a webes információáramlás egyre jelentékenyebb szövegbányászati igényt támaszt a nyelvtechnológiai fejlesztések felé, és – annak jelentős

⁴⁸ Az angol nyelvű irodalom általában a *sentiment analysis*, az *opinion mining*, az *opinion extraction*, a *sentiment mining* és a *review mining* megnevezéseket alkalmazza (vö. Liu 2012: 7).

hasznosíthatósága okán – a felhasználói szövegek különböző szempontú és célú elemzésének hatékony megoldását sürgeti (vö. Liu 2012: 118–119).

A szentimentelemzési feladat jelentőségének megértéséhez magát a nyelvi értékelés jelentőségét kell fontolóra vennünk. Ahogyan Liu (2012: 8) fogalmaz, az értékítéletek majdnem minden emberi cselekvésben kulcsszerepet töltenek be, azoknak mintegy mozgatórugóiként szolgálnak. Ez a gyakorlatban azt jelenti, hogy azokban az esetekben, amikor döntéseket kell hoznunk, igyekszünk megismerni mások véleményét, tapasztalatait az adott kérdést, problémát illetően. Ugyanakkor, az adott cél szempontjából releváns online tartalmak kimerítő manuális megkeresése és értelmezése, értékelése kivitelezhetetlen feladat, ezért ehhez hatékony automatikus támogatásra van szükség (bővebben l. Liu 2012: 118–119). Ugyanakkor, a gépi támogatás nem csupán emberi erőforrások által feldolgozhatatlan szövegmennyiség, hanem az objektivitás miatt is egyre fontosabb kérdéssé vált napjainkra. A tartalomelemzés hagyományos módszertana szerint a szövegek szemantikai információit a feldolgozók előre definiált szempontok szerint manuálisan értékelték (vö. Weber 1990, Krippendorff 2004). Tekintettel arra, hogy a manuális elemzés jobbra szubjektív ítéletekre hagyatkozik, gyakran a gyakorlott tartalomelemzők is tévesen ítélték meg a szöveges tartalmakat (vö. Hart és mtsai 2013).

A szentimentelemzés eredményeinek a jelenlegi legfontosabb, gazdasági felhasználási területei a következők: a tőzsdeindex mozgásának előrejelzése; a fogyasztói csoport benyomásai, tapasztalatai bizonyos termékek és szolgáltatások vonatkozásában; politikusokkal, politikai eseményekkel kapcsolatos attitűdök felmérése és választási előrejelzések (vö. pl. Antweiler–Frank 2004, Engelberg 2008, Li 2008, Tetlock és mtsai 2008, Loughran–McDonald 2011). A jelentős gyakorlati haszon okán ma már számos információs technológiai nagyvállalat működtet saját szentimentelemző rendszert, így például a Microsoft vagy a Google (vö. Liu 2012: 9).

A gazdasági hasznosíthatóság, valamint a feladat nyelvészeti és technológiai kihívásaival összefüggésben ma már nem csupán az angol, de számos egyéb nyelv vonatkozásában foglalkoznak szentimentelemzéssel a kutatók és fejlesztők, így például a német (vö. Kim–Hovy 2006), a japán (vö. Kanayama–Nasukawa 2006), a kínai (vö. Hu és mtsai 2005) és az arab (vö. Shoukry–Rafea 2012) nyelvben. A leggyakrabban elemzett szövegtípusok között a különböző közösségi-média szövegeket találjuk, például a tweeteket, a blog- és a Facebook-bejegyzéseket (vö. pl. Godbole és mtsai 2007, Poursepanj és mtsai 2013, Ortigosa és mtsai 2014). Közülük is leggyakrabban talán a Twitteren

publikált bejegyzéseket vizsgálják, illetve használják fel fejlesztési célokra (vö. pl. Pak–Paroubek 2010, Kouloumpis és mtsai 2011, Poursepanj és mtsai 2013).

A szentimentelemzéshez kapcsolódó kutatási-fejlesztési munkában az utóbbi néhány évtizedben számtalan olyan dolgozat született, amely nem egy már elkészült applikációt publikál, hanem a szentimentelemzés valamely részproblémájának megoldására igyekszik gyakorlatorientált javaslatot tenni (pl. Liu és mtsai 2007, Sadikov és mtsai 2009, O'Connor és mtsai 2010, Chen és mtsai 2010, Miller és mtsai 2011, Groh and Hauffa 2011, Castellanos és mtsai 2011). Fontos rámutatnunk azonban, hogy a publikációk túlnyomó többsége nyelvtechnológiai oldalról közelít ezekhez a problémákhoz; azok nyelvészeti, és különösen elméleti nyelvészeti vonatkozásait csupán csekély számú dolgozat igyekszik behatóbban tárgyalni. Amint azonban emellett az értekezés bevezetőjében is érveltem, ezeknek az automatikus értékeléselemzést érintő kérdéseknek és problémáknak az alapos, nyelvészeti szempontú vizsgálata a fejlesztői oldal számára is fontos lenne, hiszen az elméleti és alkalmazott nyelvészeti kutatások eredményei hatékonyan képesek támogatni a nyelvtechnológiai fejlesztési feladatokat.

Amint azt a disszertáció bevezetőjében ugyancsak hangsúlyoztam, a nyelvi értékelés elemzésének legalapvetőbb, elsődleges feladata kell, hogy legyen a nyelvi értékelés meghatározása. Ez adhat ugyanis kizárólag megfelelő alapot az automatikus szentimentelemzéshez is. Amint ugyanakkor azt később (1. fentebb, 2) részletesen tárgyaltam, a nyelvi értékelés mibenléte az elméleti irodalomban nem tisztázott, s ezzel összefüggésben az sem egyértelmű, hogy mely nyelvészeti kategóriaként képzelhető el. A disszertáció említett részével ezt a hiányt igyekeztem pótolni.

A fejezet további részében áttekintem a nemzetközi és a hazai, szentimentelemzéshez kapcsolódó nyelvtechnológiai kutatási és fejlesztési feladatokat. A bemutatás rendező elvéül a szentimentelemzéshez kapcsolódó legfontosabb problémaköröket választottam. Azt kívánom megmutatni, hogy melyek azok a részproblémák, amelyekkel a legtöbb szentimentelemzéssel kapcsolatos munka foglalkozik, illetve azt, hogy mindezeket a problémákat milyen módszerek és eszközök segítségével igyekeznek megoldani a téma szakértői.

5.2. A szentimentelemzéssel kapcsolatos kutatások áttekintése

5.2.1 A feladat nemzetközi vonatkozásai

Áttekintve a nemzetközi szentimentelemzéssel kapcsolatos kutatásokat, megállapítható, hogy a vizsgálatok alapvetően a következő legfontosabb két probléma köré csoportosulnak: a szentimentek targetfüggő kezelésének, valamint a szentimentshifterek hatékony feldolgozásának a feladatához. A fejezet további részében ezért e két problémakört tárgyalom részletesebben, megmutatva a legfontosabb eredményeket.

5.2.1.1 A targetek kezelésének a feladata

Amint azt a jelen fejezet bevezetőjében is ismertettem (l. 5.1), az automatikus értékeléselemzés egyik fontos részfeladata az, hogy megtalálja az értékelés tárgyát, tehát azt, hogy a nyelvi értékelés az adott esetben mire irányul. Azt, amelyre az értékítélet vonatkozik, nyelvtechnológiai terminussal *target*nek (magyarul: célpontnak) nevezzük. A szentimentelemzés targetazonosítási feladata alapvetően összefügg azzal, hogy a munka során mekkora szövegmennyiséget tekintünk az elemzés egységeként. Vizsgáljuk meg tehát ezeket a megoldásokat részletesebben is!

Az automatikus szentimentelemzésben alapvetően három szintet különböztetünk meg abból a szempontból, hogy az elemző mekkora szövegmennyiséget kezel egy egységeként (vö. Liu 2012: 10). Az elemzés történhet a teljes dokumentum szintjén (*Document-level sentiment classification*), a szöveg egyes mondatainak, esetlegesen tagmondatainak szintjén (*Sentence-level sentiment classification*), illetve az aspektusok vagy a targetek szintjén is (*Aspect-level sentiment classification*; egyes szerzőknél még: *Feature level / Feature-based opinion mining and summarization*; vö. Hu–Liu 2004).

A dokumentumszintű szentimentelemzés célja az, hogy megállapítsa, a teljes dokumentum, összességében pozitív vagy negatív értékelő véleményt fogalmaz-e meg (vö. Pang és mtsai 2002, Turney 2002). Ezt a megoldást alkalmazzák például azok az elemzések, amelyek termékvélemények szövegeit dolgozzák fel abból a célból, hogy átfogó képet adjanak az adott termékkel kapcsolatos megítélésről (vö. Liu 2012: 10). A dokumentumszintű elemzés a legegyszerűbb elemzési megoldásnak tekinthető, hiszen nem törekszik a szentimentértékek kisebb szövegegységekenkénti megadására, illetve a

szenimentek targeteinek a meghatározására sem. Az elemző rendszer gyakran szótáralapú, és szóillesztéses megoldással dolgozza fel a szövegben szereplő értékelő tartalmakat: minden olyan kifejezéshez, amelyet az elemző a szótár alapján a szövegben azonosít, a lexikonban megadott értéket rendeli, végül a kapott értékekből egy algoritmussal kiszámítja a teljes szövegre vonatkozó értéket (vö. Taboada és mtsai 2011).

Az egyszerűsége okán a dokumentumszintű szenimentelemzéssel találkozunk a legtöbb automatikus értékeléselemzésben. Ugyanakkor belátható, hogy a módszer számos hiányosságot és problémát rejt magában. Mindenekelőtt, mivel a teljes dokumentumot egyetlen összetartozó egységként kezeli, azzal az alapvetéssel él, hogy az egyetlen entitást jellemez, azaz minden benne szereplő szenimentérték ugyanahhoz a targethez kapcsolódik (vö. Liu 2012: 10). Kizárja tehát annak a lehetőségét, hogy a szöveg több entitást, és azokhoz kapcsolódóan eltérő szenimentértékeket tartalmazzon. Emellett nem teremt lehetőséget az entitások, valamint azok különböző aspektusainak az elkülönítésére sem. Vizsgáljuk meg az entitások és az aspektusok kettősségének a problémáját az automatikus értékeléselemzés szempontjából közelebbről is!

Az entitások, valamint azok különböző aspektusainak az eltérő kezelése a szenimentértékek súlyozásában jelentős szereppel bír. Egy adott szenimentérték ugyanis mind magához az entitáshoz, mind annak csupán valamely aspektusához (illetve aspektusaihoz) is kapcsolódhat (vö. Liu 2012: 11, Szabó–Vincze 2015: 222). Például, egy fényképezőgép mint entitás többek között a képminőség, a szín és az ár aspektusokkal rendelkezik. Az, hogy az értékelő az adott entitás, illetve az egyes aspektusok vonatkozásában milyen értékítéleteket közöl, nem kezelhető egyenrangú ítéletként. Tekintsük a (64) alatti példákat (a példákban a vizsgált elemeket dőlt szedéssel emelem ki)!

- (64) a. Bár a *kiszolgálás* nem olyan jó, ettől függetlenül imádom ezt az *éttermet*.
b. A *kiszolgálás* nem olyan jó, a *berendezés* viszont barátságos.

A fentebbi példák egyszerre két ellentétes értéket fogalmaznak meg. Azonban a (64a) alatti példában a pozitív szenimentérték jelentősebb a negatív szenimentértéknél, tekintve, hogy az előbbi a teljes entitáshoz, az utóbbi pedig csupán annak egyetlen aspektusához kapcsolódik. Ugyanakkor a (64b) alatti példában mind a pozitív, mind a negatív értékelés az adott entitás egy-egy aspektusára vonatkozik. (Az entitás–aspektus-kettősség problémájára a későbbiekben még visszatérek.)

A fentebbiek alapján belátható tehát, hogy a dokumentumszintű szentimentelemzés sem az egyes dokumentumon belül előforduló több entitás, sem a fentebb tárgyalt entitás–aspektus-kettősség kezelésére nem alkalmas.

A második elemzési típus, azaz a mondat szintű (esetlegesen tagmondat szintű) szentimentelemzés a dokumentumszintű elemzéssel összevetve már egy részletesebb és pontosabb feldolgozást tesz lehetővé, hiszen nem „reked meg” a globális szövegvizsgálatnál. A megoldás az elemezni kívánt szöveget mondatokra bontja, majd a szentimentértékeket az egyes mondatokra vonatkozóan határozza meg. A szentimentelemzés magját ebben a típusban is alapvetően a szentimentszótárak adják. Azokban az esetekben, ahol a program nem azonosítja a mondatban a szótárban szereplő szentimentkifejezést, a mondatot neutrálisnak tekinti, tehát olyan megnyilatkozásnak, amely nem fogalmaz meg értékelést.⁴⁹

Mivel egyetlen mondat is tartalmazhat több, külön-külön értékelt targetet, arra is találni példát, hogy az elemzők a mondatok helyett a tagmondatok szintjén igyekeznek a szentimentértékeket megállapítani. Ugyanakkor, amint arra Liu (2012: 11) is rámutat, egy tagmondat is tartalmazhat több targetet. Tekintsük a szerző alábbi példáját!

(65) *Apple is doing very well in this lousy economy.*

‘Az Apple nagyon jól teljesít ebben a pocsék gazdasági helyzetben.’

A targetszintű elemzés – a fentebb bemutatott megoldásoktól eltérően – ahelyett, hogy egy előre meghatározott szövegegységet venne alapul, közvetlenül magát az értékelés és a target kapcsolatok feltárását célozza (vö. Hu–Liu 2004, Liu 2012: 11). Alapvető feladatának tekinti tehát, hogy a szentimentértékek megtalálásán és azonosításán túl a targeteket is detektálja, és azokat a megfelelő szentimentértékekhez kösse.

⁴⁹ A mondat szintű szentimentelemzésben neutrálisaként elemzett mondatok nem azonosíthatóak az ún. szubjektivitáselemzés során objektívként kezelt mondatokkal. A szubjektivitáselemzés annak megállapítására törekszik, hogy egy adott mondat tartalmaz-e szubjektív véleményt. Amennyiben igen, a mondatot szubjektívként, amennyiben nem, objektívként klasszifikálják (vö. Wiebe és mtsai 1999). Jelentős különbség azonban, hogy amíg egy a szentimentérték szempontjából neutrális mondat egyáltalán nem fejez ki értékelést, addig egy a szubjektivitás szempontjából objektívként osztályozott mondat esetlegesen igen. Tekintsük Liu (2012: 11) alábbi, angol nyelvű példáját!

(66) *We bought the car last month and the windshield wiper has fallen off.*

‘A múlt hónapban vettünk egy autót, és az ablaktörlő már leesett.’

A fentebbi példa, bár objektív tényközlést tartalmaz, a szentimentelemzésben negatív értékelésként kezelendő.

Az objektív és szubjektív értékelés kettősségét fedi le Péter (1991: 46) már korábban tárgyalt rendszere, a racionális és az emocionális értékelés típusa (l. fentebb, 2.2).

Kétségtelen, hogy valóban hatékony szentimentelemzés csupán akkor érhető el, ha sikerül pontosan azonosítani az egyes szentimentek targeteit.

A targetszintű szentimentelemzés még hatékonyabb, s egyben még nagyobb nyelvtechnológiai kihívást jelentő változata az, amelyben kísérletet teszünk az entitások, valamint azok aspektusainak az eltérő kezelésére (l. fentebb). Ahogyan arról már szóltam, értékeléseink nem azonos jelentőséggel bírnak attól függően, hogy a teljes entitásokra, vagy csupán azok bizonyos aspektusára, illetve aspektusaira vonatkoznak.

A targetek azonosítása, különösen, amennyiben az entitásokat és az aspektusokat is különböző módon szeretnénk súlyozni, rendkívül bonyolult, és számos problémát felvető vállalkozás.⁵⁰ Nem véletlen tehát, hogy ez az elemzési megoldás még nemzetközi szinten is csupán a kísérleti szakaszában jár (vö. pl. Pontiki és mtsai 2015).⁵¹ (A szentimentelemzés szintjeire a kézzel annotált szentimentkorpuszt tárgyaló fejezetben (6.3.1) visszatérek.)

A szentiment–target-kapcsolatok pontos kezelésének az ún. targetfüggő szentimentek szempontjából is kimagasló jelentősége van. Amint azt korábban már tárgyaltam (l. 2.3.2.1), számos kifejezés kontextusbeli szentimentértéke attól függ, hogy az adott kifejezés mely targethez kapcsolódik (vö. Liu 2012: 95). Egy telefon esetében például a *halk* szó negatív, míg egy porszívó esetében pozitív értékelő véleményt jelenthet.

A tárgyalt sajátságokat figyelembe véve Ahn és munkatársai (2012: 5–6) két szentimentérték-típust tételeznek fel, s rendszerüket a koreai melléknevek példáján keresztül a következőképpen mutatják be: Amennyiben egy melléknév úgynevezett *abszolút polaritást* hordoz, úgy értékelő tartalmának minősége (pozitív vagy negatív volta) nem függ az adott doméntól, amelyben szerepel, illetve targettől, amelyet minősít. Amennyiben azonban a melléknév *relatív polaritással* rendelkezik, úgy értékelő

⁵⁰ A beható targetszintű elemzés egyik fontos részproblémájaként említjük meg a koreferencia-viszonyok kezelését (vö. Caño 2014). A szentimentelemzés gyakorlatában ez azt jelenti, hogy miután az entitásokat kinyertük, az azonos jelöléssel rendelkező jelölőket össze kell tudnunk kapcsolni. Gyakori jelenség ugyanis, hogy a nyelvhasználók nem azonos kifejezésekkel utalnak ugyanarra az entitásra. Például, a *Motorola* a *Mot*, valamint a *Moto* elemek ugyanazt az entitást jelölhetik, de természetesen névmások is utalhatnak rá a szövegben (részletesebben l. *explicit és implicit aspektuskifejezések*, vö. Hu–Liu 2004). Liu (2012: 21) alapján az *entitáskategória* terminusmagát az egyedített entitást jelöli, míg az *entitáskifejezés* a szóelőfordulásokra utal. A folyamatot, amelynek során az entitáskifejezéseket entitáskategóriákba rendezzük, *entitáskategorizációnak* nevezzük.

Természetesen a feladatot az aspektusok vonatkozásában is el kell végeznünk. Az *entitáskategória* és az *entitáskifejezés* mintájára tehát *aspektuskategóriát* és *aspektuskifejezést* különböztetünk meg (vö. Liu 2012: 21). Az aspektuskifejezések aspektuskategóriákba történő rendezését *aspektuskategorizációnak* nevezzük.

⁵¹ A kézzel annotált szentimentkorpuszban (Dívány-korpusz) – a magyar nyelvű korpuszok között egyedülálló módon – a szentimenteket a targetek szintjén annotáltuk, megkülönböztetve köztük az entitásokat és az aspektusokat (részletesen l. 6.3.1.4).

tartalmának milyensége az aktuális domén, illetve target függvényében áll (részletesebben l. 2.3.2.1, ahol e megközelítési mód problematikusságát is tárgyaltam). Itt csupán a következőre szeretném felhívni a figyelmet: A szerzők (Ahn és mtsai 2012: 6) vizsgálati eredményei szerint a koreai nyelvű anyagukban a melléknévi predikátumoknak mindössze 41,4%-a mutatott inherens értékjelentést. Ez ugyancsak azt támasztja alá, hogy az értékjelentést nem határozhatja meg pusztán a lexikai szemantika szintje, ily módon a szentimentelemzés szótáralapú megoldása önmagában nem hozhat kielégítő eredményt.

5.2.1.2 A szentimentshifterek kezelésének a feladata

A szentimentelemzéssel kapcsolatos kutatások másik, jelenleg legaktívabban kutatott területe az ún. szentimentshifterek automatikus kezelésének a problémája. A jelenség egy tágabb problémakörbe, az ún. szentimentérték-módosuláshoz tartozik (részletesen l. 3).. Egy szentimentkifejezés lexikai szintű szentimentértéke számos okból kifolyólag eltérhet a bennfoglaló megnyilatkozás értékétől. Így például egy kifejezés negálható, az értéke csökkenthető vagy növelhető, sőt például az irónia eszközével az ellentétére is fordítható (vö. Székely 2007, Pete 2009, Ruppenhofer–Rehbein 2012). A korábban (2) alatt közölt példákat itt (67) alatt megismétlem.

- (67) a. **Nem** hazudott.
b. **Valamennyire** tehetséges.
c. **Rendkívül** okos.
d. Ezt jól megcsináltad!

Az angol nyelvű nyelvtechnológiai irodalom azokat az elemeket, amelyek értékbeli eltolódást okoznak a teljes megnyilatkozás szentimentértékében a benne szereplő kifejezések értékelő tartalmához képest, *szentimentshiftereknek* nevezi (Liu 2012: 60) (l. 3). A shifterek problémáit a szerzők a legtöbb esetben kombinációs szabályokat alkalmazva igyekeznek megoldani (vö. pl. Feldman és mtsai 2011: 1643, Ruppenhofer–Rehbein 2012) (részletesebben l. 1.1).

Amint azt a (2.3.2)-ben már tárgyaltam, a kombinációs szabályok alkalmazása a szentimentelemzésben több problémát is felvet, ráadásul nem fedhető le velük minden jelenség. Egyrészt, amint arra a (67a-c) alatti példák is rámutatnak, amíg bizonyos

megnyilatkozások esetében van olyan nyelvi elem, amely azonosítható az értékmódosulási folyamatban (pl. egy tagadó operátor), addig más esetekben ilyen elemet nem vagy nem feltétlenül tudunk azonosítani (67d). Az automatikus azonosítás problémáját veti fel a disszertáció egyik fókuszpontját képező értékváltási jelenség is, amely esetben bizonyos, negatív szemantikai tartalommal rendelkező elemek – negatív tartalmuk ellenére – bizonyos kontextusokban pozitív értékelést is kifejeznek (vö. Szabó 2015a: 632, 2015c: 52) (részletesebben l. a 3. fejezetben). A korábban (20a-b) alatt közölt példákat itt (68) alatt megismétlem.

- (68) a. azt is említettük, hogy hamarosan hazánkba is ellátogat a banda egy borítékolhatóan **brutális** koncert keretében [rockstation.blog.hu]
b. **Durva** volt, amit a hegedűs művelt a hangversenyen [Nádasdy 2003]

Ugyanakkor, ezekben az esetekben a megnyilatkozásokban nincs olyan, automatikus módszerrel azonosítható elem, amely ezt a jelentésbeli változást indukálná, aminek következtében az értékváltó elemekre a szótáralapú automatikus szentiment- és emócióelemzés – a lexikai szintű szentimentérték miatt – téves elemzési eredményt ad. Emellett nem kapcsolódik hozzájuk olyan, automatikus módszerrel azonosítható elem, amely lehetővé tenné ebben az esetben valamely kombinációs szabály alkalmazását.

Másrészt, az értékelést kifejező megnyilatkozások gyakorta nem teljesen kompozicionálisak (részletesebben l. fentebb, 2.3.2.2). Ahogyan arra több szerző (vö. Israel 2004: 711, Feldman és mtsai 2010: 4) is rámutat, a szentimentérték nem kalkulálható ki minden esetben maradéktalanul kombinációs szabályok segítségével (vö. pl. a (13b) alatti példa esetében korábban).⁵² Harmadrészt, amint azt korábban már szintén tárgyaltam (l. 2.1), a kombinációs szabályokat alkalmazó megoldás szótáralapú, és ebből kifolyólag – akárcsak a pusztán szótáralapú megoldás – a lexémák inherens értékelő tartalmával dolgozik. Ennek problematikus voltáról részben már a (2.3.2.1)-ben és a megelőző alfejezetben is szóltam: a szótáralapú elemzés fontos dilemmája, hogy számos kifejezés egy adott szövegdoménben, illetve targetre vonatkozóan eltérő értékelő szemantikai tartalommal rendelkezik valamely más doménbeli vagy targetfüggő sajátosságához képest (vö. Ahn és mtsai 2012, Vázquez–Bel 2012: 3557, Vechtomova és mtsai 2014: 553–554).

⁵² A problémával kapcsolatban l. még a 3. lábjegyzetet az 1.1-es alfejezetben.

A targetfüggésről a megelőző alfejezetben, a targetek kezelési feladata kapcsán már volt szó, ezért itt a doménfüggés esetét emelem ki. Tekintsük az alábbi példákat!

- (69) a. A Tremex fuscicornis nevű fadarázs nőstény gombaspórákat hordoz, és azzal *oltja be* a fát, mikor belepetézik. [nyugattolkeletig.ipolyerdo.hu/]
b. (...) az eredmények azt is láttatták, hogy a szlovákiai pártoknak rendkívül jól sikerült “úgy *beoltani*” a választók nagy részét, hogy azok ne lássák a történések közötti összefüggéseket.⁵³ [24.hu/]

Amint látjuk, a *beolt* ige sem az orvoslás, sem a biológia doménjében alapvetően nem fejez ki értékelést (69a), azonban politikai tárgyú diskurzusban rendre negatív értékelő tartalmat hordoz (69b).

5.2.1.3 A legfontosabb elemzési módszerek és eszközök

A jelenlegi, szentimentelemzést célzó kutatásokra általánosságban jellemző (l. 2.1), hogy a feladatot az úgynevezett *szentimentkifejezések*⁵⁴ szótári formába rendezett listáival igyekeznek – legalább részben – megoldani (vö. Vázquez–Bel 2012, Liu 2012: 12–13). A szótár létrehozását követően az automatikus szentimentelemzést szótáralapon, egyszerű szóillesztéses módszerrel hajtják végre (részletesebben l. lentebb).

Ahogy azt Liu (2012: 12) is hangsúlyozza, e szótárak a legtöbb szentimentelemzési megoldásban kulcsszerepet töltenek be. Az, hogy a lexikon kiemelkedő szerepet tölt be a nyelvtechnológiában, indokolt: A számítógépes nyelvészet adatorientált terület, és a nyelvtechnológiai alkalmazások fejlesztésekor a hatékonyság a cél (vö. Gábor és mtsai 2008: 854–855). A számítógépes nyelvészetnek valódi nyelvi produktumokon működő rendszereket kell létrehoznia, és a nyelvi tudást ehhez valamilyen

⁵³ Amint azt a 4.1. alfejezetben, a (22d) alatti példa kapcsán ismertettem (l. 16. lábjegyzet), az igés szerkezetek esetében előfordulhat, hogy az adott elemzési egységben több elem is tekinthető targetnek, azonban ezt az elemzés célja egyértelműsíteni tudja. Így például a (69b) alatti példa esetében az, hogy a szlovákiai pártokat vagy a választókat tekintjük a targetnek, attól függ, hogy az aktuális tartalomelemzési feladat szempontjából mely szubjektumra nézve fontos az információ, azaz melyikről szeretnénk értékelést kinyerni a szövegből.

⁵⁴ Amint azt már az értekezés bevezetőjében tisztáztam, abból a megfontolásból használok a *szentimentkifejezés* terminust (eltérve az angol *sentiment word* terminus szó szerinti fordításától), hogy szentimentkifejezésként nem csupán egy-, de többszavas frazeológiai egységek is szolgálhatnak; például *értéktelen ~ egy fillért nem ér*.

motornak hatékonyan működtetnie kell az alkalmazásban. Ezzel összefüggésben, „a nyelvfeldolgozó rendszerek tudásának legnagyobb részét a lexikon teszi ki” (Gábor és mtsai 2008: 855). A lexikon tehát nemcsak, hogy nélkülözhetetlen szinte minden számítógépes nyelvészeti feladat megoldásában, de az alkalmazás sikere is leginkább a lexikonon múlik (vö. Gábor és mtsai 2008: 854).

A szentimentszótárak készítésének három alapvető metódusát különböztethetjük meg: a manuális, a lexikonalapú, valamint a korpuszvezérelt megoldást (vö. Liu 2012: 91).⁵⁵ A manuális eljárásmodot, mivel önmagában rendkívül időigényes és fáradtságos, ritkán alkalmazzák (vö. Stone és mtsai 1966, Tong 2001, Szabó 2014, 2015b). Mivel számos, a szinonimikus és antonimikus kapcsolatokat is prezentáló lexikon áll a rendelkezésünkre – pl. a WordNet (Miller és mtsai 1990) –, kézenfekvő, hogy a szentimentszótárak készítésekor a szerkesztők – lehetőség szerint – ezeket a forrásokat is kiaknázzák (pl. Hatzivassiloglou–McKeown 1997, Turney 2002, Turney–Littman 2003). A munka kezdeteként egy viszonylag kevés elemből álló szólistát állítanak össze kézzel a szakértők, és a lista elemeihez szentimentértékeket rendelnek. Ezt követően a listát a WordNet segítségével automatikusan bővítik úgy, hogy a meglevő elemekhez a velük szinonimikus viszonyban álló (másképpen: egy synsetben levő) elemeket hozzáadják, a kapott eredményt pedig ismét manuálisan ellenőrzik. A műveletet ezt követően többször megismétlik, addig, amíg a létrehozott szótár méretét nem ítélik megfelelőnek (vö. Hu–Liu 2004, Godbole és mtsai 2007, Liu 2012: 91). A kapcsolódó kutatásokról összefoglalót ad Liu (2012: 91–95). A harmadik megoldás szerint a kutatók nyers szövegek korpusza alapján, nem felügyelt vagy félig felügyelt módszerrel, gépi tanuló algoritmust alkalmazva állítják össze a pozitív és a negatív szentimentértékkel rendelkező kifejezések listáit. A módszer azon a hipotézisen alapul, hogy az azonos értékelő tartalmat hordozó kifejezések gyakori együttes előfordulást mutatnak (vö. Kaji–Kitsuregawa 2007: 1075). A munka első lépéseként – akárcsak a lexikonalapú megoldás esetében – a szakértők összeállítanak egy viszonylag kisméretű, pozitív és negatív elemekből álló listát, majd a felhasznált szövegtörzsben azonosítják a listákon szereplő elemeket (vö. Hatzivassiloglou–McKeown 1997, Turney 2002, Kanayama–Nasukawa 2006). Ezt követően megvizsgálják, mely elemek fordulnak elő a legnagyobb gyakorisággal az azonosított elemek környezetében. A korpuszvezérelt megoldás hatékonysága mellett és ellen gyakran ugyanazt az érvet hozzák fel: mivel a módszer egy adott szövegtörzsön alapszik, az így

⁵⁵ Magam a szentimentszótár készítésekor egy angol nyelvű szentimentlexikon anyagából indultam ki, azt javítottuk és egészítettük ki a végső verzióra (részletesebben l. 7.2).

létrehozott szótár szükségszerűen doménorientált lesz (vö. Turney 2002, Qiu 2009). Mivel számos kifejezés szentimentértéke az adott domén függvényében áll (l. fentebb, 5.2.1.2), az ezzel a módszerrel létrehozott szótár csupán az adott domén szövegein fog hatékonyan teljesíteni. Ugyanakkor e probléma kapcsán azt is érdemes fontolóra venni, hogy éppen az elmondottak okán egy hatékony elemzésre képes doménfüggetlen szentimentlexikon létrehozása – e módszertől függetlenül is – valószínűleg kivitelezhetetlen feladat (vö. Qiu 2009).

Az elkészített szentimentszótárakat legtöbbször az egyszerű, szótáralapú elemzésben használják a szentimentelemzésben (vö. Szabó 2014, 2015b). A szótáralapú elemzés során szóillesztést végeznek, ami azt jelenti, hogy minden olyan kifejezéshez, amelyet az elemző a szótár alapján a szövegben azonosít, a lexikonban megadott értéket rendeli. Végül a kapott értékek összegzése alapján megadja a teljes elemzési egységre (szövegre, mondatra stb.) vonatkozó, összesített értéket. Ezt a feldolgozási módszert nevezzük *szózsák-modell*nek (a nemzetközi terminológiában: *bag-of-words representation*), részletesebben l. pl. Boiy–Moens (2009).

Az egyszerű szótáralapú elemzést számos újabb kutatás (vö. pl. Polanyi–Zaenen 2004, Moilanen–Pulman 2007, Choi–Cardie 2008) szemantikai kombinációs szabályok alkalmazásával egészíti ki. E megoldás elméleti alapja a kompozicionalitás elve, mely szerint egy adott kifejezés jelentése kiszámítható a részek jelentéséből és azok kapcsolódási módjából (vö. Osgood és mtsai 1957, Montague 1974, Dowty és mtsai 1981). A szabályokon alapuló elemző megoldások egy példája az ún. *SO-CAL* rendszer (*Semantic Orientation CALculator*), amely a szentimentszótárban szereplő kifejezések értékeiből, az azokat módosító fokozó szerepű és negáló elemek figyelembevételével kalkulálja ki a végső értéket (vö. Taboada és mtsai 2011: 267).

A gépi tanuló algoritmusok alkalmazásáról már a szentimentszótárak korpuszvezérelt létrehozási megoldása kapcsán említést tettem. A gépi tanulás módszerét azonban a szentimentelemzésben nem csupán szótárak létrehozására, hanem egyre gyakrabban mint önálló elemzési megoldást alkalmazzák a téma szakértői (vö. pl. Pang és mtsai 2002, Dave és mtsai 2003, Finn–Kushmerick 2003, Salvetti és mtsai 2004, Aue–Gamon 2005, Bai és mtsai 2005, Wang–Wang 2007, Boiy–Moens 2009). A gépi tanuló algoritmusok alapját kézzel annotált korpuszok képezik. Az elemzők tehát első lépésben egy címkézett (másképpen, a feladatnak megfelelően annotált) adatbázist készítenek, majd

ezen tanítják és tesztelik az elemzéshez használt algoritmust (vö. Boiy–Moens 2009: 3).⁵⁶ (A gépi tanuló algoritmusok alkalmazásáról, valamint alkalmazásuk nehézségeiről részletesebben l. pl. Boiy–Moens 2009.)

A fentebb elmondottakkal összefüggésben, a manuálisan annotált szentimentkorpuszok nagyon fontos és értékes eszközei a szentimentelemzésnek, hiszen jelentős szerepet töltenek be az elemző algoritmusok létrehozásában, valamint azok hatékonyságának a tesztelésében egyaránt.⁵⁷ Emellett a szentimentelemzéshez kapcsolódó kvantitatív és kvalitatív vizsgálatokban is kulcsszerepet játszanak. Ugyanakkor létrehozásuk költségigénye miatt a számuk csekély. (Az adatbázisok, különös tekintettel a kézzel annotált szövegek korpuszok felhasználási lehetőségeiről később részletesen szólok a 6. fejezetben.)

A szentimentelemzésben még újnak tekinthető, ám egyre népszerűbbé váló megoldás a mesterséges neurális hálókra alapuló elemzés (vö. pl. Chen és mtsai 2011, Maas és mtsai 2011). A megoldás alapjaként a kutatók ugyancsak tanító korpust készítenek, amely a különböző nyelvi elemek hálózataként ábrázolja a korpusz adatait. Példaként említsük meg a Stanford Sentiment Treebank-et (vö. Socher és mtsai 2013), amely közel 12 ezer angol nyelvű mondat felépítését ábrázolja függőségi fa formátumban. Az így felépített korpuszok alapján tanított elemző algoritmusok nagy hatékonyságot mutatnak nyers szövegek szentimentelemzési feladatában, mivel képesek figyelembe venni a szentimentkifejezések kapcsolatait az adott elemzési egységen belül (a megoldásról részletesen l. Socher és mtsai 2013).

5.2.2 A feladat hazai vonatkozásai

Amint azt már az értekezés bevezetőjében is hangsúlyoztam (l. 1.3), a magyar nyelvű szövegek véleménykivonatolási feladatával kimondottan csekély számú dolgozat foglalkozik. A hazai dolgozatok egy része is az angol nyelvű szövegek elemzésével

⁵⁶ A tanító- és tesztadatok a munka során értelemszerűen nem eshetnek egybe. Az egyik megoldás szerint a kézzel annotált korpust a feladat elvégzése elején két részre, tanító- és tesztadatokra osztják, a tanító adatbázison tanítják, a tesztadatokon pedig tesztelik az algoritmus eredményességét. Más esetben ún. keresztvalidációval (angolul *cross-validation*) dolgoznak. A keresztvalidáció során a korpust két egyenlő részre osztják, majd minden részt pontosan egyszer kiválasztanak, és a maradék adaton végzik a tanítást, az elkülönített részen pedig a tesztelést (Pedregosa és mtsai 2011: 2828).

⁵⁷ Ez természetesen nem csupán a szentimentelemzés feladatkörére igaz. Általánosságban is elmondható, hogy a nyelvtechnológiai kutatási és fejlesztési munkában kimagasló jelentőséggel bírnak a kézzel annotált adatbázisok.

foglalkozik. A jelen fejezetben rövid áttekintést adok a hazai, szentimentelemzéssel kapcsolatos kutatásokról.

Először azokat a hazai dolgozatokat veszem számba, amelyek az angol nyelvű, vagy legalábbis nem kifejezetten vagy kizárólagosan a magyar nyelvű szövegek szentimentelemzési feladatával foglalkoznak.

Az *OpinHu* elemző rendszer, illetve az úgynevezett *OpinHuBank* projekt Miháltz Márton munkája (2010, 2013). Az OpinHu rendszer internetes hírportálokon, blogokon, közösségi oldalakon megjelent szövegek tartalomelemzését végzi el angol, magyar, német, arab és kínai nyelven (vö. Miháltz 2010: 14). Az eszköz alapvetően szótáralapon működik, és a szózsák-modell segítségével rendel szentimentértéket az egyes mondatokhoz. Az angol nyelvű szövegeken azonban a fentebb említett megoldás mellett szintaktikai elemzést is alkalmaz, a következőképpen: a rendszer 16 meghatározott függőségi minta segítségével keresi meg az egyes szentimentkifejezések és a targetek közötti kapcsolatokat (vö. Miháltz 2010: 17–18).

Berend és munkatársai (2011) nevéhez fűződik az a vállalkozás, amelyben angol nyelvű szövegekre fejlesztettek szentimentelemző eszközt a kulcskifejezéseket detektáló nyelvtechnológiai módszer és a SentiWordNet (vö. Baccianella és mtsai 2010) segítségével. Ez utóbbi lexikont az egyes tokenek szentimentértékének a megadására használták. A feladatban gépi tanulói megoldást alkalmaztak a kutatók. Berend és Vincze (2012) későbbi dolgozata is ehhez a fejlesztéshez kapcsolódik, amelyben a szerzők a korábbi munkában alkalmazott módszerekre építve a kiértékelés problémakörét, annak esetleges szubjektivitását veszik górcső alá.

Hangya és munkatársai (2013) a fentebb vázolt megoldáshoz hasonlóan a gépi tanulás módszerét, valamint a SentiWordNetet alkalmazza egy új domén szövegeire, tweetekre és sms-üzenetekre, ugyancsak angol nyelvű szövegeken.

Hangya és munkatársai (2014, 2017) azt vizsgálják, milyen lehetőségeket nyújt a szintaktikai elemzés a szentimentértékek helytálló automatikus kezelésében. A munkához a szerzők angol nyelvű szövegeket használnak fel, amelyek három különböző doménből származnak. Azt tesztelik, hogy amennyiben ismerjük a mondatszintű szentimentértékeket, valamint a mondatok szintaktikai elemzése is rendelkezésre áll, a szentimentkifejezések és a mondat további elemeinek kapcsolódási mintái alapján el lehet-e végezni eredményesen a gépi tanítást.

Most tekintsük azokat a dolgozatokat, amelyek kifejezetten a magyar nyelvű szövegek szentimentelemzési feladatával foglalkoznak!

Berend és Farkas (2008) munkája a kettős állampolgárság témájához kapcsolódó, magyar nyelvű szövegek gépi tanuláson alapuló automatikus feldolgozását célozza. A szerzők először összegyűjtöttek összesen 1294 fórumposztot egy a népszavazást megelőző időszakból, majd a posztokat egyenként annotáltatták. Az annotátorok azt az utasítást kapták, hogy a szövegeket a következő három tag valamelyikével lássák el: irreleváns, támogató vagy elutasító. Ezt követően, a dokumentumszinten annotált korpuszon két gépi tanuló algoritmust fejlesztettek, és azok eredményességét tesztelték (vö. Berend–Farkas 2008: 409–412).

Az OpinHu rendszer (l. fentebb) a magyar nyelvű szövegek elemzéséhez szótárat alkalmaz, amelyet a készítőik egy eredetileg angol nyelvű szentimentszótár fordításával és szinonimákkal való bővítésével hoztak létre (vö. Miháltz 2010: 16). Amint azt már említettem (l. fentebb), az elemző a szöveg szentimentkifejezéseit a szótár alapján azonosítja, majd szózsák (*bag-of-words*) algoritmussal összegzi azok értékét az egyes mondatokra vonatkozóan. Az értékelések targetfüggő kezelését azzal az alapvetéssel oldották meg, hogy amennyiben „egy érzelmi kifejezés és egy kulcsszó együtt fordul elegy mondatban, akkor az érzelem a célpontra irányul” (Miháltz 2010: 17).

Az ún. OpinHuBank korpusz (vö. Miháltz 2013) egy kutatási és fejlesztési célokra szabadon hozzáférhető, magyar nyelvű szentimentkorpusz. A korpusz anyagához az OpinHu projekt (vö. Miháltz 2010, l. fentebb) adatbázisa szolgáltatott alapot, amely 500 hazai online forrásból (híroldalak, blogok, fórumok) több millió szöveget tartalmaz a 2009 és 2012 közötti időszakból. A szövegeket első lépésben mondatokra bontották, tokenizálták, morfológiaileg elemezték, szófajilag egyértelműsítették, végül névelemfelismerést hajtottak végre rajtuk (vö. Miháltz 2013). Ezt követően az adatbázisból kiszűrtek 12 ezer mondatot a kézi feldolgozáshoz. A kiválasztás egy fontos kritériuma volt, hogy tartalmazzon legalább egy, PERSON (személynév) típusú entitást (Miháltz 2013: 344). Végül, a különböző szűréseket követően összesen 10 ezer elemezni kívánt egység maradt, és ezen végeztek el az annotátorokkal a manuális feldolgozást. A munka során a mondatok negatív, pozitív vagy semleges értéket kaphattak. Az annotáció fontos alapvetése volt, hogy a mondatok szentimentértékét az azokban szereplő PERSON (személynév) típusú entitás vonatkozásában ítéljék meg.

Hangya és mtsai (2015) az OpinHuBank adatbázisát felhasználva épített olyan elemző rendszert, amely képes különböző hírszövegekben a nyelvi értékelést a targetekkel összefüggésben azonosítani. A feladat elvégzéséhez fordítási módszerrel magyar nyelvű szentimentszótárat hoztak létre a készítőik, majd gépi tanulást alkalmaztak. Amint arra

maguk a szerzők (Hangya és mtsai 2015: 227) is felhívják a figyelmet, problémát jelentenek azok az esetek, amikor egy mondaton belül több target, és hozzájuk kapcsolódóan több szentimentkifejezés is szerepel. Ilyen esetekben ugyanis fontos volna a szentimentkifejezéseket a mondaton belül a megfelelő targetekhez kapcsolni. A problémát a szerzők úgy igyekeztek kezelni, hogy a mondatok feldolgozásához dependenciaelemzőt is építettek a modellbe(vö. Hangya és mtsai 2015: 228). Az ily módon létrehozott elemző jelentős problémáját jelentheti azonban az, hogy az OpinHuBank korpusz, amelyen a tanítást elvégezték, a targeteket illetően komolyabb hibákat tartalmaz. (Erről a sajátságról az adatbázisok létrehozását és vizsgálatát tárgyaló fejezetben részletesen szólok majd, 1.6.3.1)

Végezetül, Hangya és Farkas (2015) nevéhez fűződik az a vállalkozás, amelynek keretében doménspecifikus szentimentlexikonokat fejlesztettek magyar nyelvre, automatikus megoldások segítségével. Amint azt már több ízben tárgyaltam (l. fentebb, 2.5.2.1, 5.2.1.2), számos kifejezés a különböző domének szövegeiben eltérő szentimentértéket hordozhat. A probléma megoldásának egy lehetséges útja az, ha a fejlesztők olyan szótár létrehozására törekednek, amely az elemezni kíván domén szövegeinek sajátságai szerint készült. A szerzők a doménspecifikus szótárak létrehozására több módszert is kipróbáltak. Mindenekelőtt hangsúlyozzák, hogy ilyen specifikus szótárak a legegyszerűbben az adott doménhez tartozó, annotált szövegekből fejleszthetők, azonban ilyen korpusz nem állt a rendelkezésükre. Ezért olyan fejlesztő módszerekkel próbálkoztak, amelyekhez kézzel annotált adatbázisokra nincs is szükség. Az egyik megoldás szerint a magyar wordnet (Hungarian WordNet (HuWN), vö. Miháltz és mtsai 2008) alapján dolgoztak, azzal a feltételezéssel élve, hogy a lexikonban található elemek és azok szinonimáinak a szentimentértéke megegyezik. A másik kísérletben klaszterezési megoldást alkalmaztak a szerzők. A magyar wordnetnél alkalmazott módszerhez hasonlóan azt feltételezték, hogy az egy klaszterbe tartozó szavak szentimentértéke megegyező. A kiindulási szótárát tehát úgy terjesztették ki, hogy minden olyan szót felvettek, amely a feldolgozott szövegek alapján egy klaszterbe tartozott valamely, a kiindulási lexikonban levő szóval. Végül a szótárak hatékonyságát az OpinHuBank korpuszon, valamint egy erre a célra létrehozott, a technológiai doménhez tartozó szövegekből álló adatbázison értékelték ki (vö. Hangya–Farkas 2015: 215–217).

5.2.3 A szentimentérték módosulási jelensége a szentimentelemzés szempontjából, különös tekintettel az értékváltási és értékvesztési típusokra⁵⁸

Amint azt korábban, a szentimentelemzés legfontosabb módszereit és eszközeit áttekintő részben már részletesen tárgyaltam (l. 5.2.1.3), a szótárak a legtöbb számítógépes értékeléselemzési megoldásban kulcsszerepet töltenek be. Ugyanakkor, a szentimentszótárak a szentimentérték módosulásának a jelenségét egyáltalán nem képesek kezelni (vö. pl. Feldman és mtsai 2010, Ruppenhofer–Rehbein 2012). Ahogyan azt már a dolgozatban több ízben ismertettem (l. 1.1., 5.2.1.2), a számítógépes értékeléselemzés egy kardinális, alkalmazásszintű problémája az ún. szentimentérték-módosulás jelensége, amely alatt általában azt az esetet értik a szerzők, hogy egy nyelvi elem lexikai szintű szentimentértéke nem azonos, vagy nem teljes mértékben azonos az őt magában foglaló teljes megnyilatkozás szentimentértékével (vö. Bolinger 1972: 15–20, Feldman és mtsai 2011: 1643, Liu 2012: 36–38, Szabó 2015a: 630–633).

Bár egy szentimentkifejezés lexikai szintű szentimentértéke számos okból kifolyólag eltérhet a bennfoglaló megnyilatkozás értékétől, a legtöbb szerző az ún. szentimentshifterek módosító hatását igyekszik valahogyan kezelni. A nyelvtechnológiai irodalomban a *sentiment shifter*, a *polarity shifter* vagy a *contextual valence shifter* (*shift*: 'változás, eltolódás, elmozdulás') terminusokkal utalnak minden olyan nyelvi elemre, amely egy adott szentimentfragmentumban szerepelve eltérést okoz a szövegben szereplő szentimentkifejezések lexikai szintű értéke, valamint a teljes fragmentum szentimentértéke között (vö. Li és mtsai 2010: 635) (részletesebben l. fentebb, 3). Ezt az értékmódosulási típust tárgyalja a legtöbb, a problémakörrel foglalkozó irodalmi tétel (vö. pl. Na és mtsai 2004, Polanyi–Zaenen 2004, Kennedy–Inkpen 2006, Ikeda és mtsai 2008, Wilson és mtsai 2009, Li és mtsai 2010, 2013, Boubel és mtsai 2013), és ezzel a jelenséggel igyekeznek elszámolni a már ugyancsak tárgyalt (l. fentebb, 2.1, 5.2.1.2, 5.2.1.3) ún. szemantikai kombinációs szabályok is (vö. pl. Feldman és mtsai 2010; Ruppenhofer–Rehbein 2012).

A szentimentérték módosulását, illetve a szentimentshifterek különböző típusait illetően megállapítható, hogy a szerzők az egyes típusokat áttekintő jelleggel, behatóbb vizsgálat keretében a szentimentelemzés szempontjából nem tárgyalják. Különösen igaz ez a magyar nyelvű szövegek szentimentelemzésére vonatkozóan (részletesebben l. 1.3).

⁵⁸ A jelenségek pontos definícióját l. 3.

Különösen fontos kiemelni, hogy a szentimentérték módosulását nem minden esetben az említett shifterek okozzák (l. fentebb, 3). Lehetségesek ugyanis olyan módosulások, amelyek végbemenetelét nem lehetséges egyetlen nyelvi elemhez kötni a megnyilatkozásban. A szentimentérték módosulásának a disszertáció – a shifterek által okozott módosulásokon túl – két fontos típusát különbözteti meg: az úgynevezett *értékvesztést* és az *értékváltást*. Ezek azok az elemek, amelyek lexikai szinten negatív emotív szemantikai tartalommal rendelkeznek, azonban fokozó funkcióban képesek azt részben vagy egészben elveszíteni, sőt bizonyos esetekben arra is, hogy negatív tartalmuk ellenére pozitív értékelést fejezzenek ki (vö. Szabó 2015a: 632, Szabó 2016: 165–167). A korábban (19) és (20) alatt közölt példákat itt (70) és (71) alatt megismétlem. A vizsgált elemeket a példákban félkövérrel emelem ki.

- (70) a. Szerintetek is **borzasztó** nagy a fiam lába? [gyakorikerdesek.hu]
- b. Nagyon egyszerű leszek: **rohadt** jól vannak kitalálva a szereplők személyiségei [szeretlekmagyarorszag.hu]
- (71) a. azt is említettük, hogy hamarosan hazánkba is ellátogat a banda egy borítékolhatóan **brutális** koncert keretében [rockstation.blog.hu]
- b. **Durva** volt, amit a hegedűs művelt a hangversenyen [Nádasdy 2003]

A két jelenség mind elméleti, mind alkalmazott nyelvészeti, különösen nyelvtechnológiai szempontból figyelemre méltó, ennek ellenére mindkét területen kevés dolgozat foglalkozik velük (vö. pl. Andor 2011, Kugler 2014, Dragut–Fellbaum 2014, Szabó 2015c: 52, Szabó és mtsai 2017a: 252, Drávucz és mtsai 2017: 236) (részletesen l. 1.1, 4.2).

Ami a magyar nyelvű szövegek automatikus értékeléselemzésének a feladatkörét illeti, az értékvesztés és az értékváltás jelenségeit saját munkáimon (Szabó 2015a, 2015c, 2016, 2017, Szabó és mtsai 2017a, 2017b, 2017d) kívül más nem tárgyalja. Ezeken a publikációkon kívül kizárólag Hangya és munkatársai (2015: 212) tesznek egy rövid említést a problémáról. A szerzők a szentimentszótárak készítése kapcsán megjegyzik, hogy gondot jelentenek többek között azok a kifejezések, amelyek „két ellentétes polaritású jelentéssel rendelkeznek” (Hangya és mtsai (2015: 212). Fontos felhívni arra a figyelmet azonban, hogy a szerzők a problémát a szótár szintjén megoldhatónak látják úgy, hogy a szótár fordítása során a magyar nyelvben az aktuális doménnek megfelelő

ekvivalenst választják ki. Amint ugyanakkor magam számos dolgozatban (Szabó 2015a, 2016, Szabó és mtsai 2017a, 2017b), valamint a jelen disszertációban is több helyütt érveltem (l. fentebb, 4.5, 5.6), a probléma maradéktalanul nem oldható meg szótári szinten, hiszen túlmutat a lexikai szemantika szféráján.

A fentebb elmondottakkal összefüggésben, a fejezet zárásaként érdemes kiemelni, hogy a szentimentelemzést a szakirodalom a nyelvtechnológiai kutatások között alapvetően szemantikai vonatkozású problémaként tartja számon. Csupán elvétve találni olyan, a témában született dolgozatot, amely egyáltalán említést tenne a feladat pragmatikai vonatkozásairól (vö. pl. Honkela és mtsai 2012).

6. A szentimentérték változásának vizsgálata adatbázisok segítségével

A disszertáció jelen fejezetében részletesen ismertetem azt a három adatbázist, valamint az azok elemzésével kapott vizsgálati eredményeket, amelyek a disszertáció alapját képező kutatómunka során jöttek létre.

A fejezet elején a különböző adatbázisoknak a nyelvészeti kutatásban és fejlesztésben betöltött szerepét tárgyalom. A korpuszok esetében tisztázom a korpusz fogalmát, majd ismertetem a korpuszok típusait, feldolgozásuk módszereit és eszközeit, számos példát hozva az egyes adatbázistípusokra. Mindeközben ráirányítom a figyelmet a korpuszoknak a kutatási és fejlesztési feladatokban betöltött fontos szerepére. Végül a korpuszannotáció minőségbiztosítási megoldásait tárgyalom részletesen.

A szóasszociációs adatbázisokat illetően ismertetem a szóasszociációs adatfelvételek célját és hasznát. Ezt követően bemutatom az adatfelvételi, valamint az adatalemzési megoldásokat, valamint tárgyalom a legfontosabb nemzetközi és hazai kutatási eredményeket.

Az általános bemutatást követően ismertetem azt a három vizsgálati eszközt, amelyek a disszertáció alapját képező kutatómunka keretében jöttek létre: a fragmentum- és targetszintű annotációt tartalmazó, kézzel annotált szentimentkorpuszt, az értékvesztés és az értékváltás jelenségei szempontjából kézzel annotált korpuszt, valamint a szóasszociációs adatbázist, amely az értékvesztésre – és esetlegesen az értékváltásra is – képes elemekre adott asszociációkat tartalmazza. A bemutatás során tárgyalom az eszközök létrehozásának az okát és célját, a munkavégzés alapelveit és pontos menetét, eszközeit, valamint a minőségbiztosítási megoldásokat. Végül közlöm az alapvető és részletes statisztikai adatokat, valamint a vizsgálatoknak az értekezés szempontjából legrelevánsabb eredményeit, és az azokból levonható legfontosabb következtetéseket.

6.1 A szöveggörpuszok mint a kutatás és fejlesztés eszközei

Az alkalmazott nyelvészeti kutatások egyik legfontosabb vizsgálati eszközét az ún. szöveggörpuszok jelentik.

A *korpusz* fogalmának a meghatározását illetően a szakirodalom nem egységes. A Nyelvi fogalmak kasszótárában a következő definíciót találjuk a *korpusz* címszó alatt: „meghatározott szempontok alapján kiválasztott szövegmennyiség, amelyen a nyelvész vizsgálatát végzi” (Kugler–Tolcsvai Nagy 2000: 132). Ez a meghatározás mind a szöveg mennyiségére, mind annak az előkészítettségére vonatkozó kritériumok hiánya miatt elnagyoltnak tűnik azzal összevetve, hogy milyen szöveggyűjteményeket tekinthetünk korpusznak.

A Magyar Nemzeti Szövegtár (MNSZ2, Oravecz és mtsai 2014) honlapja(http://corpus.nytud.hu/mnsz/bevezeto_hun.html) a következőképpen definiálja a *korpuszt*: „A korpusz ténylegesen előforduló írott, vagy lejegyzett beszélt nyelvi adatok gyűjteménye. A szövegeket valamilyen szempont szerint válogatják és rendezik. Nem feltétlenül egész szövegeket tartalmaz, és nem csak tárháza a szövegeknek, hanem tartalmazza azok bibliográfiai adatait, bejelöli a szerkezeti egységeket (bekezdés, mondat)”. A definíció nem mutat rá, hogy a szöveggörpuszok a legtöbb esetben valamilyen manuális vagy automatikus feldolgozási folyamaton esnek át, s arra sem, hogy ennek a feldolgozási folyamatnak (másképpen: *annotációnak*) a sajátosságait gyakorta a korpusz jövőbeli felhasználási céljai határozzák meg. A görpuszok tehát mindig valamilyen céllal készülnek, és ennek nem mond ellent az sem, ha egy adott görpusz nem egy bizonyos jelenség vizsgálatához jön létre, hanem például egy adott nyelv reprezentatív görpusza kíván lenni (mint amilyen például a Magyar Nemzeti Szövegtár). Ilyenkor ugyanis a készítőik úgy válogatják össze a görpusz szöveganyagát, hogy az éppen ennek az igénynek tudjon megfelelni.

A görpusz sajátosságai közül kiemelendő az, amelyre az MNSZ2 oldalán közölt definíció utal: a görpusz mindenképpen olyan nyelvi adatokból áll, amelyek a nyelvhasználat folyamán keletkeztek, tehát nem a nyelvész állítja elő az adatokat a vizsgálat céljából (vö. McEnery 2005: 449).

Bár a *korpusz* definíciójába nem tartozik bele elengedhetetlen kritériumként, fontos említést tenni arról is, hogy a görpuszok legtöbbször digitalizált formájúak, ugyanis csupán ez teszi lehetővé a gépi elemzést (vö. McEnery 2005: 449). A görpusznyelvészet egyik legfontosabb célja az, hogy a nyelvi jelenségeket empirikusan, kvalitatív és kvantitatív szempontból egyaránt vizsgálja, ehhez pedig nagy mennyiségű szöveg elemzésére van szükség, ami manuálisan fáradtságos és nem költséghatékony feladat (vö. Копотев–Гурин 2006: 280). A görpusz szövegeinek tehát digitalizált formájúnak, ezáltal géppel feldolgozhatóknak és lekérdezhetőnek kell lennie. Automatikus módszerrel gyorsabban,

egyszerűbben és pontosabban végezhető el jelentős mennyiségű szöveg vizsgálata, ami egyben arra is lehetőséget ad, hogy általános érvényű, empirikusan igazolt megállapításokat tehessünk a vizsgált jelenség vonatkozásában (vö. Károly 2003: 19).⁵⁹

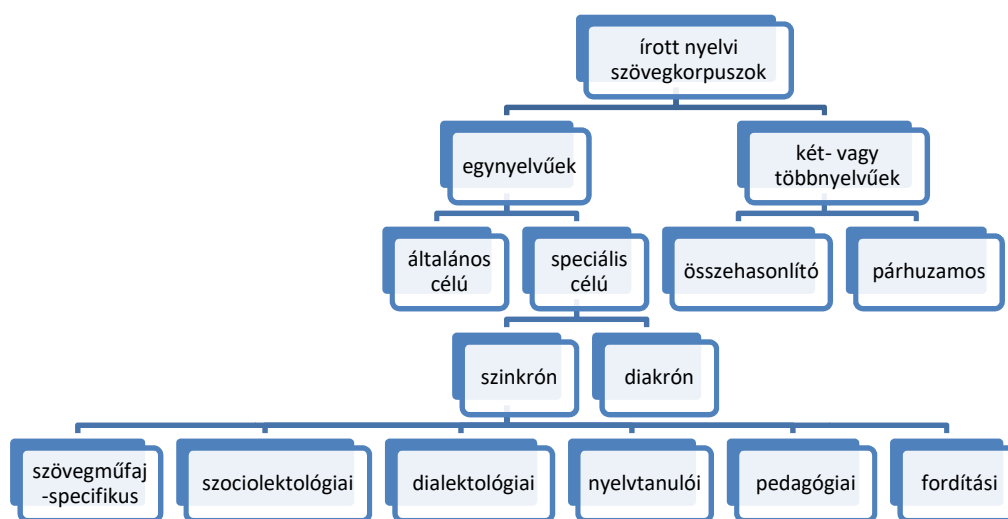
A korpuszok alkalmazása ma már nélkülözhetetlen a különböző nyelvészeti munkák során. Többek között a lexikográfia is nagymértékben épít rá, hiszen fontos segítséget ad a szavak jelentésének és kollokációs mintájának a vizsgálatához, valamint a különböző regiszterekhez kötött szógyakoriság megismeréséhez is (vö. Károly 2003: 20).⁶⁰ A lexikográfia mellett jelentős szerepet töltenek be a korpuszok a fordításkutatásban és a fordítóképzésben, az anyanyelv és idegen nyelv elsajátításának vizsgálatában, valamint az idegen- és szaknyelvoktatás területén (vö. Károly 2003: 20, Szirmai 2005: 17).

A korpuszok ugyanakkor nem kisebb szerepet játszanak a kutatás mellett a fejlesztési oldalon is. A különböző nyelvfeldolgozó algoritmusok fejlesztése és tesztelése ugyanis kézzel annotált korpuszok alapján történik (vö. McEnery 2005: 448) (részletesebben l. lentebb).

A korpuszok között beszélt nyelvi szövegtörzsöket is találunk, azonban a legtöbb törzs az írott nyelvet reprezentálja (vö. McEnery 2005: 450–451). A beszélt nyelvi korpuszok két hazai példája a HuComTech (vö. Pápay és mtsai 2011) és a HuTongue (Szabó–Galántai 2017) adatbázisok. A jelentősen csekélyebb számmal összefüggésben a beszélt nyelvi korpuszok között kevés altípust különíthetünk el, szemben az írott nyelvi korpuszokkal, amelyeknek számos fajtája létezik. Az alábbi ábra bemutatja, magam hogyan tekintem lehetségesnek az írott nyelvi szövegtörzsök osztályozását az alapján, hogy a törzs milyen szövegeket tartalmaz.

⁵⁹ Annak ellenére, hogy a törzsalapú kutatások számítógépes támogatottság hiányában nehezen kivitelezhetőek, a törzsök alkalmazásának kezdetleges formáival a számítógépek elterjedését megelőző időben is találkozunk. Már ekkor alkalmazták ugyanis szövegtörzsöket grammatikák és szótárak szerkesztésére (vö. Ильясов 2003: 7). Mivel ezeket az adatbázisokat nem elektronikus formában tárolták, számítógéppel támogatott elemzésük értelemszerűen nem is volt lehetséges. Egy szövegtörzs papíron, kézzel végzett elemzése ugyanakkor rendkívül időigényes és fáradságos feladat, s már néhány száz mondat manuális vizsgálata is csupán nehezen hajtható végre (vö. Szirmai 2005: 47). Nem véletlen tehát, hogy az elektronikus szövegtörzsök megjelenése új távlatokat nyitott az egyes nyelvek, illetve nyelvi jelenségek tanulmányozásában (vö. Krug 2000: 21, Klaudy 2005: 153, Péch 2007: 79, Szabó és mtsai 2017c).

⁶⁰ Hazánkban a Magyar Tudományos Akadémia 1984-ben határozatban döntött arról, hogy a nagyszótár munkálatainak folytatását elektronikus szövegtörzs alapján kell végezni (vö. Prószték és mtsai 2003: 571).



1. ábra: Az írott nyelvi szövegkorpuszok osztályozásának egy lehetséges módja

Az egynyelvű korpuszok egy adott nyelv írott szövegeit tartalmazzák (vö. McEnery 2005: 450). Két alapvető típusukat különböztethetjük meg: az általános célú, valamint a speciális egynyelvű korpuszokat. Az általános célú korpuszok esetében alapvető követelmény a különböző stílusú szövegek megfelelő aránya (vö. Szirmai 2005: 33). Ilyen korpusz például a magyar nyelv esetében a Magyar Nemzeti Szövegtár (MNSZ2, Oravecz és mtsai 2014) vagy a Szeged Korpusz (vö. Csendes és mtsai 2003), vagy az orosz nyelv esetében az Orosz nyelv nemzeti korpusza (Национальный корпус русского языка, НКРЯ).

A speciális célú egynyelvű korpuszok a bennük szereplő szövegek megválasztásával, illetve azok feldolgozási módjával bizonyos szövegtípusok vagy nyelvi jelenségek vizsgálatát támogatják. Közöttük mind diakrón vagy történeti, mind szinkrón, azaz a jelen nyelvállapotot tükröző korpuszokat találunk. A történeti korpuszok egy prominens példája a Magyar Történeti Szövegtár (MTSZ) (vö. Sass 2017), de több olyan adatbázis is létezik, amely egy meghatározott történelmi kor szövegeit tartalmazza, például a Rákosi-korszak jegyzőkönyveit magában foglaló korpusz (vö. Gulyás és mtsai 2018).

A szinkrón korpuszok közé tartoznak azok az adatbázisok, amelyek egy bizonyos műfajhoz tartozó szövegeket (pl. Twitter- vagy blogszövegek, termékvélemények stb.), nyelvjárási vagy rétegnyelvi szövegeket tartalmaznak (vö. Károly 2003: 18, Vincze és mtsai 2016). A nyelvtanulói korpuszok egy adott nyelvet idegen nyelvként tanulók nyelvi produktumait tartalmazzák (vö. Durst és mtsai 2014), amelynek hazai példája a

HunLearner (vö. Vincze és mtsai 2013, Durst és mtsai 2014). A pedagógiai korpusz olyan szövegek gyűjteménye, amelyekkel a nyelvtanuló egy adott kurzus során találkozhat (vö. McEnery–Wilson 1997, Szirmai 2005: 35). Végül, az úgynevezett fordításkorpusz egy adott nyelvre fordított szövegeket tartalmaz, lehetővé téve ezzel, hogy a célnyelvi szövegek sajátosságait feltárhassuk (vö. Károly 2003: 19, Szirmai 2005: 34).

A korpuszok másik nagy csoportját a két vagy többnyelvű korpuszok képezik. Közöttük összehasonlító és párhuzamos típusúakat különböztetünk meg. Az összehasonlító korpuszok hasonló struktúrájú és rendeltetésű szövegeket tartalmaznak (vö. Kohn 1999: 69, Károly 2003: 19), a párhuzamos korpuszok pedig forrásnyelvi szövegekből és azok fordításaiból állnak (vö. McEnery 2005: 450, Szabó és mtsai 2012: 2453). E típus egyik példája a HunOr magyar–orosz párhuzamos korpusz (vö. Szabó és mtsai 2012).

A fentebb bemutatott osztályozás tehát az alapján tárgyalta a korpuszokat, hogy azokban milyen szöveganyagot dolgoznak fel a készítők. A korpuszok tipizálásának azonban egy további, nem kevésbé fontos szempontja, hogy bennük a szövegeket milyen annotációval látták el a korpusz építői. Az annotáció olyan annotálási jelek (számítógépes nyelvészeti terminussal: *tagek*) alkalmazását jelenti, amelyekkel a korpuszban levő szövegek egyes részeit (pl. mondatait, többszavas kifejezéseit, bizonyos morfémaikat stb.) jelölik be.⁶¹ Ezek a *tagek* hivatottak géppel kereshetővé, feldolgozhatóvá tenni a nyelvi adatok olyan sajátosságait, amelyek az annotáció nélkül a gép számára még nem kereshető, listázható információk (vö. McEnery 2005: 453). Fontos megjegyezni, hogy az annotáció által a korpusz szövegei nem változnak meg, az annotációt azoktól elkülönítve lehetséges – és kell – kezelni. Tekintsük a korpuszok annotálási módszereit részleteiben is!

Mindenekelőtt, az annotációt tipizálhatjuk aszerint, hogy a korpuszban levő szövegekben milyen mélységig „hatol le”, tehát, hogy mit tekint a feldolgozás egységének. Eszerint megkülönböztetünk szöveg-, mondat-, valamint tokenszintű elemzést (vö. Vincze Veronika előadásjegyzete, 2016). A korpusz annotálása automatikus, félautomatikus, valamint manuális munkával valósítható meg (vö. McEnery 2005: 455–456). Számos feladat (pl. lemmatizálás vagy szófaji egyértelműsítés) ma már olyan hatékonysággal végezhető el automatikus módszerrel, hogy ezekben a feladatokban nem szükséges humán

⁶¹ Aszerint, hogy a korpuszban pontosan milyen típusú nyelvi információkat annotálnak a készítők, McEnery (2005: 458) a következő fő típusait különbözteti meg: lemmatizálás, szófaji egyértelműsítés (POS-taggelés), szintaktikai elemzés (másképpen: *parsolás*, angolul: *parsing*), szemantikai elemzés, diskurzus szintű annotálás, pragmatikai vagy stilisztikai szintű feldolgozás, valamint említést tesz egy olyan feldolgozási megoldásról is, amely specifikusan az információkinyerési feladatokat igyekszik hatékonyan támogatni. A jelen disszertációnak a korpuszannotáció pontos tipizálása nem célja, azonban azt érdemes megjegyezni, hogy a fentebbi osztályozásban az egyes típusok egymástól nem vagy nem feltétlenül választhatóak szét.

annotátorokat alkalmazni. A legtöbb annotálási munka azonban automatikusan nem végezhető el kielégítő eredményességgel, illetve a különböző elemző eszközök fejlesztéséhez is gyakorta kézzel annotált korpuszokra van szükség. Amennyiben az annotálás egy része automatikus módszerrel elvégezhető olyan hatékonysággal, hogy az már támogatni tudja a manuális munkát, úgy először géppel elemeztetik a korpusz szövegeit, amelyet humán ellenőrzés és korrekció követ. Ezt nevezzük félautomatikus megoldásnak. Amennyiben azonban az annotálás – annak jellege miatt – automatikus módszerrel egyáltalán nem támogatható, úgy a teljes munkát az annotátoroknak kell elvégezniük (vö. pl. Vincze és mtsai 2014a, 2014b).

Amennyiben a korpusz feldolgozása részben vagy egészben manuálisan történik, úgy a korpusz annotátorai között – a minőség biztosítása céljából – egyetértést mérnek a készítőik. Ez azt jelenti, hogy a korpusz egy bizonyos részét az összes annotátor feldolgozza, legideálisabb esetben a teljes korpuszt legalább két annotátor annotálja. Ezt követően az annotációt valamely, erre a célra fejlesztett algoritmus segítségével összevetik egymással. Ilyen mérési eszköz például az ún. Cohen-féle Kappa-statisztika (angolul: *Cohen's Kappa statistic*) (vö. Cohen 1960). A méréssel kapott értékek megmutatják, hogy az annotátorok hány esetben látták el taggal a korpusz bizonyos egységeit, valamint azt is, hogy hány alkalommal használták rájuk ugyanazt a taget. Ennek köszönhetően a Cohen-féle Kappa-statisztika komplex, a minőség javítására jól alkalmazható mutatónak tekinthető.

Mint már fentebb említettem, a korpuszok nem csupán a kutatási, de a fejlesztési oldalon is fontos szerepet töltenek be. A különböző nyelvfeldolgozó algoritmusok fejlesztése és tesztelése ugyanis legtöbbször kézzel annotált korpuszok alapján történik (vö. McEnery 2005: 448, Nikunen 2007: 7, Boiy–Moens 2009: 3) (részletesebben l. 5.2.1.3).

Az algoritmusok eredményességét általában az ún. F-mértékkel értékelik ki a fejlesztők (vö. Powers 2011: 37–39, Derczynski 2016).⁶² Az F-mértéket az ún. *pontosság* (angolul: *precision*) és a *fedés* (angolul: *recall*) értékeinek kombinációjából kalkuláljuk ki. A *pontosság* megadja, hogy az algoritmus által a kézzel annotált korpuszban talált adatokból hány adat valóban pozitív, azaz az eszköz helyesen találta azt meg. A *fedés* pedig azt mutatja meg, hogy hány adatot talált meg a program a valóban pozitív adatokból (szemben azokkal, amelyeket nem talált meg). Azért, hogy a két érték kombinációjának a

⁶² Az F-mérték az annotátorok közötti egyetértés mérésére is használható, l. 6.3.1.5.

fontosságát érzékeltessük, említsük meg a következő lehetőségeket: elképzelhető olyan algoritmus, amelynek minden találata helyes ugyan, azonban ezeknek a helyes találatoknak összesen nagyon kicsi a száma az összes lehetséges találathoz képest. Az algoritmus pontossága tehát ebben az esetben magas, míg a fedése alacsony. Illetve az is lehetséges, hogy az algoritmus a releváns elemek többségét megtalálja ugyan, azonban számos olyan elemet is azonosít a korpuszban, amely nem releváns, így gyakran hibázik. Az algoritmus pontossága tehát ebben az esetben alacsony, míg a fedése magas. Az F-mérték e két értéket kombinálva informatív mutatója az algoritmusok hatékonyságának.

Végezetül, érdemes rámutatni arra a tendenciára, amely a szövegtárházak és a szövegfeldolgozó algoritmusok keletkezését jellemzi. Azt láthatjuk ugyanis, hogy minél több, nagyobb méretű és jobb minőségű annotált tárház áll a rendelkezésünkre, annál több és jobb minőségű feldolgozó eszköz fejlesztésére nyílik lehetőség, ez pedig újabb, még nagyobb méretű és jobb minőségű annotált tárház létrehozását segíti elő (vö. McEnery 2005: 459). A folyamat tehát öngerjesztő, és a nyelvtudomány egyre nagyobb ütemben zajló fejlődését eredményezi.

6.2 A szóasszociációs adatbázisok mint a kutatás és fejlesztés eszközei

Amint azt a (4.6.2)-ben kapcsán igyekeztem megvilágítani, a szóasszociációs adatbázisok a mentális lexikon modellezésének fontos kutatási és fejlesztési eszközeiként szolgálnak. A szóasszociációs mintázatok feltárása ugyanis a mentális lexikon hálózatos működésébe engedhet betekintést (vö. Lengyel 2008: 18). Például, ezek segítségével elemezhetjük a nyelvi elemek szemantikai szerveződésének, a jelentésviszonyok hálózatának a sajátosságait (vö. Pléh 2008: 833).

Az asszociációkat illetően gyakran elemzik a válaszok számát, grammatikai sajátosságait (pl. szófaját), valamint fogalmi elemzésnek is gyakran alávetik őket (vö. Pléh 2008: 799, Lengyel 2010: 195–198). Az adatbázisok vizsgálatának fontos hozadéka lehet, hogy segítségével azt is feltárhatjuk, milyen tág az egyes nyelvi elemekre adott asszociációk köre, illetve, megismerjük az egyes nyelvi elemekhez tartozó asszociatív átfedéseket is, ezáltal pontosabb képet kapva a hálózat szerveződéséről (vö. Pléh 2008: 833, 835). Amennyiben feltárjuk az asszociációk közötti ún. átfedési mátrixot, azokat klaszterelemzéssel vagy többdimenziós skálázással tovább elemezhetjük, még mélyebb betekintést nyerve ezzel a mentális lexikon struktúrájába (vö. pl. Oláh-Nagy 2012).

Az asszociációs adatbázisok további haszonnal szolgálhatnak olyan vizsgálatoknál, amelyekben már a tervezési szakaszban fontos az asszociációs mintázatok megismerése (vö. Pléh és mtsai 2008: 833–834). Így például, fontosak lehetnek olyan pszicholingvisztikai kísérletek tervezésénél, mint például az előfeszítési jelenség (*priming*) vagy a mondatfeldolgozási mechanizmusok vizsgálata, ahol „az ingeranyag összeállításánál figyelembe kell venni az asszociatív viszonyokat” (Pléh és mtsai 2008: 833).

Az elmondottakon túl, a hálózatos modellek hozzájárulhatnak a számítógépek számára értelmezhető és kezelhető fogalmi rendszerek kialakításához (vö. Kovács 2013: 201).

A fentebbiekkel összefüggésben, a szóasszociációs vizsgálatok száma folyamatosan növekszik. Cramer például már az 1968-as, a szóasszociációkról írt dolgozatában mintegy 300, 1950 és 1965 között keletkezett szóasszociációs kísérlet eredményét veszi számba (vö. Kovács 2013: 34).

A szóasszociációs adatbázisok elkészítésének több módszerét különböztethetjük meg aszerint, hogy az adatfelvétel hogyan történik, milyen instrukciókat kapnak az adatközlők. Ezeket a módszereket többféleképpen is csoportosíthatjuk. Gósy (2005) alapján Kovács (2013: 36) a következőképpen osztályozza az adatfelvétel típusait: szabad- és szűkített asszociáció, valamint az ún. nyitott teszt. Szabadnak tekinti az asszociációt, ha a válaszadó egyetlen egyszavas választ adhat minden további megkötés nélkül, ha meghatározott időn belül több választ adhat, vagy ha mindenféle megkötés nélkül azt és annyit közöl, amennyit szeretne. Szűkített az asszociáció, ha az adatközlő egy a számára eleve megadott kategóriából „választhat” asszociátumot (pl. ellentét, hasonló hangzás stb.). És végül, nyitott tesztnek nevezzük az adatfelvételt akkor, ha az asszociációt – a fentebbi típushoz hasonlóan – egy meghatározott jelentéskategórián belül kell ugyan az adatközlőnek elvégeznie (pl. állatok csoportja), azonban a hívószóra tetszőleges számú választ adhat.

Pléh és mtsai (2008: 835) fluenciateszteknek nevezi az adatfelvétel azon típusát, amikor az adatközlőnek meghatározott ideig kell asszociátumokat adni a hívószóra. Fluenciateszt elképzelhető úgy is, hogy az alanyok nem szavakat kapnak ingeranyagként, hanem betűket, és ezzel mint kezdőbetűvel kell szavakat produkálnia. (Ez utóbbi megoldással a jelen disszertációban részletesen nem foglalkozom, l. pl. Pléh (2008: 835).)

Bár az egyválaszos szóasszociációs adatfelvétel a mentális lexikon kutatásának nem új keletű módszere, az utóbbi néhány évtizedben jelentős változáson ment keresztül (vö.

Lengyel 2008: 18). Ennek az az oka, hogy a mentális lexikon hálózatos modellezésében egyre nagyobb teret kap a kisvilágok elmélete és a többdimenziós szemlélet (részletesebben l. 4.6.2). A szóasszociációs adatbázisban a szavak közötti kapcsolatok a kisvilágok általános elméleti keretében is elemezhetőek, értelmezhetőek (vö. Lengyel 2008: 18). A mentális lexikon többdimenziós felépítésével összefüggésben pedig az egyes szavak, valamint szócsoportok között egyszerre több és többféle kapcsolat is létezhet. Ezek a kapcsolatok „részben lingvisztikai (fonetikai, grammatikai, szemantikai) tartalmúak [...], részben használatnyelvészeti (pragmatikaiak, szociolingvisztikaiak, szociokulturálisak)” (Lengyel 2008: 18).

A szóasszociációs adatbázisok vizsgálati módszereit illetően a következőket mondhatjuk el: a hívószókra kapott asszociációkon kvantitatív és kvalitatív vizsgálatok egyaránt végezhetőek (vö. Pléh 2008: 799, Kovács 2013: 40). Kvalitatív szempontú vizsgálatot végzünk például akkor, ha azt vesszük górcső alá, hogy melyek a leggyakrabban adott első asszociációk. Kvalitatív megállapításokat tehetünk akkor is, ha azt vizsgáljuk, hogy a hívószó és a válaszszó közötti kapcsolat milyen típusú (pl. paradigmatis, szintagmatis, alá-, fölérendelt, ellentét, szinonima, fogalmi közelség stb.). Fontos eredményeket hozhat, ha több stimulust vetünk össze egymással, és a rájuk adott válaszok hasonlóságait, illetve eltéréseit vizsgáljuk. Kvantitatív szempontból vizsgáljuk az adatbázist például akkor, ha megnézzük, hogy egy-egy hívószóra a fluenciatesztben összesen hány válasz érkezik, vagy ha azt tárjuk fel, hogy a különböző ingerszók összesen hány különböző válaszszót hívnak.

Kovács (2013: 40) alapján, az asszociátumok osztályozása legtöbbször a következő szempont szerint történik: fonetikai, szemantikai, paradigmatis vagy szintagmatis (vö. pl. Fitzpatrick 2006, Gósy 2005). Fitzpatrick (2006) a jelentésen (szinonimák, fogalmi kapcsolat, alá-, fölérendelt viszony), a mondatbeli pozíción (kollokáció), valamint a formán alapuló (ragozás, alaktani hasonlóság), továbbá az ún. hibás asszociációkat (nem magyarázható kapcsolat vagy nincs válasz) különbözteti meg. Gósy és Kovács (2001) ettől a megoldástól némileg eltérően hangzási (homonímia vagy részleges fonetikai azonosság), szerkezeti (szóképzés, szóösszetétel vagy szószerkezet), szemantikai (koordináció, alárendeltségi, ill. fölérendeltségi kapcsolat), valamint szófaji azonosságon alapuló kapcsolatokról beszél.

Lengyel Zsolt (2010: 196–198, 2012: 18–22) a hívószó és az asszociátum közötti kapcsolatokat a következőképpen rendszerezi és magyarázza: Alapvetően szófajmegőrző és szófajváltó kapcsolatokról beszél. A szófajmegőrzőek tagjai paradigmatis viszonyban

állnak egymással, és közöttük három altípust különböztet meg, szemantikai alapon: alá-, fölé- és mellérendelő viszonyt. A szófajváltó típusban levő párok legtöbbször szintagmatikus viszonyt képeznek (vö. Lengyel 2012: 18–22). (A kapcsolattípusokra a szóasszociációs adatbázis létrehozása és vizsgálata kapcsán még visszatérek, l. 6.3.3).

Ugyanakkor Szalay és Deese (1978: 145), valamint Chaffin és Herrmann (1988) tapasztalatai alapján Kovács (2013: 40) azt is megjegyzi, hogy a kapcsolatok besorolása nem mindig problémamentes, továbbá a kategorizálást maga a kutatási cél is befolyásolhatja, illetve meghatározhatja (vö. Kovács 2013: 40). Mindemellett az asszociációk típusokba sorolása nem feltétlenül automatizálható (vö. Szalay–Deese 1978: 145).

A továbbiakban, javarészt Kovács (2013: 37–38)-ra támaszkodva röviden összefoglalom a legfontosabb szóasszociációs adatfelvételi és vizsgálati eredményeket, különös tekintettel a hazai kutatásokra.

Az asszociációt befolyásolhatja a hívószó kategóriája (például, hogy pozitív vagy negatív szemantikai tartalmú szó, illetve köznyelvi szó vagy szakszó), gyakorisága, valamint ismertsége is (vö. Cramer 1968). Emellett a válaszokat illetően eltéréseket figyelhetünk meg többek között a válaszadó kora (vö. Libárdi 2001, Gósy 2000, Gósy 2001), neme (vö. Cser 1939, Kegyesné 2002, Rosenzweig 1970), képzettsége (a képzettség szintje és szakterülete) (vö. Cramer 1968, Rosenzweig 1970), valamint a nyelvismeret szintje (vö. Södermann 1993). Mindemellett, már a korai szóasszociációs vizsgálatok is foglalkoznak az ún. szemantikai előfeszítés problémájával, tehát azzal, hogy a teszt elvégzése előtti bármilyen cselekvés, esemény vagy információ befolyásolhatja a válaszadást (vö. Kovács 2013: 37).

A magyar szóasszociációs kutatások egy korai hazai példája Cser János (1939) kutatása, amelyben 10–14 évesek szabad asszociációit vizsgálta összesen 4483 gyermek esetében (vö. Lengyel 2012: 29). A szerző a kapott adatokból összeállította a 10–14 éves gyermekek szótárát, külön-külön feldolgozva és elemezve a fiúk és lányok által adott válaszokat. Papp Ferenc (1984/2006) magyar és orosz szóasszociációk sajátosságait veti össze egymással (vö. Kovács 2013: 38). Lengyel Zsolt (2012) magyar anyagon végzett szóasszociációs vizsgálatok eredményeit közli, azokat több szempontból (például a paradigmikus/szintagmatikus és a fonetikus-fonológiai kapcsolatok motiváltsága szerint) elemezve. Gósy Mária és Kovács Magdolna (2001) a 12–13 évesek körében végeztek adatfelvételt, és a kapott adatokat összevetették Cser (1939) azonos korosztályúaktól nyertadataival (vö. Kovács 2013: 38–39). Az összevető vizsgálat egy figyelemre méltó

eredménye, hogy amíg 1939-ben a vizsgálatban résztvevő gyerekek 37 912 szót aktiváltak, addig a 60 évvel későbbi kísérletben ez a szám 52 225-re emelkedett. Navracsics (2007, 2011) kétnyelvűek körében végzett egyválaszos szóasszociációs vizsgálatot. A létrehozott adatbázis nyelvtipológiai tárgyú kutatások fontos vizsgálati anyagául szolgálhat (vö. Lengyel 2012: 29).

Lengyel Zsolt 2000-től megismételte a Cser-féle kutatást. Az adatfelvétel eredményeit 2008-ban, szótári formába rendezve publikálta. Lengyel (2008) adatait elemezve Csomai (2012) arra is rámutat, hogy nemcsak az asszociációk száma módosult a két adatfelvétel között eltelt idő során, hanem az egyes fogalmi kategóriákon belül a fiúk és a lányok aktiválása is nagymértékben megváltozott az 1939-es adatokhoz képest. Az újabb adatokban ugyanis már nem olyan egyértelműen mutathatók ki a fiú és a lány kategóriák (vö. Kovács 2013: 39).

Folyamatos asszociációk kapcsán megfigyelt jelenség, hogy a válaszokat eltérően kell súlyoznunk. Szalay és Deese (1978) egy megismételt kísérlettel rámutatott ugyanis, hogy amíg az első kísérlet első helyén érkező válaszoknak a 61%-át a második alkalommal is aktiválták az adatközlők, addig az első kísérlet 11. helyén aktivált szavaknak csupán a 12,5%-a aktiválódott újra az ismétlésben.

Kovács (2013: 41–42) kiemeli Postman és Keppel (1970) szóasszociációs normákat tartalmazó kötetét, amely 15 év vizsgálataiból mutat be asszociációs normákat. A kötetben ismertetett kutatások több nyelven zajlottak, lehetőséget teremtve ezzel az asszociációs normák nyelvek közötti összevető elemzésére. A tesztek eredményeit többek között Kovács (2009) is összehasonlítja, és rámutat, hogy még a „leghétköznapiabb” szavakat is (pl. *ház*) más asszociációs normák jellemzik a különböző nyelveken. Malt, Sloman és Gennari (2003) alapján Kovács (2013: 41–42) szerint „a különböző nyelvek esetében a fogalmak megnevezését több (pl. kulturális, nyelvspecifikus) tényező befolyásolja, ezért a megnevezések nem fedik egymást”. Amint arra Eco (1999: 335–336) és Aitchison (1997: 68) alapján Kovács (2013: 42) rávilágít, bizonyos ekvivalensek eltérő képzetekkel kapcsolódnak az eltérő nyelvekben, és ezek az eltérések az asszociációs adatbázisokban tetten érhetőek. Ez alapján az asszociációs adatbázisok a fordításkutatásban is sikerrel alkalmazhatóak.

Kovács (2013: 42) véleménye szerint a Postman és Keppel kötetében (1970) publikált kutatások három legfontosabb megállapítása a következő: a válaszok többnyire a hívószóval kapcsolatos tárgykörből származnak, az esetek nagy többségében a párok egymást hívják elő, valamint, hogy a felnőttek esetében a válaszsónak legtöbbször

ugyanaz a szófaja, mint a hívószónak. Ugyanakkor, a kötetben bemutatott kísérletek túlnyomó többségében a hívószók toldalék nélküli alakok voltak, amely sajátság jelentősen befolyásolta az asszociációs vizsgálatok kimenetelét (vö. Kovács 2013: 42). A toldalékolt forma ugyanis „kontextust feltételez, illetve kontextusba ágyazott tárolási módot” (Lengyel 2008: 24).

Végezetül említsük meg az ún. Agykapocs-rendszert, amelyet szóasszociációs adatok gyűjtésére, valamint a mentális lexikon hálózatos szerveződésének vizsgálatára hozott létre Kovács László (2008). Az Agykapocs-kutatást részletesen tárgyalja Kovács (2013: 99–131). A projekt egy online felületen szabadon elérhető, az asszociációk megadásában bárki részt vehet (vö. Kovács 2013: 103–105). Az adatközlő szabad asszociációs tesztet végez, és az ingerszókra írásban ad választ. A rendszer jelenleg 11 nyelven (angol, francia, holland, lengyel, lett, magyar, német, olasz, orosz, spanyol, szlovák) fogad asszociációkat. Az Agykapocs-kutatás fontos ismérve, hogy a hívószók bármilyen nyelvtani alakban szerepelhetnek, tekintve, hogy a korábbi válaszszókból folyamatosan bővül a hívószók köre (vö. Kovács 2013: 42). Kovács (2013: 42) erre vonatkozóan, valamint Aitchison (2003) fentebb ismertetett megállapítását illetően a következőkre hívja fel a figyelmet: az Agykapocs-adatbázis adatai alapján „a szótári alakú hívószók legtöbb esetben szintén szótári alakú, azonos szófajú válaszsavakat hívnak elő. Toldalékolt alakok viszont nagyobb százalékban aktiváltak más szófajú szavakat, többnyire szintagmatikus kapcsolatok formájában”, például *boltban – vagyok, könyvet – olvas*. (A jelen fejezetben tárgyalt vizsgálati módszerekre és eredményekre visszatérek a disszertáció egy későbbi, a szóasszociációs adatfelvétel módszerét, hipotéziseit és eredményeit tárgyaló részében, l. 6.3.3.)

6.3 A kutatáshoz létrehozott adatbázisok és a vizsgálati eredmények bemutatása

A disszertációban tárgyalt vizsgálatokhoz szükséges adatbázisok a jelen kutatómunka keretében készültek. A kutatáshoz három vizsgálati eszköz jött létre: a szentimentek szintjén kézzel annotált korpusz, az értékvesztés és az értékváltás jelenségei szempontjából annotált korpusz, valamint a szóasszociációs adatbázis, amely az értékvesztésre és az értékváltásra képes elemekre adott asszociációkat tartalmazza.

A jelen fejezetben e vizsgálati eszközöket mutatom be. Ismertetem létrehozásuk okát és célját, valamint a korpuszok esetében a bennük feldolgozott szövegtípusok sajátosságait. Mind a korpuszok, mind az asszociációs adatbázis kapcsán ismertetem a munka alapelveit és pontos menetét, eszközeit, valamint a minőségbiztosítási megoldásokat. Ezt követően közlöm az alapvető és részletes statisztikai adatokat, végül pedig az elemzéseknek az értekezés szempontjából legrelevánsabb eredményeit.

6.3.1 A szentimentkorpusz

A korpusz hiánypótló magyar nyelvű kutatási és fejlesztési eszköz. Tudomásunk szerint ez az egyetlen olyan magyar nyelvű szentimentkorpusz, amely fragmentum-, entitás- és aspektusszintű annotációval rendelkezik, emellett az annotáció keretében számos további, a szentimentelemzés szempontjából releváns nyelvi elem is önálló taget kapott (vö. Szabó–Vincze 2015: 221–222, Szabó és mtsai 2016a: 176–177). A korpusz sokrétű annotációja számos, a nyelvi értékelés elemzésében kardinális jelenség (pl. a negáló vagy a fokozó elemek hatása, az entitások és az aspektusok összefüggései, a targetek és a szentimentkifejezések kapcsolódási módjai) elemzésére alkalmassá teszi az adatbázist. Így módon a korpusz fontos szerepet tölthet be a szentimentelemzéssel kapcsolatos kutatási és fejlesztési feladatokban egyaránt.

6.3.1.1 A korpusz létrehozásának az oka és célja

A disszertációban tárgyalt szentimentérték-változási jelenség beható vizsgálatához nélkülözhetetlen volt egy a kutatási céloknak megfelelő annotációval rendelkező szentimentkorpusz létrehozása.

Amint arról már korábban említést tettem (l. 5.2.2), a disszertációhoz készített szentimentkorpuszt megelőzően egyetlen olyan magyar nyelvű korpuszról volt tudomásunk (vö. Miháltz 2013), amelyben a korpusz építői a munka során a szentimentek annotálását célozták. Az OpinHuBank kutatási és fejlesztési célokra szabadon hozzáférhető eszköz, amelynek anyagához az ún. OpinHu projekt (vö. Miháltz 2010) adatbázisának 500 hazai online forrásból (híroldalak, blogok, fórumok) származó, több millió szövege szolgáltatott alapot. A készítők a szövegeket első lépésben mondatokra bontották,

tokenizálták, morfológiailag elemezték, szófajilag egyértelműsítették, végül névelemfelismerést hajtottak végre rajtuk. Ezt követően az adatbázisból kiszűrtek 12 ezer mondatot a kézi feldolgozáshoz.

A kiválasztás egy fontos kritériuma volt, hogy tartalmazzon legalább egy, PERSON (személynév) típusú entitást, ugyanis a korpusz létrehozásának céljai között szerepelt a szentimentérték targetfüggő kezelésének a gépi tanítása is. Azt az esetet, amikor a korpusz egy adott mondata egyszerre több entitást is tartalmazott, a következő módszerrel igyekeztek kezelni a készítők: „minden mondat minden különböző entitás-előfordulása egy-egy különböző annotációs egység lett” (Miháltz 2013: 344).

Mivel a fejlesztők olyan korpuszt kívántak létrehozni, amelyben a PERSON típusú entitások a szentimentek targeteinek a szerepét töltik be, automatikus módszerrel kivették azokat az eseteket, ahol az entitás nagy valószínűséggel a mondatban megjelenő vélemény forrása volt, majd kézi ellenőrzéssel kiszűrték azokat a példákat, ahol az automatikus megoldás hibát vétett.

A különböző szűréseket követően megmaradó egységekből kiválasztották az első 10 ezer adatot, és ezen végeztették el öt annotátor bevonásával a manuális feldolgozást (vö. Miháltz 2013: 344). A feldolgozás során az egységek negatív, pozitív vagy semleges értéket kaphattak. Az annotáció fontos alapvetése volt, hogy az egységek szentimentértékét az azokban szereplő PERSON típusú entitás vonatkozásában ítélik meg.

Ugyanakkor, az elkészült korpusz több lényegi sajátsága miatt elemzési és tesztelési célokra csupán korlátozottan alkalmazható (vö. Szabó–Vincze 2015: 219–220). Tekintsük ezeket a sajátságokat részleteiben is! Mindenekelőtt, a szövegekben a szentimentkifejezéseket, a targeteket és a probléma szempontjából lényeges egyéb jelenségeket önálló tagekkel nem annotálták a korpusz építői. Az adott elemzési egység szentimentértékét (vagy annak semlegességét) ugyanis a teljes egység szintjén határozták meg, az azon belüli további elemzés nélkül. Mindez azért problematikus, mert az annotáció így nem tárja fel a nyelvi értékelő kifejezések belső szerkezetét, valamint az értékelő tartalmak és a targetek viszonyát az egységen belül. Másrészt – amint azt ismertettem – az annotátoroknak a szentimentértékről az adott egységben szereplő tulajdonnévi entitás viszonylatában kellett döntést hozniuk: arra kérték őket, hogy ítélik meg, vajon pozitív vagy negatív ítéletet fejez-e ki az elemzett mondat a bennfoglalt személynévi entitás vonatkozásában. Ez a korpuszfeldolgozási megoldás két jelentős problémát is eredményez: Mindenekelőtt, egy nyelvi értékelő kifejezés targetének szerepét a személynév típusú entitáson kívül számtalan elem (pl. hely, esemény, termék vagy akár a termék egy

aspektusa is) betöltheti (részletesebben l. fentebb, 5.2.1.1). Az alapvetés tehát, miszerint a korpuszban kizárólag személynév töltheti be a target szerepét, nyilvánvalóan jelentősen korlátozza az eszköz alkalmazhatóságát. A helyzetet az építőknek az a törekvése sem oldja meg, hogy kiszűrjék azokat az eseteket, ahol az entitás nagy valószínűséggel a mondatban megjelenő vélemény forrása volt. Egyrészt azért nem, mert attól, hogy ezeket az eseteket kiszűrjük, még számtalan olyan eset lehetséges, ahol a személynév nem a vélemény forrása ugyan, ennek ellenére mégsem tölti be a target funkcióját. Tekintsük az alábbi példát (a példában a tulajdonnévi entitásokat dőlt szedéssel emelem ki)!

- (72) És mivel állandóan megmorogtam a magyar "Románia-politika" ügyetlenségét, *Karikás Péter* egyik barátja végül bevitt *Aczél György*höz.
[MNSZ2, szépirodalom, erdélyi, article]

A fenti példa két tulajdonnévi entitást is tartalmaz, azonban a mondatban negatív értékelése egy harmadik entitás, *a magyar "Románia-politika"* viszonylatában fogalmazódik meg. A példához hasonló adatokat tehát az OpinHuBank korpuszhoz használt alapvetéssel nem lehet kiszűrni.

A másik sajátság, amiért az alapvetés nem oldja meg az említett problémát, az az, hogy a korpuszból a törekvések ellenére sem sikerült kiszűrni minden olyan mondatot, ahol a vélemény forrása a tulajdonnévi entitással azonos. Bár a korpusz készítői törekedtek arra, hogy ezeket az eseteket kivegyék az anyagból, a korpusz sajnálatos módon számos ilyen esetet tartalmaz. Tekintsük az alábbi, az OpinHuBank-ből származó példát (a példában a tulajdonnévi entitást itt is dőlt szedéssel emelem ki)!

- (73) *Martonyi János* leszögezte: noha a jelenlegi szlovák kormánykoalíció egyik pártjának vezetői gyakran elfogadhatatlan kijelentéseket tesznek, a magyar kormány nem ilyen stílusban fog reagálni [...]

A fentebb idézett, a korpuszból származó példában tehát a tulajdonnév nem az értékelés targete, hanem a forrása. Tekintettel arra, hogy a készítők az annotátorokat arra kérték, hogy az egységek szentimentértékét az azokban szereplő tulajdonnévi entitás vonatkozásában ítélik meg, a fentebb elmondott sajátságok sok annotációs hibának is teret engedhetnek.

Az OpinHuBank feldolgozásának a mélységét, valamint az említett problémákat megfontolva azt a lehetőséget, hogy a korpuszt a szentimentelemzéssel kapcsolatos kutatási feladatokban kiaknázzuk, elvettem. Ugyanakkor, mivel más magyar nyelvű szentimentkorpusz nem állt a rendelkezésünkre, nélkülözhetetlenné vált egy olyan manuálisan annotált szentimentkorpusz létrehozása, amely képes a szentimentelemzés problémaköréhez kapcsolódó elméleti nyelvészeti és nyelvtechnológiai kutatásokat és fejlesztéseket hatékonyan támogatni.

6.3.1.2 A szövegválasztás indoklása, a vizsgált szövegtípus jellemzése

A magyar nyelvű szentimentkorpuszt termékvélemény-szövegekből hoztuk létre kutatási és fejlesztési céllal (vö. Szabó–Vincze 2015, Szabó és mtsai 2016a, 2016b). A korpusz szöveganyagát a [divany.hu/] honlap termékvéleményeiből állítottuk össze. Amint már utaltam rá (l. 5.1), a szentimentelemzésben a leggyakrabban feldolgozott szövegtípusok között a különböző közösségimédia-szövegeket találjuk, így például a tweeteket, a blog-, valamint a Facebook-bejegyzéseket (vö. pl. Godbole és mtsai 2007, Poursepanj és mtsai 2013, Ortigosa és mtsai 2014). A globális kommunikáció korában a legnépszerűbb közlési formának az online információátadás tekinthető (vö. Елькина 2009). Jelkina (Елькина 2009) az online kommunikációnak két alapvető típusát különbözteti meg: a szinkrón és az aszinkrón közlést. A két kommunikációs formát eltérő verbális és nonverbális eszközök használata jellemzi. A szinkrón internet-kommunikáció során a tagok azonos időben vesznek részt a diskurzusban. E típushoz tartozik például a chat vagy a különböző audio- és videokonferenciák. Ezzel szemben az aszinkrón kommunikáció során az információ közlése és fogadása között lehetséges az időbeli eltérés. Itt említhetjük például a fórum-hozzászólásokat, az elektronikus levelezést, az audio- és videoüzeneteket vagy a különböző blogbejegyzéseket.

A szentimentkorpuszunk összeállítása szempontjából kiemelkedő szerepű kommunikációs forma volt a blogok, azaz az online naplók, amelyekben tetszőleges hosszúságú és témájú szövegek tehetők közzé.⁶³ A naplókra jellemzően kronológiai sorrendben közli a bejegyzéseket, azokat a publikálás dátumával együtt tárolva.

⁶³ Maga a *blog* szó az angol nyelvű *web log* 'online napló' kifejezésből származik (vö. Елькина 2009).

A blogok egy része egyetlen szerzőhöz tartozik, és általában személyes tartalmakat közöl. Ugyanakkor saját blogot gyakran nagyobb csoportok is működtethetnek, például azonos érdeklődésű emberek vagy különböző vállalatok és intézmények (vö. Елькина 2009). Ebben az esetben a blog tartalma természetesen nem személyes jellegű, hanem a közösségi érdeklődést, illetve érdekeket kívánja kiszolgálni.

A blogok lehetnek nyitottak és zártak (vö. Елькина 2009). Az előbbi esetben a bejegyzések bárki számára szabadon olvashatóak, míg az utóbbi esetben az elérés regisztrációhoz kötött. A legtöbb blogoldalon, az olvasók számára adott a kommentelés lehetősége is, ami azt jelenti, hogy az egyes bejegyzésekhez kapcsolódóan, rövid szövegekben megfogalmazhatják véleményüket, észrevételeiket. Népszerű blogok esetében a kommentekben egész vitafolyamok is kialakulhatnak, amelyekben a blog olvasói egymás között vagy a blog szerzőjével diskurzust folytatnak (vö. Елькина 2009).

A blogok egyre nagyobb népszerűségnek örvendenek, s ezzel összefüggésben ma már hatalmas mennyiségű blogszöveg áll a kutatók és a fejlesztők rendelkezésére.

Az egyik legnépszerűbb blogtípus a tematika szempontjából a termékvéleményeket megosztó oldal. Az online elérhető termékvélemények növekvő szerepét Hu és Liu (2004) a következőképpen magyarázza: Az online kereskedelem egyre nagyobb mértékű terjedésével a termékvélemények is egyre jelentősebb befolyásoló erővel bírnak. Az online vásárlások száma növekszik, és ezzel összefüggésben mindinkább fontossá válik, hogy a vásárlók tapasztalatot tudjanak cserélni egymással a különböző termékek vonatkozásában. Ezzel összefüggésben egyre több olyan online oldal jelenik meg, amely specifikusan különböző gyártók termékeinek a véleményezése, összehasonlító értékelése kíván platformot teremteni (pl. a *kremmania.hu* vagy a *testbike.hu*).

A termékvélemények különösen hasznosak lehetnek akkor, ha a kivonatolásukat automatikus eszköz támogatja az olvasó számára (vö. Hu–Liu 2004). Ez is egy jelentős oka annak, hogy a szentimentelemzéssel kapcsolatos kutatási és fejlesztési munka során gyakorta termékvélemény-szövegek feldolgozásával foglalkoznak a téma szakértői.

A fentebb elmondottakkal összefüggésben, a jelen korpusz szöveganyagának a megválasztása a szentimentelemzéshez kapcsolódó kutató és fejlesztő munka szempontjából helytállónak tekinthető. Egyrészt, a *divany.hu* oldal szövegei a blog műfajához tartoznak, másrészt, az oldalon közölt tartalmak a szentimentelemzés céljainak megfelelő termékvélemény-szövegek.

6.3.1.3 A nyers korpusz alapvető adatai

Adivany.hu honlap készítői időközönként bizonyos termékcsoportokat tesztelnek, s közzéteszik a tesztelők véleményét. (74) alatt közlök két, egy negatív (74a) és egy pozitív (74b) értékelést megfogalmazó részletet a nyers korpuszból.

(74) a. Elképesztően kellemetlen íze van, előfordult, hogy valaki nem is ismerte fel benne a csokoládé ízt. Legtöbbször az hangzott el, hogy mellékíze van, édes és bővli termék. Sajnáljuk azokat a gyerekeket, akik ezt kapják. Igaz, a csomagolása elég kecsegtető, de ez ne tévesszen meg senkit!
[divany.hu/2013/karacsony/adventinaptarteszt]

b. Nem maradt kellemetlen, zsíros érzés a bőrömön használat után. Az illata nagyon kellemes, nem kell nagy mennyiség belőle és úgy érzem, jól tisztít. Nem szárítja a bőröm, így használata után nem is kell hidratálót felkennem, üde, jó érzés. Ráadásul nagyon olcsó, az 50 milliliteres mini kiszerelés csak 199 forint!

[divany.hu/tejbenvajban/2014/08/28/arclemoso_teszt_tej_vagy_gel]

A korpusz két annotálási fázisban készült. Az első szakaszban az adatbázis mintegy negyed részét annotáltuk, majd megvizsgáltuk, hogyan képes támogatni az elkészült korpusz a tervezett kutatási és fejlesztési célokat. Ezt követően átdolgoztuk az annotálási elveket, majd a korpusz teljes szöveganyagát két annotátor segítségével feldolgoztuk kézzel (a korpusz annotációjáról részletesen l. 6.3.1.4).

Az elkészült adatbázis összesen 154 véleményszöveget, 17 059 mondatot és 251 202 tokenet tartalmaz.

6.3.1.4 A korpusz feldolgozásának a módszere és eszköze⁶⁴

A korpusz annotálását két annotátor hajtotta végre, a korpusz annotálásához készített, annak elveit részletező annotálási útmutató alapján. Ez azt jelenti, hogy a munka

⁶⁴ A korpusz annotálási útmutatóját l. 4. számú melléklet.

megkezdése előtt a korpusz szöveganyagát felosztottuk az annotátorok között, azonos korpuszrészletet tehát csupán az egyetértésméréshez annotáltunk (részletesen l. 6.3.1.5).⁶⁵ Az annotálás eszközeként egy a Szegedi Tudományegyetem Természettudományi és Informatikai Karának Informatikai Intézete által fejlesztett annotáló program szolgált.

Amint azt korábban már részletesen tárgyaltam (l. 5.2.1.1), a szentimentelemzésben a feldolgozási egységek terjedelme szerint alapvetően a következő három szintet különböztetjük meg: a dokumentum-, a mondat-, valamint az entitás- és aspektusszintű annotációt (vö. Liu 2012: 10–11). A dokumentumszintű elemzés célja, hogy a teljes szöveg viszonylatában megadja a szentimentértéket, tehát azt, hogy a szöveg összességében pozitív vagy negatív értékelést fejez-e ki. A mondat- vagy tagmondatszintű elemzés a szöveg egyes mondatainak vagy tagmondatainak a vonatkozásában kívánja meghatározni ezt az értéket. Bár a mondat-, illetve tagmondatszintű osztályozás kétségtelenül árnyaltabb feldolgozást tesz lehetővé a dokumentumszintű osztályozásnál, igazán hatékony szentimentelemzést egyik megoldással sem érhetünk el. Tudniillik, egyik fentebbi eljárás móddal sem tárhatjuk fel azt, hogy a szövegben megfogalmazott értékelés pontosan mire, mely targetre irányul, emellett további problémákat eredményez, ha az egyetlen egységként kezelt szövegrész (dokumentum, mondat vagy tagmondat) több targetet és/vagy több szentimentet is tartalmaz (vö. Szabó és mtsai 2016a: 175, 2016b: 2873).

A fentebb elmondottakkal összefüggésben, a leghatékonyabb elemzési megoldást az ún. aspektusszintű feldolgozás kínálja. Ez a megoldás ahelyett, hogy a szöveg valamely szerkezeti egységét venné az elemzés alapjául, kifejezetten magukra az értékelésekre, az értékelést kifejező nyelvi elemekre koncentrál, és a feldolgozás alapegységét egy target, valamint az arra vonatkozó szentiment kapcsolatában határozza meg (vö. Liu 2012: 11, Szabó és mtsai 2016a: 175–176).

Az aspektusszintű annotálás még sokrétűbb, ezáltal még több kutatási és fejlesztési feladatot lehetővé tevő változata az, ha az entitásokat és azok különböző aspektusait is eltérően kezelik a készítők. Ezt a módszert azonban a korpusz feldolgozásának első szakaszában még nem, csupán a másodikban alkalmaztuk. Tekintsük a korpusz feldolgozásának folyamatát részleteiben is!

A korpusz első annotálási fázisában a teljes értékelő kifejezést (másképpen: szentimentfragmentumot), azon belül a pozitív és a negatív értékű szentimentkifejezéseket, azok targeteit, valamint esetleges shiftereit jelöltük be a korpuszban (vö. Szabó–Vincze

⁶⁵ A kutatás anyagi keretei nem tették lehetővé, hogy a teljes korpuszt végigannotáltassuk a két annotátorral.

2015: 222–223).⁶⁶ A *fragmentum*, a *target*, a *szenimentkifejezés*, valamint a *shifter* fogalmát a disszertáció 3. fejezetében közölt meghatározások szerint értelmeztük, egyetlen olyan eltéréssel, amelyet a korpusz későbbi vizsgálati lehetőségei szempontjából fontosnak tartottam. Eszerint nem csupán az inherens, a szótár szintjén szenimentértékkel rendelkező elemeket annotáltuk szenimentkifejezésként, hanem mindazokat, amelyek az aktuális kontextusban értékelő szemantikai tartalommal rendelkeztek (részletesebben l. lentebb).

*Szenimentfragmentum*ként annotáltunk minden olyan szövegrészt, amely egyetlen meghatározott (pozitív vagy negatív) értékelést fejezett ki egyetlen (esetleg több, egymással mellérendelő szintaktikai viszonyban álló) target vonatkozásában (vö. Szabó és mtsai 2016a: 176).⁶⁷ E megoldás következtében a fragmentumok az esetek túlnyomó többségében mondatnyi, vagy a mondatnál kisebb egységek, azonban ritkán az is előfordul, hogy egy fragmentum átível a mondat határán.

A szenimentfragmentum annotálását követően, azon belül több, a szenimentelemzés szempontjából releváns nyelvi elemet eltérő taggel láttunk el.

A nemzetközi számítógépes nyelvészeti terminológia alapján *target*nek jelöltük azokat a nyelvi elemeket, amelyekre az adott értékelés irányult (vö. Szabó–Vincze 2015). A feldolgozott szövegek sajátsága okán úgy döntöttünk, hogy a tesztelt termékek címszerű elnevezéseit <topic> címkével látjuk el, míg az egyes szenimentekhez kapcsolódó targetek <target> címkét kapnak.

A *szenimentkifejezések* esetében eltértünk az általam korábban megadott definíciótól. Szenimentkifejezésként kezeltük azokat az egy szóból álló, vagy állandósult többszavas szókapcsolatokat, amelyek értékítéletet fejeznek ki valamely target vonatkozásában (vö. Szabó és mtsai 2016a: 176). Ugyanakkor ezzel a taggel nem csupán azokat a nyelvi elemeket annotáltattuk, amelyek lexikai szinten egyetlen meghatározott szenimentértékkel (pozitív vagy negatív) rendelkeznek (pl. *jó*, *rossz*, *szép*, *csúnya*) (vö. Szabó 2014, 2015c), hanem arra kértük az annotátorokat, hogy amennyiben egy adott elem az adott targetre vonatkozóan vagy az adott kontextusban meghatározható módon pozitív vagy negatív tartalmat hordoz, úgy azt is annotálják a megfelelő taggel. Ezt a megoldást azért tekintettem célravezetőnek, mert ezáltal a korpusz a kontextusfüggő szenimentértékek vizsgálatára is lehetőséget ad.

⁶⁶ Köszönöm Simkó Katalinnak és Varga Viktornak a korpusz annotálásának elvégzését.

⁶⁷ A mellérendelések annyira kis számban fordultak elő az előzetes próbaannotációk során, hogy úgy véltük, különleges kezelésük nem indokolt.

A nyelvtechnológiai irodalom terminológiájával (vö. pl. Polanyi–Zaenen 2004, Ikeda és mtsai 2008, Li és mtsai 2010, 2013, Boubel és mtsai 2013) összhangban *shifterként* annotáltattunk minden olyan nyelvi elemet, amely egy adott szentimentfragmentumban szerepelve eltérést okoz a szövegben szereplő szentimentkifejezések lexikai szintű értéke és a fragmentum értéke között. A szentimentshifterekben belül a negáló, a növelő és a csökkentő értelmű fokozó, valamint az ún. irreáló elemeket különböztettük meg (vö. Szabó–Vincze 2015: 221).⁶⁸ Negálóként annotáltunk minden olyan elemet, amely vagy az ellenkezőjére változtatta az adott szentimentkifejezés értékét (75a), vagy pedig törölte azt (75b) (vö. Kiefer 2008: 513) (A negáció hatásairól a (2.3.2.2)-ben már részletesen szóltam.) A példákban a vizsgált elemeket dőlt szedéssel emelem ki.

- (75) a. Mari *nem* szép. ('Mari csúnya')
- b. A béka *nem* gusztustalan. (nem jelenti azt, hogy 'gusztusos, tetszetős')

A fentebbi meghatározással összefüggésben negálóként kezeltük a tagadószókat (pl. *ne*, *sem*, *dehogy*), a létige tagadó alakjait (*nincs*, *nincsen*, *sincs*, *sincsen*), tagadó névutókat (pl. *hiányában*, *nélkül*), valamint bizonyos módosítószókat (pl. *aligha*, *látszatra*) is (vö. Pete 1999: 306). (Az annotált korpusz vizsgálatakor kiderült, hogy a negáló elemek funkcióját ezeken kívül sokkal több elem betöltheti, erről a fejezet korpuszelemző részében részletesen szólok majd, l. lentebb, 6.3.1.6.)

A szentimentértékek ún. fokozóiként annotáltuk azokat az elemeket, amelyek a szentimentkifejezések értékét erősítették (76a) vagy csökkentették (76b).

- (76) a. A hangminőség *nagyon* jó.
- b. A hangminőség *aránylag* jó.

A szentimentértékek fokozására és csökkentésére számtalan elem alkalmas lehet, pl. *rendkívül*, *rendkívüli módon*, *borzasztóan*, *elképesztően*, *valamennyire*, *valamelyest*, *feliből-nagyjából*, *részben*, *kevésbé* stb. (vö. Moilanen–Pulman 2007: 378, Székely 2007).

Az ún. irreáló elemek a szentimentelemzés szempontjából azért fontosak, mert képesek arra, hogy lehetetlenné tegyék az értékelést megfogalmazó szövegrész faktív

⁶⁸ Ez tehát egyúttal azt is jelenti, hogy az itt felsorolt elemeket kizárólag a fragmentumokon belül jelöltük. Ha a szöveg tartalmazott például egy negáló elemet, amely nem képezte fragmentum részét, nem annotáltuk.

olvasatát (vö. Pete 2002: 189, Brooke és mtsai 2009: 51). Másképpen, az irreálók megakadályozzák, hogy az adott értékelést tényszerű információként fogadjuk el. Az irreáló elemek kezelése azért fontos a szentimentelemzés folyamatában, mert a szentimentérték irreálása következtében az adott értékelő tartalmat nem kezelhetjük az értékeléselemzésben teljes értékű adatként. Brooke és munkatársai (2009) az angol nyelvben a modális segédigéket (pl. *would, could*), néhány igét (*expect, doubt*), valamint bizonyos központoszási jeleneket (pl. kérdőjel) tekintenek az irreálók közé tartozóaknak.

Vessük össze az alábbi faktív (77) és nem faktív (78) olvasatú példákat!

(77) A hangminőség jó.

- (78) a. A hangminőség *valószínűleg* jó.
b. *Lehet*, hogy a hangminőség jó.
c. Jó a hangminőség?
d. *Nem tudom*, hogy a hangminőség jó-e.
e. A hangminőség jó *lehet*.

A (77) alatti példában az értékelés megfogalmazója elkötelezi magát a proposíció igazsága iránt, addig a (78) alatti példák esetében nem. Minden olyan elemet tehát, amely azt jelöli, hogy az értékelés proposíciós tartalmát a beszélő nem tényként tekinti, külön taggel láttunk el a korpuszban. (vö. a jelenséget a (2.2)-ben és a (2.3.1)-ben tárgyalt modalitással.)

A jelen disszertációban vizsgált negatív emotív elemek ebben az annotációban értékvesztés esetén a fokozó értelmű intenzifikáló elemek között, értékváltás esetén pedig szentimentkifejezéseként jelennek meg.

A korpusz második feldolgozási szakaszában bizonyos mértékben eltérő annotálási megoldásokat alkalmaztunk a korábbi módszerhez képest (vö. Szabó és mtsai 2016a: 176–177). Ennek az volt az oka, hogy az első munkafázis, valamint a kezdeti korpuszvizsgálati tapasztalatok alapján úgy véltük, a targetek kezelésében kidolgozottabb, még részletesebb elemzést lehetővé tevő annotációra van szükség.

A korpusz első feldolgozási fázisában, a fragmentumszintű feldolgozásnak köszönhetően már a targetekkel összefüggésben kezeltük a szentimenteket. Ugyanakkor mélyebb, entitás- és aspektusszintű annotációt nem tudtunk megvalósítani, tekintettel arra, hogy az entításokat, valamint azok egyes aspektusait nem különböztettük meg az annotáció szintjén egymástól. Az újabb annotációs alapelv szerint a korpuszban előforduló

entitásokat és azok aspektusait eltérő annotációs taggel <target 1-20> láttuk el, következésképpen alkalmazva azokat az egyes dokumentumokon belül. A <target 1> címkével rendre az adott termék, azaz maga az entitás szövegbeli előfordulását jelöltük, míg a többi <target> címkével annak különböző aspektusait annotáltuk. A címszerűen előforduló terméknevek jelölésére, akárcsak a korpusz első annotálási fázisában, a <topic> címkét alkalmaztuk.

A bemutatott feldolgozási megoldás kettős hasznót hozott. Egyrészt, a módosítás eredményeképpen olyan annotáció jött létre, amely lehetőséget ad az entitás-aspektus-összefüggések gépi feltárására, segítve ezzel egy hatékonyabb automatikus szentimentelemző rendszer kidolgozását. Amint azt korábban már részletesen tárgyaltam (l. 5.2.1.1), az entitások és az aspektusok megkülönböztetése a szentimentelemzés szempontjából kardinális probléma. Nagy jelentőséggel bír ugyanis, hogy egy adott értékelés egy bizonyos entitáshoz vagy csupán annak valamely aspektusához kapcsolódik, illetve, hogy a szövegben megjelenő aspektusok közül pontosan melyikre vonatkozik (vö. Liu 2012: 11). A korábban (64) alatt közölt példákat itt (79) alatt megismétlem (a példákban a vizsgált elemeket dőlt szedéssel emelem ki).

- (79) a. Bár a *kiszolgálás* nem olyan jó, ettől függetlenül imádom ezt az *éttermet*.
b. A *kiszolgálás* nem olyan jó, a *berendezés* viszont barátságos.

A fentebbi példák egyszerre két ellentétes értéket fogalmaznak meg. Azonban a (79a) alatti példában a pozitív szentimentérték jelentősebb a negatív szentimentértéknél, tekintve, hogy az előbbi a teljes entitáshoz, az utóbbi pedig csupán annak egyetlen aspektusához kapcsolódik. Ugyanakkor a (79b) alatti példában mind a pozitív, mind a negatív értékelés az adott entitás egy-egy aspektusára vonatkozik. A korpuszban alkalmazott annotációs megoldást a jövőben az entitás-aspektus-kettősség automatikus vizsgálatában és feldolgozásában is ki szeretnénk aknázni.

Az új annotációs megoldás másrészt a targeteket érintő koreferenciaviszonyokat érintő kutatási és fejlesztési feladatokat is támogatja, hiszen egy dokumentumon belül egy adott entitást, valamint egy adott aspektust (és az esetlegesen azokat helyettesítő

névmásokat) rendre ugyanazzal a <target> taggel láttunk el (vö. Szabó és mtsai 2016a: 177).⁶⁹

Az alábbi részlet bemutatja a korpuszban alkalmazott végső annotálási megoldást:

```
Negyedik helyezett:
<topic>
    Kolios goat's cheese
</topic>
„<NegFrag>
    <target2>
        Állagra
    </target2>
    olyan, mint a
    <NegSent>
        gumi
    </NegSent>
</NegFrag>
,
<NegFrag>
    <target3>
        ízre
    </target3>
    pedig
    <NegSent>
        fanyar
    </NegSent>
</NegFrag>.
<NegFrag>
    Nekem
    <ShiftNeg>
        nem
    </ShiftNeg>
    <PosSent>
        jön be
    </PosSent>
</NegFrag>.”
```

2. ábra: Példa a szentimentkorpusz annotálási megoldására

⁶⁹ A targetek vonatkozásában alkalmazott feldolgozási megoldás tehát minden szempontból kielégíti a dolgozat bírálójának, Szécsényi Tibornak az igényeit arra vonatkozóan, hogy az entitások és az aspektusok, valamint azok egymáshoz képesti viszonyai megjelenjenek a korpusz annotációjában.

6.3.1.5 A korpusz annotálásának egyetértés-mérési eredményei

A korpuszannotáció minőségének biztosítása céljából – a korpuszok építésére vonatkozó nemzetközi sztenderdnek megfelelően (l. 6.1) – a teljes szöveganyag feldolgozása előtt annotátorok közötti egyetértésmérést végeztünk (vö. Szabó és mtsai 2016b). Ehhez a korpusz egy kisebb részét mindkét annotátor feldolgozta, majd az F-mértéket alkalmazva kiszámítottuk az egyetértést az annotátorok között (vö. Derczynski 2016).⁷⁰

alapján az annotátorok között 65,02%-os egyetértési értéket számoltunk. (Az F-mértékről a (6.1)-ben részletesen szóltam.)

Az annotátorok közötti egyezés messze a <topic> címke alkalmazásában volt a legmagasabb, tekintettel az ezzel a taggel annotált elemek szövegstruktúra-beli sajátosságaira (címszerű elhelyezkedésére). A legrosszabb ráta az irreáló elemek feldolgozásában mutatkozott, aminek kapcsán úgy véljük, hogy az eltérést az irreálók lexikai és strukturális sokszínűsége, nehezebben behatárolható kategóriája okozta.

6.3.1.6 Az annotált korpusz alapvető statisztikai adatai

Az alábbi táblázat összefoglalja az annotált korpusz alapvető statisztikai adatait.

⁷⁰ Jelen esetben a következők miatt döntöttük az F-mérték mellett a Cohen-féle kappa statisztika helyett: Amíg az utóbbi kiszámításához szükség van a valós negatív értékekre is (azaz arra, hogy ténylegesen hány olyan elem volt, amely helyesen nem kapott semmilyen annotációt), addig az F-mérték alkalmazása esetében nincs szükség erre. Ennek megfelelően például névelem-felismeréskor is ezt alkalmazzák, hiszen ebben az annotálási feladatban sem ismert előre, hogy a szövegben pontosan hány névelem fordul elő. Jelen esetben sem volt előre adott, hogy mennyi szentimentkifejezés lesz a korpuszban, így utóbbit választottuk.

A kalkulációt a következő alapon végeztük el: valós pozitívnak tekintettük azt az esetet, amikor amit mindkét annotátor ugyanolyan kategóriaként jelölt az adott elemet vagy szövegrészt (a határokat is beleszámítva). Fals pozitívnak vettük, amikor az egyik annotátor jelölte az adott kifejezést, míg a másik nem. Végül, fals negatívnak tekintettük az előbbi eset fordítottját, tehát, amikor az egyik annotátor nem jelölte az adott kifejezést, de másik igen.

Az egyetértés eredményeit az alábbi képletek segítségével számítottuk ki (vö. Derczynski 2016: 262):

$$P = \frac{|\text{true positives}|}{|\text{true positives}| + |\text{false positives}|}$$

$$P = \frac{|\text{true positives}|}{|\text{true positives}| + |\text{false negatives}|}$$

$$F_\beta = \frac{(1 + \beta^2) PR}{\beta^2 P + R}$$

annotált elem	összes előfordulás	pozitív fragmentumban	negatív frag- mentumban
PosFrag	6693		
NegFrag	8053		
PosSent	7554	5800	1754
NegSent	7698	1272	6426
Topic	1365		
Target	7827	3731	4096
Negation	3516	1460	2056
IntPlus	4973	2405	2568
IntMinus	1080	301	779
Irreal	991	296	695
OtherShifter	1134	642	492
ÖSSZES:	50921	15907	18866

3. táblázat: A korpuszannotáció alapvető statisztikai adatai

Az annotáció kvantitatív adatai alapján a következőket mondhatjuk el: A korpuszban összesen 14 746 szentimentfragmentumot annotáltunk, ebből 6693-at pozitívként és 8053-at negatívként. A fragmentumokon belül 7554 szentimentkifejezés kapott pozitív, 7698 negatív taget. <topic> címkét 1365 esetben, <target> címkét 7827 esetben annotáltunk a korpuszban. Összesen 11 694 elemet dolgoztunk fel shifterként, ebből 3516 negációt, 6053 növelő és csökkentő értelmű intenzifikálót, 991 irreálót és 1134 olyan elemet jelöltünk, amely az előző shifter-kategóriákba megítélésünk szerint nem volt besorolható.

6.3.1.7 A korpusz vizsgálati eredményei

Az annotált korpusz lehetőséget teremtett a különböző, a nyelvi értékeléselemzés szempontjából lényeges sajátságok és összefüggések feltárására a magyar nyelvű szövegek vonatkozásában. Így tehát, megismerhettük a különböző annotált elemek gyakorisági

sajátságait, szófaji megoszlását, egyedi elemeinek a számát (lexikai diverzitását), valamint azok pozitív és negatív fragmentum-beli megoszlási arányait.

A targetekre vonatkozóan a következő megállapításokat tehetjük: a <target1> címkét 2773 esetben annotáltuk. Ezzel jelöltük a korábban ismertetteknek megfelelően magukat az entitásokat. A többi, összesen 5054 <target> címkével az egyes entitások aspektusait jelöltük. Az entitások és az aspektusok gyakorisági előfordulása, különösen ez utóbbiak magas frekvenciája a korpuszban ugyancsak arra mutat, hogy egy termékvélemény-szövegek hatékony szentimentszintű feldolgozásához nélkülözhetetlen az entitások és az aspektusok megfelelő kezelése, eltérő súlyozása.

A szófaji megoszlási arányok tekintetében különösen figyelemre méltónak találtuk a negáló elemek sajátságait (vö. Szabó és mtsai 2016a: 179–180). Az annotált korpusz szófaji szempontú vizsgálata előtt azt feltételeztük (l. 6.3.1.4), hogy negáló elemekként alapvetően tagadószókat (pl. *ne, sem, dehogy*) és a létige tagadó alakjait (*nincs, nincsen, sincs, sincsen*), emellett kisebb számban tagadó szerepű névutókat (pl. *hiányában, nélkül*), valamint néhány egyéb módosítósót (pl. *aligha, látszatra*) fogunk annotálni (vö. Szabó–Vincze 2015: 221, Pete 1999: 306). A korpuszvizsgálat eredményei azonban megmutatták, hogy a negáló szerepű elemeknek nagyobb a szófaji változatossága a kezdeti hipotézisünkénél. A korpuszban annotált összesen 3516 negáló elem legnagyobb részét határozósók (2587), igék (468), névmások (145), valamint mellérendelő kötőszók (93) adták.

A negálóként annotált elemek lexikai diverzitása is figyelemre méltó. A korpuszban összesen 97 különböző, negáló taggel ellátott elemet találtam.⁷¹ Annak céljából, hogy érzékeltessem a negáló elemek változatosságát, közülük (80) alatt közlök néhány példát.

- (80) hiánya, elillant, nélkülözi, bizarr lenne azt állítani, helyett, hiányoltam, semmi köze sincs, nulla, nem érezni benne, lespórolták

Megvizsgálva a csökkentő és a növelő értelmű intenzifikáló elemek szófaji megoszlását megállapítottuk, hogy amíg az előbbi szerepét legtöbbször melléknév tölti be mind a pozitív, mind a negatív fragmentumokban, addig az utóbbi funkciójában az esetek többségében határozósót találunk (vö. Szabó és mtsai 2016a: 179). Ugyanakkor azt az

⁷¹ Az adat a különböző szóalakok számára vonatkozik. A nagy szám nem meglepő, ugyanis a negáció fogalmát az annotálás során tágran értelmeztük, a szentimentértékre gyakorolt hatásukra fókuszáltunk. Ennek megfelelően például a *tévedés volna azt állítani, hogy kifejezés is negáló taget* kapott.

érdekes megfigyelést tettük, hogy a csökkentő értelmű intenzifikáló elemek szerepében a melléknevek jóval ritkábbak (31,44%) a pozitív fragmentumokban a negatív fragmentumoknál (52,08%). A csökkentő értelmű intenzifikáló elemek esetében tehát a szófaji megoszlási arányok nem azonosak a pozitív és a negatív szentimentértékű fragmentumok között. Ezzel szemben a növelő értelmű intenzifikáló elemek szerepét a pozitív és a negatív fragmentumokban közel azonos gyakorisággal tölti be határozószó (52,32% és 47,69%).

Az irreálók szófaji vizsgálata megmutatta, hogy körülbelül kétharmadukat ige vagy határozószó teszi ki. A határozószók (pl. *állítólag, valószínűleg, talán*) párhuzamot mutatnak az úgynevezett episztemikus bizonytalanságot jelölő lexikai elemekkel (vö. Szarvas és mtsai 2012: 339, Vincze 2013: 385). Az igeik esetében vagy a fő szemantikai tartalma (pl. *tűnik, hasonlít, érez, gondol*), vagy valamely, a feltételeesség jelölésére szolgáló morfológiai eszköz segítségével jut kifejezésre az irreáló tartalom. Ilyen például a *-hAt* toldalék, amely a korpusz irreáló elemeinek a 11%-át teszi ki (vö. Szabó és mtsai 2016a: 180). A *ha* és a *mintha* kötőszavak is gyakori irreáló elemek a korpuszban, összesen 228 esetben (22%) annotáltuk őket.⁷²

A pozitív (PosSent) és a negatív szentimentkifejezések (NegSent) gyakorisági megoszlását illetően a következő legfontosabb megállapításokat tehetjük: A pozitív és a negatív szentimentkifejezések közel egyenlő arányban képviseltetik magukat a korpuszban, ugyanakkor ez nem eredményezi azt, hogy a pozitív (PosFrag) és a negatív fragmentumok (NegFrag) is hasonló megoszlási arányt mutatnának (vö. Szabó és mtsai 2016a: 178). A negatív fragmentumok ugyanis többségben vannak a pozitív fragmentumokkal szemben. Az eredmény – legalább részben bizonyosan – a shifterek szentimentelemzés-beni szerepére mutat rá, amelyek gyakorta a szentimentkifejezés, valamint az azt magában foglaló fragmentum szentimentértéke között eltérést okoznak.

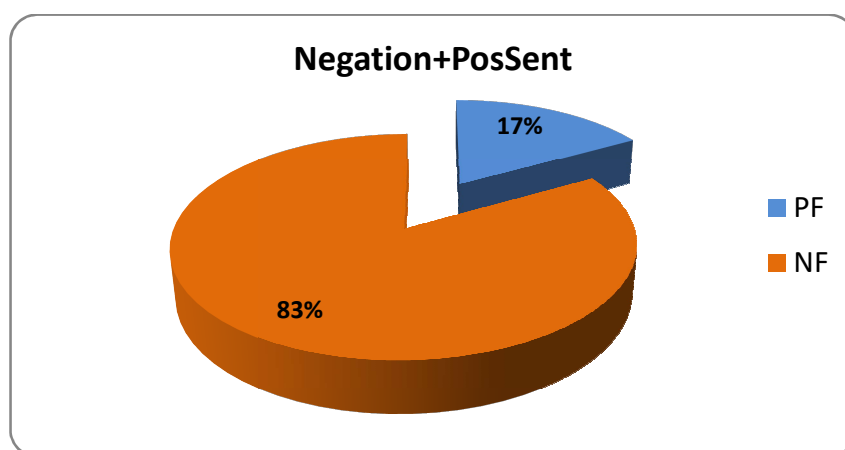
A fentebbi eredményekkel összefüggésben, a pozitív és a negatív szentimentkifejezések megoszlása a velük azonos szentimentértékű fragmentumokban nem egyforma (vö. Szabó és mtsai 2016a: 178). Amíg ugyanis a negatív szentimentkifejezések 83,47%-a azonos értékű fragmentumban szerepel, azaz csupán 16,52%-uk található pozitív fragmentumban, addig a pozitív szentimentkifejezések magasabb megoszlási aránnyal

⁷² A korpuszban megjelenő irreáló elemek szoros átfedést mutatnak az ún. nyelvi bizonytalanságot jelölő elemekkel, ezért célul tűztük ki, hogy a két elemcsoportot egy ezt követő kutatási lépésként összevessük. E kutatási folyamatba kapcsolódik az a vizsgálatunk, amelynek keretében a bizonytalanságjelölő elemek szótáraival az itt bemutatott szentimentkorpuszt elemeztük, és a különböző típusú szemantikai tartalmak gyakorisági összefüggéseit igyekeztünk feltárni (a kutatásról, valamint a vizsgálati eredményekről részletesebben l. Drávucz–Szabó 2017).

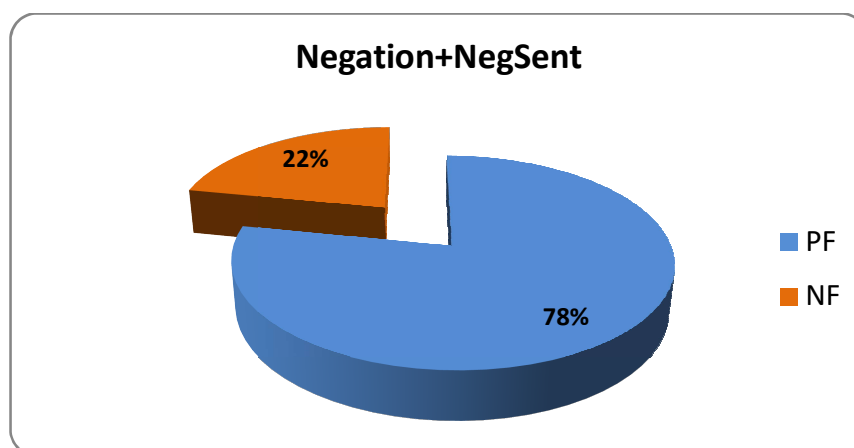
(23,21%) fordulnak elő a velük ellentétes értékű fragmentumokban. A kapott eredmények alapján, egy negatív értékelést negált pozitív szentimentkifejezéssel (pl. *nem jó*) gyakrabban fogalmazunk meg, mint egy pozitív értékelést negált negatív szentimentkifejezéssel (pl. *nem rossz*). A tapasztalatok illeszkednek az ún. Pollyanna-hipotézishez (vö. Boucher–Osgood 1969), amely nyelvi univerzáléként tételezi a pozitív töltetű kifejezések magasabb használati arányát a negatív töltetű elemekkel szemben (részletesebben l. 4.4).

A szentimentelemzés eredményességét az egyik legnagyobb mértékben befolyásoló, ugyanakkor az egyik legnehezebben kezelhető nyelvi jelenség a negáció. Hatékony automatikus feldolgozását két sajátosság is nehezíti: Egyrészt – amint arra korábban már rámutattam (l. 2.3.2.2), a negáló elemekkel alkotott szerkezetek gyakorta nem kompozicionálisak (vö. Israel 2004: 711, Feldman és mtsai 2010: 4). Ebből következően nem okoznak rendre eltérést a szentimentkifejezés értéke és a bennfoglalt fragmentum értéke között. Másrészt, imént láthattuk, a negáló elemek magas fokú lexikai diverzitást mutatnak, ami jelentősen megnehezítheti annak kezelését.

Azért, hogy pontosabb képet kaphassunk a negálók szerepéről, megvizsgáltam, hogy amennyiben jelen vannak a fragmentumban, vajon hány esetben okozzák ténylegesen a szentimentérték módosulását (vö. Szabó–Vincze 2016). Az elemzés eredményeit a 3. és a 4. ábrán mutatom be.



3 ábra: A negálók és a pozitív szentimentkifejezések együttes előfordulásának gyakorisági megoszlása a pozitív és a negatív fragmentumokban



4. ábra: A negálók és a negatív szentimentkifejezések együttes előfordulásának gyakorisági megoszlása a pozitív és a negatív fragmentumokban

A vizsgálat eredményei rámutatnak, hogy amennyiben egy adott fragmentum negáló elemet tartalmaz, az esetek túlnyomó többségében (83 és 78%-ában) a módosulás bekövetkezik, azaz a fragmentum szentimentértéke ellentétes lesz a benne szereplő szentimentkifejezés értékével (vö. Szabó–Vincze 2016: 276). A tapasztalatok alapján – bár a negáló elemek elméleti szinten csupán részben okozzák az érték megváltozását a fragmentumban (l. 2.3.2.2, valamint Israel 2004: 711, Feldman és mtsai 2010: 4) – ez a változás a korpuszban kimagasló rendszerességgel következik be. Mindez arra mutat, hogy a negáló elemek és a szentimentkifejezések kapcsolatainak kombinációs szabályokba foglalt alkalmazása – a vizsgált szövegtípusban legalábbis – jelentős mértékű javulást eredményezhet a kizárólagosan szótáralapon működő elemző algoritmushoz képest.

A korpuszannotáció azt is lehetővé tette, hogy megvizsgáljam a pozitív szentimentkifejezéseknek a negáló elemhez képesti pozícióját a negatív fragmentumokban (81a-b), valamint minden negatív szentimentkifejezésnek a negáló elemhez képesti pozícióját a pozitív fragmentumokban (82a-b) (vö. Szabó 2015a, Szabó–Vincze 2016: 277–279). Amint ugyanis azt már a korpusz vizsgálatának első tapasztalatai is feltárták, a negáló elemek nem feltétlenül a szentimentkifejezést közvetlenül megelőző pozícióban helyezkednek el, sőt, nem is feltétlenül a szentimentkifejezést megelőző pozícióban. Tekintsük az alábbi példákat!⁷³

⁷³ A jelen példákban a targetek nem szerepelnek, ezek rendre azok a termékek, vagy azoknak a termékeknek valamely aspektusai, amelyeket az annotált szövegekben aktuálisan értékelték a tesztelők. Ennek megfelelően automatikus azonosításukra az annotáció alapján két lehetőség kínálkozik: a topic-annotáció alapján a szentimenteket a termékek címszerű elnevezéseivel kapcsolhatjuk, vagy pedig azonosíthatjuk azt

- (81) a. mégis valahogy **nem**<negation> kapott **jó**<PosSent> pontokat⁷⁴
 b. éppen csak **jó**<PosSent> pontokat **nem**<negation> kapott
- (82) a. **nincsen**<negation> benne **hozzáadott cukor**<SentWordNeg>.
 b. **mellékíz**<SentWordNeg> **nélkül**<negation>

Amint látjuk tehát, a (81a) és a (82a) alatti példákban a negáló elem nem közvetlenül a szentimentkifejezés előtti pozíciót foglalja el, a (81b) és a (82b) alatti példák pedig azt mutatják meg, hogy esetlegesen a szentimentkifejezés után is kerülhetnek.

Az elemzés eredményeit az alábbi táblázat foglalja össze:

	pozíciók szerinti gyakorisági megoszlás							
	-4.	-3.	-2.	-1.	1.	2.	3.	4.
A negáló elem pozíciója a PosSent-hez képest NF-ban:	4,03%	11,21%	27,35%	43,49%	6,72%	4,48%	0,89%	0,44%
A negáló elem pozíciója a NegSent-hez képest PF-ban:	–	7,14%	23,07%	62,08%	3,29%	1,09%	1,09%	0,54%

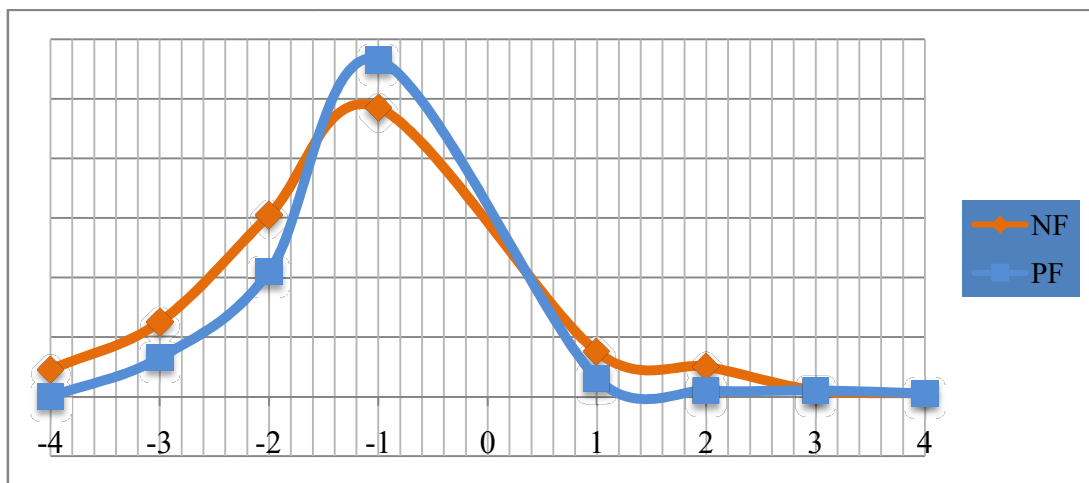
4 táblázat: A negáló elemeknek a szentimentkifejezésekhez képesti pozíciója az ellentétes értékű fragmentumokban

A vizsgálat eredményei alapján megállapítható, hogy a negáló elem a legtöbb esetben a szentimentkifejezést közvetlenül megelőző pozíciót foglalja el, azonban a -3.-ig nem elhanyagolható a többi, szentimentkifejezés előtti pozíció sem (vö. Szabó–Vincze 2016:

az aktuális fragmentumot megelőző fragmentumban szereplő targettel. Olyan jellegű vizsgálatot, amely feltárná, hogy mely megoldás vezet helytállóbb eredményre, eddig nem végeztünk.

⁷⁴ A dolgozat bírálója, Szécsényi Tibor megjegyzi, hogy ebben a példában „a tagadószó az értékelésben részt nem vevő ígét módosítja”. A szentimentkorporusz annotálásakor a módosító elemek szerepét a teljes értékelő fragmentum szempontjából vizsgáltuk. Ennek következtében nem feltétlenül kellett a negálónak a szentimentkifejezést negálnia, minden esetben a teljes szemantikai tartalomra tett hatását vizsgáltuk. Mindemellett, számomra ez a sajátság az aktuális probléma szempontjából nem volt fontos: a jelen esetben azt vizsgáltam, hogy a fragmentumon belül a negáló elem a szentimentszótárral megtalálható szentimentkifejezéshez képest hol helyezkedik el. Tekintettel arra, hogy nem szintaktikai, hanem memóriaalapú elemzéssel kívántam megoldani az automatikus elemzést (l. 7), számomra nem volt fontos a kapcsolódási mód, csupán az egymáshoz képesti pozíció.

278). Ugyanakkor a negálók nagyon ritkán töltik be a szentimentkifejezést követő pozíciókat a pozitív és a negatív fragmentumokban egyaránt. Figyelemre méltó, hogy a pozitív fragmentumokban a szentimentkifejezést közvetlenül megelőző pozíció jobban preferált összevetve a negatív fragmentumoknál tapasztalt gyakorisági megoszlásokkal: amúgy a pozitív fragmentumokban a negáló elemek 62,08%-ban ebben a pozícióban helyezkednek el, addig a gyakoriság ugyanebben a pozícióban, a negatív fragmentumok esetében csupán 43,49%. . Ugyanakkor, a negatív fragmentumokban a különböző pozíciók egyenletesebb megoszlási aránnyal jutnak szerephez, a szentimentkifejezést megelőzően és követően egyaránt (-4. pozíció 4,03%, -3. 11,21%, -2. 27,35%, 1. 6,72%, 2. 4,48%). Ugyanezeknek a pozícióknak a kitöltési aránya a pozitív fragmentumokban alacsonyabb. A tapasztalatokat az 5. ábrával igyekszem szemléletesebbé tenni.



5. ábra: A negálók és a szentimentkifejezések egymáshoz képesti pozíciós sajátosságai az ellentétes értékű fragmentumokban

A fenti eredmények arra mutatnak, hogy a negáló kifejezések automatikus kezelése során nagyobb figyelmet kell fordítanunk majd a pozitív szentimentkifejezések mellett szereplő negáló elemekre, ugyanis azoknak nagyobb a pozíciós variabilitása. (A fejezetben tett megállapítások gyakorlati alkalmazását l. az implementációt tárgyaló 7. részben)

A negáló elemek sajátosságainak elemzésén túl megvizsgáltam, vajon mutatkozik-e valamilyen megoszlási eltérés a növelő és csökkentő értelmű intenzifikáló elemeket illetően a pozitív és a negatív fragmentumok között (vö. Szabó és mtsai 2016a: 178). Megállapítottam, hogy a növelő értelmű intenzifikáló elemek arányaikban közel azonos gyakorisággal fordulnak elő a pozitív (2405) és a negatív (2568) értékű fragmentumokban

(pl. *nagyon jó, nagyon rossz*). Ugyanakkor, a csökkentő értelmű intenzifikálók előfordulási gyakorisága nem azonos a két fragmentumtípusban: a negatív fragmentumokban jóval (779) gyakrabban fordulnak elő a pozitív fragmentumoknál (301). A jelenségnek első megközelítésben azt az okot tulajdonítottam, hogy amíg a növelő értelmű intenzifikáló elemek nem képesek a szentimentérték ellentétesre változtatására, addig a csökkentő értelműek igen, és ez utóbbiak jelenléte nem azonos módosulást eredményez a szentimentérték tekintetében a pozitív és a negatív értékű szentimentkifejezések mellett. Amíg ugyanis a pozitív szentimentkifejezésekkel együtt negatív értékelést fejeznek ki (pl. *kevésbé jó*), addig a negatív szentimentkifejezések mellett nem jelölnek feltétlenül pozitív véleményt (pl. *kevésbé rossz*) (vö. Szabó és mtsai 2016a: 178). Azonban, a korpusz vizsgálata rávilágított, hogy ez a feltételezés nem állja meg a helyét. Azt tapasztaltam ugyanis, hogy a csökkentő értelmű intenzifikáló elemek sem a negatív, sem a pozitív szentimentkifejezések mellett az esetek többségében nem okoznak teljes szentimentérték-beli módosulást, egyszerűen csökkentik azok pozitív vagy negatív értékelő tartalmát.

Végezetül, az irreáló elemek esetében ugyanazt a tendenciát figyeltem meg, mint ami a csökkentő értelmű intenzifikáló elemeknél mutatkozott. Ezek az elemek szintén gyakoribb előfordulásúak a korpuszban a negatív értékű fragmentumokban (pl. *jó volna*) (vö. Szabó és mtsai 2016a: 178). Ugyanakkor, akárcsak a csökkentő értelmű intenzifikáló elemek esetében, a legtöbbször ezek sem idéznek elő értékváltozást, sem a pozitív, sem a negatív szentimentkifejezések mellett.

A csökkentő értelmű intenzifikáló, valamint az irreáló elemekre vonatkozó fentebbi korpuszvizsgálati eredmények – akárcsak a negáció kérdését illetően – ismételten arra mutatnak, hogy a szentimentshifterekre vonatkozó előzetes feltételezéseinket nélkülözhetetlen nagy mennyiségű korpuszadat alapján ellenőrizni. Amíg ugyanis, például a negáció esetében nem tételeztem olyan jelentősnek a tendenciát, mint azt a korpusz vizsgálata alapján tapasztaltam, addig a csökkentő értelmű intenzifikáló, valamint az irreáló elemek esetében a vizsgálati eredmények épp ellenkezőleg, cáfolták a kezdeti hipotéziseimet. Azt tapasztaltam ugyanis, hogy egyik elemtípus sem okoz az esetek többségében teljes szentimentérték-beli módosulást, sem a negatív, sem a pozitív szentimentkifejezések mellett.

Bár a korpusz annotációját nem specifikusan az értékvesztés és az értékváltás vizsgálatára alakítottam ki, a két jelenség közül a különböző tagok szűrésével lehetőségem

volt az értékvesztési eseteket is megvizsgálni a korpuszban.^{75,76} Az alábbiakban ezeket az elemzési tapasztalatokat ismertetem.

A korpusz összes fokozó értelmű intenzifikáló eleme közül 133 olyan előfordulást találtam, ahol az ezzel a taggel jelölt elem szerepét negatív fokozó elem töltötte be. A 133 esetből összesen 37-szer pozitív, 96 alkalommal negatív értékűnek annotált fragmentumban szerepel.

Az elemek lexikai diverzitása magas, ugyanis a 133 előfordulás összesen 40 különböző negatív emotív fokozó elemet tartalmaz. Amennyiben az azonos tövű elemeket együtt tekintjük, a leggyakrabban az *elképesztő(-en)* (31), a *borzasztó(-an)* és a *borzalmas(-an)* (25), valamint a *baromi(-ra)* (15) fordulnak elő (az *elképesztő(-en)* elem lexikai szintű emotív szemantikai tartalmát érintő dilemmáról már szoltam, l. 4.4).

Az egyes negatív fokozó elemek gyakorisági megoszlása a negatív és a pozitív fragmentumokban részben eltér egymástól. A negatív szentimentértékű fragmentumokban 24 különböző fokozót találunk, közülük a legnagyobb frekvenciával a *borzasztó(-an)* és a *borzalmas(-an)* (24) szerepel, a második és a harmadik leggyakoribb az *elképesztő(-en)* (14) és az *iszonyú(-an)*, *iszonyat(-os)(-an)* (8). A pozitív fragmentumokban 9 különböző elem fordul elő, és közülük a három leggyakoribb az *elképesztő(-en)* (17), a *baromi(-ra)* (8), valamint az *eszméletlen(-ül)* (5). A kapott adatok alapján – szem előtt tartva a vizsgált lexémák viszonylagosan csekély korpuszbeli számát – a következő legfontosabb megállapítást tehetjük: A korpuszban előforduló leggyakoribb negatív fokozó elemek közül csupán egy, az *elképesztő(-en)* az, amely közel azonos alkalommal szerepel a negatív és a pozitív kontextusokban. Ettől eltekintve azt látjuk, hogy a negatív és a pozitív fragmentumokban teljesen eltérő kifejezések szerepelnek a leggyakrabban. Így például a *borzasztó(-an)* és a *borzalmas(-an)* kifejezetten a negatív fragmentumokra jellemző, ahol ez az elem a negatív fokozók közül a leggyakoribb, szemben a pozitív fragmentumokkal, amelyekben mindössze egyszer fordul elő.

A korpusz vizsgálati eredményei fontos alapot szolgáltatnak a szabályalapú szentimentelemző eszköz létrehozásához (l. 7).

⁷⁵ A jelenségek pontos definícióját l. 3.

⁷⁶ Az értékvesztő elemeket a következőképpen nyertem ki a korpuszból: először lekérdeztem a fokozó értelmű intenzifikáló elemeket, majd azokat manuális egyenként megvizsgáltam, leszűkítve ezzel a lekérdezés eredményét a negatív emotív fokozó elemekre.

6.3.2 A NegEmotív korpusz

A jelen fejezetben egy, kifejezetten az értékváltás és az értékvesztés jelenségeinek vizsgálatát támogató szövegkorpusz létrehozásáról, valamint vizsgálati eredményeiről számolok be részletesen.⁷⁷

6.3.2.1 A korpusz létrehozásának oka és célja

A korpusz létrehozásának elsődleges célja az volt, hogy olyan vizsgálati anyagot teremtsék, amelynek segítségével az értékvesztéshez és az értékváltáshoz kapcsolódó szemantikai-pragmatikai jelenségeket az annotáció alapján, kvalitatív és kvantitatív szempontból egyaránt elemezni lehet. Amint azt Andor (2011: 33–34) is hangsúlyozza, „a polaritásváltások azonosítása, konstatálása, a megfigyelések validitásának biztosítása napjaink nyelvtechnológiai eszköztárát igénybe véve kizárólagosan korpusznyelvészeti eszközökkel, elsősorban konkordancia és kollokációs vizsgálatokkal, valamint a korpuszokból nyert eredmények empirikus, anyanyelvű alanyok intuitív megítélésére támaszkodó kontrollteszteléses vizsgálatok révén érhető el”. Ugyanakkor a Díványkorpusz az említett részproblémák kimerítő vizsgálatára nem alkalmas (l. 6.3.1.7), hiszen nem specifikusan az értékvesztés és az értékváltás jelenségeinek vizsgálata céljából készült. Ezzel összefüggésben, a korpusz egyrészt nem tartalmaz a probléma vizsgálatához elegendő mennyiségű adatot, másrészt nem rendelkezik olyan annotációval, amely a jelenségeket könnyen lekérdezhetővé tenné az adatbázisból. Azt a lehetőséget, hogy a kutatást a nagyméretű és sokrétűen annotált Magyar Nemzeti Szövegtár újabb verziója (MNSZ2, vö. Oravecz és mtsai 2014) alapján végezzem el, a korpusz több lényegi jellemzője miatt ugyancsak el kellett vetnem. A döntésem legfőbb oka az volt, hogy a két probléma meglátásom szerint általános céllal készült korpuszok segítségével kielégítő módon nem vizsgálható.⁷⁸

⁷⁷ A jelenségek pontos definícióját l. 3.

⁷⁸ A korpusz a szentimentértékek szempontjából nincs annotálva, így módon nem ad lehetőséget az értékváltási esetek lekérdezésére. Emellett az értékvesztő és értékváltó elemek specifikus annotációval nem rendelkeznek benne, így lekérdezésük sem lehetséges. Végül, a korpusz személyes alkorpusza nagyon sok korpuszhibát tartalmaz, amely szövegtípusban az értékváltás és értékvesztés jelenségei különösen magas számban fordulnak elő. A korpuszhiba ebben az esetben azt jelenti, hogy számos tweet többször megismétlődik a korpuszban, amelynek valószínűleg az az oka, hogy a korpusz létrehozásakor, a tweetszövegek gyűjtésekor a készítők nem szűrték ki az ún. retweetekben szereplő duplikátumokat. A Twitteren minden, válaszként közölt üzenet esetében az előző üzenet megismétlődik a válasz előtt, ennek

A rendelkezésre álló adatbázisok problémáit megfontolva úgy döntöttem, létrehozok egy olyan korpuszt, amely kimondottan az értékváltási és az értékvesztési jelenség sajátosságainak a feltárását támogatja.

6.3.2.2 A szövegválasztás indoklása, a vizsgált szövegtípus jellemzése

A NegEmotív, kézzel annotált korpuszt egy nagyméretű, nyers, Twitter-szövegeket tartalmazó korpuszból hoztuk létre.

Amint azt korábban már tárgyaltam (l. 5.1), a szentimentelemzésben a leggyakrabban feldolgozott szövegtípusok között a különböző közösségimédia-szövegeket (pl. tweetek, blog-, és Facebook-bejegyzések) találjuk (vö. pl. Godbole és mtsai 2007, Poursepanj és mtsai 2013, Ortigosa és mtsai 2014), és közülük is leggyakrabban a Twitteren publikált bejegyzéseket vizsgálják, illetve használják fel fejlesztési célokra (vö. pl. Pak–Paroubek 2010, Kouloumpis és mtsai 2011, Poursepanj és mtsai 2013). Részben ez a sajátosság, részben pedig az, hogy a tweetekben elsődleges vizsgálati eredményeim (l. fentebb, 4.5) alapján – más szövegtípusok viszonylatában (vö. Szabó 2015c: 55–59) – magas az értékváltó és értékvesztő elemek száma, indokolja a korpuszom szövegének a megválasztását.

Tekintsük röviden a tweetek legfontosabb jellemzőit!

A Twitter egy ingyenes ismeretségi hálózat és mikroblog-szolgáltatás, amely ma egyike a legnépszerűbb online tömegkommunikációs eszközöknek. Lehetőséget ad arra, hogy a felhasználók rövid üzeneteket osszanak meg egymással, illetve hogy egymás megosztásaira röviden reagáljanak.⁷⁹ Az üzenetek hossza nem haladhatja meg a 140 karaktert, ugyanakkor azt nem korlátozza a rendszer, hogy egy adott felhasználó hány tweetet (vagy ahogyan néha magyarul említik: *csiripet*) tesz közzé.

következtében egész diskurzusfolyamok alakulnak ki. Amikor korpuszépítési céllal, automatikus módszerrel tweeteket gyűjtünk, nélkülözhetetlen az anyagot ezektől a duplikátumoktól megtisztítani, hogy az ismétlődések ne torzíthassák a későbbi vizsgálatok eredményeit.

⁷⁹ A Twitter rendszerét 2006 októberében a San Francisco-i Obvious Corporation létesítette (vö. Tolnai 2011). Akárcsak a legtöbb online közösségi háló esetében, itt is lehetősége van a szolgáltatáshoz már csatlakozott embereknek, hogy mások felé a kapcsolódási szándékukat jelezzék, vagy, amennyiben azok még nem a Twitter regisztrált tagjai, úgy nekik meghívót küldjenek a csatlakozáshoz (vö. Lavallee 2007). Minden Twitter-tag saját oldallal rendelkezik, amely az adott felhasználó összes korábbi bejegyzését logolja, azaz tárolja. A felhasználó – alapbeállításként – azt is meghatározhatja, hogy csupán a baráti köréhez tartozók, vagy pedig bárki láthassa azokat.

Mára a Twitter kiemelkedően fontos közösségi térré vált. A szolgáltatás hivatalos oldalának adatai alapján a Twitter 2016 januárjában 332 millió aktív felhasználóval rendelkezett. E hatalmas mennyiségű, elektronikus formájú szöveg a nyelvészeti, szűkebben a számítógépes nyelvészeti kutatási és fejlesztési feladatok számára is fontos anyagként szolgál, emellett a közösségi tartalmak folyamatos monitorozására a gazdasági oldal is jelentős igényt támaszt. Nem véletlen, hogy a nyelvtechnológiai kutatásban és fejlesztésben ma már ez az egyik leggyakrabban vizsgált szövegtípus (vö. pl. Go és mtsai 2009, Kouloumpis és mtsai 2011, Bollen és mtsai 2011, Jiang és mtsai 2011).

6.3.2.3 A kiinduló (nyers) korpusz mérete és alapvető statisztikai adatai

A nyers Twitter-korpusz szöveganyagát összesen 37818 automatikus módszerrel gyűjtött, magyar nyelvű tweet alkotja. Az adatok a Twitter magyar oldaláról származnak. Az összes tweetből 18236 darab férfi, 19559 pedig női szerzőségű. A nyers korpusz alapvető statisztikai adatait az alábbi táblázat mutatja be:

	nő	férfi	összes
A tweetekben lévő mondatok átlagos száma (eszköz: magyarlanc, vö. Zsibrita és mtsai 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

5. táblázat: A nyers Twitter-korpusz alapvető statisztikai adatai

A fenti statisztikai adatok alapján a nők és a férfiak korpusza az alapvető kvantitatív sajátságokat tekintve nem mutat érdemi eltérést. A férfi felhasználók átlagosan több mondatra bontják tweetjeiket, míg a női felhasználók átlagosan több szót használnak fel tweetenként, ugyanakkor egyik eltérés sem szignifikáns.

A nyers Twitter-korpusz szolgált alapul a NegEmotív, kézzel annotált korpusz létrehozásához úgy, hogy abból gyűjtöttük ki a korpuszépítéshez szükséges adatokat.⁸⁰

⁸⁰ Végeztünk egy szótáralapú emócióelemzést is a teljes Twitter-korpuszon. Azt szeretnénk volna feltárni, hogy a különböző emotív tartalmak tekintetében van-e szignifikáns eltérés a férfi és a női szerzőségű tweetek között. Ennek céljából a két teljes, nyers alkorpuszt automatikus módszerrel elemeztük a hat és a nyolc

Minden olyan tweetet megfelelőnek tekintettünk, amely tartalmazott legalább egy értékvesztésre képes elemet.^{81,82} A munka első fázisában tehát kézzel összeállítottam egy listát, amely az értékvesztésre képes elemeket tartalmazza (amelyek közül bizonyosak értékváltásra is képesek). A lista elkészítése során egy fokozó értelmű kifejezésekből álló szótárra (vö. Tukacs 2015), a Magyar Nemzeti Szövegtár adataira (MNSZ2, vö. Oravecz és mtsai 2014), valamint a korábban, a szemantikai motivációs vizsgálatok kapcsán feltárt tapasztalatokra és adatokra támaszkodtam (l. 4.5). A létrehozott szólista 106 szóalakot tartalmaz (l. 1. számú melléklet).⁸³ Ezt követően, a szólista alapján automatikusan kigyűjtöttük a Twitter-korpuszból az összes megfelelő adatot. A létrehozott vizsgálati korpusz összesen 610 tweetből áll (vö. Szabó és mtsai 2017a: 253).

6.3.2.4 A NegEmotív korpusz létrehozásának a menete és eszköze⁸⁴

A nyers korpusz kézi feldolgozásához a Brat nevű, online elérhető annotáló programot használtuk. Az eszköz bármilyen annotálási feladatra alkalmas, a felhasználó maga konfigurálhatja a használni kívánt annotációs tageket, azok csoportjait, kapcsolatait, és további, bevinni kívánt információkat. Teljesen személyre szabható és egyszerűen kezelhető. A config fájlt és az annotálandó szövegeket txt formátumban töltöttük fel a programba, és az annotációs fájlokat is ebben a formátumban kaptuk vissza. Az output

érzelemtípus nyelvi elemeit tartalmazó emóciólexikonunkkal. A hat kategóriából álló szótárunk az ekmani alapérzelmekekre (vö. Ekman–Friesen 1969) épül, míg a nyolc kategóriás rendszerezésünkben két további érzelemtípus, a FESZÜLTség és a VONZALOM is helyet kapott (a két emóciószótárról, annak tartalmáról, létrehozásának elméleti alapvetéseiről és menetéről, valamint a szótárak adatairól részletesen l. Szabó–Morvay 2015, Szabó és mtsai 2016c).

A két nem közötti különbségre kétoldali t-próbát számoltunk, hogy lássuk, statisztikailag szignifikánsak-e az eltérések. Az eredmények p-értéke az ún. bizonytalansági faktor, amely akkor utal szignifikáns eltérésre, ha az értéke kisebb 0,05-nél. A hat elemből álló szótár esetében az ÖRÖM és a MEGLEPŐDÉS emóciók esetében a két csoport közötti eltérés szignifikáns volt. A MEGLEPŐDÉS emóció a férfiak tweetjeiben, míg az ÖRÖM emóció a nők tweetjeiben jelent meg többször. A nyolc elemből álló szótár eredményei alapján a DÜH, az ÖRÖM, a FÉLELEM, a FESZÜLTség és a VONZALOM esetében tapasztaltunk érdemi eltérést. Ezek az emóciók a nők tweetjeiben jelentek meg gyakrabban.

⁸¹ Tekintettel a megállapításomra (l. 4.5), mely szerint minden értékváltásra képes elem értékvesztőként is szerepelhet, az értékváltást itt külön nem említem.

⁸² A válogatás következtében nem olyan korpusz jött létre tehát, amely minden adat esetében értékvesztést (vagy értékváltást) tartalmazott, hanem szerepelt benne legalább egy olyan elem, amely erre a nyelvi sajátosságra képes.

⁸³ A listát a dolgozat bírálója, Sass Bálint megjegyzései alapján a vizsgálatok elvégzése után 116 elemből állóra bővítettem. Ennek megfelelően a disszertációban elvégzett vizsgálatokban a 106 elemmel dolgoztam. A mellékletben az újonnan felvett elemeket dőlt szedéssel emelem ki.

⁸⁴ A korpusz annotálási útmutatóját l. 5. számú melléklet.

fájlban az annotált tagek lokációi és a létrehozott kapcsolatok szerepelnek egyszerű listaként.⁸⁵

Az adatbázis feldolgozását három annotátor végezte el. Ebben az esetben is, akárcsak a Dívány-korpusz munkálatai során, az annotálás megkezdése előtt a három annotátor között felosztottam a korpusz szöveganyagát. Azonos korpuszrészre az annotátorok csupán az egyetértésmérési feladathoz annotáltak. A feldolgozói munkájához előzetesen elkészítettem egy útmutatót, amely részletesen, példák segítségével bemutatta az annotálási feladatot (vö. Szabó és mtsai 2017a: 253).

A korpusz feldolgozását az annotátorok a következő annotációs szempontok alapján végezték el manuálisan: Először bejelölték azokat a kifejezéseket, amelyek esetében a lexikai szintű szentimentérték eltért a kontextusbeli értéktől. Még ebben a lépésben döntést hoztak arról is, hogy ez az eltérés az aktuális kontextusban értékváltságnak vagy pedig értékvesztésnek tekinthető-e. Értékváltás (83a) esetén jelölték az értékelés targetét is, értékvesztés (83b) esetében pedig azt az elemet, amelyet módosít, tehát a frázis alaptagját.⁸⁶ (A vizsgált elemeket kövér, a további annotált kifejezéseket dőlt szedéssel jelölöm.)

- (83) a. ismét **brutális** *koncertet* adott az énekes
b. Valljátok be, **rohad**t jó az időérzésem!

Amint a korpusz fenti példái mutatják, a (83a) esetében a vizsgált elem szentimentkifejezés funkcióját tölti be, és a *koncert* targetre vonatkozóan fejez ki pozitív értékelést, értékváltó elem funkciójában. Ettől eltérően, a (83b) alatti példában a vizsgált elem a szerkezeti alaptag módosítójának szerepét tölti be, azaz értékvesztő elem szerepében áll (vö. 3. fejezet).

A bemutatott annotálási megoldásnak az volt a célja, hogy segítségével a jövőben vizsgálni lehessen egyrészt az értékváltó elemeknek és azok targeteinek, másrészt az értékvesztő elemeknek és az általuk módosított alaptagoknak a kapcsolatait.

⁸⁵ Köszönöm Morvay Gergőnek, hogy létrehozta a config fájlt, majd az alapján beállította a munkához az annotáló programot.

⁸⁶ A dolgozat bírálója, Sass Bálint felhívta a figyelmet arra, hogy a targeteket az értékvesztés esetében is érdemes lett volna jelölni. A korpusz elkészítésekor ez nem merült fel, hiszen a korpusz alapján, az értékvesztést illetően kimondottan a fokozó elem és az alaptag szemantikai-pragmatikai viszonyát szerettem volna megvizsgálni. Azt sem annotáltuk ebben az esetben, hogy a teljes fragmentum milyen értéket hordoz. Ebben az annotációban a szentimentértékek módosulását nem ilyen jellegű feldolgozási megoldással jelölték az annotátorok, más módszereket alkalmaztunk. Ugyanakkor köszönöm a megjegyzéseket, egy jövőbeli munka során ezeket is szem előtt fogom tartani.

A feldolgozás során az értékváltásra képes elemek közül csupán azokat annotáltuk, amelyek az adott megnyilatkozásban valóban értékváltást mutattak. Amennyiben tehát az elem az adott kontextusban negatív értékelést fejezett ki, nem annotáltuk, például

(84) Egy **rohadt** kukásautó miatt áll már vagy tíz perce a troli.

A bemutatott feldolgozási megoldásokon túl az értékvesztésre vonatkozóan további tageket is alkalmaztunk, a vizsgált elemek további szemantikai-pragmatikai viselkedése alapján (vö. Szabó és mtsai 2017a: 254). E megoldás alkalmazása mellett korábbi vizsgálati eredményeimre (l. 4.5) támaszkodva döntöttem, melyek szerint a negatív fokozó elemek semleges alaptagok mellett a szemantikai motiváció szempontjából nem viselkednek egységesen. Annak céljából, hogy a korpusz annotációja ezeket a sajátságokat is lekérdezhetővé tegye, az annotátorok az értékvesztés esetében a következőket jelölték: egyrészt, a vizsgált elem által módosított alaptag pozitív, negatív vagy semleges szentimentértékű-e, másrészt, a fokozó elem az aktuális kontextusban semleges (deszemantizált), negatív vagy esetleg pozitív-e.

Annak céljából, hogy az annotátoroknak ne kelljen mindenképpen elfogadniuk az előre meghatározott típusokat és szemantikai sajátságokat, a target kategóriáján kívül az összes annotált taghez létrehoztam egy “egyéb” kategóriát is. Ezáltal arra is lehetőséget adtam, hogy a munka során esetlegesen felmerülő, előre nem várt jelenségeket is jelölni tudjuk a korpuszban.

Az ismertetett annotálási rendszert az alábbi táblázat foglalja röviden össze:

	Az annotált tagek és szemantikai-pragmatikai sajátságok (ez utóbbit a program technikai sajátsága okán viszonyként jelöltük):
Értékváltás esetén:	értékváltó elem (vagy egyéb)
	target
Értékvesztés esetén:	értékvesztő elem, amely a szemantikai tartalma alapján lehetett: deszemantizált, negatív, pozitív (vagy egyéb) ⁸⁷
	alaptag, amely a szemantikai tartalma alapján lehetett: negatív, pozitív, semleges (vagy egyéb)

6. táblázat: Az annotálási rendszer rövid összefoglalása

Az alábbi ábra egy részletet közöl a korpusz annotációjából a Brat programban:

100	xnuffx_625896339566669824.txt,Tegnap durván	évesztő	deszemant	alapneg	leégett az a jóllakott óvodás fejem a szoliban..
101	(nagymosoly) (2 hete nem voltam) [[durván]],durván				
102	MrSuperEgo_587687911019118593.txt,A váltásérzés a múlté.				
103	Ez ma kiderült edzés közben (halk juhé).				
104	Az is kiderült hogy három hónap alatt elképeszt?en	évesztő	deszemant	alapsem	sokat romlott a formám!
105	[[elképeszt?en]], elképeszt?en				
106	wasandras_526364731100397568.txt,"épp az el?bb baszott le az Öcsém, hogy öregszem, mert nem értek a Snapchat-hez...				
107	(nagymosoly) [[baszott]]".baszott				
108	BaracskaiGreta_627620118747607040.txt,itt szól a szomszédba	évesztő	negatív	alapsem	hangosan a mulatós zene 01:22-kor.. gratula.. nem lehet aludni (semleges)?
109	[[brutál]],brutál				
110	szokeptr_578180711032713216.txt,"fuu, kurva	évesztő	deszemant	alapneg	szár lett az OS X wifi kezelése 10.10 után [[kurva]]",kurva
111	szilagyi_vivien_584998216049000448.txt,amikor 2 nap alatt alszol durván 10 órát (mosoly) felbecsülhetetlen [[durván]],durván				
112	kockasfalu_570244277810421760.txt,"nézd a jó oldalát: aki ennyiért marad, baromira	évesztő	deszemant	alappoz	elhivatott lehet.
113	[[baromira]]".baromira				

6. ábra: Részlet a korpusz annotációjából a Brat nevű programban

6.3.2.5 A korpusz annotálásának egyetértés-mérési eredményei

Ahhoz, hogy az annotátorok közötti konszenzust mérni tudjuk, a teljes korpusz annotálása előtt elvégeztünk egy egyetértésmérést a korpusz egy kisebb részletén (vö. Szabó és mtsai 2017a: 255–256). Összesen 100 tweetet annotáltunk, a korpusz kis mérete miatt ugyanis

⁸⁷ Amint azt a szemantikai motivációs vizsgálatoknál megmutattam (l. 4.5.1), az általam *értékvesztés*nek nevezett jelenség esetében a vizsgált elem fokozó szerepbe lép, ugyanakkor szemantikai tartalma nem feltétlenül lesz ebben a pozícióban semleges, hiszen negatív, ritkán pozitív tartalmat is hordozhatnak. A 6. táblázatban ismertetett annotálási megoldással ennek jelölésére kívántam lehetőséget adni. Meg kell jegyezni tehát, hogy a feltárt szemantikai sajátság okán az általam alkalmazott terminus nem szerencsés. Fontos volna egy olyan fogalmat bevezetni, amely pontosabban és kevésbé félrevezető módon utal a jelenségre, azonban ilyen kifejezést a dolgozat leadásáig nem találtam.

már ez a mennyiség is a teljes adatmennyiség hatod részét tette ki. A méréshez a Cohen-féle Kappa-statisztikát alkalmaztuk (vö. Cohen 1960) (részletesen l. 6.1). 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

7. táblázat: A Cohen-féle Kappa-statisztika eredményeinek értelmezési tartományai (vö. Landis and Koch 1977)

A pilot-annotálás megmutatta, hogy az annotátorok a Kappa-érték szerint, összesítve 0,489, azaz a Kappa-sávok alapján közepes szintű átlagos egyetértéssel dolgoztak.⁸⁸ 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 újraannotáltattuk a korpuszrészlet, majd megismételtük az annotátorok közötti egyetértés mérését. Az újabb adatok alapján, a Kappa-sávok szerint az annotátorok közötti egyetértés 0,597-re 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.^{89,90}

⁸⁸ Köszönöm Balogh Kittinek az egyetértésmérés elvégzését.

⁸⁹ Köszönöm Ilyés Virágnak a minőségbiztosítás átgondolásában és az ellenőrzés lefolytatásában nyújtott hathatós segítségét, valamint az egyetértésmérés elvégzését.

⁹⁰ Az egyetértés eredményét azért fogadtuk el, mert magát a feladatvégzést a szemantikai-pragmatikai szempontok miatt jelentősen szubjektívebbnek ítéltük egy egyszerűbb, pl. morfológiai vagy névelem-annotálási feladatnál. Ugyanakkor próbáltunk a minőségen újraannotálással és az annotálási útmutató újraírásával, pontosításával javítani.

6.3.2.6 Az annotált korpusz alapvető statisztikai adatai

A 610 tweetből összesen 289-et annotáltunk. A maradék 321 tweetet azért nem dolgoztuk fel, mert nem tartalmaztak egyetlen olyan elemet sem, amely megfelelt volna az annotációs alapelveknek (l. 6.3.2.4).

A 289 tweetben összesen 383 értékvesztést vagy értékváltást jelöltünk be. Ebből összesen 308 esetben jelöltünk értékvesztést és 75 esetben értékváltást (vö. Szabó és mtsai 2017b).

6.3.2.7 A korpusz részletes vizsgálati eredményei

A korpusz alapján először kvantitatív és kvalitatív szempontból elemeztem az annotált adatokat.

Amint ismertettem (l. 6.3.2.6), a 383 annotált esetből összesen 75 esetben jelöltünk értékváltást, és 308 esetben értékvesztést, tehát fokozó funkciót. Az értékváltás jelensége tehát az értékvesztésnél jóval ritkább előfordulásúnak mutatkozik a korpusz adatai alapján.

Megvizsgálva az értékváltás eseteit, a következő megállapításokat tehetjük: A két leggyakoribb elem ebben a szerepben a *durva(-n)* (28) és a *kemény* (17) volt. Számos további elemet is annotáltunk értékváltóként, azonban ezek lexémánként átlagosan mindössze egy vagy két alkalommal fordultak elő a korpuszban.

Szerettem volna megtudni, hogy vajon mennyire jellemző e két leggyakoribb elemre az értékváltás, azaz feltehető-e, hogy ezek az elemek szentimentkifejezés funkciójában legtöbbször pozitív aktuális szentimentértékkel rendelkeznek. Megvizsgáltam tehát ezeknek a kifejezéseknek az összes, nem annotált előfordulását is a korpuszban (vö. Szabó és mtsai 2017a: 256). Azt tapasztaltam azonban, hogy mindkét elem gyakori előfordulása negatív értékelés kifejezőjeként is. Ez a sajátság nyilvánvalóan megnehezíti az értékváltásra képes elemek aktuális szentimentértékének a helyes automatikus kezelését (a problémáról részletesebben l. lentebb).

A 238 értékvesztési eset adatait az alábbi táblázat mutatja be (vö. Szabó és mtsai 2017b).

		a vizsgált elem aktuális szemantikai tartalma				ÖSSZ.:
		deszem.	negatív	pozitív	egyéb	
az alaptag szemantikai tartalma	pozitív	128	1	1	1	131
	negatív	86	1	-	8	95
	semleges	42	37	3	-	82
	egyéb	-	-	-	-	-
	ÖSSZ.:	256: 83,11%	39: 12,66%	4: 1,29%	9: 2,92%	308

8. táblázat: Az annotált értékvesztési esetek részletes statisztikai adatai

Az eredmények alapján a következő megállapításokat tehetjük (vö. Szabó és mtsai 2017b): A korpusz annotátorai az összesen 308 esetből 256 alkalommal vélték úgy, hogy a vizsgált elem teljesen elvesztette negatív szemantikai tartalmát, és az alaptag mellett pusztán fokozó szerepet töltött be. Ez az összes eset 83,11%-át tette ki.

A vizsgált negatív fokozó elemek a deszemantizálódás tekintetében – összhangban korábban bemutatott vizsgálati eredményeimmel (l. 4.5) – a pozitív és a negatív alaptagok mellett mutatkoztak a legegységesebbnek. A szentimentértékkel rendelkező alaptagok módosítóiként ugyanis szinte kivétel nélkül elvesztették negatív tartalmukat. (A vizsgált elemet kövér, a szerkezeti alaptagot dőlt szedéssel jelölöm.)

(85) nem tudom ti hogy vagytok vele, de a legtöbb fan annyira cuki, hogy engem **iszonyat boldoggá** tesz ez az egész összefogás

Ugyanakkor, semleges alaptagok módosítóiként változatos szemantikai-pragmatikai viselkedésüket látjuk. Az összesen 82 esetnek megközelítőleg csupán a felében (51,21%) jelöltek az annotátorok deszemantizáltságot. (A vizsgált elemet kövér, a módosított tagot dőlt szedéssel jelölöm.)

(86) Tegnap találtam egy **baszott nagy** gyíkot

A többi esetben az annotátorok úgy ítélték, hogy a vizsgált elem nem vagy nem teljes mértékben deszemantizált: vagy negatív (37: 45,12%), vagy – nagyon ritkán (3: 3,65%) – pozitív értéket ad a megnyilatkozáshoz.

- (87) a. **Szörnyen meleg** van még így az éjszaka közepén is.
b. 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 a korpuszannotáció alapján a szemantikai tartalom nem merül ki a fokozásban.

A vizsgálat fenti tanulsága figyelemre méltó, hiszen a kevés számú, vonatkozó dolgozat szerzője a szemantikai motiváció kérdését illetően inkább a negatív fokozó elemek deszemantizáltsága mellett teszi le a voksát (részletesebben l. 4.5). Így például, Balogh (2009) úgy gondolja, hogy amennyiben „az ilyen, másodlagos fokozó elemeket egy-egy megfelelő kulcsszóhoz kapcsoljuk, elveszítik elsődleges, azaz lexikális jelentésüket és átveszik a fokozó értelmű «nagyon» adverbium jelentését.” Hozzá hasonlóan, Jing-Schmidt (2007: 434) szerint ezek az elemek az elsődleges szemantikai tartalmuktól teljesen eltávolodnak, és az a pusztán intenzitásra redukálódik. Ezzel szemben azonban azt látjuk, hogy nem csupán a (4.5)-ben tárgyalt eredményeim, de a NegEmotív korpusz elemzésének tanulságai is a vizsgált elemek árnyaltabb nyelvi viselkedésére mutatnak rá.

Szerettem volna feltárni, hogy hány egyedi lexéma szerepel a korpuszban értékvesztő funkcióban, valamint azok gyakorisági megoszlását (vö. Szabó és mtsai 2017b). A korpusz lekérdezésével 54 szóalakot kaptam, amely összesen 39 szótőre volt visszavezethető.

A fokozó elemek közül az öt leggyakoribb elem a *kurva(-ra)* (112), a *rohad(-ul)* (53), a *(ki-)baszott(-ul)* (31), a *baromi(-ra)* (27), valamint az *iszonyat(-os)(-an)/iszonyú(-an)* (19) voltak.

Azt is megvizsgáltam, hogy a korpusz értékvesztő elemei milyen kollokációs sajátságokkal rendelkeznek. A kollokációs potenciál a szemantikai motiváció problémája szempontjából is fontos sajátság. Partington (1993: 183) alapján ugyanis minél inkább delexikalizálódott egy negatív fokozó elem, annál többféle kollokációban fordul elő (a delexikalizációról bővebben l. 4.5). Ugyanakkor valószínűleg ez a tendencia az ellentétes irányban is hat: minél többféle kollokációban szerepel egy negatív fokozó elem, annál

többet veszít lexikai szintű szemantikai tartalmából.⁹¹ A kollokációs potenciált illetően Partington (1993: 179) megállapítja, hogy bizonyos elemek kifejezetten csekély számú alaptagot módosíthatnak.

Annak céljából, hogy a kollokációs potenciált a korpusz segítségével megvizsgálhassam, áttekintettem az öt leggyakoribb értékvesztő elemre vonatkozó kvantitatív adatokat, és kiszámítottam az összes előfordulás, valamint azok kollokációinak az arányát.⁹² Arra a kérdésre kerestem tehát a választ, hogy a leggyakoribb elemek között melyek azok, amelyek több különféle alaptagot módosíthatnak, és azok szemantikai sajátása, tudniillik a negatív emotív tartalom elhomályosulásának a mértéke tükrözi-e ezt a sajátást. Az eredményeket az alábbi táblázatban közlöm:

vizsgált fokozó elem	összes előfordulás	kollokációk száma	arány
<i>kurva(-ra)</i>	112	33	3,39
<i>rohad(-ul)</i>	53	22	2,40
<i>(ki-)baszott(-ul)</i>	31	14	2,21
<i>baromi(-ra)</i>	27	19	1,42
<i>iszonyat(-os)(-an)/iszonyú(-an)</i>	19	14	1,53

9. táblázat: Az öt leggyakoribb elem összes előfordulásának és kollokációinak a száma, valamint azok aránya

A táblázat első oszlopa megmutatja, hogy hány alkalommal fordul elő a korpuszban az adott elem, a második oszlopban pedig azt, hogy hány egyedi kollokációban szerepel. Az utolsó oszlopban azt látjuk, hogyan aránylik egymáshoz a két adat, tehát, hogy az egyes elemek az előfordulásukhoz képest hány esetben jelentek meg egyedi kollokációban. Másképpen, hányféle egyedi alaptaggal kollokálódott az adott elem. Az arány értékét úgy kell tehát értelmeznünk, hogy az minél nagyobb, annál kevesebb különböző alaptaggal szerepel együtt.

Az adatok meglepő módon azt a tendenciát tárják fel, hogy minél gyakoribb egy elem a korpuszban, annál kevesebb egyedi alaptagot módosít. A leggyakoribb fokozó eleme a gyakoriságához viszonyítva jelentősen kevesebb kollokációban szerepel, mint például a 4. és az 5. leggyakoribb *baromi(-ra)* és az *iszonyat(-os)(-an)/iszonyú(-an)*. Az eredmény összhangban áll Partington (1993: 183) fentebb tárgyalt megállapításával,

⁹¹ A sajátosság más nyelvi jelenségek esetében is megfigyelhető, így például az ún. *félig kompozicionális szerkezetek* igei tagjainak esetében (vö. Vincze 2011).

⁹² A kollokáció tehát itt kizárólag az alaptaggal való kapcsolatot jelenti.

amennyiben meggondoljuk, hogy a NegEmotív korpusz három leggyakoribb elemének lexikai szintű szemantikai tartalma a düh emócióval hozható összefüggésbe, és szitokszónak tekinthető. Ennek következtében, összes előfordulásukhoz arányítva kevesebb különböző elemmel kollokálódnak, mint például a *baromi* fokozó elem, amely jelentősen gyengébb affektív értékkel rendelkezik. Ugyanakkor elgondolkodtató és a fentebbieknek valamelyest ellentmond, hogy ezek az elemek a leggyakoribb negatív emotív fokozók a korpuszban. (A vulgáris fokozó elemek társadalmi megítélését a nemek közötti nyelvhasználati eltérések kapcsán Jing-Schmidt (2007: 432) is említi, részletesebben l. lentebb).

Tekintettel a NegEmotív korpusz méretbeli korlátaira, az öt elem kollokációs potenciálját az MNSZ2 (vö. Oravecz és mtsai 2014) segítségével is megvizsgáltam, a következőképpen: az elemeket lekérdeztem a teljes korpuszból, majd a találatokat azokra az esetekre szűkítettem, amelyekben a vizsgált elemet melléknév követte. Ezután lekérdeztem a találatok kollokációs adatait. A gyakoriságot illetően azonban ugyanabba a komoly problémába ütköztem, mint amely alapján a (6.3.2.1)-ben a NegEmotív korpuszlétrehozásának szükségessége mellett érveltem: az MNSZ2 (vö. Oravecz és mtsai 2014) személyes alkorpusza nagyon sok ismétlődő adatot tartalmaz.

A probléma érzékeltetése céljából alább bemutatom, milyen eredményeket kapunk a korpuszban akkor, ha például a *baromi* elem előfordulásaira, majd kollokációira keresünk rá.

Keresés		Kollokáció-jelöltek megtekintése				
		oldalszám: 1		Ugrás	következő >	
			<u>Freq</u>	<u>T-score</u>	<u>MI</u>	<u>logDice</u>
		p/n tavaszias	47	6.855	13.582	8.297
		p/n ronda	46	6.780	11.784	7.454
		p/n Borús	21	4.582	13.545	7.405
		p/n unalmas	75	8.656	10.946	7.044
		p/n meglepő	115	10.715	10.200	6.465
		p/n ciki	20	4.470	10.749	6.351
		p/n randa	10	3.161	11.679	6.160
		p/n drága	118	10.851	9.863	6.155
		p/n drágák	12	3.462	10.408	5.832
		p/n céltudatos	9	2.998	10.920	5.820

7. ábra: A *baromi* elem melléknévi kollokációi,
a korpuszbeli frekvenciájukkal együtt az MNSZ2-ben

A kapott gyakorisági adatok alapján a vizsgált elem például a *meglepő* melléknévvel együtt összesen 115 alkalommal szerepel. Azonban, ha lekérdezzük az összes előfordulást egyenként, azt látjuk, hogy mind a 115 kollokáció pontosan ugyanabban a megnyilatkozásban szerepel. Az alábbi ábra ebből a lekérdezésből mutat meg egy részletet.

Keresés	korpusz: MNSZ2
?	találatok száma: 115
Mentés	doc#1920
Megjelenítés	doc#1920
KWIC/mondat	doc#1920
Rendezés	doc#1921
bal jobb	doc#1921
találat	doc#1921
infó	doc#1921
keverés	doc#1921
Véletlen minta	doc#1921
Szűrés	doc#1922
Gyakoriságok	doc#1922
szóalakok	doc#1922
szótövek	doc#1922
elemzések	doc#1922
dokumentumok	doc#1922

8. ábra: Részlet a baromi meglepő kollokáció 115 előfordulásából az MNSZ2-ben

A fenti tapasztalatok miatt az MNSZ2 felhasználását a tárgyalt probléma vizsgálatában el kellett, hogy vessem.

A jelen alfejezetben tárgyart korpuszvizsgálati eredményeket felhasználtam az implementációban a szabályok létrehozásakor (l. 7.1).

6.3.2.8. A korpusz vizsgálati eredményei a nemek közötti nyelvhasználati eltérésekre vonatkozóan

A korpusz metaadatai lehetőséget adtak arra, hogy az értékvesztés és az értékváltás nyelvhasználati sajátságait a nemek közötti eltérések szempontjából is megvizsgáljam. A jelen alfejezetben a problémát ebből a szempontból tárgyalom.

Mindenekelőtt, tekintsük át az ezekkel a nyelvi jelenségekkel kapcsolatos szakirodalmi megállapításokat! Mivel az irodalom a nemek közötti esetleges nyelvhasználati eltéréseket illetően nem szól az értékváltás, csupán az értékvesztés jelenségéről, ezért az alábbi megállapítások az értékvesztésre vonatkoznak.

Székely (2006: 330–331) véleménye szerint a magyar nyelv női és férfi beszélői a fokozásra nem azonos elemeket használnak: 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*. Székely (2006: 331) megállapításaira reflektálva Balogh (2009) hozzáteszi, hogy bár a férfiak nyelvhasználatában is előfordulnak hasonló kifejezések, „mégis inkább a nőknél gyakoribbak”.

Székely (2006: 331) alapján 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 felhívja a figyelmet a vulgáris fokozó elemeket érintő nemek közötti nyelvhasználati eltérésekre. Dolgozata alapján, a német anyanyelvű beszélők elfogadhatóbbnak találják, amikor férfi beszélő használja a német *verdammt* ’átkozott’ vulgáris 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) szerint a fenti megfigyelések, még ha csupán indirekt módon is, de abból a kulturális sajátságból fakadnak, hogy amíg a férfi beszélőktől a düh érzellemmel kapcsolatos nyelvi elemek elfogadottabbak, addig a nőktől kevésbé. Másképpen, a düh emóció verbális kifejeződéseit a két nem részéről a társadalom nem egyformán tolerálja. A szerző felhívja a figyelmet azokra a szociológiai tárgyú kutatásokra is, amelyek a düh kifejezésére irányulnak, és a nemek között aszimmetriát állapítanak meg. Úgy véli, ez a szocializációs eltérés összefüggésben állhat a szemantikai tartalmukban a düh érzelemhez kapcsolódó fokozó elemek használati sajátságaival is.

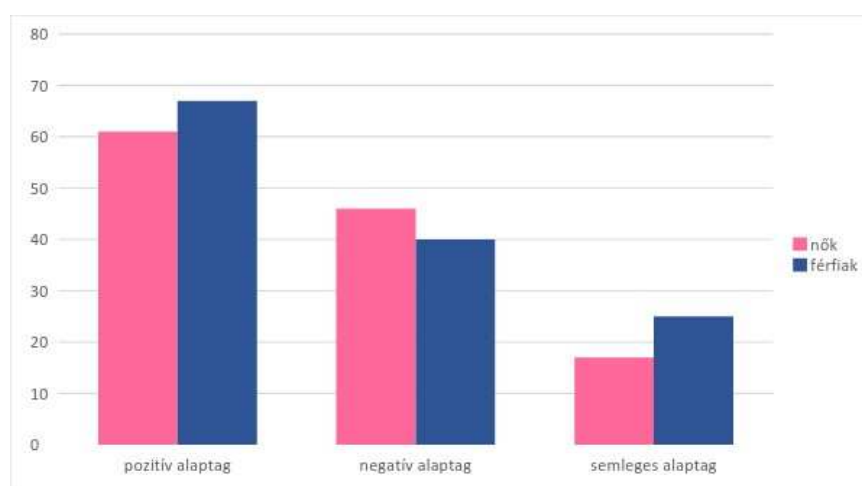
A NegEmotív korpusz alapján, a nemek közötti nyelvhasználati elérések szempontjából az értékváltásra vonatkozóan a következő megállapításokat tehetjük (vö. Szabó és mtsai 2017b): Az összesen 75 értékváltási eset közül 40 a férfi szerzőségű tweetekben, 35 pedig a női szerzőségűekben szerepelt. Ennek a típusnak az eloszlásában tehát a nemek között nem mutatkozik érdemi eltérés. Az egyes elemek vizsgálata alapján ugyancsak az látszik, hogy a két nem között jelentős nyelvhasználati különbség 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. Érdeki különbséget csupán a *brutál(-is)* elemnél látunk. Ez ugyanis a férfiak korpuszában három alkalommal megjelenik, míg a nőkben egyszer sem. Emellett mind a három adat a technológia témájához kapcsolódik. (A vizsgált elemet kövérszedéssel emelem ki.)

(88) Ezek a John Deere-ek forgácsolják szét a kivágott fák föld alatti részeit, **brutális** fűrőfejekkel.

Az eredmény összevág azzal a korábbi korpuszvizsgálati tapasztalatommal, amely szerint 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 2015c: 58–59).

A korpuszban annotált, összesen 308 értékvesztési eset a nemek közötti nyelvhasználati eltérések szempontjából az alábbi legfontosabb sajátosságokat mutatja (vö. Szabó és mtsai 2017b): A negatív fokozókat a férfiak valamivel gyakrabban (160) alkalmazzák, mint a nők (148). Megvizsgáltam, hogy az annotátorok által deszemantizáltak ítélt fokozó elemek mutatnak-e eltérést abban a tekintetben, hogy milyen szemantikai tartalmú alaptag mellett használják őket. Azt tapasztaltam, hogy a nők és a férfiak eltérő gyakorisággal élnek velük deszemantizált fokozóként a pozitív, a negatív és a semleges alaptagok mellett. Az eredményeket az alábbi ábra közli.



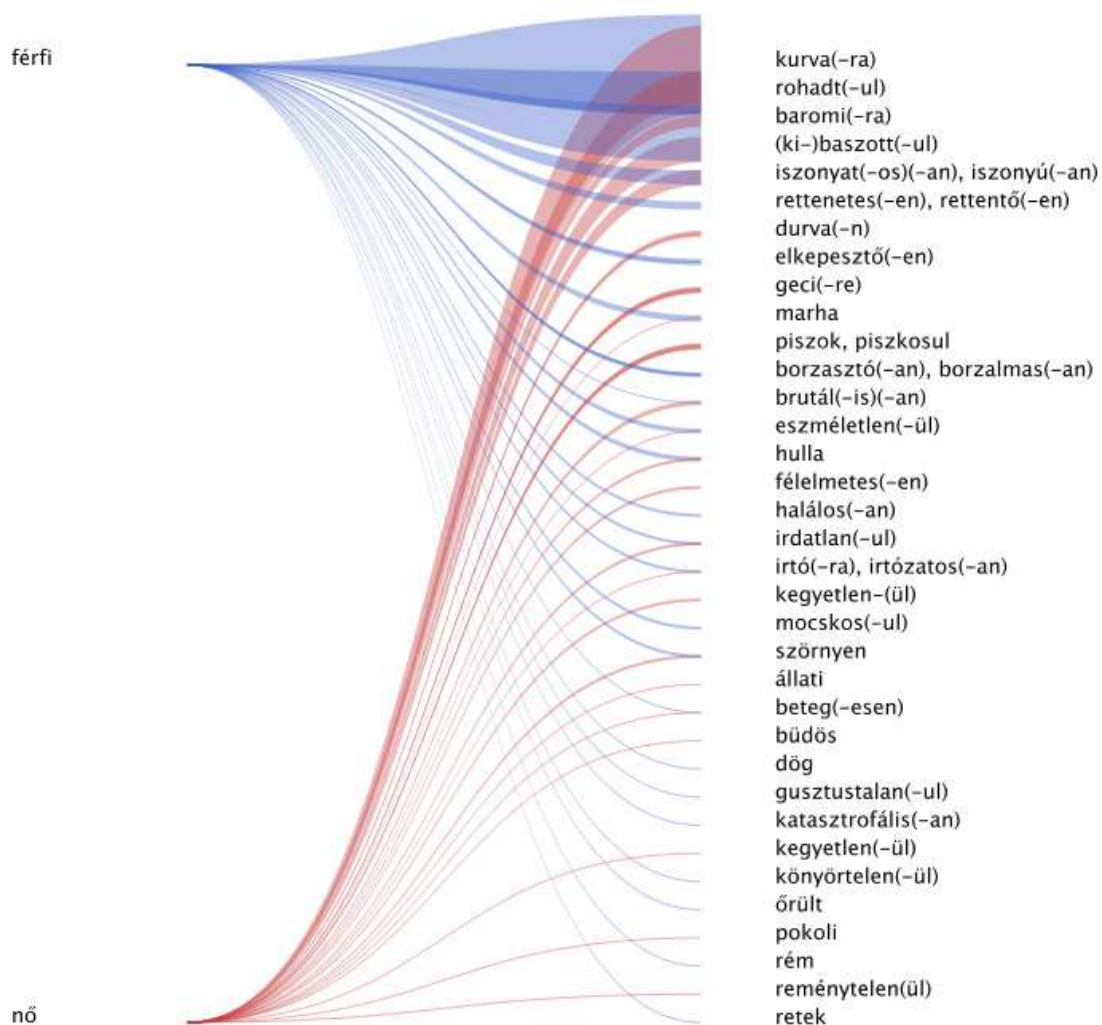
9. ábra: A deszemantizált fokozó elemek gyakorisága a két nemnél a különböző szemantikai tartalmú alaptagok mellett

Azt látjuk, hogy bár a két nemnél a deszemantizált fokozók arányai hasonlóak (a pozitív alaptagok módosítóiként a leggyakoribbak, utána a negatív, végül a semleges alaptag következik), egymáshoz viszonyítva azonban nem azonos gyakorisággal alkalmazzák őket a különböző alaptagok mellett. Az eltérések statisztikailag nem tekinthetőek szignifikánsnak, ugyanakkor elgondolkodtató, hogy a férfiak a pozitív (89) és a semleges (90) alaptagok módosítójaként valamivel gyakrabban élnek a vizsgált elemekkel, míg a nők a negatív alaptagok (91a-b) melletti negatív fokozókat preferálják jobban. (A vizsgált elemeket itt is kövér, a módosított elemeket dőlt szedéssel emelem ki.)

- (89) Azért veszik sokan a DW-t, mert **rohadt** jó kis óra! Én arra szavaznék
- (90) Hát így azért olyan **kurva** sok oroszlán nem futkosna.
- (91) a. Ma **rettentően** *haszontalan* vagyok, nem mentem messzebb a konyhánál és oda is csak kv-t főzni
- b. az amúgy honnan jött? **iszonyú** *idegesítő* szó

A vizsgálati tapasztalatok arra mutatnak, hogy a férfiak szívesebben aknázzák ki az ezekben az elemekben rejlő pragmatikai lehetőséget olyankor is, amikor nem negatív ítéletüket szeretnék kifejezni, azaz pusztá hatáskeltés céljából. A nők ugyanakkor inkább negatív kontextusokban élnek velük, erősítve a negatív tartalmat. Mindezt az adatok mennyisége, valamint a statisztikai eredmények fényében hipotézisnek tekintem, amelyet érdemes lenne nagyobb mennyiségű adat segítségével további vizsgálatoknak alávetni.

Végezetül megvizsgáltam, hány egyedi lexémát használnak a férfiak és a nők, valamint, hogy azoknak milyen a gyakorisági megoszlása. Az eredményeket az alábbi ábra mutatja be.



10. ábra: Az értékvesztő elemek gyakorisági megoszlása
a férfiak és a nők korpuszában⁹³

A nők 31, a férfiak 24 különböző elemet használnak ebben a funkcióban. Mindkét nem esetében ugyanaz az öt elem a leggyakoribb (az azonos tövűeket itt is összevontam): a *kurva(-ra)*, a *rohad(-ul)*, a *baromi(-ra)*, a *(ki-)baszott*, valamint az *iszonyat(-os)(-an)/iszonyú(-an)*. Ugyanakkor, amint azt a 10. ábra érzékelteti, amíg a férfiak korpuszában ez az öt elem kiteszi az összes előfordulás 78%-át, addig a nőkben az egyes elemek előfordulási gyakorisága kiegyenlítettebb képet mutat; náluk az öt leggyakoribb elem csupán az előfordulások 63%-át adja.

⁹³ Köszönöm Katona Eszternek az ábra elkészítését.

A leggyakrabban használt negatív fokozó elemek mindkét nem esetében szitokszavak. Ami a nemek közötti megoszlási arányt illeti, a női korpusz összes adatának 54,72%-a, a férfiak korpuszának pedig 63,12%-a szitokszó. Úgy vélem, hogy ezek az adatok összhangban állnak Székely (2006), Jing-Schmidt (2007) és Balogh (2009) megállapításával, mely szerint 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. Ugyanakkor az is látható, hogy az eltérés – a vizsgált szövegtípusban legalábbis – nem kimagasló.

A NegEmotív korpusz vizsgálatának tapasztalatait a következőképpen összegezhettük, kiemelve a legfontosabb eredményeket: A semleges alaptagok módosítóiként a negatív fokozó elemek jelentős szemantikai változatosságot mutatnak, és csupán az eseteknek valamivel több mint a felében deszemantizálódnak. Az eredmény egybevág a szemantikai motivációs vizsgálati eredményeimmel (l. 4.5), és figyelemre méltó, hiszen számos, a negatív fokozó elemekkel foglalkozó dolgozat alapján azok szemantikailag rendre kiüresednek, elveszítik lexikai szintű negatív tartalmukat (vö. pl. Jing-Schmidt 2007, Balogh 2009). Elemzési eredményeim azonban a negatív fokozók árnyaltabb nyelvi viselkedésére mutatnak rá.

A vizsgálat fontos hozadéka az, hogy kvantitatív és kvalitatív hasonlóságokat, illetve eltéréseket tárt fel a férfi és a női szerzőségű tweetek között a kutatási kérdés tekintetében. Amint azt a szakirodalmi áttekintésben megmutattam, a problémához kapcsolódó megállapítások száma csekély, és azok sem nagyobb mennyiségű valós példa elemzésén alapulnak. A jelen kutatás, specifikusan a jelenség vizsgálatára készített korpusz alapján pontosabb képet adott a jelenségről, annak kvantitatív és kvalitatív sajátosságairól, emellett a szakirodalmi megállapítások egy részét is cáfolta.

A nyelvtechnológiai kezelés problémáit illetően a következők mondhatók el: Tekintettel arra, hogy a negatív fokozó elemek pozitív és negatív alaptagok módosítóiként szinte kivétel nélkül elveszítették lexikai szintű szemantikai tartalmukat, kielégítő automatikus kezelésüket egy egyszerű pozíciós szabály bevezetésével megvalósíthatónak tartom a szótár- és szabályalapú szentimentelemzésben. E megoldás helytállóságát a kutatómunka eredményeire építő implementációban tesztelem (l. 7). Ugyanakkor az is megállapítható, hogy a semleges alaptagok melletti, valamint az érték váltó elemek ilyen jellegű kezelése nem lehetséges. Bár erre az értekezés keretei között már nem vállalkozhatom, a jövőben azt is meg szeretném vizsgálni, hogy milyen egyéb nyelvi sajátosságok (pl. a tweetek szövegtípusa esetében az emoji és az emotikonok

figyelembevétele, vö. Carvalho és mtsai 2015) tudnák-e az értékváltás, valamint a semleges alaptagok melletti értékvesztő elemek automatikus kezelését támogatni.

6.3.3 A negatív emotív elemek szóasszociációs adatbázisa

A harmadik elemzési eszköz, amelyet a disszertációhoz kapcsolódó kutatási és fejlesztési munka keretében létrehoztam, egy szóasszociációs adatbázis volt, amelyben a felhasznált hívószókat bizonyos szempontok alapján válogatott negatív elemek, valamint – az összehasonlíthatóság céljából – néhány egyéb, fokozó funkciójú nyelvi elem adta. A jelen fejezetben az adatbázis létrehozásának az okát és célját, az adatfelvétel anyagát és alapelveit, a munka eszközét, az adatok gyűjtésének, kezelésének és feldolgozásának a módszerét, végül pedig az adatbázis részletes adatait és vizsgálati eredményeit ismertetem.

6.3.3.1 Az adatbázis létrehozásának az oka és célja

Az asszociációs adatbázisokkal kapcsolatban korábban tárgyaltakat (l. 6.2) megfontolva úgy véltem, hogy az értékvesztésre – és esetlegesen az értékváltásra is – képes elemek szóasszociációs adatbázisa fontos adalékul szolgálhat a jelenség szemantikai-pragmatikai szempontú vizsgálatához, ugyanis a hívószókra adott asszociációk típusai betekintést engedhetnek ezeknek az elemeknek a mentális lexikonban levő, különböző erősségű kapcsolataiba (a szóasszociációs adatbázisok kutatásbeli hasznát részletesen tárgyaltam a 6.2-ben, a vizsgálat hipotéziseit részletesen l. 6.3.3.4).

Mindenekelőtt arra a kérdésre kerestem a választ, hogy a vizsgált negatív emotív elemek esetében melyek azok a szemantikai komponensek, amelyek kontextus nélkül aktiválódni látszanak a rájuk adott válaszok tükrében. A szóasszociációs vizsgálatot tehát elsősorban a lexikai reprezentációs elemzéseimmel (l. 4.6) összefüggésben tartottam fontosnak és hasznosíthatónak.

6.3.3.2 A tesztjellegű adatfelvétel anyaga, eszköze és módszere

A szóasszociációs adatbázis építésének megkezdése előtt elvégeztünk egy kisebb adatfelvételt, tesztelési céllal.⁹⁴ Ingerszóként értékvesztésre képes elemeket (*iszonyú, iszonyúan, durva, durván, borzasztó, borzasztóan, brutális, brutálisan, rohadt, ruhadtul, kurva, kurvára, embertelen, embertelenül*), nem negatív fokozó elemeket (*különösen, meglehetősen, végtelen, végtelenül, abszolút, igazán*), valamint egy semleges fokozó elemet (*nagyon*) választottunk. Mindez összesen 21 szót tett ki.

Egy adatközlő három hívószót kapott, amelyeket előzetesen magunk rendeztük hármas csoportokba. A hívószók között figyelemelterelő feladatokat alkalmaztunk, az előfeszítés minimalizálása céljából (részletesebben l. 6.3.3.3). A kérdőívet a Google Forms nevű eszközzel készítettük el, és online tettük elérhetővé a kitöltők számára.

Az adatfelvétel során 75 adatközlő válaszait gyűjtöttük össze, ami összesen 204 asszociátumot eredményezett. A tesztjellegű adatfelvétel legfontosabb tanulságairól l. a (6.3.3.3)-ban az új adatfelvétel előkészítésével összefüggésben.

6.3.3.3 Az új adatfelvétel anyaga, eszköze és módszere⁹⁵

Az adatfelvételt – akárcsak a tesztelési szakaszban, a végső változatban is – egy online kérdőív segítségével (Google Forms) végeztük (vö. Szabó és mtsai 2017).⁹⁶ Összesen 36 ingerszóval dolgoztunk, tehát 15-tel több elemmel, mint a tesztelés során. Ennek az volt az oka, hogy az azóta a szemantikai motivációs vizsgálati eredményeim (l. 4.5) és a NegEmotív korpusz elemzésének tanulságai (l. 6.3.2) alapján további elemeket is be szerettem volna vonni a vizsgálati körbe: olyanokat, amelyek a kutatás szempontjából valamiért érdekesek voltak, például azt vártam az esetükben, hogy a grammatikalizáció jelét fogják mutatni, vagy kíváncsi voltam, mennyire dominál a negatív tartalmuk vagy a grammatikai formájuk a rájuk adott válaszokban. Ugyanakkor itt nem alkalmaztam minden elemet azok közül, amelyek a tesztjellegű adatfelvételben szerepeltek (l. 6.3.3.2). Ennek az volt a legfőbb oka, hogy a tesztben az ott választott nem negatív fokozó elemekre (pl.

⁹⁴Köszönöm Régeni Annának, hogy elkészítette a kérdőívet, valamint segített az adatok felvételében és feldolgozásában.

⁹⁵A kérdőív felépítését és szövegét l. 6. számú melléklet.

⁹⁶Köszönöm Ilyés Virágnak a segítségét az adatfelvétel megtervezésében, valamint a kérdőív elkészítését, az adatok kezelését és statisztikai feldolgozását.

igazán) nem a várt eredményeket kaptuk minden hívószó esetében: olyan asszociátumokat eredményeztek, amelyek nem vagy nehezen lettek volna összevethetőek a negatív elemekre adott asszociátumokkal. Ebből a szempontból is át kellett gondolni és dolgozni tehát a szóanyagot az újabb adatfelvétel előtt. Bizonyos szavakat tehát, amelyeket a vizsgálat szempontjából ígéretesnek találtam, megtartottam a végső adadfelvétel ingeranyagában is, másokat új elemekkel helyettesítettem. A szavak grammatikai formáját illetően lehetőleg minden lemmához adtam melléknévi és határozószerű alakot is azért, hogy a grammatikai alak jelentőségét is megvizsgálhassam.

Az ingerszók közül egyenként 4-4 elemet kaptak az adatközlők. Bár a tesztjellegű adatfelvétel során csupán három ingerszót kaptak (l. fentebb, 6.3.3.2), az adatközlői visszajelzések alapján úgy véltük, valamelyest emelhetjük az ingeradatok mennyiségét. A négyes adatsortokat véletlen sorsolással alakítottuk ki. A sorsolásnál figyeltünk arra, hogy az egyes csoportokba ne kerülhessen azonos tövű szó.

A 36 ingerszó a lexikai szintű szemantikai tartalmat illetően a következő megoszlási arányt mutatja: 29 negatív elem, amely értékvesztésre – és esetlegesen értékváltásra is – képes, négy olyan fokozó elem, amely a meglepetés emócióval kapcsolható össze, két semleges, valamint egy pozitív fokozó elem.⁹⁷ Emellett fontos szempont volt, hogy a negatív emotív tartalmat illetően mindhárom negatív emóció elemeit szerettem volna bevinni az ingerszók közé.

Az ingeranyag megválasztásánál további két fontos sajátsgot is szem előtt tartottam: Egyrészt, a negatív elemeket illetően szerettem volna, ha a lexikai szintű szemantikai tartalmukat illetően a három negatív emócióhoz, azaz a félelem, a düh és az undor érzelmekhez kapcsolódó elemek egyaránt szerepelnek a hívószók között (a lexikai szintű emotív szemantikai tartalom kérdéséről részletesen l. 4.4). Meg szerettem volna vizsgálni ugyanis, hogy az eltérő emóciókhoz kapcsolódó negatív elemek asszociatívai mennyiben mutatnak átfedést, illetve eltérést. Másrészt – akárcsak az Agykapocs-kutatásban – a hívószók többféle nyelvtani alakban szerepelhettek (vö. Kovács 2013: 42). Ez adott lehetőséget ugyanis annak feltárására, hogy mennyiben befolyásolja a válaszokat, ha eltérő szintaktikai funkciók betöltésére alkalmas alakokat adunk az adatközlőknek. Ennek a megoldásnak köszönhetően például megtudhattam, az asszociáció során a hívószó grammatikai szerepe vagy a szemantikai tartalma jut nagyobb szerephez.

⁹⁷ A kutatáshoz további értékes adalékul szolgálhatna, ha fordított irányban kérnénk az adatközlőktől az asszociációkat, tehát melléknemekhez vagy határozószerűkhöz várnánk fokozó elemeket válaszként. Ezt a jelen disszertáció keretei már nem tették lehetővé, de fontos jövőbeli kutatási feladatnak tartom.

A hívószókat a szemantikai tartalom szerinti bontásban az alábbi táblázat közli:

félelem	negatív		meglepetés	semleges	pozitív
	düh	undor			
<i>borzasztó, borzasztóan, brutál, brutális, brutálisan, durva, durván, embertelen, embertelenül, eszméletlenül, félelmetesen, kegyetlen, kegyetlenül, rettenetesen, szédületesen, szörnyen</i>	<i>átkozott, átkozottul, kibaszott, kibaszottul, kurva, kurvára, rohadt, rohadtul</i>	<i>iszonyat, iszonyatosan, iszonyú, iszonyúan</i>	<i>hihetetlenül, bámulatosan, elképesztően, megdöbbenően</i>	<i>nagyon, meglehető sen</i>	<i>fantaszti kusan</i>

10. táblázat: A hívószók listái, szemantikai tartalom szerinti bontásban⁹⁸

Amint azt a (6.2)-ben már ismertettem, a szóasszociációs adatbázisok elkészítésének több módszerét különböztethetjük meg aszerint, hogy az adatfelvétel hogyan történik, milyen instrukciókat kapnak az adatközlők. A jelen adatfelvétel a szabad asszociációs típushoz tartozik, amelyben a válaszadóknak semmilyen előzetes kategóriát vagy szempontot nem adtunk meg. Ugyanakkor, bár a tesztjellegű adatfelvétel során (l. 6.3.3.2) a hívószókra egyetlen választ adhattak az adatközlők, úgy véltem, informatívabb lehetne az adatbázis egynél több asszociátum lehetősége esetén. Az adatközlők tehát több, de legfeljebb öt választ adhattak egy-egy ingerszóra.

Az asszociációs adatfelvételeknél nagyon fontos foglalkozni az ún. szemantikai előfeszítés (*priming*) problémájával, tehát azzal, hogy a válaszadás előtti bármilyen cselekvés, esemény vagy információ befolyásolhatja az eredményt (vö. Kovács 2013: 37) (l. 6.2). Annak céljából, hogy az adatközlés során a hívószók és az azokra adott válaszok minél kisebb hatással lehessenek az azokat követő asszociátumokra, a hívószók közé különböző feladatokat tettük figyelemelterelés céljából. A feladatok nem nyelvi, hanem vizuális megoldást követeltek meg az adatközlőktől, ezzel is kerülve a válaszok lehetséges befolyásolását. A feladatok jellegét illetően klasszikus, az adatfelvételek során alkalmazott

⁹⁸ Az itt közölt listák már a végső adatfelvétel ingeranyagát mutatják be.

matematikai-logikai feladatokat választottunk.⁹⁹ Mindemellett azért, hogy előfeszítést a hívószót megelőző instrukció se okozhasson, a feladat leírását még a kérdőív elején közöltük. Ily módon az egyes hívószók már csupán önmagukban, a feladatok között álltak.

A szakirodalmi adatok alapján az asszociátumokat illetően eltéréseket figyelhetünk meg a válaszadó kora (vö. Libárdi 2001, Gósy 2000, Gósy 2001), neme (vö. Cser 1939, Kegyesné 2002, Rosenzweig 1970), képzettsége (szintje és szakterülete) (vö. Cramer 1968, Rosenzweig 1970), valamint a nyelvismeret szintje (vö. Söderman 1993) szerint (l. 6.2). Annak céljából, hogy az adatbázist e szempontok alapján is vizsgálni lehessen, a kérdőív végén az adatközlőkre vonatkozóan a következő adatokat is felvettük: neme, kora, iskolai végzettsége, a magyar-e az anyanyelve, valamint Magyarország területén, a határon túlön vagy külföldön élő magyar-e. Emellett – mivel ezt a vizsgált nyelvi jelenséggel összefüggésben fontosnak tartottuk – megkértük, hogy jelölje be az aktuális hangulatát egy ötfokozatú skálán.

Amint arról a (6.2)-ben már szintén volt szó, folyamatos asszociációk esetében a válaszokat eltérően kell súlyoznunk. A jelen adatbázis adatait tehát magunk is úgy tároltuk, hogy az adatokat azok „beérkezési” sorrendje szerint külön tudjuk kezelni, mind a vizsgálatokban, mind az adatok esetleges jövőbeli vizualizációjában.

6.3.3.4 A vizsgálat hipotézisei

A vizsgálattal kapcsolatos hipotézisek megfogalmazásához a szakirodalom megállapításaira támaszkodtam (l. 6.2), figyelembe véve a pszichológiai és szemantikai motivációra vonatkozó, valamint a különböző korpuszvizsgálati eredményeimet (l. 4.4, 4.5, 6.3.1.7, 6.3.2.7) és a tesztjellegű adatfelvétel tapasztalatait is (l. 6.3.3.2). Tekintsük a hipotéziseket részleteiben is!

Mindenekelőtt, az eredményeket befolyásolni fogja az, hogy a hívószók között toldalékolt és toldalékkal nem rendelkező szavak egyaránt előfordulnak. Kovács (2013: 42) alapján a szótári alakú hívószók legtöbbször szintén szótári alakú, azonos szófajú válaszszókat hívnak. A toldalékolt alakok ugyanakkor Lengyel (2008: 24) alapján nagyobb százalékban aktiválnak más szófajú szavakat, többnyire szintagmatikus kapcsolatok

⁹⁹Köszönöm Ilyés Virágnak a feladatok gondos kiválasztását és a kérdőívbe illesztését.

formájában. Ez a sajátosság meg fog jelenni az asszociatív átfedésekre vonatkozó vizsgálatok eredményeiben is.

Azt is feltételezem, hogy a negatív hívószók esetében valamelyest a lexikalizációra, illetve a lexikalizáltság fokára utaló jeleket is mutat majd az asszociátumok szófaji megoszlása. Ahogyan azt a mentális lexikon hálózatos modellezése kapcsán már tárgyaltam (l. 4.6.2), a többdimenziós felépítéssel összefüggésben az egyes szavak, valamint szócsoportok között egyszerre több és többféle kapcsolat is létezhet (vö. Lengyel 2008: 18). Ezek a kapcsolatok „részben lingvisztikai (fonetikai, grammatikai, szemantikai) tartalmúak [...], részben használatnyelvészeti (pragmatikaiak, szociolingvisztikaiak, szociokulturálisak)” (Lengyel 2008: 18). Részben ennek a sajátásnak köszönhetően a különböző hálózati komponensek között a kapcsolatok erőssége nem azonos. Mindezek alapján azt tételezem, hogy, mivel a negatív hívószók lexikalizáltsági foka nem azonos, azoknak a mentális hálózatban levő eltérő erősségű kapcsolatai az asszociátumokban is meg fognak mutatkozni. Úgy vélem, hogy minél lexikalizáltabb egy elem fokozó szerepben, annál gyakoribb, hogy melléknévi választ hív, amellyel kollokálódva ezt a fokozó szerepet betöltheti (vö. pl. a *borzasztó ember* és a *borzasztó nagy* kollokációkat). Az elképzelést támogatja Kugler (2014: 137) észrevétele is, mely szerint a negatív fokozó elemek, „bár intenzifikáló szerepüket hatásosságuknak köszönhetik, minél kiterjedtebbé válik az alkalmazásuk [...], annál ártalmatlanabbak is lesznek az érzelmi működésünk szempontjából”.

A fentebb elmondottakkal összefüggésben, a lexikai szintű szemantikai tartalom, valamint a grammatikai forma befolyásoló ereje beláthatóan bizonyos esetekben egymás ellen ható tendenciaként is jelentkezhetnek.

A szentiment- és az emócióelemzés eredményeit illetően azt feltételezem, hogy minden típusú szemantikai tartalom esetében jelentős összefüggést fogunk látni a hívószó és a válaszszó értékelő, illetve emotív szemantikai tartalma között. Mivel a hívószókat önmagukban látják az adatközlők, úgy vélem, hogy annak inherens szentimentértéke (amennyiben van) meghatározza majd a válaszok szentimentértékét is. Ehhez hasonlóan a hívószók emotív tartalma (amennyiben van) is korrelálni fog a válaszok emotív tartalmával. Ez a sajátosság – a grammatikai tulajdonsághoz hasonlóan (l. fentebb) – ugyancsak meg fog jelenni az asszociatív átfedésekre vonatkozó vizsgálatok eredményeiben is.

Azt is valószínűsítem, hogy a fentebbi sajátságoknak (a hívószók grammatikai sajátságai, valamint szentiment- és emócióértéke) megfelelően fognak az egyes hívószókra kapott asszociációk asszociatív átfedéseket mutatni (vö. Pléh 2008: 835).

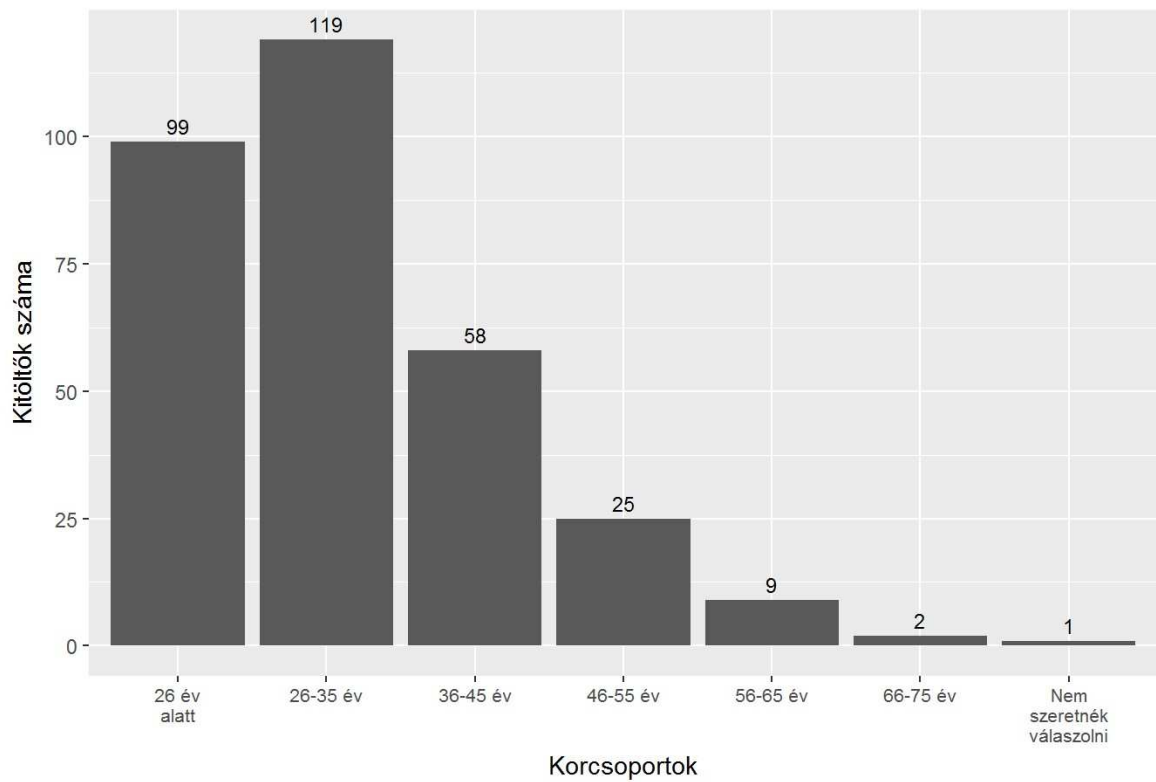
Az adatok negatív emotív elemek listájával történő elemzését illetően úgy vélem, hogy a legtöbb negatív emotív elemet hasonló tartalmú és funkciójú elem hívja majd.

Végezetül, a férfi és a női asszociátumok hasonló, illetve eltérő sajátságaira vonatkozó hipotézisemet a NegEmotív korpusz vizsgálati eredményei (l. 6.3.2.8), valamint Kovács (2013: 39) megállapításai alapján fogalmaztam meg. Kovács (2013: 39) rámutat, hogy az 1939-ben felvett vizsgálati eredményekhez képest az Agykapocs-kutatás adatai szerint a női és a férfi adatközlők aktiválása nagymértékben megváltozott. Az újabb adatokban ugyanis már nem olyan egyértelműen mutathatók ki a nemek kategóriái. A NegEmotív korpusz vizsgálata ugyancsak azt tárta fel, hogy az értékvesztő és értékváltó elemek használatában a két nem között nincs igazán kimagasló (pl. a gyakorisági megoszlás tekintetében tapasztalható) eltérés. Mindezek alapján azt feltételeztem, hogy az adatbázis vizsgálata során, bár megfigyelhetünk majd némi eltérést, nem fogunk jelentősebb különbséget tapasztalni a nemek között.

6.3.3.5 Az adatbázis alapvető statisztikai adatai

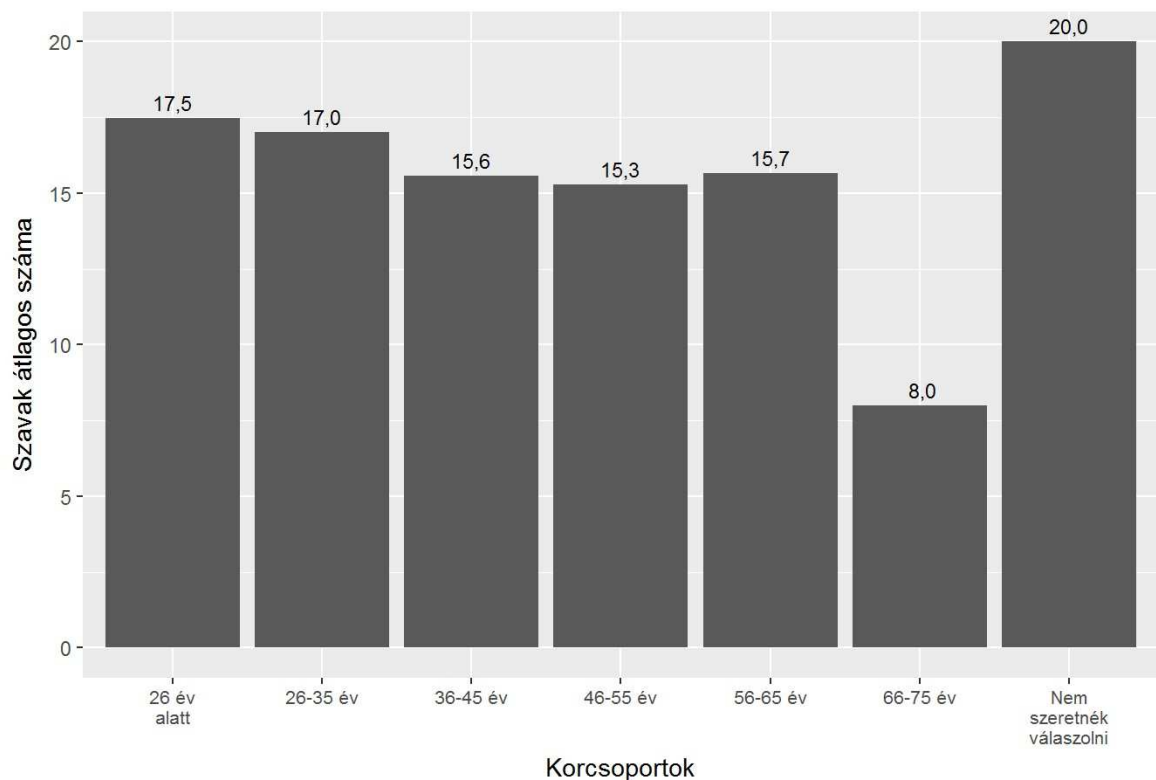
A szóaszociációs kérdőívünket összesen 313 adatközlő töltötte ki, ebből 252 nő és 61 férfi volt. Kitöltőnként a 4 ingerszóra átlagosan 16,6, összesen a teljes adatfelvétel során 5198 asszociátum érkezett. Amennyiben a mennyiségi adatokat nemek szerinti bontásban tekintjük, azt látjuk, hogy amíg a női adatközlők átlagosan 17, addig a férfiak csupán 15,5 választ adtak.

A kitöltők száma az egyes korcsoportokhoz tartozóan az alábbi megoszlást mutatja:



11 ábra: A kitöltők száma az egyes korcsoportokban

A korcsoportokat illetően az adatfelvétel egy érdekes tapasztalata az, hogy a kor növekedésével az adatközlők kevesebb asszociátumot adtak. Az eredményeket az alábbi diagram mutatja be úgy, hogy azon minden korcsoportnál az adatközlők válaszainak átlagos számát látjuk.



12. ábra: A válaszadók asszociátumainak átlagos száma, korcsoportok szerinti bontásban¹⁰⁰

Az iskolai végzettséget illetően a kitöltők túlnyomó többsége (222: 70,92%) a felsőfokú (egyetemi vagy főiskola, alap- vagy mesterszak) végzettséget jelölte meg. A 2. és a 3. leggyakoribb a gimnázium (50: 15,97%), valamint a szakközépiskola, technikum (30: 9,58%) voltak. Az adatközlők közül öten (1,59%) PhD-fokozattal rendelkeztek.

A lakóhely tekintetében a kitöltők közül 128 (40,89%) fővárosi lakos volt, 74-en (23,64%) a megyei jogú várost, 73-an (23,32%) a várost, 35-en (11,18%) pedig a községet adták meg. 3 adatközlő erre a kérdésre nem szeretett volna válaszolni (0,95%).

Arra a kérdésre, miszerint Magyarországon, vagy pedig a határon túli területen, illetve külföldön élő magyar-e, 295 adatközlő (94,24%) jelölte meg az első válaszlehetőséget.

A hangulatot illetően az ötfokozatú skálán a legtöbben (151: 48,24%) a 4. fokot jelölték meg, amelyet a 3. (94: 30,03%), majd az 5. fokozat (38: 12,14%) követett.

¹⁰⁰ Tekintettel arra, hogy a 66 és 75 év közé eső korcsoportban összesen két válaszadó volt, az asszociátumok átlagos számából valódi következtetéseket nem vonhatunk le.

6.3.3.6 Az adatbázis vizsgálati eredményei

Amint azt a (6.2)-ben már tárgyaltam, a hívószókra kapott asszociátumokon kvantitatív és kvalitatív vizsgálatok egyaránt végezhetőek (vö. Kovács 2013: 40). Kvalitatív szempontú vizsgálatot végzünk például akkor, ha a leggyakrabban adott első asszociátumok grammatikai sajátságait (pl. szófajt) elemezzük, vagy pedig szemantikai elemzésnek vetjük alá őket (vö. Pléh 2008: 799). Kvantitatív szempontból vizsgáljuk az adatbázist például akkor, ha azt tárjuk fel, hogy a különböző ingerszók összesen hány különböző válaszsót hívtak.

A következőkben részletesen bemutatom, hogy adatbázisunk mely legfontosabb információkkal szolgál a kutatási kérdés tekintetében. Annak céljából, hogy a különböző szemantikai tartalommal rendelkező stimulusok asszociátumait összevethessük egymással, és a rájuk adott válaszok hasonlóságait, illetve eltéréseit megvizsgálhassuk, az eredményeket több alkalommal az ingerszók hatos csoportosítása (l. fentebb, 10. táblázat) szerint mutatom be.

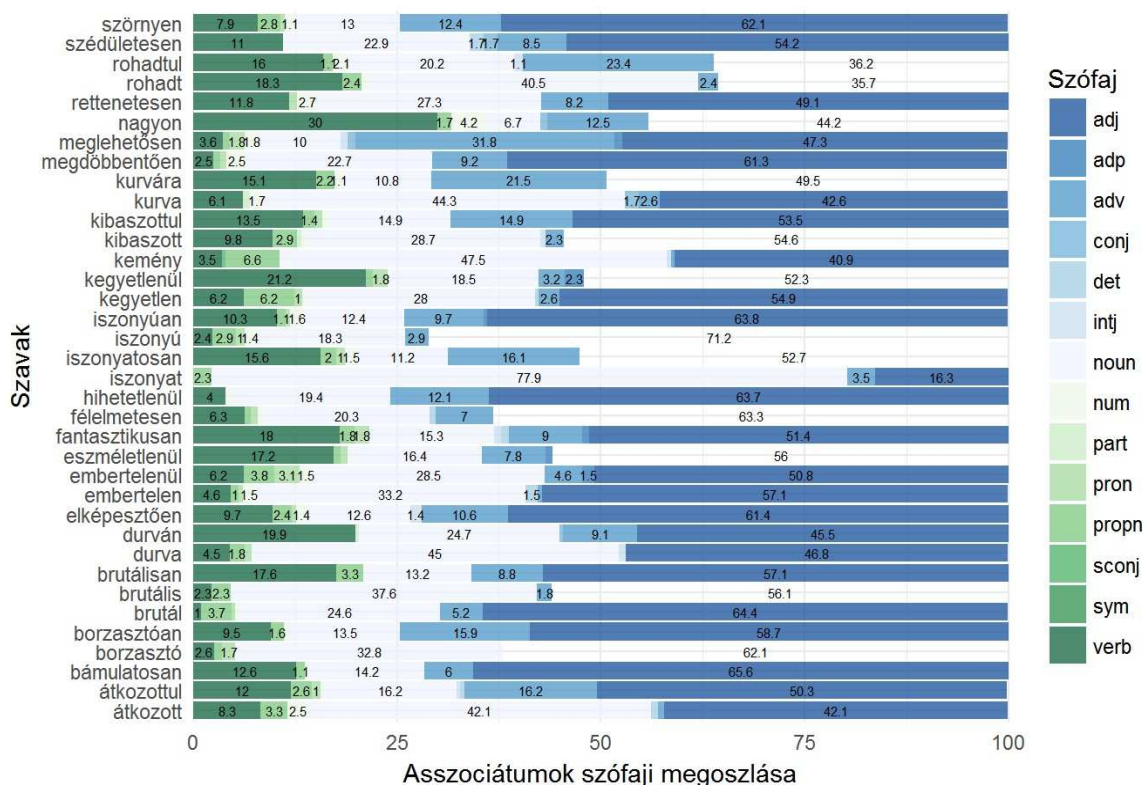
Első lépésben nézzük meg, mit tapasztalunk, ha az asszociátumokat a szófaji megoszlások szempontjából vesszük górcső alá! (A szófaji elemzéseket minden esetben a magyarul elemző eszközzel (vö. Zsibrita és mtsai 2013) végeztük, és az eszközben definiált szófajokra hagyatkoztunk. A szófajok rövidítéseit, amelyeket az ábrákon feltüntetünk, a *Rövidítések jegyzékében* oldom fel.)

A hívószó és az asszociátum közötti szófaji kapcsolat lehet szófajmegőrző és szófajváltó (vö. Lengyel 2010: 196–198, 2012: 18–22). A szófajmegőrzőek tagjai paradigmatis, míg a szófajváltó típusban levő párok legtöbbször szintagmatikus viszonyban állnak egymással. Kovács (2013: 42) alapján a hívószók toldalékolta, illetve toldalék nélküli volta jelentősen befolyásolja az asszociációs vizsgálatok kimenetelét, és a szótári alakú hívószók legtöbb esetben szintén szótári alakú, azonos szófajú válaszokat hívnak. A toldalékolta szavak ugyanakkor nagyobb százalékban aktiválnak más szófajú szavakat, többnyire szintagmatikus kapcsolatok formájában.

A fentebbi sajátságoknak a mentális lexikon hálózatos elrendeződése szempontjából van nagy szerepe. A szemantikai hálót ugyanis az egyes fogalmak közötti szinonim, antonim, hiperonim, holonim, meronim kapcsolatok építik fel, míg a kollokációs hálózatot a gyakori együttes előfordulású szavak, illetve az állandó kifejezések kapcsolatai alkotják (vö. Kovács 2013: 188–190). Ily módon a válaszok szófaji sajátságai betekintést

engedhetnek e hálózatok felépítésébe, és arra is rámutathatnak, hogy egy-egy hívószó esetében mely hálózat jut domináns szerephez.

Az adatbázisom asszociátumainak szófaji megoszlását a 13. ábra mutatja be. A könnyebb értelmezhetőség és az átláthatóság céljából az adatokat százalékokban adom meg, és csupán az 1% feletti értékeket jelenítem meg.



13. ábra: Az asszociátumok szófaji megoszlása, az összes válaszsóra nézve¹⁰¹

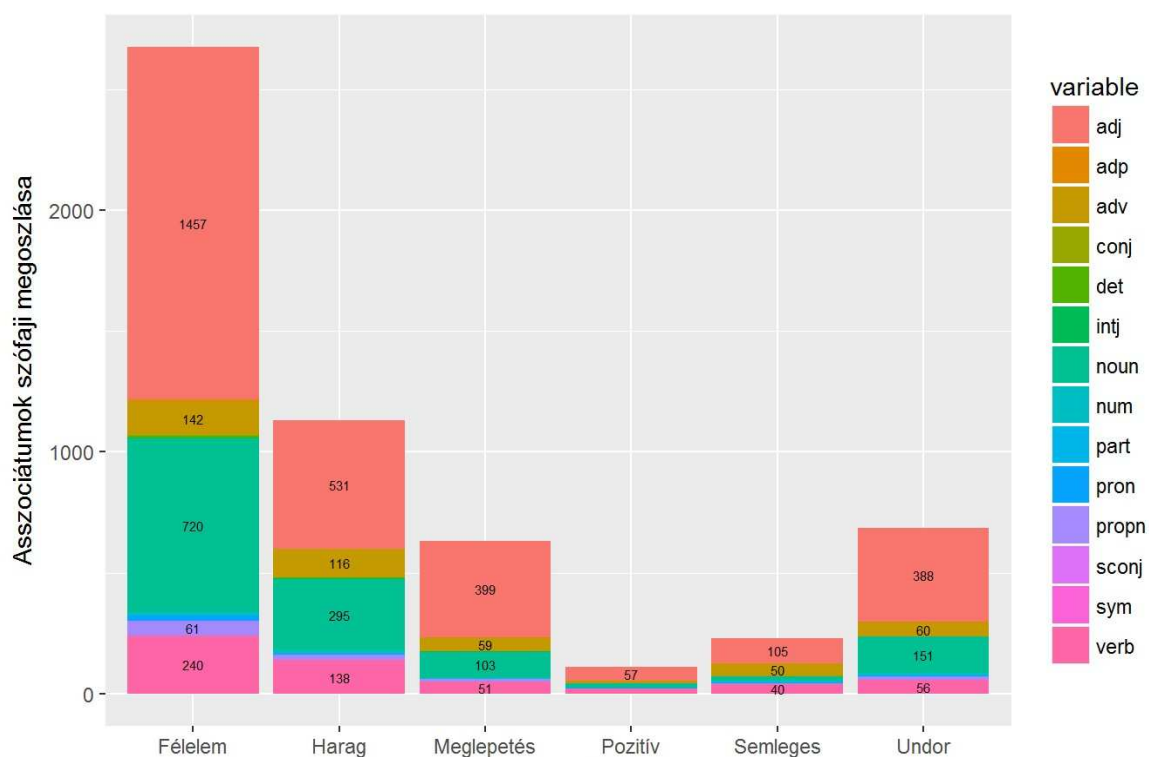
A legtöbb hívószó esetében az asszociátumok többsége melléknévi szófajú, majd azt a főnévi, végül az igei és a határozószói válaszok követik. Ez utóbbi két kategóriában a gyakoriság kiegyenlített képet mutat. A válaszok szófaját valószínűleg alapvetően a hívószó szófaja befolyásolhatja: így például a főnévi alakú *iszonyat* a legtöbb esetben főnévi asszociátumot hív. Ugyanakkor, szerényebb mértékben ugyan, de a negatív elemek lexikalizáltsági foka is kirajzolódni látszik azokban az esetekben, ha a hívószót nem határozószói alakban adtuk meg. Hasonlítsuk össze például a *borzasztó–borzasztóan* és az *átkozott–átkozottul* hívószók asszociátumait! Amíg a határozószói alakok esetében körülbelül ugyanolyan a válaszok szófaji aránya (a melléknév túlnyomó többségben van mindkét hívószó esetében), addig az előbbieket asszociátumainak az aránya a szófaji

¹⁰¹ A hívószók az ábrákon és a táblázatokban ékezetek nélkül jelennek meg, mert az adatok feldolgozását gépi elemzéssel végeztük, és az elemző eszköz ékezetek nélküli adatokon működik.

megoszlás tekintetében eltérő. A *borzasztó* válaszainak szófaji arányai a *borzasztóan* hívószóéhoz közelítenek, az *átkozott* esetében ugyanakkor a főnévi és a melléknévi válaszok gyakorisága azonos. A *borzasztó* tehát lexikalizáltabb fokozónak mutatkozik, és ez az eredmény összességben a NegEmotív korpusz vizsgálati tapasztalataival (l. 6.3.2.7), melyek szerint ez az elem gyakrabban fordul elő fokozó szerepben az *átkozott*-nál.

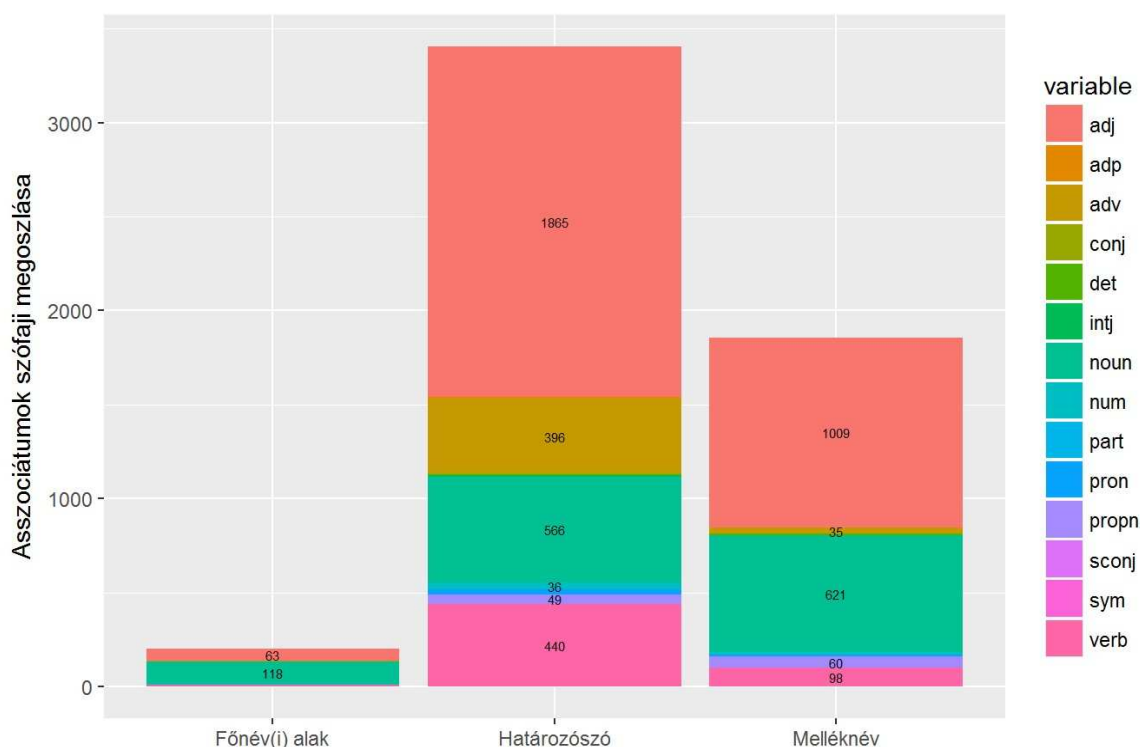
Hasonló szempontból érdekes a *durván* hívószó is, amely több mint kétszer annyi melléknevet hív, mint igét. Véleményem szerint ez arra mutat, hogy az elem egyre lexikalizáltabbá válik fokozó funkciójú elemként (vö. pl. *durván jó*), ezáltal egyre távolodik módhatározószoói funkciójától (vö. pl. *durván viselkedik*).

A szófaji megoszlást befolyásoló tényezők további vizsgálatához készítettünk egy-egy ábrát a hívószók szemantikai, valamint grammatikai sajátosságai szerinti bontásban is.



14. ábra: Az asszociátumok szófaji megoszlása
a hívószók szemantikai tartalma szerinti bontásban¹⁰²

¹⁰² A disszertáció bírálója, Szécsényi Tibor megjegyzi, hogy szerencsésebb lett volna a 14-17. ábrákon a válaszcikkek számát és a szófaji megoszlást külön kezelni. Sajnos az ábrák ezen verzióit a dolgozat leadásáig kapacitás hiányában nem állt módunkban elkészíteni.

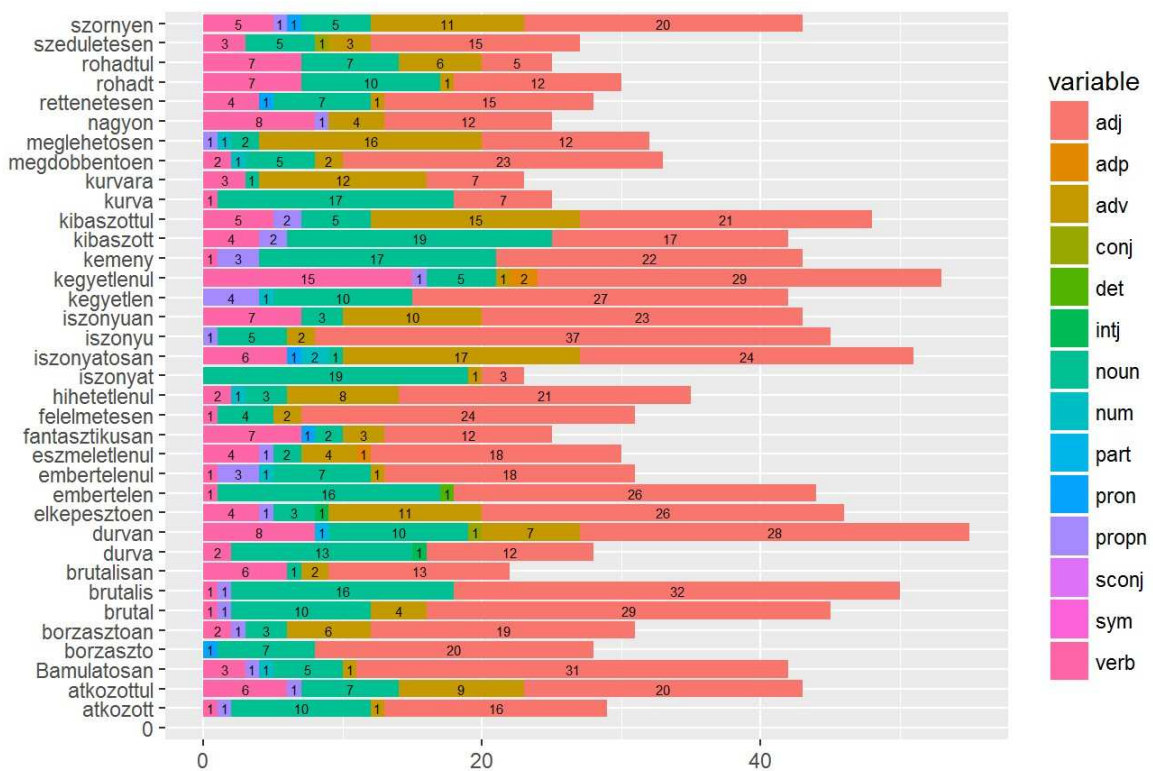


15. ábra: Az asszociátumok szófaji megoszlása
a hívószók grammatikai sajátysága szerinti bontásban

A vizsgálat megmutatta, hogy a válaszok szófajának gyakorisági megoszlása a hívószók grammatikai sajátyságával szoros összefüggést mutat. A főnévi alakú elemek legtöbbször főnevet, ritkábban melléknevet hívnak, míg ige szinte sohasem. A határozószók és a melléknevek esetében egyaránt a melléknévi asszociátum a leggyakoribb, és mindkét esetben azt a főnévi szófaj követi.¹⁰³ Jelentősebb eltérés azonban, hogy amíg a határozószók esetében az ige és a határozószó is gyakori asszociátum, addig ugyanezek a melléknévi hívószók esetében kifejezetten ritkák.

Az alábbi ábrán azt látjuk, hogyan alakul az asszociátumok szófaji megoszlása, ha kizárólag az első helyen érkező válaszciklusokra készítjük el a 13. ábrán bemutatott statisztikát:

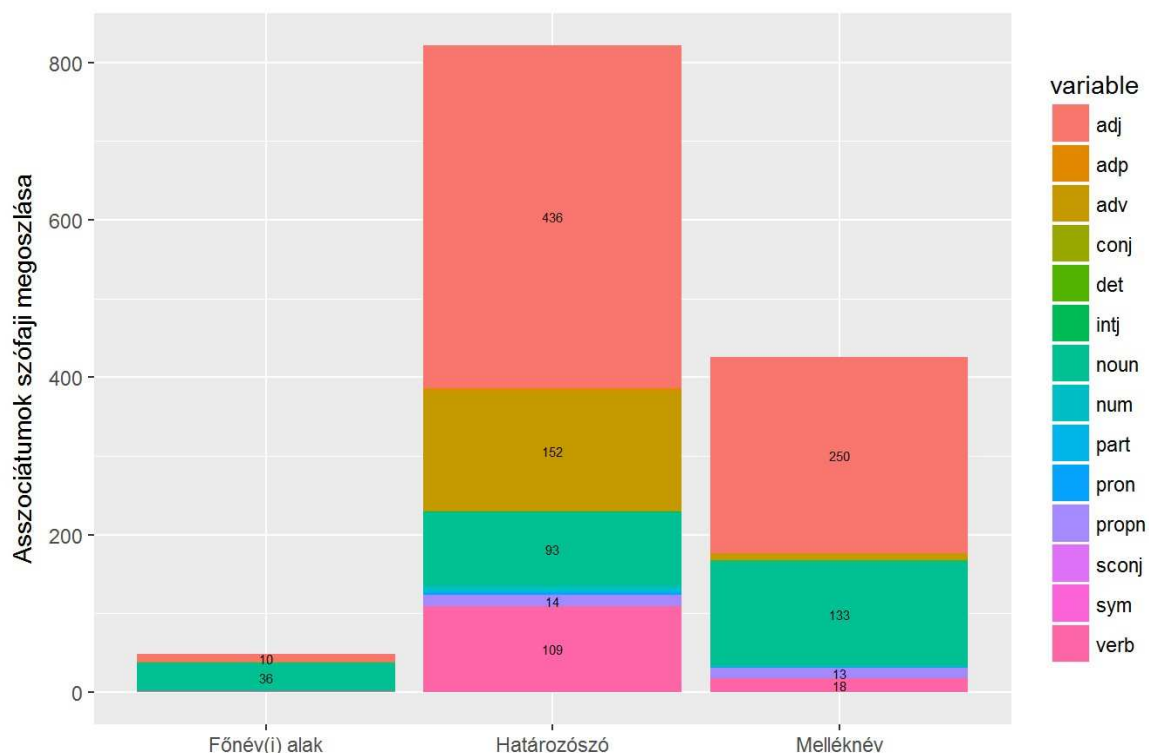
¹⁰³ A határozószóra adott főnévi asszociátumok esetében egyrészt a határozószó szemantikai tartalma alapján érkeztek a válaszok, például *kegyetlenül – kínzás, hóhér; iszonyúan – rettegés, borzalom*, másrészt kaptunk főnévi alakú választ a határozószó szótővével is, például *iszonyúan – iszony*.



16. ábra: Az első helyen adott asszociátumok szófaji megoszlása

Összevetve a 13. és a 16. ábra adatait, a legszembeötlőbb sajátosság talán az, hogy amíg az első helyen adott asszociátumok között viszonylag gyakoribbak a határozószói szófajú válaszok, addig a további asszociációkban azok száma lecsökken, elsősorban a melléknévi asszociátumok javára.

Ha az első helyen adott asszociátumokat a hívószók szófaja szerinti bontásában tekintjük, még látványosabbá, és egyben árnyaltabbá válik a tendencia:



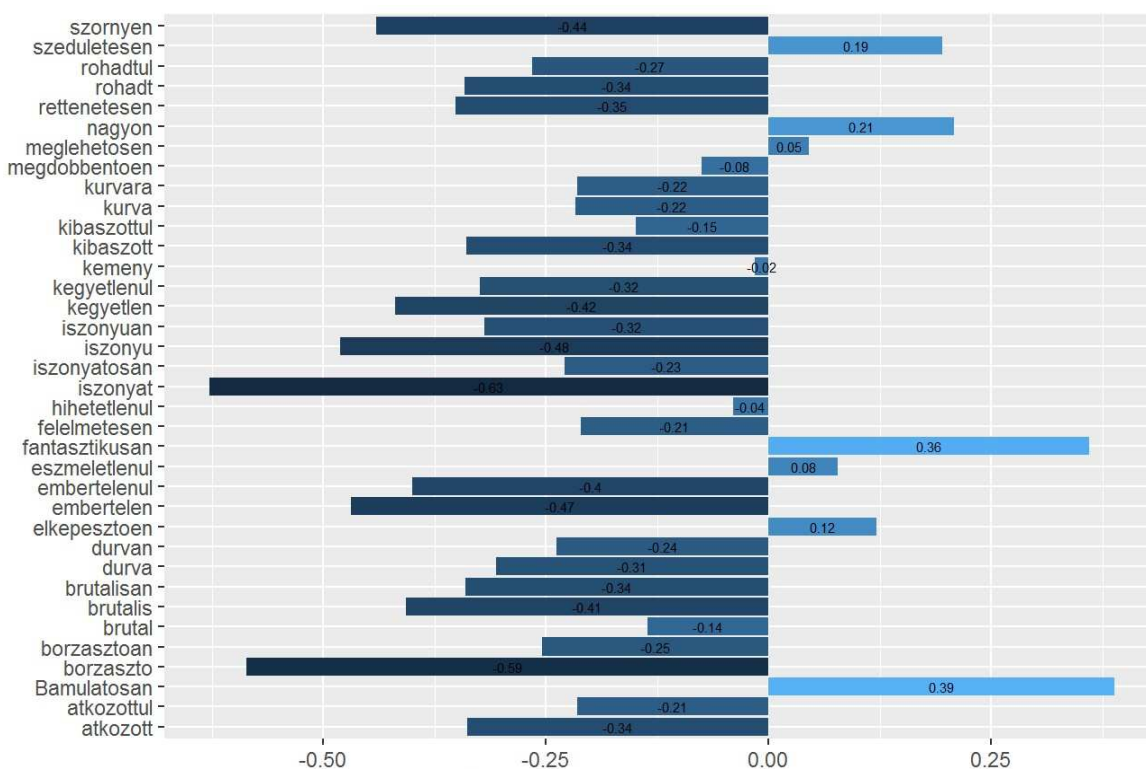
17. ábra: Az első helyen adott asszociátumok szófaji megoszlása a hívószók grammatikai sajátsága szerinti bontásban

A 15. és a 17. ábrákat összevetve azt látjuk, hogy a főnévi alakokra főnévi, a határozószói alakokra határozószói alakok gyakrabban érkeznek az első, mint a további helyeken, és mindkét szófaj esetében a további asszociátumok adásával a melléktípusok javára tolódik el a gyakorisági arány. A melléktípusi hívószók esetében ilyen aránybeli változást nem tudunk azonosítani. Amíg tehát az első asszociátumok grammatikai sajátságai gyakran azonosak a hívószó grammatikai sajátságaival, addig a további asszociátumokban többször jelennek meg eltérő grammatikai sajátságokkal rendelkező, esetlegesen szemantikai vagy kollokációalapú válaszok. Ezt a megállapítást természetesen több és más sajátságokkal rendelkező hívószó bevonásával lehetne igazolni vagy cáfolni.

Az asszociátumok szemantikai vonatkozásait illetően megvizsgáltam, vajon milyen sajátságokat tárhatunk fel a szentiment- és az emóciósztáraink segítségével. Szentimentsztárunk a jelen kutatómunka keretében készült, és két listában tartalmazza a negatív és a pozitív szemantikakifejezéseket (vö. Szabó 2014, 2015b) (részletesebben l. 7.2). Az emóciósztárat Ekman és Friesen (1969) érzelmekategorizálási rendszere alapján

hoztuk létre szövegalapú tartalomelemzési céllal (vö. Szabó–Morvay 2015, Szabó–Vincze–Morvay 2016c). A lexikonokat tehát a válaszszók csoportjára illesztettük.

A szentimentszótárral végzett elemzés összesített eredményét az alábbi ábra mutatja be.



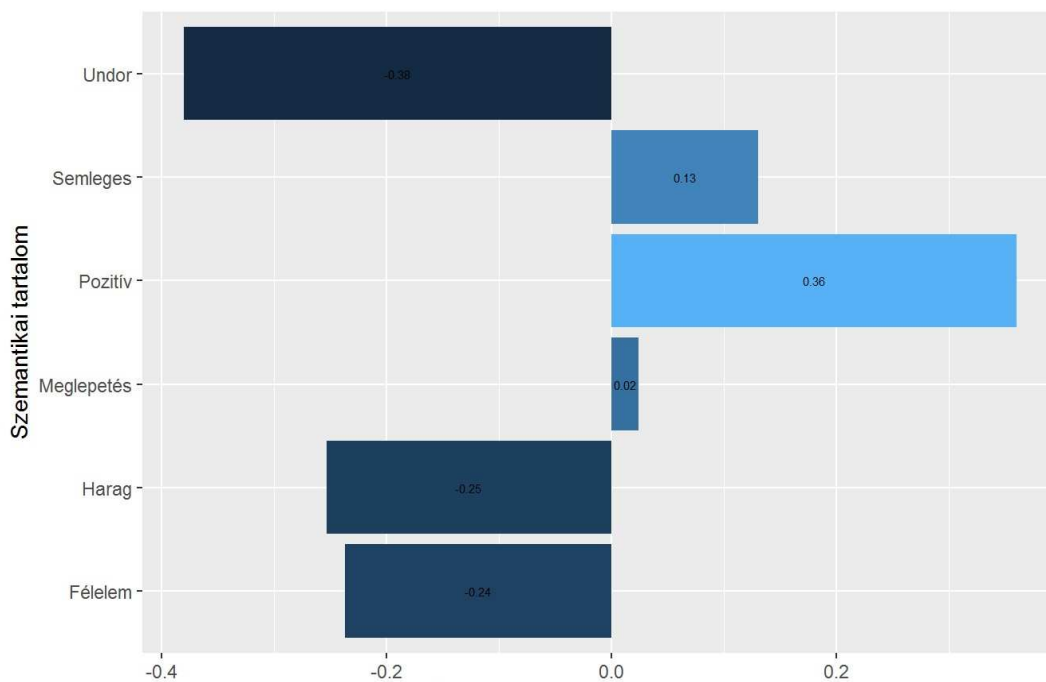
18. ábra: Az asszociátumok összesített szentimentelemzési eredményei

Az ábrán látható értékeket a következőképpen kaptuk: minden ingerszó összes választát elemeztük a szótárral, és minden, a szótár alapján pozitívként azonosított szó +1, minden negatív -1 értéket kapott. Végül az egyes ingerszókra érkezett válaszok szentimentértékeit átlagoltuk.

Az adatok alapján az asszociátumok szentimentértékei a hívószók szemantikai tartalmával szoros összefüggést mutatnak. Még azoknak az elemeknek a válaszai is, amelyek a NegEmotív korpusz elemzése alapján (l. 6.3.2.7) gyakori fokozó elemnek mutatkoznak és viszonylag magas kollokációs potenciállal rendelkeznek (mint pl. az *iszonyat*), szentimentértéküket tekintve a hívószók szemantikai tartalmától függenek. Hasonlóan, a NegEmotív korpusz alapján láttuk, hogy például a *durva* nagyon gyakran szerepel értékváltóként pozitívan értékelő szerepben, az asszociációs vizsgálatban, kontextus híján mégis a negatív tartalma határozza meg a rá adott válaszok értékét.

Érdekes tapasztalat, hogy amennyiben az azonos tövek melléknévi és a határozószói alakjainak adatait vetjük össze (pl. *borzasztó–borzasztóan, iszonyú–iszonyúan* stb.), az előbbire tendenciaszerűen több negatív válasz érkezik.

Amennyiben a hívószókat azok szemantikai tartalma szerinti hatos bontásban vizsgáljuk, a következő ábrát kapjuk.



19. ábra: Az asszociátumok összesített széntimentelemzési eredményei, a hívószók hatos csoportosítás szerint

A 18. és a 19. ábrák adatai alapján a következő legfontosabb megállapításokat tehetjük: A szemantikai tartalmukat tekintve semleges és pozitív elemekre adott asszociátumok átlagos széntimentértéke rendre pozitív. A meglepődés emóció elemei összesítve minimális mértékben, de pozitív széntimentértékű asszociátumokat eredményeztek. A meglepődés hívószói közül a *bámulatosan* és az *elképesztően* pozitív, a *hihetetlenül* és a *megdöbbentően* csekély mértékben ugyan, de negatív átlagos széntimentértékű válaszokat hív. Az adatfelvételben alkalmazott negatív elemek, toldalékoltságtól függetlenül átlagosan rendre negatív értékű asszociátumokat eredményeztek. A legnegatívabb asszociátumokat az *iszonyat*, a *borzasztó*, az *iszonyú*, az *embertelen* és a *szörnyen* hívták.

A széntimentelemzést elvégeztük úgy is, hogy csupán az első helyen adott asszociátumokat vettük figyelembe, ugyanakkor az első, valamint az összes helyen adott asszociátumok a széntimentértékek tekintetében semmilyen érdemi eltérést nem mutattak.

félelem	düh	undor	meglepődés	semleges és pozitív
---------	-----	-------	------------	---------------------

Hívószó	Legnagyobb átfedés	Legkisebb átfedés
embertelenul	brutális	iszonyat
hihetetlenul	elkepesztoen	iszonyat
megdobbentoen	elkepesztoen	kemény
meglehetosen	atkozottul	iszonyat
atkozott	kibaszott	iszonyat
borzasztoan	iszonyatosan	iszonyat
felelmetesen	borzasztoan	iszonyat
rohadtt	kegyetlen	meglehetosen
atkozottul	elkepesztoen	iszonyat
elkepesztoen	atkozottul	iszonyat
embertelen	kegyetlen	fantasztikusan
kegyetlen	embertelen	meglehetosen
brutal	kibaszottul	meglehetosen
iszonyu	iszonyuan	kemény
kemény	brutal	megdobbentoen
kibaszottul	elkepesztoen	iszonyat
borzaszto	iszonyu	kemény
durva	durvan	szeduletesen
eszmeletlenul	elkepesztoen	atkozott
rettenetesen	iszonyatosan	megdobbentoen
brutalisan	brutális	meglehetosen
iszonyat	iszonyu	meglehetosen
kurvara	kibaszottul	iszonyat
rohadtul	kibaszottul	kemény
brutális	durvan	meglehetosen
durvan	brutális	iszonyat
iszonyatosan	iszonyuan	kurva
kegyetlenul	durvan	megdobbentoen
Bamulatosan	elkepesztoen	embertelen
iszonyuan	iszonyatosan	durva
kibaszott	kibaszottul	durva
szornyen	atkozottul	kemény
fantasztikusan	elkepesztoen	embertelen
kurva	iszonyu	iszonyat
nagyon	elkepesztoen	iszonyat
szeduletesen	elkepesztoen	durva

11. táblázat: Az egyes hívószók asszociátumainak legnagyobb és legkisebb átfedései

A táblázat adatai alapján a következő három legfontosabb tendenciát látjuk kirajzolódni: Egyrészt, a legtöbb hívószó vele azonos grammatikai sajátosságú hívószóval mutatja a legnagyobb átfedést (a 32-ből 23 esetében). Másrészt, az azonos tövek gyakran egymással

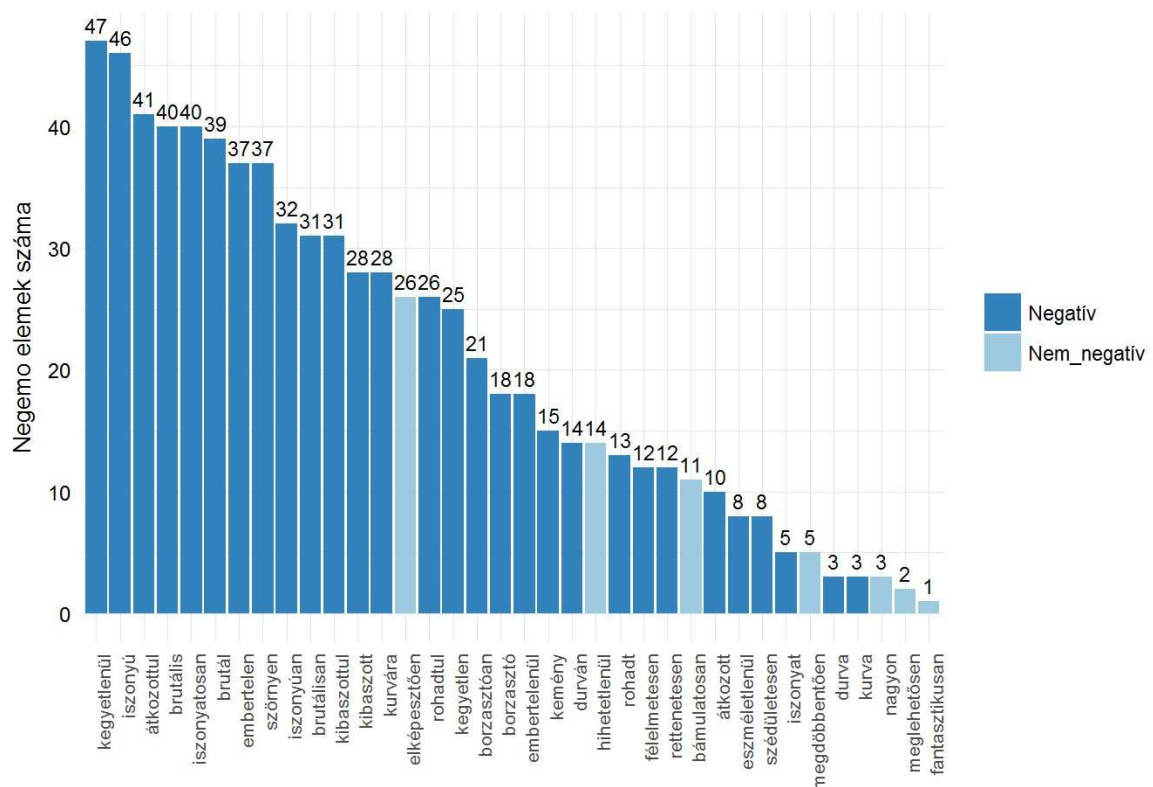
fednek át a legnagyobb mértékben (pl. *durva* és *durván*, *brutálisan* és *brutális*). Ez utóbbi tendencia az, amely javarészt az első tendencia alóli kivételek eseteit képezi. Harmadrészt, a negatív, valamint a semleges és a pozitív fokozó elemek az átfedéseket illetően erősen elkülönülnek egymástól. Egyetlen olyan hívószópárt találunk (*meglehetősen* és *átkozottul*), amely kivételt képez ez alól.

Megfigyelhető továbbá, hogy a meglepetés emócióhoz köthető szavak asszociátumai gyakorta a legnagyobb átfedést valamely negatív fokozó elem asszociátumaival mutatják. A negatív elemek csoportján belül a félelem emóció elemei a legtöbb esetben egy ugyancsak a félelem emócióhoz kapcsolódó elemmel fednek át a legnagyobb mértékben (a 17-ből 10 esetben). Az undor emócióhoz köthető elemek hasonlóan viselkednek (a 6-ból 5 esetben). A düh hívószói azonban nem ezt a tendenciát mutatják: a 7-ből mindössze három esetben fednek át a legnagyobb mértékben hasonló szemantikai tartalmú hívószóval. Ez utóbbi tapasztalatra még nem találtam megfelelő magyarázatot, ezért további vizsgálatokat tartok szükségesnek.

Azt látjuk tehát, hogy kontextus hiányában rendre a hívószó negatív emotív tartalma dominál. Ez a tapasztalat véleményem szerint különösen fontos a vizsgált elemek szótári reprezentációját illetően (l. 4.6). Amint azt a lexikai pragmatika elméleti keretét alkalmazva tárgyaltam (l. 4.6.1), a negatív emotív fokozó elemeknek önálló lexikai tételt kellene feltennünk, amelynek a [félelem/düh/undor] szemantikai komponens még opcionális elemként sem lehet a része. Ugyanakkor az asszociációs eredmény azt indokolja, hogy ezzel a komponenssel feltétlenül számolnunk kellene a jelentésreprezentációban. Véleményem szerint a tapasztalat az aktivációterjedéses hálózati modellben megmutatott reprezentáció (l. 4.6.2) létjogosultságát támogatja.

A szentiment- és az emóciószótárakon kívül az asszociátumokat az értékvesztésre – és esetlegesen az értékváltásra is – képes elemek listájával is elemeztük (a listáról részletesebben l. 6.3.2.4, a listát l. 1. számú melléklet).¹⁰⁴ Azt szerettem volna feltárni, hogy a vizsgált negatív elemekre hány esetben érkezik értékvesztésre – és esetlegesen értékváltásra is – képes elemválaszként, összevetve a többi hívószóra kapott válaszokkal. Az összesített eredményeket az alábbi ábra mutatja be. Az áttekinthetőség céljából a negatív és a nem negatív ingerszókat eltérő színnel jelöltük.

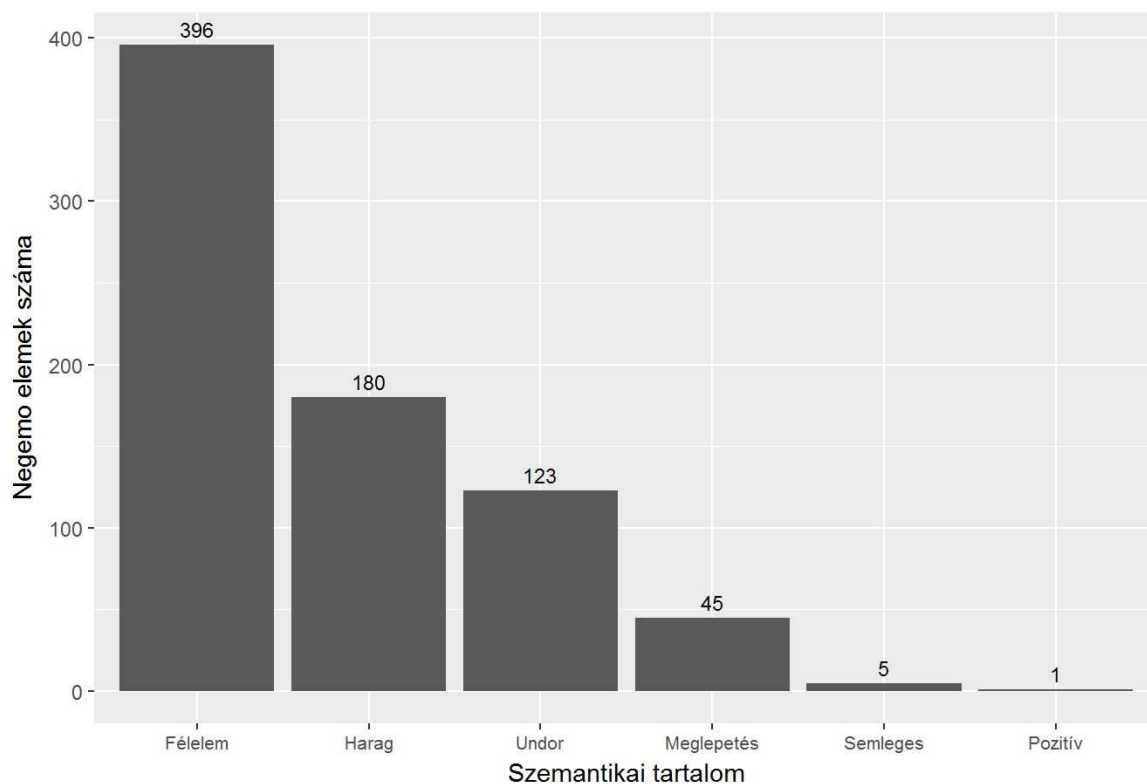
¹⁰⁴ A listát a dolgozat bírálója, Sass Bálint megjegyzései alapján a vizsgálatok elvégzése után 116 elemből állóra bővítettem. Ennek megfelelően a disszertációban elvégzett vizsgálatokban a 106 elemmel dolgoztam. A mellékletben az újonnan felvett elemeket dőlt szedéssel emelem ki.



21. ábra: Az értékvesztésre – és esetlegesen értékváltásra is – képes asszociátumok gyakorisági megoszlása a negatív és a nem negatív hívószók szerinti bontásban

A kapott eredmények alapján a legtöbb értékvesztésre képes asszociátum tendenciaszerűen a hasonló elemekre érkezett. Viszonylagosan gyakrabban adtak még értékvesztésre képes elemet válaszként a meglepetés emócióhoz köthető ingerszókra is. Közülük az *elképesztően* a legkiemelkedőbb ebből a szempontból, amelyre a *rohadtul* hívószóval azonos számú értékvesztő válasz érkezett. A semleges, valamint a pozitív fokozó hívószókra elenyésző esetben érkezett csupán értékvesztésre képes asszociátum.

Az alábbi ábra a hat kategória szerinti bontásban közli a fenti adatokat.



Ábra 22. Az értékvesztésre – és esetlegesen értékváltásra is – képes asszociátumok gyakorisági megoszlása a hívószók hat kategóriája szerinti bontásban

Amint látjuk, a legtöbb értékvesztésre képes elem a félelem emócióhoz kapcsolódó ingerszóra érkezett, amelyet a düh, majd az undor emóció hívószói követnek.

Végezetül minden tárgyalt szempont alapján megvizsgáltam, hogy mutatkozik-e valamilyen eltérés a női és a férfi adatközlők asszociátumainak sajátosságai között. A korábban elmondottaknak megfelelően (l. 6.3.3.4), azt feltételeztem, hogy nemek szerinti bontásban az eredményekben nem lesz jelentősebb különbség.

Vizsgáljuk meg itt is első lépésben az asszociátumok szófaji megoszlási arányait!

Annak céljából, hogy a hasonlóságokat, illetve eltéréseket számszerűsíthessem, kétoldali kétmintás t-próbát alkalmaztunk a női és a férfi asszociátumok összevetésére. Az alábbi táblázat összefoglalja a teszt eredményeit. A táblázat 4. és 5. oszlopa tartalmazza a jelen probléma szempontjából számunkra legrelevánsabb mutatókat. Amennyiben a p-értéke kisebb mint 0,05, akkor statisztikai bizonyítékunk van a kétoldali kétmintás t-próba nullhipotézise ellen. A nullhipotézis értelmében nincs szignifikáns különbség a nők és férfiak eredményei között. Azokban az esetekben tehát, ahol a p-érték kisebb 0,05-nél, az eltérés szignifikánsnak tekinthető. Emellett az is elmondható, hogy minél kisebb a p-értéke,

az eltérés annál jelentősebb mértékű. A táblázatban szürke színnel kiemelem a statisztikailag szignifikáns adatokat.

POS-tag	Férfiak átlaga	Nők átlaga	p-érték	t-érték (szabadságfokok)	szórás	standard hiba
adj	2,34	2,361	0,8529	-0,1856 (371,1)	2,256	0,1138
adp	0	0,01098	0,002245	-3,064 (1001)	0,1134	0,003583
adv	0,3607	0,3483	0,7896	0,2671 (368,6)	0,9143	0,04625
conj	0,0123	0,003992	0,2592	1,13 (282,7)	0,1272	0,007345
det	0	0,00998	0,01835	-2,362 (1001)	0,1337	0,004225
intj	0,004098	0,008982	0,3537	-0,9282 (599)	0,1225	0,005261
noun	0,8852	1,087	0,02948	-2,184 (424,5)	1,928	0,09228
num	0,03279	0,04291	0,4904	-0,6903 (407,8)	0,302	0,01467
part	0,004098	0,001996	0,628	0,485 (303)	0,07805	0,004334
pron	0,02049	0,02894	0,4421	-0,7693 (493)	0,2414	0,01098
propn	0,04508	0,0998	0,009325	-2,609 (561)	0,4793	0,02097
sconj	0	0,000998	0,3176	-1 (1001)	0,03159	0,000998
sym	0	0	NA	NA (NA)	0	0
verb	0,291	0,4731	0,0007197	-3,401 (554,3)	1,219	0,05353

*12. táblázat: Az asszociátumok szófaji megoszlásának összevetése
a nők és a férfiak adatai között*

A mérési eredmények alapján a férfi és a női adatközlők asszociátumai öt szófaj tekintetében jelentős gyakorisági eltérést mutatnak, ezek a névutó, a névelő, a főnév, a tulajdonnév és az ige. Meglepő módon mind az öt említett szófaj a nők adataiban fordul elő jóval nagyobb frekvenciával. A férfiak néhány egyéb szófajt valamivel gyakrabban használtak a nőknél (pl. a határozószót), ugyanakkor ezek esetében a p-érték nem bizonyult statisztikailag szignifikánsnak.

Megvizsgáltam azt is, hogy van-e szignifikáns különbség nők és férfiak között abban a tekintetben, hogy milyen szentiment- és emócióértékű válaszokat adtak az ingerszókra. A mérési eredményeket az alábbi táblázatban közlöm. Itt is, akárcsak fentebb, az adatokat úgy kell értelmezni, hogy amennyiben a p értéke kisebb 0,05-nél, statisztikai bizonyítékunk van a kétoldalú kétmintás t-próba nullhipotézise ellen. A táblázatban a szignifikáns adatokat itt is szürke színnel emelem ki.

szen- timente k és emóciók	fér- fiak átlaga	nők átlaga	p-érték	t-érték (szabadság- fokok)	sd	standard hiba
pozitív szentiment	0,09231	0,1016	0,3675	-0,9014 (1471)	0,4185	0,01031
negatív szentiment	-0,361	-0,2977	0,0001753	-3,762 (1383)	0,6634	0,01684
összesített szentiment	-0,2687	-0,1961	0,000846	-3,345 (1403)	0,8616	0,02172
emóció / bánat	0,02256	0,02295	0,9415	-0,07342 (1436)	0,211	0,005257
emóció / düh	0,02359	0,02184	0,7422	0,329 (1393)	0,2108	0,00533
emóció / félelem	0,06154	0,05147	0,2296	1,202 (1354)	0,3266	0,008377
emóció / meglepődés	0,02769	0,02362	0,4769	0,7115 (1360)	0,2236	0,005726
emóció / öröm	0,08103	0,07843	0,7875	0,2697 (1414)	0,3832	0,009621
emóció / undor	0,05949	0,04434	0,06424	1,852 (1313)	0,3137	0,008178

13. táblázat: Az asszociátumok szentiment- és emócióértékeinek összevetése
a nők és a férfiak között

A két nem adataiban mindössze a szentimentértékek szempontjából mutatkozik eltérés. A férfiak ugyanis gyakrabban adnak negatív szentimentértékű válaszokat a hívószókra. A hívószókat a szentimentértékek tekintetében egyenként is megvizsgáltam. Az adatok arra mutatnak, hogy a férfiak gyakrabban adnak negatív asszociátumokat még az olyan ingerszók esetében is, amelyek szemantikai tartalmukat tekintve nem a negatív elemek csoportjába tartoznak. Így például, a *szédületesen* és a *hihetetlenül* a férfiaknál átlagosan negatív, a nőknél átlagosan pozitív szentimentértékű válaszokat eredményezett.

Az adatokat úgy is elemeztem, hogy csupán az első helyen adott válaszszókat vettem figyelembe. Azt tapasztaltam, hogy a negatív szentimentértékek éppúgy a férfiak

asszociátumaiban jelennek meg szignifikánsan gyakrabban. További érdekes eredmény, hogy ebben az esetben a félelem érzelem is szignifikáns eltérést mutat, amely ugyancsak a férfiak adataiban jelenik meg többször.

A kutatás eredményeit az alábbiakban röviden összefoglalom.

A szófaji megoszlásokat illetően a válaszokat alapvetően a hívószó szófaja befolyásolja. E tapasztalatok nem mondanak ellent Kovács (2013: 42) már tárgyalt megállapításának (l. 6.3.3.3), mely szerint a hívószók toldalékolt, illetve toldalék nélküli volta jelentősen befolyásolja az asszociációs vizsgálatok kimenetelét. Ugyanakkor, szerényebb mértékben ugyan, de a negatív fokozók esetében a lexikalizáltsági fokra utaló jeleket is látunk. .

Az asszociátumok szentimentértékei a hívószók szentimentértékével szoros összefüggést mutatnak. A negatív elemek például, toldalékoltságtól függetlenül rendre negatív értékű asszociátumokat eredményeznek. Ily módon, még azok az elemek is, amelyek a NegEmotív korpusz elemzése alapján gyakori fokozó elemnek mutatkoznak és viszonylag magas kollokációs potenciállal rendelkeznek, valamint a gyakori értékváltó esetében is a negatív szemantikai tartalom érvényesül a válaszokban. A hívószók lexikai szintű szemantikai tartalmával az asszociátumok emóciótartalma ugyancsak szignifikáns összefüggést mutat. Azt látjuk tehát, hogy a vizsgált negatív emotív elemeknél, kontextus nélkül, az esetek túlnyomó többségében a negatív emotív szemantikai tartalom dominál. Úgy is fogalmazhatunk, hogy ezek a szemantikai komponensek a szóasszociáció során mindenképpen aktívak (vö. a korábbi vizsgálatok a lexikai pragmatika és a mentális lexikon hálózatos modelljének segítségével, l. 4.6). A tapasztalatok összhangban állnak kezdeti hipotézisemmel, amelyben azt feltételeztem, hogy minden típusú lexikai szintű szemantikai tartalom esetében jelentős összefüggést fogunk látni a hívószó és a válaszcsoport értékelő, illetve emotív szemantikai tartalma között.

Az asszociatív átfedések szempontjából az adatok a következő három legfontosabb tendenciát mutatták: Egyrészt, a legtöbb hívószó azonos grammatikai sajátságú hívószóval mutatja a legnagyobb átfedést. Másrészt, az azonos tövek gyakran egymással fednek át a legnagyobb mértékben. Ez utóbbi tendencia képezte az első tendencia alóli kivételek legtöbb esetét. Harmadrészt, a negatív, valamint a semleges és a pozitív elemek az átfedéseket illetően erősen elkülönültek egymástól, és a negatív elemek csoportján belül az átfedések jelentősen korreláltak a hívószók emotív szemantikai tartalmával (félelem, düh

vagy undor).¹⁰⁵ Ebben az esetben is azt látjuk tehát, hogy – a grammatikai sajátságok mellett – a szemantikai tartalomnak kiemelkedő jelentősége van, és a negatív emotív elemek esetében ez a sajátság az emotív tartalom szerint rendezi az átfedési mátrixot. Ez az eredmény is arra mutat, hogy a vizsgált elemek esetében, kontextus nélkül, a negatív emotív tartalom kiemelkedően domináns.

A vizsgálat egy érdekes tapasztalata volt, hogy az asszociátumokat az értékvesztő elemek (amelyek közül néhány értékváltó szerepet is betölthet) listájával elemezve a legtöbb negatív emotív válasz hasonló elemekre érkezett. Azt látjuk tehát, hogy a negatív emotív elemek nem egyszerűen negatív tartalmú válaszokat hívnak, hanem olyan elemeket, amelyek azonos (fokozó) funkciót képesek betölteni. Ugyancsak figyelemre méltó eredmény, hogy viszonylagosan gyakrabban adtak még hasonló választ az adatközlők a meglepetés emócióhoz köthető elemekre is, annak ellenére, hogy a meglepetés érzelem nem negatív. Valószínűleg ez is azt mutatja, hogy az asszociációkban érvényesült a fokozó funkció hatása is. .

Végezetül, a tárgyalt szempontok alapján azt is megvizsgáltam, mutatkozik-e valamilyen eltérés a női és a férfi adatközlők válaszainak a sajátságai között. A vizsgálati tapasztalatok eltértek attól, mint amit az elemzés elején feltételeztem. Úgy véltem ugyanis, hogy az asszociációs adatbázis vizsgálata során, bár megfigyelhetünk majd némi eltérést, nem fogunk jelentősebb különbségeket detektálni a férfiak és a nők adatai között. Ez a hipotézis azonban sem a grammatikai, sem a szemantikai sajátságok tekintetében nem igazolódott.

Bár a jelen munka keretében erre már nem vállalkozhattam, a kutatás egy következő lépéseként a létrehozott adatbázist a magyar nyelvű WordNet (vö. Miháltz és mtsai 2008) segítségével is fel szeretném dolgozni. Azt gondolom, hogy az elemzés újabb összefüggéseket tárhatna fel a hívószók és a válaszsók közötti kapcsolatok (pl. paradigmatis, szintagmatis, alá-, fölérendelt, ellentét, szinonima, fogalmi közelség stb.) vonatkozásában.

¹⁰⁵ Amint azt korábban is megjegyeztem, a düh emóció elemei gyakran a legnagyobb mértékben más emóció elemeivel fedtek át, szemben a másik két negatív emócióval. Erre a tapasztalatra azonban még nem találtam megfelelő magyarázatot, ezért további vizsgálatokat tartok szükségesnek.

7. Egy szótár- és szabályalapú szentimentelemző eszköz létrehozása

Amint azt az értekezésnek mind a kompozíciós szemantika lehetőségeit (l. 2.3.2.2), mind a szemantikai motivációt tárgyaló részében (l. 4.5) ismertettem, a nyelvi értékelelemzésben a szótárak és kombinációs szabályok alkalmazása kétségkívül korlátozott érvényű, és nem képes maradéktalanul kezelni több fontos sajátságot. Ugyanakkor az adatbázisok elemzési tapasztalatai (l. 6.3.1.7, 6.3.2.7, 6.3.3.5) arra mutatnak, hogy a módszer költséghatékony, mégis eredményes megoldást jelenthet legalább bizonyos részfeladatok vonatkozásában a szentimentelemzés számára.¹⁰⁶ Amint ugyancsak kiemeltem (l. 2.1), a számítógépes nyelvészet adatorientált terület. A nyelvtchnológiai alkalmazások fejlesztésekor a hatékonyság a cél (vö. Gábor és mtsai 2008: 854–855), és a számítógépes nyelvészeti fejlesztéseknek a valódi nyelvi produktumok sajátságaihoz kell igazodniuk.

Ebben a fejezetben bemutatok egy olyan, a jelen kutatómunka keretében létrehozott, saját fejlesztésű szentimentelemző eszközt, amely szótárak és kombinációs szabályok alapján rendel összesített szentimentértéket az egyes mondatokhoz.¹⁰⁷ Összegzem mindazokat a vizsgálati tapasztalatokat, amelyekre támaszkodva a kombinációs szabályokat létrehoztam, ismertetem a szabályokat, majd bemutatom az elemző eszköz alapját képező szótárakat. Ezt követően tárgyalom annak a kézzel, mondatszinten annotált szentimentkorpusznak az anyagát, feldolgozási elveit és alapvető adatait, amelyet az eredményesség mérésére hoztam létre. Végül ismertetem az eszköz hatékonyságát a humán annotációval, valamint az egyszerű szótáralapú (szószak-modell alapján működő) elemzési eredményekkel összehasonlítva.

¹⁰⁶ Azt, hogy költséghatékonyak tekintem a szótár- és szabályalapú elemzési megoldást, a következőképpen indoklom: Igaz, hogy a szabályok létrehozásához korpuszok vizsgálatára, majd programozási munkára van szükség, a kódolt szabályok a különböző elemzési feladatokban újra és újra alkalmazhatóak. Amennyiben például tanító adatok segítségével szeretnénk hasonló feladatot végezni, lehetséges, hogy az egyes szövegdomének esetében külön-külön tanító adatbázist kellene készítenünk. A szótár- és szabályalapú elemzési megoldásnál az szótár anyagát valószínűleg módosítunk kell az aktuális feladatnak megfelelően, és a szabályok hatékonyságát is ellenőriznünk kell egy kisebb annotált anyagon, nagyobb manuális munka azonban nem szükséges.

¹⁰⁷ Az elemző eszközt szerencsésebb lett volna úgy elkészíteni, hogy a targeteket is kezelni tudja, illetve a szentimentértékeket a targetek függvényében legyen képes meghatározni. A dolgozatban magam is hangsúlyoztam a targetek jelentőségét és a targetfüggés-beli elemzés szerepét. A jelen implementációban alkalmazott megoldás oka az, hogy a Dívany-korpuszon, a targetekkel kapcsolatosan végzett mérési eredményeink, miszerint hogyan tudnánk automatikusan összekapcsolni a szentimentkifejezéseket a targetekkel, egyelőre nem vezettek eredményre.

Úgy gondolom, hogy ebben a tekintetben a szintaktikai előfeldolgozás hozhatna megoldást, de ennek a módszernek az alkalmazására ebben a disszertációban már nem volt lehetőségem. Ugyanakkor véleményem szerint a jelen kutatásban feltárt eredmények és elkészített korpuszok egy további, szintaktikai elemzésre építő munkához is hasznos alapot nyújthatnak.

7.1 A kombinációs szabályok bemutatása

A létrehozott kombinációs szabályok a szentimentszótárban foglalt pozitív és negatív szentimentkifejezések, valamint bizonyos egyéb nyelvi elemek egymáshoz képesti pozícióit kódolják. Amint már ismertettem (l. 1.1), a szabályok azt határozzák meg, hogy amennyiben egy adott elemzési egységen (jelen esetben mondaton) belül egy szentimentkifejezés környezetében, egy a szabályban rögzített pozícióban egy bizonyos elemtípus megtalálható, úgy az elemző rendszer az adott szentimentkifejezéshez a szótár alapján rendelt értéket a szabályban foglaltak szerint módosítja. A szabályokat, valamint az azok kialakításához felhasznált vizsgálati tapasztalataimat az alábbiakban ismertetem.

Amint azt a Dívány-korpusz vizsgálati eredményeinél tárgyaltam (l. 6.3.1.7), bár a negáló elemek elméleti szinten csupán részben okozzák az érték megváltozását a fragmentumban (l. 2.3.2.2, valamint Israel 2004: 711, Feldman és mtsai 2010: 4) – ez a változás az elemzett szövegekben kimagasló rendszerességgel következik be. Ez alapján a negáló elemek és a szentimentkifejezések kapcsolatainak kombinációs szabályokba foglalt alkalmazása – a vizsgált szövegtípusban legalábbis – jelentős javulást eredményezhet a kizárólagosan szótáralapon működő elemző algoritmushoz képest. Megvizsgáltam szentimentkifejezéseknek a negáló elemhez képesti pozíciós sajátosságait is (vö. Szabó–Vincze 2016: 277–279). Megállapítottam, hogy a pozitív fragmentumokban a szentimentkifejezést közvetlenül megelőző pozíció kiemelkedően preferált, de a hatékony elemzéshez figyelembe kell vennünk a -2., -3. és a -4., kevésbé gyakori pozíciókat is (vö. Szabó–Vincze 2016: 278). A negatív fragmentumokban az említett pozíciók egyenletesebb megoszlási aránnyal jutnak szerephez, a pozitív fragmentumokkal szemben azonban a -4. pozíció betöltése igen ritka. Mindemellett a negálók kifejezetten ritkán fordulnak elő a szentimentkifejezést követő pozíciókban, a pozitív és a negatív fragmentumokban egyaránt. Az ezeknek megfelelően kialakított pozíciós szabályaim lehetővé teszik, hogy a pozitív szentimentkifejezések előtti 4, valamint a negatív szentimentkifejezések előtti 3 pozíció valamelyikében szereplő negáló elemet figyelembe vegyük az automatikus elemzésben (l. 1. és 2. szabály a 14. táblázatban).¹⁰⁸

A negáló elemek sajátosságainak elemzésén túl megvizsgáltam a növelő és csökkentő értelmű intenzifikáló elemek hatását is (vö. Szabó és mtsai 2016a: 178) (l.

¹⁰⁸ A negációra vonatkozó szabályok nem vezetnek rossz elemzési eredményre a magyar nyelvre vonatkozó negatív egyeztetés miatt, pl. *nincs semmi baj*. A hasonló szerkezetekben ugyanis az elemző csak a szentimentkifejezéshez közelebbi negáló elemet veszi figyelembe, a távolabbit nem.

6.3.1.7). Tekintettel arra, hogy a növelő értelmű intenzifikáló elemek mind a pozitív, mind a negatív kifejezések mellett fokozó szereppel bírnak, a kombinációs szabályokba azok működési mechanizmusát is ennek megfelelően vettem fel (l. lentebb, 14. táblázat). Bár a csökkentő értelmű intenzifikálókat illetően mást feltételeztem a vizsgálat előtt (részletesen l. 6.3.1.7), megállapítottam, hogy azok sem a negatív, sem a pozitív szentimentkifejezések mellett nem változtatják az esetek többségében szentimentértéket az ellentétére (l. 3-6. szabályok a 14. táblázatban).

A növelő és a csökkentő értelmű intenzifikáló elemeket illetően megvizsgáltam azok kombinációit a negáló elemek és a különböző értékű szentimentkifejezések együttes előfordulásaival is (pl. *mégsem kifejezetten jó, nem annyira rossz* stb.). A vizsgált elemek négy lehetséges kombinációja közül egyetlen esetben mutatkozott a szentimentérték módosulása tendenciaszerűnek, abban, amelyben a negáló elem fokozó értelmű intenzifikálóval és negatív szentimentkifejezéssel szerepelt együtt (pl. *nem kimondottan kellemetlen*). Ezekkel a megnyilatkozásokkal a beszélő valamivel negatívabb értékelést fogalmaz meg annál, mint amit a negált negatív szókapcsolatokkal kifejezésre juttat (ez utóbbival az esetek többségében pozitív értékelést fejez ki). Úgy döntöttem tehát, hogy erre a lehetőségre egy további szabályt alkalmazok az elemzőben, lehetőséget adva az eredményesség további javítására (l. 7. szabály a 14. táblázatban)..

Az értékvtáltás eseteit a jelen megoldással nem állt módomban kezelni.¹⁰⁹ Bár a dolgozat egyik célja az volt, hogy mind az értékvesztés, mind az értékvtáltás tekintetében feltérképezze azok helytálló automatikus elemzésének a lehetőségeit, a vizsgálati eredmények megmutatták, hogy az értékvtáltást illetően erre, az itt alkalmazott szótár- és szabályalapú elemzéssel nem lesz lehetőség.¹¹⁰ A NegEmotív korpuszban (l. 6.3.2.7) a két leggyakoribb értékvtáltó elem a *durva(-n)* és a *kemény* volt. Ugyanakkor, ezen kifejezések összes, nem annotált előfordulásának az elemzése azt is megmutatta, hogy mindkét elem gyakori nem értékvtáltó funkcióban, azaz negatív értékelés kifejezőjeként is. Mivel nem találtam olyan negatív emotív elemet, amely kifejezetten értékvtáltó szerepben volna gyakori, ezért annak helyes elemzését szótári szinten nem tudtam megoldani, azaz a vizsgált elemek közül egyet sem tehettem a pozitív szentimentszótárba. Ugyanakkor nem sikerült feltárnom olyan pozíciós sajátságot sem, amely alapján kombinációs szabályt adhattam volna meg a kezelésükre. Egy megoldást mégis alkalmaztam, amellyel

¹⁰⁹ A jelenség pontos definícióját l. 3.

¹¹⁰ A jelenség helytálló kezeléséhez további vizsgálatok szükségesek, amelyek például a targetek és az értékvtáltó elemek közötti kapcsolatokra, a szintaktikai viszonyokra és az értékvtáltó elemek környezetében levő elemek szemantikai, illetve pragmatikai sajátságaira irányulnak.

tulajdonképpen elkerülhető az értékváltó elemek nem helyes automatikus elemzése: az általam összegyűjtött értékváltásra képes elemeket nem szerepeltetem a negatív szótárban. Ennek köszönhetően, bár néhány szentimentértéket elveszíthetünk az elemzés során, téves feldolgozásukat bizonyosan elkerüljük.¹¹¹

Az értékvesztést illetően, mind a szemantikai motivációs vizsgálataim (l. 4.5), mind a NegEmotív korpusz elemzési eredményei (l. 6.3.2.7) alapján a negatív fokozó elemek a szentimentértékkel rendelkező alaptagok módosítóiként szinte kivétel nélkül elveszítik negatív tartalmukat.¹¹² Ezt a megállapítást érvényesnek tekintettem a pozitív és a negatív alaptagok melletti fokozó elemekre egyaránt. Ugyanakkor, semleges alaptagok módosítóiként változatos szemantikai-pragmatikai viselkedést mutatnak. Ezeknek az eseteknek csupán a felében jelöltek az annotátorok a korpuszban deszemantizáltságot. Mindezek alapján azt a sajátságot állt módomban szabályba foglalni, hogy amennyiben egy negatív emotív elemet közvetlenül egy pozitív vagy egy negatív szentimentkifejezés követ, úgy a vizsgált elem pusztán fokozó funkciójában vegyen részt az automatikus elemzésben (l. 8. és 9. szabály a 14. táblázatban). Említsük meg, hogy ez a szabály, bár nagyon egyszerű, és csupán az esetek egy bizonyos típusának a kezelésére képes, összhangban áll azokkal a humán interpretációs mechanizmusokkal, amelyeket a lexikai pragmatika, valamint a mentális lexikon aktivációterjedéses hálózati modellje alapján tárgyaltam részletesen korábban (l. 4.6). A jelen szabályokban tulajdonképpen a lexikai reprezentációban megadott intenzitási magjelentést tesszük fel (l. 4.6.1).¹¹³

A kialakított kombinációs szabályokat, példákkal együtt az alábbi táblázat összegzi. A példákban a szabályok szempontjából releváns elemeket kövér szedéssel emelem ki.

¹¹¹ Ugyan – amint azt a (6.3.2.7)-ben tárgyaltam – az értekezés keretei között erre nem vállalkozhatom, a jövőben azt is meg szeretném vizsgálni, hogy milyen egyéb nyelvi sajátságok (pl. az emojik és az emotikonok figyelembevétele, vö. Carvalho és mtsai 2015) tudnák a hasonló esetek automatikus feldolgozását hatékonyan támogatni.

¹¹² A jelenség pontos definícióját l. 3.

¹¹³ A reprezentáció opcionális jegyei ebben az algoritmusban nem kaphattak szerepet, ahhoz ugyanis komplexebb automatikus szemantikai elemzést kellett volna végezni legalább a fokozó elemek és az alaptagok közötti kapcsolatok vonatkozásában. Ebben a dolgozatban erre nem volt lehetőségem.

	Kombináció	A módosítás	Példák
1.	negáló elem (Neg) (-3 pozícióig) + negatív szentimentkifejezés (NegSent)	→ pozitív érték	- nem is teljesített rosszul -A csoki nem bántó
2.	Neg (-4 pozícióig) + pozitív szentimentkifejezés (PosSent)	→ negatív érték	- a csokija nem valami jó - Nem mondom, hogy finom
3.	fokozó értelmű int. (IntPlus) + PosSent (megelőző pozícióban)	→ a pozitív érték duplázása	- Nagyon szuper kollekció! - határozottan jó termék
4.	IntPlus + NegSent (megelőző pozícióban)	→ a negatív érték duplázása	- én is ezt mondanám magamnak, csak már annyira elegendem van - nagyon rossz
5.	csökkentő értelmű int. (IntMinus) + PosSent (megelőző pozícióban)	→ a pozitív érték felezése	- egész finom - viszonylag jó
6.	IntMinus + NegSent (megelőző pozícióban)	→ a negatív érték felezése	- kicsit átverős szalámi - picit műízű
7.	Neg + IntPlus + NegSent	→ a negatív érték felezése	- nem aktívan rossz - nem kimondottan kellemetlen, de nem is jó
8.	negatív emotív elem (NegEmo) + PosSent	→ a pozitív érték duplázása	- piszok jól néz ki a Technics éjjekete lemezjátszója - 10 brutálisan szép partszakasz a Costa Bravan
9.	NegEmo + NegSent	→ a negatív érték duplázása	- ez a Conanes GRRM videó iszonyat béna - rohadt bűdös gyógyszer szaga van

14. táblázat: A kombinációs szabályok, példákkal

A kialakított szabályokat ezt követően forráskódba kellett foglalni. Azért, hogy a programozó a műveletet könnyebben elvégezhesse, a szabályokat BNF-formátumban írtam fel (*Backus–Naur form* vagy *Backus normal form*, vö. Ingerman, 1967), amely a kontextusfüggetlen nyelvtanok esetében alkalmazható kódolási rendszer, és amelyet gyakran használnak többek között a programozási nyelvek szintaxisának a leírására.

Az alábbi példa a fokozó értelmű intenzifikáló elemek és a szentimentkifejezések kapcsolatának a kezelési szabályát mutatja be BNF-formátumban.

(92) word* intPlus senti => 2 x senti

ahol a felsorolt elemek sorrendje a szövegbeli pozíciójukként értelmezendő, és ahol
* = nulla vagy több (azaz tetszőleges számú)

7.2 Az elemző rendszerben alkalmazott szótárak bemutatása

A szentimentelemző program elkészítéséhez különböző szótárakra volt szükség, hogy a nyelvi anyagot szótáralapon elemezni, ezáltal a kombinációs szabályokat alkalmazni lehessen. Az elkészített szoftver ezek alapján detektálja a megfelelő nyelvi elemeket a szövegekben, és végzi el rajtuk a szabályokban kódolt műveleteket. A jelen fejezetben a szótárak tartalmáról, létrehozásának menetéről, valamint alapvető adatairól számolok be.

A pozitív és a negatív szentimentkifejezéseket saját készítésű, magyar nyelvű szentimentszótárral (vö. Szabó 2014, 2015b) dolgozzuk fel. A szótárat részben automatikus, részben manuális módszerrel hoztam létre, amit egy angol nyelvű szentimentlexikon alapján végeztem.¹¹⁴ A szótár szóanyagát első lépésben automatikus módszerrel magyar nyelvre fordítottuk. Ezt követően kézzel ellenőriztük, javítottuk, illetve kiegészítettük az automatikus fordítás eredményét.¹¹⁵ Mindehhez egy angol–magyar kétnyelvű szótárat (MTA SZTAKI online szótár), valamint két szinonimaszótárt (Szinonimaszotar.hu és Poet szinonimaszótár) is segítségül hívtunk.

A szótár létrehozásakor olyan elemek felvételére törekedtem, amelyek doméntól és kontextustól függetlenül, lexikai szinten meghatározott értékelő szemantikai tartalommal rendelkeznek. A szótárba nem csupán melléknevek, hanem főnevek, határozószók és igék is kerülhettek, amennyiben úgy ítéltam, hogy az adott elemnek inherens szentimentértéke van (a munka alapelveiről és dilemmáiról bővebben l. Szabó 2014: 6–15, Szabó 2015b). Végül az elkészített szótárat a minőség javítása érdekében két humán annotátor ellenőrizte, és megtisztította a kérdésesnek tekinthető esetektől.¹¹⁶ Az így kialakított lexikon összesen

¹¹⁴ A szentimentszótár létrehozása lehetséges lett volna egyéb módokon is, például korpusz alapján, ezt azonban két okból elvettem. Egyrészt, egyetlen, már kész szentimentkorpusz állt a rendelkezésemre (a Dívány-korpusz), amely szentimentkifejezés-szintű annotációjának köszönhetően megfelelő lett volna szótárépítésre. Ezt azonban nem szerettem volna erre a célra felhasználni, mivel a korpuszt a szótárak hatékonyságának a mérésére is alkalmazni kívántam. Másrészt, a korpuszalapú szótárépítés módszere szükségszerűen doménfüggő lexikon építést eredményezi, magam azonban doménfüggetlen listákat szerettem volna létrehozni a hatékonyság növelése céljából. A korpuszalapú módszert ugyanakkor a további szólisták készítésénél alkalmaztam (l. lentebb).

¹¹⁵ Köszönöm Morvay Gergelynek, hogy a szótár létrehozásában hathatós segítséget nyújtott.

¹¹⁶ Köszönöm Nyíri Zsófinak és Hajnal Zsoltinak, hogy elvégezte a szótár minőségének a javítását.

9338 kifejezést tartalmaz, ebből 6639 a negatív és 2699 a pozitív szentimentértékű nyelvi elem.¹¹⁷ A pozitív és a negatív szólistából közlök egy-egy mintát a disszertáció mellékletében (1. 2. és 3. számú melléklet). A szótár kutatási célokra szabadon hozzáférhető a következő online felületen: [<http://opendata.hu/dataset/hungarian-sentiment-lexicon>].

A szentimentelemző eszközhöz a szentimentszótár mellett szükség volt további kifejezéslisták elkészítésére is. Ezek a negáló, a fokozó és a csökkentő értelmű intenzifikáló, valamint az értékvesztésre képes elemek voltak. Közülük a negáló, valamint a fokozó és a csökkentő értelmű intenzifikáló elemek szótárait a Divány-korpusz segítségével magam készítettem el, az abban alkalmazott annotációs tagek szűrésével. Az előbbi 106, az utóbbiak 787 és 229 számú elemet tartalmaznak. Az értékvesztésre képes elemeket illetően ugyanazt, az összesen 106 elemből álló listát használtam fel ebben a feladatban is, mint amelyet korábban, a NegEmotív korpusz létrehozásához készítettem (1. 6.3.2.4 és 1. számú melléklet). Ezeknek az elemeknek egy része – mint már említettem (1. 7.1) – átkerült a negatív szentimentkifejezések szótárába is.

7.3 A kézzel annotált tesztkorpusz bemutatása

A kombinációs szabályok tesztelése céljából létrehoztunk egy összesen 3392 mondatból álló, mondatszinten kézzel annotált szentimentkorpuszt.¹¹⁸ A jelen fejezetben e munka anyagát, menetét, eszközét és alapelveit tárgyalom részletesen. Ezt követően ismertetem az annotátori egyetértésmérés eredményeit, valamint a korpusz alapvető statisztikai adatait.

¹¹⁷ A disszertáció bírálója, Szécsényi Tibor felveti, hogy nem hozna-e jobb elemzési eredményt egy nem kizárólag bináris értékeket tartalmazó szentimentszótár. Egyetértek az elképzeléssel és szorgalmaznám, hogy a meglevő szótárat ennek megfelelően átstrukturáljuk, majd a hatékonyságát a bináris szótárral összemérjük. Ugyanakkor fontosnak tartom kiemelni, hogy a szentimentértékek skálázása kontextus nélkül kifogásolható vállalkozás, amint azt röviden tárgyaltam is a dolgozat 2. fejezetében. Ugyanezzel a felvetéssel él a bíráló a további listák elkészítését illetően is (1. lentebb), amire hasonló egyetértéssel, ugyanakkor elméleti problémák felvetésével tudok reagálni.

¹¹⁸ Amint azt a 7.3.1-ben látni fogjuk, a tesztkorpusz minden mondata tartalmaz legalább egy szentimentkifejezést. Ezzel a megoldással az annotátori munkavégzést szeretnénk volna redukálni, de fontos hangsúlyozni, hogy vele nem torzíthattuk az eredményeket: a szentimentkifejezéseket ugyanis mindkét elemzési módszernél ugyanazzal a szentimentszótárral találjuk meg. A jelen fejezetben végzett összevetés tétje az, hogy mennyiben javíthatunk a szabályok alkalmazásával az egyszerű szótáras elemzési megoldás eredményességén. Ahhoz, hogy az elemzőt szentimentkifejezéseket tartalmazó és nem tartalmazó mondatokon egyaránt ki tudjuk értékelni, egy újabb és nagyobb méretű korpusz annotálására, majd az azon történő tesztelésre lesz szükség a jövőben. A munka jelen fázisában ez nem valósulhatott meg.

7.3.1 A korpusz létrehozásának a menete és módszere

A korpusz anyagának nagyobb részét a Dívány-korpuszból nyertük, ez adta a tesztkorpusz 3002 mondatát.¹¹⁹

Az, hogy a szabályok megalkotásának alapját képező vizsgálati tapasztalataim részben a Dívány-korpusz alapján keletkeztek, majd a teszteléshez a Dívány-korpuszt is felhasználtuk, kifogásolható. A megoldás létjogosultsága melletti fontos érvnek tekintem ugyanakkor azt, hogy a korpusz csupán a két szentimentérték-típust tartalmazza, a pozitívat és a negatívát. A létrehozott elemző megoldás esetében azonban egy ötértékű szentimentosztályozó rendszert kívántunk tesztelni, és ennek megfelelően a korpusz itt felhasznált részeit más feldolgozókkal, az új szempontok szerint újra is annotáltattuk.

További fontos eltérés a korpusz két felhasználási módja között az, hogy amíg a Dívány-korpusz fragmentumszintű annotációt tartalmaz, addig a tesztelési feladatban annak részletét mondat szinten annotáltuk, majd az utóbbi változaton mértük a mondat szintű automatikus elemzés eredményességét.

Mindemellett, a Dívány részletén az értékvesztő és értékváltó elemek kezeléséhez megfogalmazott szabályok tesztelésére is lehetőségem nyílt, tekintve, hogy a korpuszban reprezentált szövegtípus bővelkedik értékvesztésre, illetve értékváltásra képes elemekben.

Azt is érdemes kiemelni, hogy nem állt rendelkezésemre másik olyan magyar nyelvű korpusz, amely ilyen részletes, a munkát megkönnyítő és meggyorsító, specifikusan a szentimentelemzéshez készült annotációt tartalmazott volna. Ez az annotáció tette lehetővé ugyanis a teszteléshez szükséges adatok hatékony szűrését, aminek köszönhetően egyrészt bizonyosan nem került a tesztkorpuszba irreleváns adat, másrészt nem volt szükség nagyméretű kiinduló adatbázisra sem. E megoldás elutasításával a tesztelés a kutatás jelen fázisában nem valósulhatott volna meg.¹²⁰

A fentebb elmondottakat figyelembe véve úgy vélem, hogy a Dívány-korpusz részletén végzett mérési eredmények fontos kiindulópontként szolgálhatnak a további fejlesztői munkához, azzal a kitételrel, hogy a mérést egyéb szövegeken (és ezzel együtt egyéb szövegtípusokon) is el kell végezni a jövőben.

A Dívány-korpusz esetében annak annotált változatát használtuk, és a következő annotációs tagekre szűrtük a mondatokat: Neg, IntPlus és IntMinus (a tagekről l. fentebb,

¹¹⁹ Köszönöm Bors Bencének a tesztkorpusz összeállítását.

¹²⁰ Az ideális megoldás az lett volna, ha a Dívány-korpusz annotációs megoldásával elkészítettünk egy más szövegekből álló tesztkorpuszt, ez azonban olyan jelentős további idő- és költségigényt támasztott volna, amelyre a munka ezen szakaszában nem voltak erőforrásaink.

14. táblázat). A 3002 mondat a három elemtípust megközelítőleg 1/3-os bontásban tartalmazza, tehát a szűrt korpusz 1001 negáló, 1000 fokozó értelmű intenzifikáló, valamint 1001 csökkentő értelmű intenzifikáló elemmel rendelkező mondatra oszlik. Ezekre a korpuszrészekre a továbbiakban a Negation, az IntPlus és az IntMinus elnevezésekkel utalok.

A tesztkorpusz további 390 mondatát tweetek adták, amelyeket olyan szempontok alapján gyűjtöttünk, hogy azok lehetővé tegyék az értékvesztő és az értékváltó elemekre megfogalmazott szabályok hatékonyságának az ellenőrzését. Ehhez ugyanazt a módszert alkalmaztuk, mint amit a NegEmotív korpusz anyagának összeállításához használtunk: olyan tweeteket kerestünk, amelyek tartalmaztak legalább egyet a 106 értékvesztésre – és esetlegesen értékváltásra is – képes elememből álló listámról (l. 6.3.2.4, valamint 1. számú melléklet).¹²¹

Ehhez a munkához azért nem a NegEmotív korpusz anyagából indultam ki, mert a jelen feladatban ugyanazok az annotátorok vettek részt, akik a NegEmotív korpuszon is dolgoztak. El szerettem volna kerülni tehát, hogy az itt alkalmazott döntéseket esetlegesen befolyásolhassák a korábbi annotálási tapasztalatok.

A tweetekből ezután kiszűrtünk minden olyan mondatot, amelyben a vizsgált elemeket közvetlenül egy olyan elem követte, amely szerepelt a szentimentszótár valamely, azaz pozitív vagy negatív listáján. A szűrési feltételeknek a nyers adatok közül összesen 390 mondat felelt meg.

A mondatra bontás, illetve általánosságban a mondatszintű munka ebben a feladatban azért volt elengedhetetlen, mert az elemző eszközt úgy szerettem volna elkészíteni, hogy az a jövőben mondatra bontott szövegekre alkalmazható legyen.

A nyers, összesen 3392 mondatból álló korpusz létrehozását követően két annotátor elvégezte a feldolgozói munkát. Ebben az esetben is, akárcsak a Dívány és a NegEmotív korpusz esetében, a munka megkezdése előtt a két annotátor között felosztottuk a korpusz szöveganyagát, és azonos korpuszrészt csupán az egyetértés-mérési feladathoz annotáltak a készítőik. A munka során a korpusz egyes mondataihoz szentimentértéket rendeltek egy ötfokozatú skála alapján, és a következő lehetőségek közül választhattak: 1: nagyon negatív, 2: negatív, 3: semleges vagy összességében közepesnek tekinthető, 4: pozitív, 5: nagyon pozitív.

¹²¹ A listát a dolgozat bírálója, Sass Bálint megjegyzései alapján a vizsgálatok elvégzése után 116 elemből állóra bővítettem. Ennek megfelelően a disszertációban elvégzett vizsgálatokban a 106 elemmel dolgoztam. A mellékletben az újonnan felvett elemeket dőlt szedéssel emelem ki.

7.3.2 Az annotátorok közötti egyetértésmérés eredményei

Ahhoz, hogy az annotátorok hatékonyságát mérni tudjuk, első lépésként elvégeztük az annotátorok közötti egyetértés mérését. A méréshez a két annotátor feldolgozta a korpuszból ugyanazt a 300 mondatot (megközelítőleg a korpusz 10%-át), majd a két adathalmaz értékeit összevetettük egymással.¹²² Az egyetértési adatokat – a NegEmotív korpusz esetében alkalmazott megoldáshoz hasonlóan (l. 6.3.2.5) – az ún. Cohen-féle Kappa-statisztikával (1960) számoltuk ki.

A jelen mérés során – az annotáció sajátosságait figyelembe véve (részletesebben l. lentebb) – két típusú kalkulációt alkalmaztunk: a súlyozatlan és a súlyozott megoldást. Az előbbi nem veszi figyelembe az eltérés mértékét, csupán annak tényét az annotátorok között. Ennek értelmében minden eltérést, amely az annotátorok között mutatkozik, egyforma súlyú eltérésként kezel. Ezzel szemben a súlyozott megoldás nem azonos értékű problémaként kezeli a különböző eltéréseket. Például, amennyiben a jelen munka esetében az egyik annotátor pozitívnak, a másik pedig negatívnak értékeli egy adott mondatot, az súlyosabb eltérésnek minősül, mint az, ha a másik annotátor ugyanazt nagyon pozitívnak jelöli. Tekintettel arra, hogy ebben a feladatban az annotátorok skálaértékekkel dolgoztak, az utóbbi statisztika informatívabbnak és használhatóbbnak bizonyult.

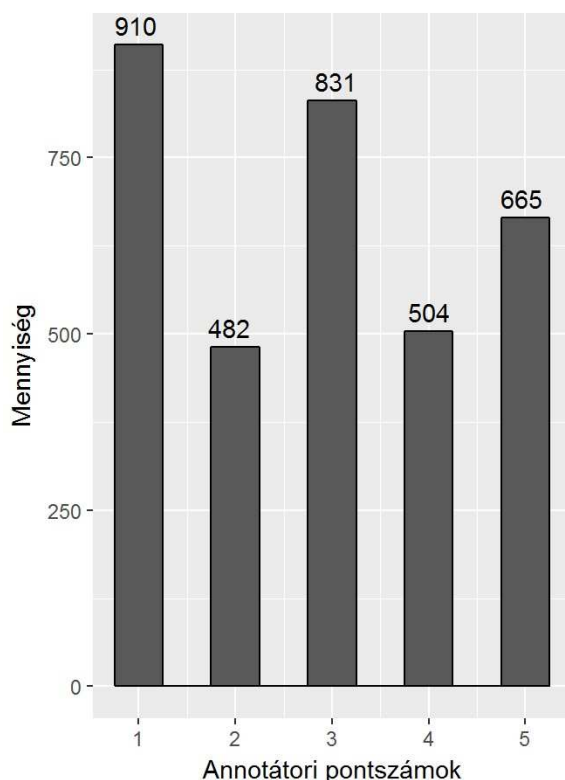
A súlyozott Kappa-statisztika alapján az annotátorok kiváló (0,874) egyetértési eredménnyel dolgoztak. Még a súlyozatlan mérési módszerrel is a Kappa értelmezési tartományai alapján jó (0,59) minőséget értek el. Azt is megállapítottuk, hogy bár az esetek egy bizonyos részében (a 300-ból összesen 98 esetben) nem azonos szentimentértéket rendeltek az adott mondathoz, az eltérések átlagos értéke mindössze 1 volt. Ez azt jelenti, hogy az ötfokozatú skálán adott eltérő értékek átlagosan 1 távolságra helyezkedtek el egymástól.

Mivel az egyetértésmérés eredménye nem adott okot a korpuszannotálási alapelvek átdolgozására vagy kiegészítésére, a mérést követően az annotátorok elvégezték a teljes korpusz feldolgozását.

¹²² Köszönöm Nyíri Zsófinak és Lázár Bernadettnek az annotálási feladat, Ilyés Virágnak pedig az egyetértésmérés elvégzését.

7.3.3 A kézzel annotált korpusz alapvető statisztikai adatai

A nyers korpuszt két annotátor kézzel dolgozta fel úgy, hogy a korpusz egyes mondataihoz szentimentértéket rendeltek egy ötfokozatú skála alapján: 1: nagyon negatív, 2: negatív, 3: semleges vagy összességében közepesnek tekinthető, 4: pozitív, 5: nagyon pozitív. Az alábbi ábra bemutatja az annotált korpusz alapvető statisztikai adatait.

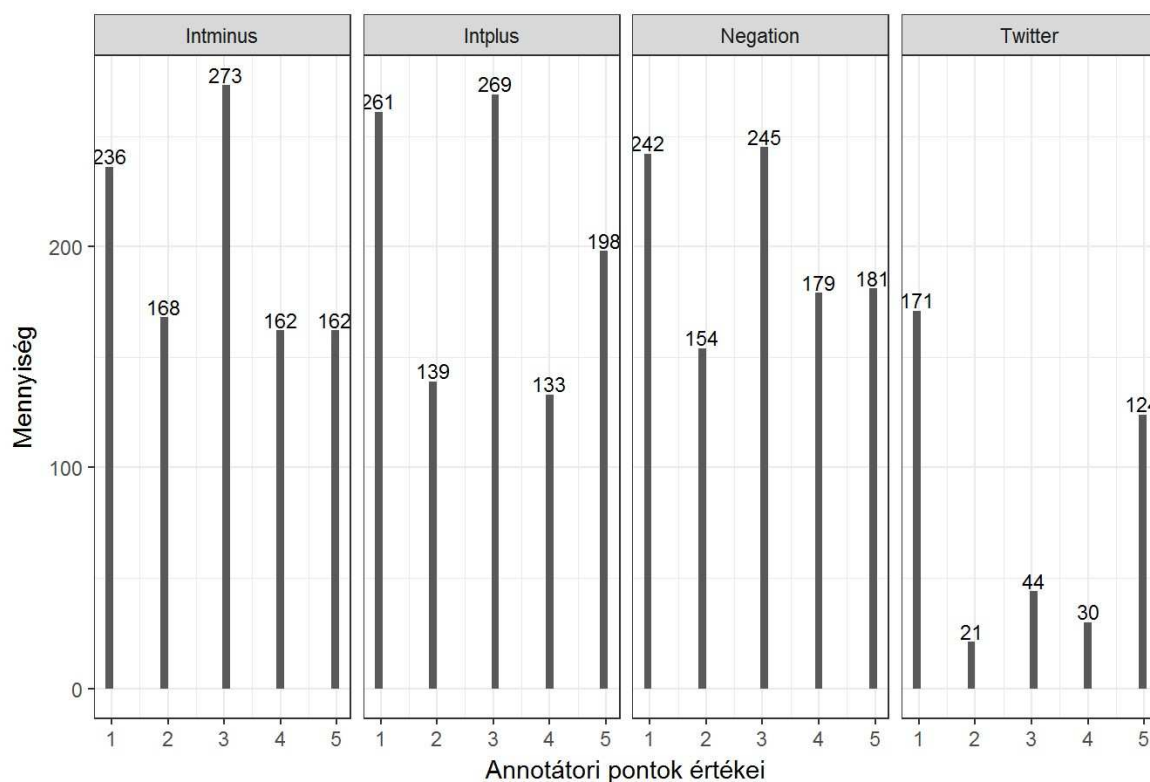


23. ábra: Az annotált korpusz alapvető statisztikai adatai

Az annotátorok manuális elemzése egyenletesnek mondható abban a tekintetben, hogy a semleges, a negatív és a pozitív értékek egyaránt gyakran szerepelnek az annotációban. Összesen 1392 (41,03%) esetben rendeltek negatív, 831 (24,49%) esetben semleges, és 1169 (34,46%) esetben pozitív értéket a mondathoz. Amennyiben az öt értéket külön-külön vizsgáljuk, azt állapíthatjuk meg, hogy a leggyakrabban alkalmazott a „nagyon negatív” (910: 26,82%) volt, amelyet a „semleges” (831: 24,49%), majd a „nagyon pozitív” (665: 19,6%) érték követ.

Szerettem volna feltárni, vajon mutatkozik-e szignifikánsabb eltérés az egyes alkörpuszokat (Negation, IntPlus, IntrMinus és az értékvesztésre képes elemeket

tartalmazó tweetek) illetően. Az alábbi ábra az eredményeket az alkorpuszok szerinti bontásban mutatja be.



24. ábra: A kézi annotálás eredménye a négy alkorpusz szerinti bontásban

Az adatok alapján azt látjuk, hogy a fokozó és a csökkentő értelmű intenzifikálók, valamint a negáló elemek alapján szűrt mondatok esetében nincs érdemi eltérés az annotátori szentimentértékek gyakorisági megoszlásában. Az értékvesztésre képes elemek alapján szűrt mondatoknál azonban teljesen eltérőek az eredmények: a legtöbb mondat „nagyon negatív” vagy „nagyon pozitív” értékelést kapott. A jelenség oka valószínűleg – legalább részben – az, hogy az értékvesztő elemek használata által a megnyilatkozásban kifejezett tartalom intenzívebbé válik, a pozitív vagy a negatív értékelő jelentés felerősödik. Ugyanakkor számolnunk kell a tényezővel is, hogy amíg az előbbi három alkorpusz egy blogoldal szövegeiből származik, addig az utóbbit tweetekből gyűjtöttük. Az eltéréshez tehát bizonyosan a szövegek műfaji különbsége is hozzájárult.

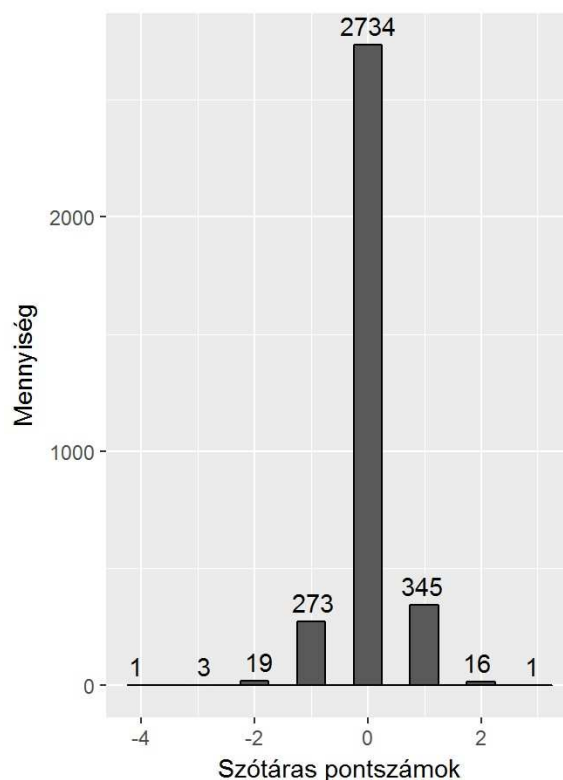
7.4 A korpusz szótáralapú elemzése és az eredmények kiértékelése

A szabályalapú elemző hatékonyságát szerettem volna összevetni az egyszerű szótáras elemzés hatékonyságával is. Ehhez szükséges volt, hogy a 3392 mondatból álló tesztkorpusz mondataihoz szótáralapon, automatikus feldolgozási módszerrel szentimentértéket rendeljünk.

A szótáras elemzéshez a pozitív és negatív listából álló szentimentszótárat alkalmaztuk (l. 7.2), amelyek segítségével a 3392 mondatnak megadtuk a szótáralapú, összesített szentimentértékét. Ehhez a szózsák-modellt (*bag-of-words*) alkalmaztuk (l. 5.2.1.3), ami a következő megoldást jelentette: minden szóhoz, amelyet az adott mondat tartalmazott és benne volt valamelyik (a pozitív vagy a negatív) szótárban, az adott szólistának megfelelően +1 vagy -1 értéket rendeltünk.^{123,124} A pozitív szótár alapján azonosított szavak tehát +1, a negatív szótárban szereplők -1 értéket kaptak. Az egyes mondatok szentimentértékét ezeknek az értékeknek az egyszerű összegéből számoltuk ki (vö. Hu–Liu 2004, Liu 2012: 49–50). A kapott eredmények statisztikai adatait az alábbi ábra foglalja össze.

¹²³ Köszönöm Bors Bencének, hogy a korpuszt a szózsák-moddellel elemezte.

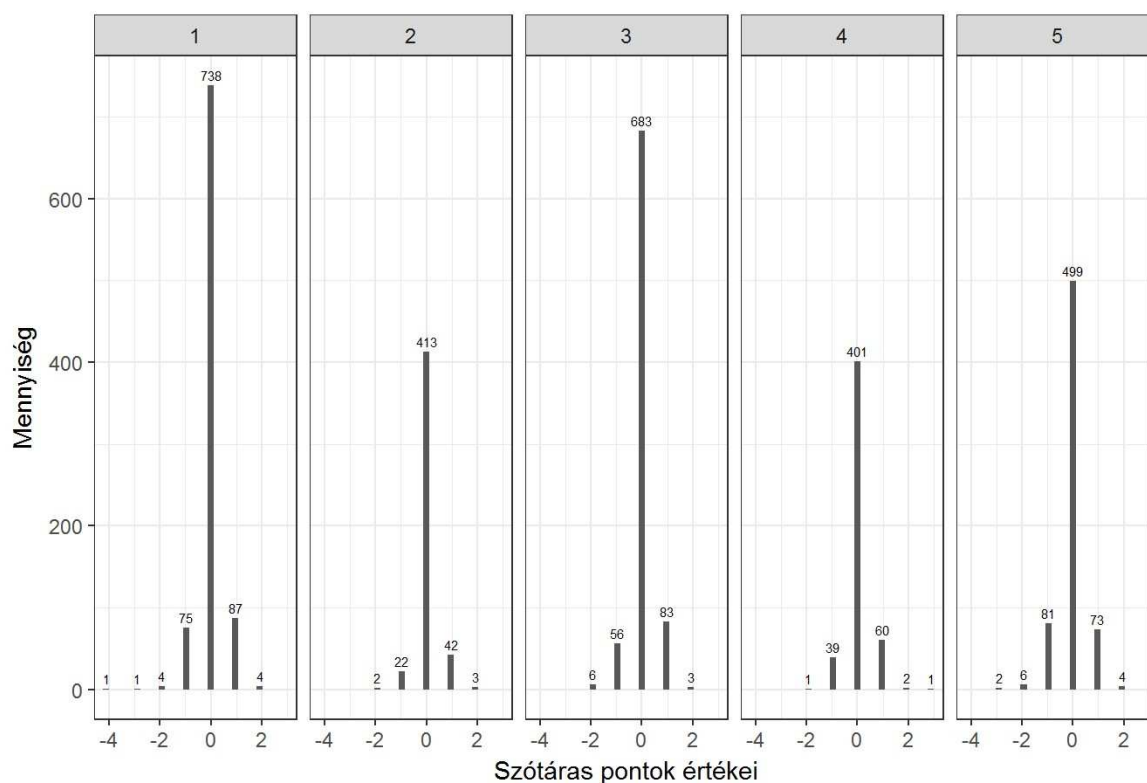
¹²⁴ A szentimentszótár elkészítésének egy fontos alapelve volt, hogy benne szóalakokat szerepeltessünk. Ez azon a megfigyelésen alapult, hogy gyakorta két, azonos szótővel rendelkező szóalak eltérő szentimentértékkel rendelkezik (vagy az egyik nem rendelkezik szentimentértékkel), pl. *Isten – isteni, haláli – halálos* stb. Ezzel összefüggésben a szótáralapú korpuszelemzést is szótővezés nélkül végeztük el. Az elváló igekötőket – szintaktikai elemzés hiányában – ebben a feladatban nem állt módunkban kezelni, az igekötős igék esetében is a szótári szóalakokat kerestük a korpuszban.



25. ábra: A szótáras elemzési eredmények statisztikai adatai

A 23. és a 25. ábrák adatait összevetve – még a pontszámok átskálázása nélkül is (l. lentebb) – azt láthatjuk, hogy az annotátori értékelések és a szótáras elemzési eredmények teljesen más típusú eloszlást mutatnak. Amíg a manuális annotációban viszonylag egyenletesebben oszlanak meg a negatív, a semleges és a pozitív értékek (l. 7.3.3), addig a szótáras feldolgozás esetében a nulla értékgyakorisága kiugróan magas (80,6%). A negatív és a pozitív értékek közül a pozitívak valamivel gyakrabban fordulnak elő. Az összevetés tapasztalata ezen a ponton arra mutat, hogy az egyszerű szótáras elemzés mind a negatív, mind a pozitív értékelő tartalmak megtalálása szempontjából problematikus, hiszen a legtöbb értékelést ezzel a módszerrel nem azonosítottuk. (Erre a következtetésre a szabályalapú elemzési megoldás eredményességének értékelése kapcsán még visszatérek, l. 7.5.)

Annak céljából, hogy a két elemzési megoldás eredményeinek az összefüggéseit alaposabban is megvizsgálhassam, megnéztük a szótáras elemzés eredményeit úgy is, hogy azokat a humán annotációban adott értékek szerint öt részre bontottuk. A kapott eredményeket az alábbi ábra mutatja be.



26. ábra: A szótáras elemzés eredményei a humán annotáció öt értéke szerinti bontásban

Az ábra alapján a következő megállapításokat tehetjük: Egyrészt, a szótáralapú elemzési eredményekben az annotátori értékelésektől függetlenül rendre a semleges érték kerül jelentős túlsúlyba. Másrészt, semmiféle korreláció nem rajzolódik ki az annotátori értékelések és szótári eredmények között (például nem látjuk azt, hogy minél pozitívabbnak értékelte az annotátor az adott mondatokat, annál magasabb átlagos pozitív érték jelenne meg a szótáras elemzés oldalán is).

Az alábbi táblázat megmutatja a pontosság, a fedés és az F-mérték értékeit a szótáras elemzés esetében, a kézi annotáció öt értéke szerint bontva:

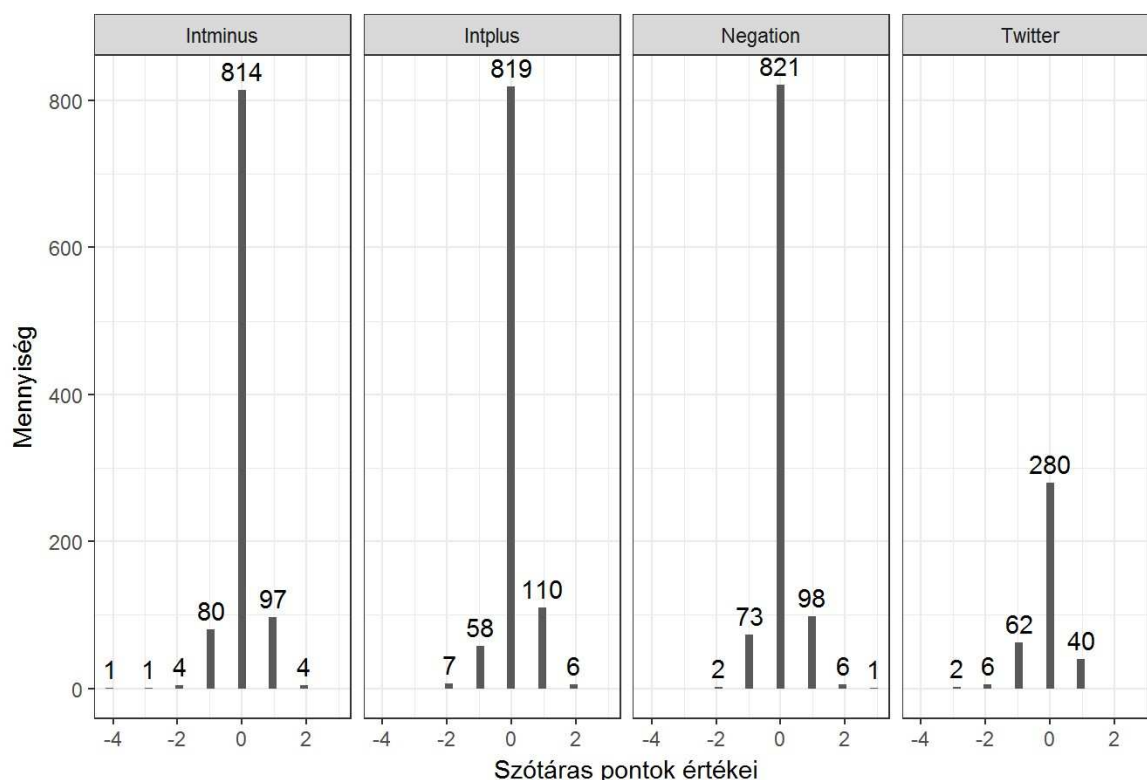
	Precision	Recall	F1
1	0,4783	0,0120	0,0235
2	0,1648	0,0933	0,1192
3	0,2524	0,8303	0,3871
4	0,2377	0,1627	0,1932
5	0,0588	0,0015	0,0029
össz:	0,2384	0,2199	0,2287

15. táblázat: A pontosság, a fedés és az F-mérték a szótáras elemzés esetében, a kézi annotáció öt értéke szerinti bontásban és összesítve

A táblázat adatai ugyancsak azt mutatják, hogy a legnagyobb probléma a szélső értékek esetében jelentkezik. A pontosság az 1-es annotátori értéknél a legnagyobb, gyakorlatilag ebben a csoportban 47,83% az esélye, hogy a humán annotációval egyező eredményt kapjuk a szótár segítségével (amennyiben az eszköz megtalálja a szövegben az értékelést). Ugyanakkor a fedés minden szélső érték, így az 1-es esetében is nagyon alacsony. A 3-as értéknél ezzel szemben magas a fedés – az algoritmus az annotátori értékek 83%-ában adta ugyanazt az értéket –, a pontosság ugyanakkor gyenge. A két érték kombinációjából számolt F-mérték megmutatja az alacsony hatékonyságot minden érték esetében, külön-külön és összesítve is.

Amennyiben súlyozott Cohen-féle Kappa-statisztikát készítünk az öt kézi annotációs értékre együttesen, azt látjuk, hogy nagyon gyenge az egyezés az annotátorok és a szótári elemzési eredmények között, pontosan 0.07927.

Azt is szerettem volna feltárni, milyen eredményt ért el a szótáralapú elemzés az egyes alkorpuszokat (Negation, IntPlus, Int Minus és az értékvesztésre képes elemeket tartalmazó tweetek) illetően. Az alábbi diagram a szótáras elemzés eredményét e bontásban mutatja be.



27. ábra: A szótáras elemzés eredménye a négy alkorpusz szerinti bontásban

A szótáralapú elemzéssel az egyes szentimentértékek hasonló gyakorisági megoszlását kapjuk minden alkorpusz esetében. Ennek problematikus volta különösen akkor szembeötlő, ha összevetjük a tweetekre vonatkozó adatokat a kézi annotálás tweetekre vonatkozó adataival (l. fentebb, 24. ábra). Amíg ugyanis a kézi feldolgozás során az annotátorok merőben eltérő gyakorisággal alkalmazták a különböző értékeket ebben az alkorpuszban a másik három alkorpuszhoz viszonyítva, addig ehhez hasonló tendencia a szótáras elemzési eredményekben egyáltalán nem mutatkozik meg.

Végezetül, az alábbi táblázat megvizsgálhatóvá teszi, hogy a humán és a szótáralapú automatikus elemzés három alapvető értékének (pozitív, negatív és semleges) különböző variációjú együttes előfordulása milyen gyakorisági megoszlást mutat.¹²⁵ A táblázatban azokat az eseteket, ahol a két elemzési megoldásban eltérés mutatkozott, szürke színnel kiemeltem.

Kézi annotáció	Szótáras elemzés	Mondatok száma
-1	-1	163
-1	0	1139
-1	1	90
0	-1	53
0	0	690
0	1	88
1	-1	80
1	0	905
1	1	184

16. táblázat: A humán és a szótáralapú elemzés három alapvető értékének különböző variációjú együttes előfordulása, azok gyakorisági adataival együtt¹²⁶

Az alapvető szentimentérték összesen 1037 esetben azonos a két elemzési megoldásban. Ez a teljes korpusznak mindössze a 30,57%-át teszi ki. Az egybeesések közül a leggyakoribb eset az, amikor a humán és a gépi annotálás egyaránt semleges értéket rendelt

¹²⁵ E vizsgálat végrehajtásához az annotátori értékeléseket a következőképpen bontottuk, a magunk definiálta csoporthatárok alapján: A semleges (3-as) kategória intuitíve beláthatóan a szótáras és kombinációs elemzés 0-s értékelése. Ettől balra helyezkedik el az 1-es és 2-es kategória, jobbra a 4-es és 5-ös. Tehát az 1-es, 2-es a negatívnak, a 4-es, 5-ös pedig a pozitívnak felel meg. Köszönöm Ilyés Virágnak a munka elvégzését.

¹²⁶ A három érték szerinti bontást azért készítettük, hogy a különböző skálákon adott értékek jobban összevethetőek legyenek egymással.

a mondathoz (690: 20,34%).¹²⁷ Az eltérések közül a legdominánsabb az, amikor a humán annotáció alapján a mondat pozitív vagy negatív szentimentértékkel rendelkezik, azonban a szótáralapú elemzés ugyanehhez semleges értéket rendel. Ez az eltéréstípus az összes esetből 2044 alkalommal fordult elő, ami a teljes korpusz 60,25%-át teszi ki.

7.5 A korpusz szabályalapú elemzése, a módszer kiértékelése és összevetése a szótáralapú módszerrel

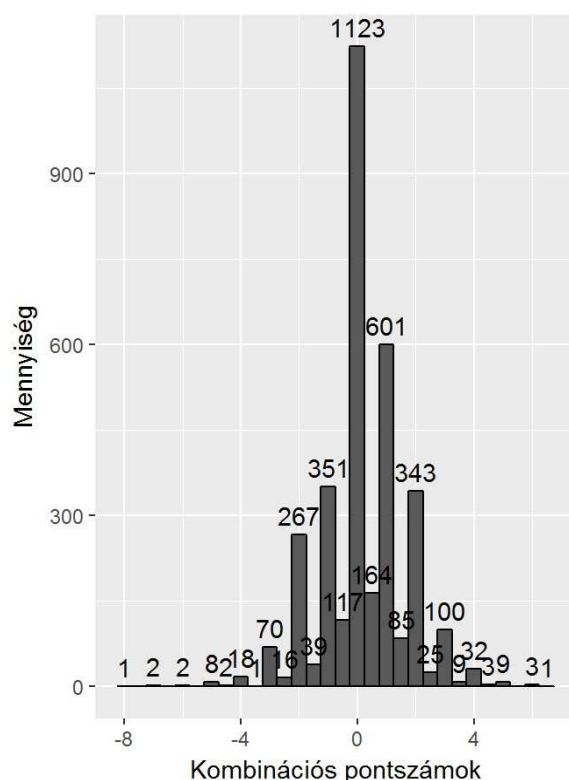
A kézi és a szótáralapú elemzést követően elvégeztük a korpusz feldolgozását a kombinációs szabályok segítségével is.¹²⁸

Az elemzési megoldás a következő alapvető lépéseket alkalmazta a korpuszon: Először, a szentimentszótárakban, azaz a pozitív és a negatív listán szereplő nyelvi elemeket azonosította, és hozzájuk – akárcsak a szótáras elemzésnél láttuk (l. 7.4) – az adott listának megfelelően +1 vagy -1 értéket rendelt. A pozitív szótár alapján azonosított szavak tehát itt is +1, a negatív szótárban szereplők -1 értéket kaptak. Ezt követően, amennyiben az adott mondatban talált egy további, a szabályokban foglaltaknak megfelelő, és az ott meghatározottak szerinti pozíciós sajátssággal rendelkező elemtípust, akkor a szentimentkifejezés értékét a szabályokban foglaltaknak megfelelően módosította. Végül, a kombinációkkal kapott értékeket mondatonként összesítette, így rendelte hozzá a korpusz egyes mondataihoz a mondatszintű szentimentértéket.

Az alábbi ábra megmutatja a szabályalapú elemzéssel kapott mondatszintű szentimentértékek gyakorisági megoszlását.

¹²⁷ Ez olyankor fordult elő, amikor az annotátor két, egymással ellentétes szentimentértéket azonosított a mondatban, és azokat körülbelül azonos súlyúnak gondolta, például: *leégés, fájó derék és lábak, de legalább alakul a kert.*

¹²⁸ Köszönöm Sipos Balázsnak a kombinációs szabályok programkódba foglalását, majd a korpusz feldolgozását a szabályalapú elemzési megoldással.



28. ábra: A szabályalapú elemzési eredmények statisztikai adatai

Amit a 23. és a 25. ábrák adatai alapján legfőbb kritikaként megfogalmaztunk a szótáras elemzésre vonatkozóan (l. 7.4), az az volt, hogy amíg a manuális munkával kapott értékek viszonylag egyenletesebben oszlanak meg a negatív, a semleges és a pozitív értékek között, addig a szótáras értékelés esetében a nulla érték gyakorisága kiugróan magas. Ez arra mutatott, hogy a szótáras elemzéssel a legtöbb értékelést ezzel a módszerrel nem találtuk meg. Amennyiben a 25. ábra adatait az itt bemutatott, 28. ábra adataival vetjük össze, azt látjuk, hogy a szótáralapú elemzéstől eltérően a szabályok alkalmazásával már jelentősen több értékelést sikerült megtalálni a mondatokban: amíg a szótáras elemzéssel 2734 mondatot (80,6%) semlegesnek ítélt az elemző eszköz (l. 7.4), addig ez az arány a szabályalapú megoldással 33,1%-ra, szám szerint 1123 esetre csökkent. E tapasztalat alapján a szótáralapú elemzésre vonatkozó első megállapítást a következőképpen módosítom: Valószínűleg a szótáras megoldással is gyakorta talál az elemző szentimentkifejezéseket a mondatokban (csakúgy, mint a szabályokat alkalmazó megoldás, amely ugyancsak a szótárra épül). Azonban ha egy adott mondatban két, egymással ellentétes érték is szerepel, mivel azokat nem tudja súlyozni, a mondatot semlegesnek

ítéli.¹²⁹ Ezzel szemben a szabályokat alkalmazó elemzési megoldásban a súlyozás lehetővé teszi, hogy két ellentétes érték közül valamelyiket nagyobb súllyal vegye figyelembe, ezáltal képes a teljes mondathoz valamely, nem semleges értéket rendelni. Mindez megmagyarázza a szabályokat alkalmazó megoldás jelentősen jobb teljesítményét, és egyúttal rámutat az egyszerű szózsák-modell komoly hiányosságára is.¹³⁰

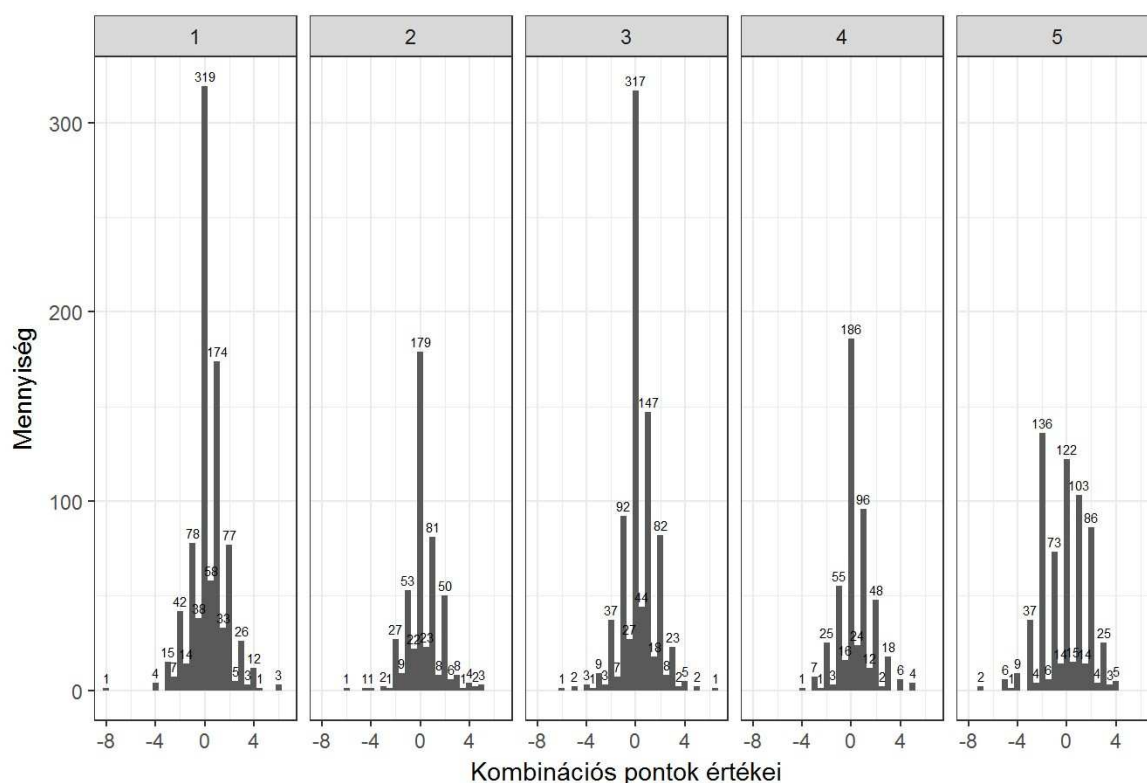
Mindemellett az is szembeötlő, hogy az annotátorok által adott öt értéket a 28. ábrán látható automatikus eredmények jelentősen jobban tükrözik, mint a 25. ábrán láthatóak, amelyen a pozitív és a negatív értékek fokozatisága gyakorlatilag – a módszer sajátosságából fakadóan érthető módon – alig jelenik meg.¹³¹

Vizsgáljuk meg a szabályalapú elemzés eredményeit úgy is, hogy azokat a humán annotáció értékei szerint öt részre bontjuk. A kapott eredményeket az alábbi ábra mutatja be.

¹²⁹ Annak feltárására, hogy a semleges ítéletek közül hány volt az egyik, és hány a másik lehetőség következménye, a disszertáció leadásáig nem tudtunk sort keríteni, ugyanis további fejlesztői munkára lett volna hozzá szükség.

¹³⁰ Ugyanezzel az észrevétellel tudok reagálni a disszertációt bíráló Szécsényi Tibor megjegyzésére, amely a következő volt: A szabályalapú elemzéssel kapott eredmények eloszlása is haranggörbe alakú lett, csak kevésbé csúcsos, jobban elterülő. Ez alapján a bíráló felveti, hogy ez a sajátosság nem csupán annak a következménye-e, hogy több lehetséges értéket oszthatott ki az elemző. Egyetértek a bírálóval abban, hogy a szabályalapú elemzés, mivel képes volt a szótárral kapott értékek további manipulálására, több értéket oszthatott ki, amelynek következtében bizonyosan csökkent a nullához közelebbi értékek gyakorisága az attól távolabbi értékek javára. Ugyanakkor ez a sajátosság semmiképpen sem indokolja a szótáras elemzéssel kapott nulla értékű adatok nagy számát, amely a szabályok alkalmazásával jelentősen csökkent. A két elemzési megoldás hatékonyságbeli eltéréséről talán informatívabban közöl a 29. ábra (l. lentebb) a 26. ábrával összevetve.

¹³¹ A szótáras elemzési megoldástól a fokozatiság kezelését értelemszerűen nem is vártam el. Itt a következő globális problémákra szerettem volna feltétlenül rámutatni: Mindenekelőtt, ahogy azt fentebb részleteztem, a szótáras módszer az ellentétes értékek egyszerű összegzése által gyakran vezet semleges értékek „kikalkulálásához”. Másrészt, semmiképpen sem érhet el jelentős korrelációt egy humán annotációval, amely árnyaltabb értékekkel dolgozik.



29. ábra: A szabályalapú elemzés eredményei
a humán annotáció öt értéke szerinti bontásban

Amint azt korábban a 26. ábra adatai megmutatták, a szótáralapú elemzési eredményekben az annotátori értékelésekről függetlenül rendre a semleges érték került jelentős túlsúlyba. Emellett azt is megállapítottuk, hogy semmiféle korreláció nem mutatkozott az annotátori értékelések és a szótári elemzési eredmények között. Ezzel szemben, a szabályalapú módszer esetében azt látjuk, hogy minél magasabb annotátori értékeléseket vizsgálunk, annál gyakoribbak a gépi elemzéssel kapott magasabb értékek is. Megállapítható továbbá, hogy az elemző eszköz a pozitív értékelések azonosításában teljesít a legkiemelkedőbben, de eredményes a semleges mondatokat illetően is.

Az alábbi táblázat a pontosságot, a fedést, valamint az F-mértéket mutatja a szabályalapú elemzés esetében, a kézi annotáció öt értékére számolva:

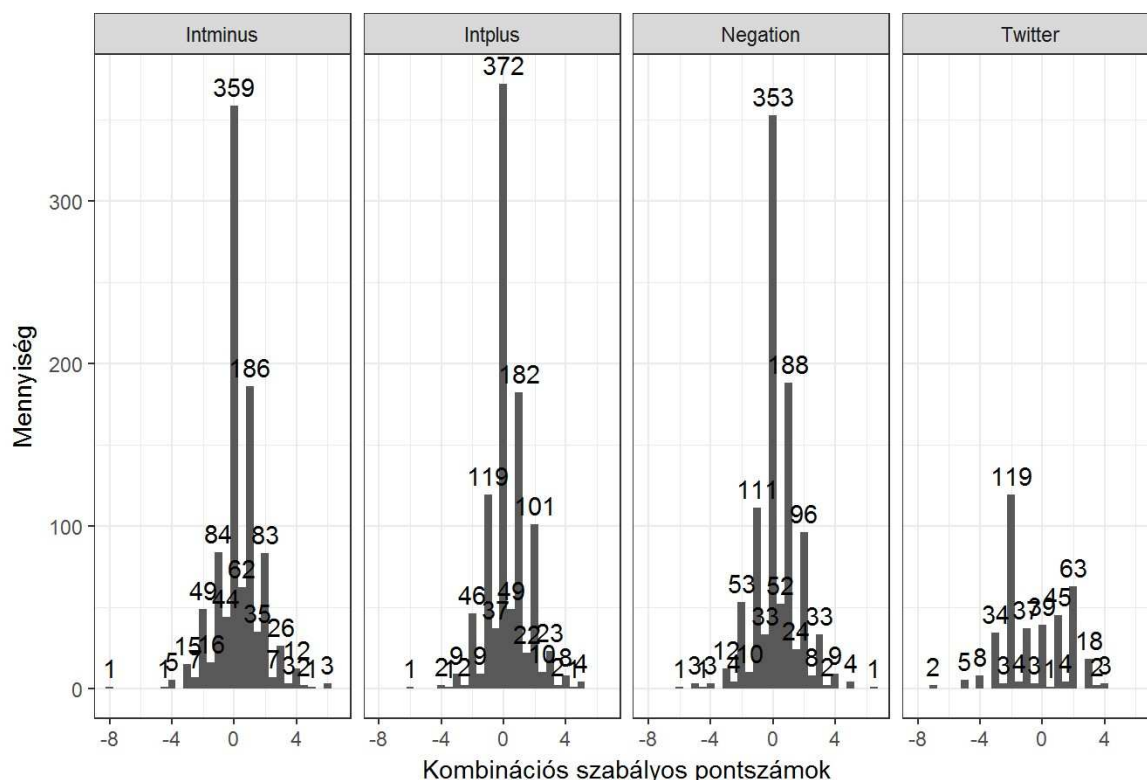
	Precision	Recall	F1
1	0,5281	0,2681	0,3557
2	0,1858	0,168	0,1765
3	0,2797	0,3911	0,3261
4	0,2215	0,2897	0,2511
5	0,3952	0,4	0,3976
össz:	0,322	0,3033	0,31236

17. táblázat: A pontosság, a fedés és az F-mérték a szabályalapú elemzés esetében, a kézi annotáció öt értéke szerinti bontásban és összesítve

A szabályalapú elemzésnél, akárcsak a szótáralapúnál, a pontosság az 1-es érték esetében a legnagyobb (itt 0,5281), azonban a szélső értékek esetében is rendre jobbak az eredmények. A fedés ebben az esetben is a 3-as értéknél a legjobb, ugyanakkor a többi érték esetében (a 4. kivételével) növekedést látunk. Az F-mérték rámutat a javulásra a szélső értékek azonosításában (0,2287-ről 0,31236-ra).

Amennyiben súlyozott Cohen-féle Kappa-statisztikát készítünk az öt kézi annotációs értékre együttesen, 0.3771 értékű egyezést látunk az annotátorok és a szótári elemzési eredmények között, ami jelentős javulás a szótárral elért 0.07927-hez képest.

Tekintsük az eredményeket az egyes alkorpuszok szerinti bontásban is!



30. ábra: A szabályalapú elemzés eredménye a négy alkorpusz szerinti bontásban

Amint azt a szótáralapú elemzést illetően tárgyaltuk (l. 7.4), az egyes szentimentértékeknek hasonló gyakorisági megoszlását kaptuk minden alkorpuszban, és ennek problematikus volta különösen szembeötlő volt akkor, amikor a tweetekre vonatkozó eredményeket vizsgáltuk. Ha összevetjük a 24., a 27. és a 30. ábrák adatait, a következőket állapíthatjuk meg: Egyrészt, a negatív és a pozitív értékek, illetve azok különböző fokozatai jelentősen gyakrabban jelennek meg a szabályalapú elemzéssel kapott adatokban minden alkorpusz esetében a szótáralapú elemzéshez képest, és ez a változás jelentősen közelebb hozza az automatikus megoldás eredményét a humán annotációhoz. Másrészt, amíg a tweetek esetében a szótáras és a humán elemzés eredménye merőben eltért, addig a szabályok alkalmazásával ez a különbség is jelentősen csökkent, és a semleges, a pozitív, valamint a negatív eredmények megoszlása gyakorlatilag korrelál egymással.¹³²

Végezetül, vessük össze a szabályalapú elemzés eredményeit is a humán annotációval aszerint, hogy a két elemzés három alapvető értékének (pozitív, negatív és semleges) különböző variációjú együttes előfordulása milyen gyakorisági megoszlást

¹³² Érdekes lett volna a szabályok alkalmazási gyakoriságát egyenként is megvizsgálni, ezáltal felmérni azok hasznosságát. E részfeladatra sajnos a munka keretében nem került sor, utólagos elvégzését pedig a disszertáció benyújtásáig nem tudtuk kivitelezni.

mutat.¹³³ A táblázatban azokat az eseteket, ahol a két elemzési megoldásban eltérés mutatkozott, szürke színnel kiemeltem.

Kézi annotáció	Szabályalapú elemzés	Mondatok száma
-1	-1	550
-1	0	514
-1	1	328
0	-1	201
0	0	314
0	1	316
1	-1	143
1	0	295
1	1	731

18. táblázat: A humán és a szabályalapú elemzés három alapvető értékének különböző variációjú együttes előfordulása, azok gyakorisági adataival együtt¹³⁴

Amint azt a szótáras elemzés és a humán annotáció eredményeit összevetve megállapítottuk (l. 7.4), 1037 esetben volt azonos az alapvető szentimentérték, ami az összes mondatnak (3392) mindössze a 30,57%-át tette ki. Az eltérések közül a legdominánsabb eset az volt, amikor a humán annotáció alapján a mondat valamilyen szentimentértékkel rendelkezett, azonban a szótáralapú elemzés ezeket nem azonosította. Ez az összes eset 60,25%-ában következett be. A szabályalapú módszer a 16. táblázat adatai alapján 1037-ről 1595-re növelte azoknak az eseteknek a számát, amikor a humán és a gépi elemzés azonos alapvető szentimentértéket adott meg, ami így az összes mondat 47,02%-át teszi ki.

A jelen fejezetben bemutatott vizsgálati eredmények alapján megállapítható, hogy a kombinációs szabályok az egyszerű szótáralapú elemzési megoldásnál egy jelentősen jobb minőségű feldolgozást tesznek lehetővé.¹³⁵ Emellett ez a módszer (különböző szabályok

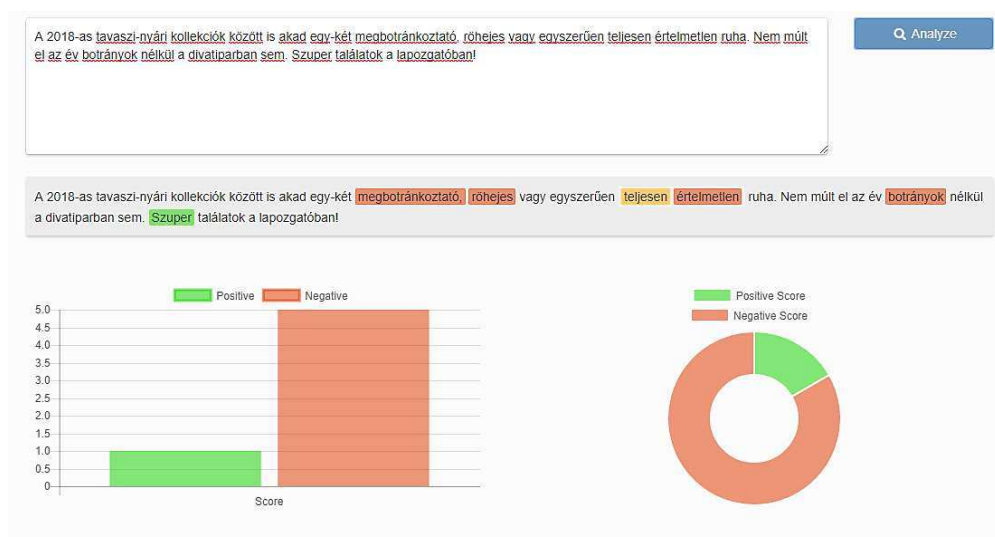
¹³³ Köszönöm Ilyés Virágnak a munka elvégzését.

¹³⁴ A három érték szerinti bontást azért készítettük, hogy a különböző skálakon adott értékek jobban összevethetőek legyenek egymással.

¹³⁵ A bemutatott elemző eszköz hatékonyságát nehéz összevetni más, magyar nyelvű szövegekre készített eszközökkel, ugyanis számuk csekély, és azokat is rendre más módszerekkel, illetve más célokat szem előtt tartva hozták létre. Az OpinHu rendszer (vö. Miháltz 2010) például 65,94%-os F-mértéket produkált egy versenyfeladatban, azonban a fejlesztők három (pozitív-negatív-semleges), illetve két kategóriát (pozitív-negatív) használtak. Hangya és mtsai (2015) ugyancsak, három, illetve két kategóriával dolgoztak, és a gépi

kódolása és alkalmazása) költséghatékony eljárás mód a nyelvtechnológiai feladatokban, hiszen például nagyméretű, kézzel annotált korpuszra sincs szükség ahhoz, hogy gépi tanulással elemző algoritmust fejlesszünk.¹³⁶ Ugyanakkor, amint arra több ízben is felhívtam a figyelmet, a kombinációs szabályokat alkalmazó szentimentelemzés bizonyos jelenségeket nem tud, és szükségszerűen nem is tudhat helytállóan kezelni (így például az értékváltást). Úgy vélem, hogy ezeknek a jelenségeknek a megfelelő automatikus feldolgozásában más jellegű, esetlegesen a szabályalapú megoldással kombinált módszer vezethetne érdemi eredményre.

Annak céljából, hogy az elkészített automatikus elemzési eszközt további – akár más doménhez tartozó szövegeken is – tesztelni lehessen, a munka záró lépéseként létrehoztunk egy olyan programot, amely egy tetszőlegesen bevitt szövegrészt a szótárak és az elemző algoritmus segítségével feldolgoz, majd ahhoz szentimentértéket rendel. Az alábbi ábra bemutatja az elemző eszköz felhasználói felületét.¹³⁷



31. ábra: A létrehozott szentimentelemző eszköz felhasználói felülete

tanítás módszerét alkalmazva próbálták hozzákapcsolni a szentimenteket a megfelelő entitásokhoz. 80%-os pontosságot meghaladva sikerült javítani az egyszerű unigramalapú osztályozás eredményein.

A disszertációban bemutatott eszközhöz hasonló, szótár- és szabályalapú szentimentelemzőt alkalmaz a magyar alapítású Neticle Technologies nevű cég (<https://neticle.hu/hogyan-mukodik-a-neticle-social-listening.html#>). Ugyanakkor a program fejlesztői üzleti érdekeket szem előtt tartva nem adnak részletes információt az elméleti megfontolásokról és a működési sajátságokról, valamint az eredményességre vonatkozóan sem közölnek mért adatokat. Emiatt a disszertációban a Neticle elemzőjének a működését nem áll módomban összevetni a saját eszközömben alkalmazott megoldásokkal és azok hatékonyságával.

¹³⁶ A részletesebb indoklásért l. a 7.1-ben az 53. lábjegyzetet.

¹³⁷ Köszönöm Bors Bencének és Sipos Balázsnak a fejlesztői munka elvégzését.

Úgy vélem, hogy a fenti elemző felület fontos segítséget nyújthat abban, hogy a szentimentelemzés szempontjából kardinális nyelvi jelenségeket vizsgálni, ezáltal pedig a jelenlegi feldolgozó módszer hatékonyságát fejleszteni lehessen.

8. Összegzés

Doktori értekezésemben a számítógépes értékeléselemzés, másképpen szentimentelemzés egy kardinálisnak tekinthető problémáját, a szentimentérték módosulásának a jelenségét tárgyaltam részletesen. Az értékmódosulás problémakörén belül az általam értékvesztésnek és értékváltásnak nevezett típusokkal foglalkoztam a legbehatóbban.

A szentimentelemzés feladata számos komoly elméleti és alkalmazott nyelvészeti kihívást rejt magában. Bár a téma különböző aspektusaival számtalan nemzetközi publikáció foglalkozik (vö. pl. Feldman és mtsai 2010, Loughran–McDonald 2011, Liu 2012, Ruppenhofer–Rehbein 2012, Young–Soroka 2012, Socher és mtsai 2013), és megjelentek a magyar nyelvű szövegek szentimentelemzésére fókuszáló dolgozatok is (vö. Berend–Farkas 2008, Miháltz 2010, 2013, Hangya és mtsai 2015) a publikációk túlnyomó többsége nyelvtechnológiai oldalról közelít ezekhez a problémákhoz; azok alkalmazott és különösen elméleti vonatkozásait csupán csekély számú dolgozat igyekszik tárgyalni.

A jelen disszertáció legfontosabb célkitűzése az volt, hogy a szentimentérték módosulását úgy tegye beható és sokrétű vizsgálat tárgyává, hogy az elemzés közben mind az elméleti, mind az alkalmazott nyelvészeti és nyelvtechnológiai vonatkozásokat egyaránt szem előtt tartja.

Az értekezés központi problémáját képező értékvesztési és értékváltási jelenségek elméleti, alkalmazott nyelvészeti és nyelvtechnológiai szempontból egyaránt figyelemre méltóak (vö. pl. Andor 2011, Kugler 2014, Szabó 2015a: 52, Szabó és mtsai 2017a: 252, Drávucz és mtsai 2017: 236). Ennek ellenére a jelenségekkel kevés dolgozat foglalkozik (vö. Tolcsvai Nagy 1988, Partington 1993, Nemesi 1998, Wierzbicka 2002, Jing-Schmidt: 2007, Laczkó 2007, Andor 2011, Kugler 2014, Dragut–Fellbaum 2014), és ezek többsége is csupán az értékvesztést tárgyalja. Az értékváltással Andor dolgozatán (2011), valamint saját munkáimon (Szabó 2015a, 2015c, 2016, 2017, Szabó és mtsai (2017a, 2017b, 2017d) kívül más nem foglalkozik. A nyelvtechnológiai oldalt illetően, egyetlen nemzetközi dolgozat (Dragut–Fellbaum 2014) elemzi a két probléma közül legalább az egyiket, az értékvesztést. A hazai nyelvtechnológiai irodalomban az értékvesztés és az értékváltás jelenségeit – egy rövid megjegyzésen kívül (vö. Hangya és mtsai 2015: 212) – kizárólag saját dolgozataim (Szabó 2015a, 2015c, 2016, 2017, valamint Szabó és mtsai 2017a, 2017b, 2017d) veszik figyelembe és vizsgálják.

Az értekezés öt nagyobb részből épült fel.

Az **első egységben** (2. fejezet) a nyelvi értékelés mibenlétének kérdését tárgyaltam részletesen, elméleti és alkalmazott nyelvészeti szempontból egyaránt. Az értékeléselemzés legalapvetőbbnek tekinthető dilemmája ugyanis – akár az elméleti, akár az alkalmazott megoldások oldaláról közelítjük meg – az, hogy pontosan mit tekinthetünk nyelvi értékelő jelentésnek.

Annak céljából, hogy feltárhassam, vajon megfelelő alapot tud-e nyújtani az elméleti irodalom a nyelvtechnológiai alkalmazások számára, áttekintettem annak megközelítési módjait (l. 2.2). Megállapítottam, hogy a legtöbb szerző a szó szemantikai komponenseként tételezi a nyelvi értékelést (vö. Berényi 1988, Péter 1991b, Szilágyi 1996) és a modalitás fogalmához kapcsolja (vö. Kiefer 1986, Berényi 1988, Murvai 2001, Pete 2002, Péter 2008). Emellett a beszélői szubjektivitással hozzák azt összefüggésbe (vö. Kiefer, 1986, Berényi 1988, Szilágyi 1986), s egyetlen kivételtől (Péter 1991b) eltekintve nem szólnak az objektív nyelvi értékelés lehetőségéről. A nemzetközi gyakorlatban a szentimentelemzést gyakran a szubjektivitáselemzés fogalmának ekvivalenseként használják (vö. pl. Wilson és mtsai 2005, Liu 2010: 627).

Rámutattam, hogy a modalitás elméleti kerete az értékeléshez úgy közelít, mint a kifejezett tartalom bizonyossági fokáról tett beszélői ítélethez (l. 2.3.1). Ez azonban nem szerencsés a számítógépes értékeléselemzés szempontjából, hiszen a nyelvi értékelés és a beszélői bizonyosság két önálló dimenziót alkotnak. Emellett különböző operátorok meglétét is feltételezi az értékelő megnyilatkozásban, ami ugyancsak – nem szerencsés módon – leszűkítené a számítógépes értékeléselemzés számára a vizsgálat tárgyát. Emellett a modalitás elméleti kerete azért is szűk a nyelvi értékeléselemzés számára, mert nem teszi lehetővé az indirekt beszédaktusokban kifejezett értékelő nyelvi tartalmak kezelését sem.

Annak céljából, hogy megállapítsam, vajon tekinthető-e az értékelés a szó jelentéskomponensének, a következő problémákat vizsgáltam meg részletesebben: a domén, a target, illetve a kontextus szerepét, valamint a kombinációs szabályok alkalmazhatóságának a lehetőségét (l. 2.3.2). Több olyan nyelvi sajátságra is rámutattam, amely alapján a nyelvi értékelést – legalább a nyelvi elemek egy csoportja esetében – bizonyosan nem tekinthetjük szemantikai komponensnek. Arra is felhívtam a figyelmet, hogy a nyelvtechnológiai irodalom – bár felfigyel ezekre a problémákra – nem kezeli azokat kellő körültekintéssel, illetve helytállóan.

Végezetül, azt is tárgyaltam, hogy a nyelvi értékelés elemzése nem korlátozódhat kizárólagosan a beszélő szubjektív ítéletét megfogalmazó értékeléstípus vizsgálatára (l.

2.3.3). Ugyanitt arról is szóltam, hogy a beszélői szubjektivitás, valamint a beszélői bizonytalanság fogalmai – bár részben összefüggenek – nem azonosíthatóak egymással.

A tapasztalatokat összegezve úgy véltem, hogy az elméleti nyelvészeti irodalomnak a nyelvi értékelésről tett megállapításai nem nyújthatnak megfelelő alapot a jelenségének sem az elméleti, sem a nyelvtechnológiai szintű kezeléséhez, és a pragmatikai sajátságok jelentősebb figyelembevételére volna szükség egy pontosabban működő automatikus elemző rendszer létrehozásához.

Az értekezés **második nagy részében** (3-4. fejezet) az értékvesztési és az értékváltási jelenségeket tárgyaltam.

Első lépésben (3. fejezet) elvégeztem a nyelvi értékelés témaköréhez kapcsolódó fogalmak tisztázását, valamint a terminológiai hiányok pótlását. Közülük itt csupán a kiegészítéseket, illetve újításokat emelem ki. *Szentimentkifejezésnek* (*sentiment words*) neveztem azokat a minimális nyelvi egységeket, amelyek lexikai szinten egyetlen meghatározott, pozitív vagy negatív értékkel rendelkeznek (pl. *szép*, *csúnya*) (vö. Szabó 2014: 7), magára az értékre pedig *szentimentértékként* utaltam. *Szentimentfragmentum*nak neveztem minden olyan szövegrészt, amely egyetlen meghatározott szentimentértékű (pozitív vagy negatív) értékelést fejez ki egyetlen target vonatkozásában (vö. Szabó és mtsai 2016: 176). A szentimentfragmentumban levő elemek lexikai szintű értéke nem feltétlenül azonos a teljes, bennfoglaló szentimentfragmentum értékével. A módosulás lehet fokozati, de akár eredményezheti azt is, hogy a két érték teljesen ellentétes lesz egymással. Azelőbbit például fokozás, az utóbbit negáció esetén tapasztalhatjuk. Emellett azt az esetet is értékmódosulásnak tekintettem, amikor egy adott elem lexikai szintű értékelő tartalma és annak kontextuális tartalma között eltérés van. Ezt példázza az értékvesztés jelensége, amikor egy a szótár szintjén negatív elem fokozó szerepbe kerül (pl. *brutálisan jó*), valamint az értékváltás is, amikor egy ugyanilyen szótári szintű szemantikai sajátsággal rendelkező elem pozitív értéket is kifejezhet (pl. *brutális koncert*). Ez a megközelítési mód azért szerencsésebb az általános gyakorlatnál, mert lehetőséget ad az olyan módosulások figyelembe vételére és vizsgálatára is, ahol a szentimentkifejezés és a bennfoglaló megnyilatkozás értéke a módosulás ellenére is azonos (pl. az általam *értékvesztésnek* nevezett jelenség bizonyos eseteiben). Amennyiben a módosulás jelenségét a nyelvtechnológiai gyakorlat szerint tekintjük, úgy az ehhez hasonló eseteket nem vehetjük figyelembe. A definíció kiegészítése emellett azért is szerencsés, mert lehetségessé teszi az olyan módosulások figyelembe vételét is, amelyekben további módosító elemeknek, azaz shiftereknek nincs szerepük. Az értékmódosulással foglalkozó

nyelvtechnológiai irodalom ugyanis nem számol azzal, hogy bizonyos esetekben egyazon lexéma hordoz eltérő értékelő jelentést a szótár és a kontextus szintjén.

A szentimentérték-módosulási jelenségein belül megkülönböztettem az úgynevezett *értékvesztést* és az *értékváltást*. Bizonyos elemek szótári szinten negatív szemantikai tartalommal rendelkeznek ugyan, azonban fokozó funkcióban képesek azt részben vagy egészben elveszíteni, sőt bizonyos esetekben arra is, hogy negatív tartalmuk ellenére pozitív értékelést fejezzenek ki (vö. Szabó 2015a: 632, 2015c: 52, 2016: 165–167). Az előbbi jelenséget *értékvesztésnek*, az utóbbit *értékváltásnak* neveztem (vö. Szabó és mtsai 2017a: 251–252).

A 4. fejezetben áttértem az értékvesztés és az értékváltás leíró nyelvészeti, nyelvtechnológiai, valamint pszichológiai szempontú vizsgálatára.

A leíró nyelvészeti megállapításokat illetően a következő legfontosabb megállapításokat tettem (l. 4.2): Az értékváltás jelenségével az elméleti oldalon részben az *enantioszémia* fogalom alatt foglalkoznak, azonban érvelésem szerint e keretben ezt a problémát nem, vagy legalábbis kimerítően nem lehetséges tárgyalni. Az ezeken kívüli dolgozatok azonban, amelyek az értékvesztés és az értékváltás jelenségét vizsgálják, ugyancsak nem írják le a jelenségeket teljes körűen, nem nagy mennyiségű, valós szövegből származó adat alapján teszik a megállapításaikat, emellett gyakorta homályosan fogalmaznak és nem is helytállóak (vö. Tolcsvai Nagy 1988, Nemesi 1998, Székely 2007, Jing-Schmidt 2007, Balogh 2009, Andor 2011, Kugler 2014).

Ami a nyelvtechnológiai vonatkozásokat illeti, egyetlen olyan dolgozatról (Dragut–Fellbaum 2014) van tudomásom, amely legalább az egyik jelenséget, az értékvesztést vizsgálja. A fejezetben részletesen bemutattam az említett kutatást (l. 4.3).

A jelenségek pszichológiai motivációját illetően részletesen tárgyaltam azt a három jelenséget (*negatív torzítás*, *pozitív torzítás* vagy *Pollyanna-jelenség* és *kongruencia*), amelyekkel a vizsgált elemek használata összefüggésbe hozható (l. 4.4). Emellett elemeztem a problémát az ún. *szemantikai prozódia* szempontjából is. Törekedtem arra, hogy rávilágítsak a fenti tendenciák dinamikus kölcsönhatására. Végezetül, az értékvesztést és az értékváltást mutató elemeket Jing-Schmidt (2007) alapján aszerint rendszereztem, hogy azok a szemantikai tartalmuk szerint az alapemóciók (vö. Ekman–Friesen 1969) közül mely típushoz kapcsolhatóak. Azok a magyar nyelvre vonatkozó dolgozatok, amelyek a negatív fokozó elemek emotív szemantikai tartalmával foglalkoznak, az érzelmek közül kizárólag a félelmet említik (vö. Nemesi 1998, Kugler 2014). Ugyanakkor vizsgálataim azt mutatták, hogy ezek az elemek a félelem mellett a düh

és az undor érzelmekhez is kapcsolódhatnak. A bánat érzelmét által motivált, értékvesztésre és/vagy értékváltásra képes elemet a magyar nyelvben – akárcsak Jing-Schmidt (2007) az angolban – nem találtam, amely jelenséget a negatív torzítás működési mechanizmusával hoztam összefüggésbe (vö. Öhman és mtsai 2001: 392–394).

Ezt követően az értékvesztést és az értékváltást mutató nyelvi elemeket a szemantikai motiváció szempontjából vizsgáltam meg részletesen, korpuszból (MNSZ2, Oravecz és mtsai 2014), valamint egyéb, valós szövegekből származó példaanyag alapján (l. 4.5). Az irodalmi tételek erre vonatkozó megállapításai ugyancsak nem egységesek, gyakran az egyes szerzők véleménye is ellentmondásos (vö. fentebb).

Az értékvesztésre vonatkozóan megállapítottam, hogy a negatív fokozó elemek a szemantikai motivátlanság tekintetében csupán inherens (pozitív vagy negatív) szentimentértékű alaptagok mellett viselkednek egységesen. Ilyenkor a vizsgált elemek deszemantizálódnak, valamint bizonyos esetekben, a pozitív alaptagok mellett a szemantikai tartalmukon interperszonális funkció alapul. Ettől eltérően, ha a vizsgált elemek inherens értékkel nem rendelkező alaptagot módosítanak, a fokozók szemantikai szempontból nem lesznek azonos viselkedésűek: negatív tartalmat hordoznak, deszemantizálódnak, vagy pedig emotív tartalmuk segítségével interperszonális funkciót töltenek be. Ráműtattam, hogy az eredményeim a vizsgált negatív emotív fokozó elemeknek egy árnyaltabb viselkedését tárták fel, ezáltal cáfolva a szakirodalmi tételek többségét, amelyek a rendszeres szemantikai kiüresedés mellett érvelnek.

A teljes és rendszeres szemantikai kiüresedés ellen szóló sajátságként ismerttettem egy korlátozási jelenséget, mely szerint semleges alaptagok mellett, pozitív értékelésben a negatív fokozó elemek használata korlátozott, ha az alaptaggal kifejezett sajátság az adott target minősítőjeként pozitív és negatív egyaránt lehet. A *szörnyen hideg víz* vagy a *rohadt hideg sör* típusú kifejezések például csak negatív értékelésként értelmezhetőek. Amennyiben az adott target vonatkozásában a megjelölt sajátság mindig pozitív tulajdonság, ilyen korlátozás nincs a pozitívan értékelő megnyilatkozásokban, például *brutálisan éles kép* vagy *marha gyors internet*.

A negatív fokozó elemeket a skalaritás szempontjából is elemeztem, és megállapítottam, hogy azok a skaláris módosító elemekkel mutatnak viselkedésbeli egyezést. Azt is adatoltam, hogy a vizsgált elemek egyes beszélők számára egymáshoz képest eltérő fokot is képesek jelölni.

Az értékvtó elemekkel kapcsolatosan feltártam, hogy velük a beszélő két típusú információt közölhet: pozitív értékelést fejezhet ki, vagy pedig az adott tartalom figyelemreméltó, meglepő voltát hangsúlyozhatja.

A vizsgálatok alapján tett fontos megállapításom, hogy amíg bizonyos, értékvesztésre képes elemek betölthetik az értékvtó funkciót (pl. *durva* és *brutális*), addig más elemek nem (pl. *átkozott* és *rohadt*). Minden általam vizsgált elem állhat értékvesztő funkcióban, azonban jóval kevesebb értékvtóként. Olyan elemet, amely éppen ellenkezőleg, tehát értékvtó lehet, de értékvesztő nem, nem találtam. A jelenséget az elemek emotív szemantikai tartalmával hoztam összefüggésbe. Azok az elemek például, amelyek az undor vagy a düh érzelemhez kapcsolódnak, tipikusan nem szerepelhetnek értékvtó, csupán értékvesztő funkcióban.

Az értekezés következő alfejezeteiben (l. 4.6) a lexikai pragmatika elméleti keretében, valamint a mentális lexikon hálózatos modelljének a segítségével vizsgáltam az értékvesztés (kis részben az értékvtás) jelenségét. Azt a problémát jártam körül, hogyan lehetne az értékvesztő elemeknek a megelőző fejezetekben feltárt sajátásaival helytállóan, ugyanakkor a számítógépes nyelvészeti alkalmazások szempontjából is hasznosítható módon elszámolni az elméleti oldalon.

A lexikai pragmatika keretét alkalmazva (l. 4.6.1) az alulspecifikált szemantikai reprezentáció elvét követtem, és alapvetően Bibok (2014, 2016) érvelésére és reprezentációs megoldására támaszkodtam. Részletesen, példákon keresztül ismételtén megvizsgáltam a szemantikai motivációt illetően korábban tett megállapításaim, és azok alapján felírtam az értékvesztésre képes elemek lexikai-szemantikai reprezentációját. Az alkalmazott keretben (vö. Bibok 2014, 2016) a reprezentációban feltüntetett opcionális jegyek a sztereotipikus struktúrából erednek. A reprezentációt a következőképpen magyaráztam: A vizsgált elemek intenzitási jegye minden megnyilatkozásban jelen kell, hogy legyen, így ezt tekinthetjük ezeknek az elemeknek a lexikai tételében a magjelentésnek. Megállapítottam, hogy bizonyos esetekben kizárólag ez a komponens vesz részt a megnyilatkozás interpretálásában (vö. Szabó 2016: 168). Az, hogy az emotív tartalomból fakadó affektivitás vagy a negatív értékelő tartalom szerephez jut-e, az aktuális használat függvénye. Bizonyos esetekben az affektív komponens aktív, és azon interperszonális funkció alapul, más esetekben a negatív szemantikai komponens (is) fontos szerepet játszik. A [félelem/düh/undor] komponens a fokozó elemek lexikai tételében nem lehet benne, kizárólag ott tételezhető, amikor a negatív elemek nem értékvesztő (és nem is értékvtó) szerepűek. A vizsgált jelenség esetében tehát ezt a

komponenst a reprezentációban opcionálisként sem jelenítem meg. Felvettem ugyanakkor, hogy hogyan volna lehetséges az értékvesztő, az értékváltó, valamint a nem értékvesztő és értékváltó használatokhoz, egyazon lexémához ugyanazt a lexikai tételt megadni. Bár a lexikai pragmatika elméleti keretében erre egyelőre nem találtam módot, megmutattam, hogyan tudna ebben a tekintetben is segíteni az aktivációterjedéses hálózati modell alkalmazása (l. lentebb).

Arra is rámutattam, hogy a vizsgált elemek interpretációs mechanizmusát további tényezők is befolyásolhatják. Ennek megfelelően például lehetséges, hogy valamely távolabbi kontextuális sajátosság ellehetetleníti, hogy a fokozó elem negatív komponense a magjelentésre épüljön, annak ellenére, hogy azt a szókapcsolat alaptagja nem akadályozná.

A lexikai pragmatika keretét illetően kiemeltam, hogy a különböző típusú jelentések megértése szükségszerűen eltérő feldolgozási időt kellene, hogy igényeljen. Ugyanakkor, ahogyan arra több szerző (vö. pl. Wilensky 1987: 20, Gibbs 1994, Nemesi 2007, Forgács és mtsai 2012) is felhívja a figyelmet, például a metaforikus és a nem metaforikus jelentések vagy az indirekt és a direkt beszédaktusok megértése a mérési eredmények alapján hasonló hosszúságú feldolgozással történik. Úgy véltem, hogy a mentális lexikon aktivációterjedéses működésének a feltételezése lehetőséget adhat a lexikai pragmatika elméleti alapvetéseinek a továbbfejlesztéséhez, ily módon a vizsgált jelenségek adekvátabb leírásához (l. 4.6.2). A modell alapján azt feltételezhetjük ugyanis, hogy metaforikus nyelvhasználat esetében az aktiváció csupán a megengedett irányba terjed tovább (vö. Caramazza (1999), idézi Kovács (2013: 187), Huszár (2005: 63–64)), esetleg egyszerre több jelentés aktivációja is bekövetkezik, és azok között nincs feltétlenül sorrendi hierarchia. Amint azt a vizsgálat során igyekeztem megmutatni, a két megközelítési mód összeegyeztethető, sőt bizonyos problémák vonatkozásában hatékonyan kiegészítheti egymást.

A mentális lexikon hálózatos reprezentációjának lehetőségét a jelen disszertáció kutatási témájával összefüggésben, több szempontból is különösen figyelemre méltónak találtam, ezért a szótári reprezentációs vizsgálatokat ebben a keretben is elvégeztem (l. 4.6.2).

Levelt (1989: 182), Baayen (2007: 97) és Kovács (2013: 26, 194) megközelítését elfogadva, a mentális lexikon egyes pontjait, amelyek a szavakat reprezentálják, „mikrohálózatként” képzeltem el. E hálózat részeit különböző típusú információk, köztük az adott szó szemantikai predikátumai teszik ki. A negatív fokozó elemek esetében az [intenzív], az [affektív], valamint a [negatív] jelentéskomponenseknek kell itt

csomópontok formájában megjelenniük. A mentális lexikont Lengyel (2012: 15) és Kovács (2013: 188) alapján komplex, többdimenziós hálózatként képzelem el, amelyben a mikro- és makrohálózatok pontjai számos hasonló alhálózati ponthoz kapcsolódnak. Ennek megfelelően tehát a vizsgált fokozó elemek bekapcsolódnak – többek között – a kollokációs hálózatba.

Az aktivációterjedéses modell alapján a hálózati kapcsolatok alapvetően meghatározzák, hogy az aktiváció mely hálózati komponensekre terjedhet ki. Amikor a vizsgált negatív elemek pozitív vagy negatív szentimentértékkel rendelkező alaptagot módosítanak, e kollokációs sajátság miatt az [intenzív], és esetlegesen az [affektív], szemantikai hálózati komponensek aktiválódnak. Amennyiben a fokozó elem a szentimentérték szempontjából semleges alaptagot módosít, a negatív minőségi jegy is aktiválódhat (vö. *rohadt hideg víz*).

Az interpretációs mechanizmust további kontextuális sajátság felülírhatja, amely ebben az elméleti keretben az aktivációterjedést gátló tényezőként jelenik meg.

Amint azt a lexikai pragmatikai elméleti keret kapcsán felvetettem, ígéretes volna a vizsgált elemek fentebbi különböző használatai esetében egyetlen lexikai tételt feltenni, majd ebből az egyetlen magjelentésből létrehozni az egyes funkciókat. Megmutattam, hogy az aktivációterjedéses hálózati modell lehetőséget ad erre, tehát minden módosítói használatnál feltehetőek az [intenzív], az [affektív], a [negatív] és a [félelem/düh/undor] komponensek a lexéma mikrohálózatában. Azt, hogy ezek közül mely komponensek jutnak szerephez az adott interpretációs folyamatban, mindig a terjedő aktiváció fogja meghatározni az aktuális kontextus függvényében.

Bár a fejezet célja az volt, hogy az értékvesztő elemek interpretálási mechanizmusát vizsgáljam, az értékváltást illetően megemlítettem, hogy ahhoz kapcsolódóan az aktuális doménnek is kiemelkedő szerep jut. Bizonyos domének ugyanis tipikusan megengedik az értékváltást, míg mások nem.

A disszertáció **harmadik nagyobb egységében** (5. fejezet) a számítógépes értékeléselemzés feladatkörével foglalkoztam részletesen.

Először részleteztem a feladat tárgyát és fontosságát (l. 5.1). Feltártam, hogy melyek azok a tudományos és technológiai kihívások és (javarészt) gazdasági hasznosítási lehetőségek, amelyek a szentimentelemzéssel kapcsolatos nemzetközi és hazai kutatásokat generálják. Tekintettel a hazai kutatások csekély számára, a legfontosabb problémákat, módszereket és eszközöket a nemzetközi kutatások bemutatásánál tárgyaltam (l. 5.2.1). A magyar nyelvű szövegekkel foglalkozó dolgozatok legfontosabb eredményeinek

áttekintését igyekeztem kimerítő jelleggel elvégezni (l. 5.2.2). A szentimentérték-módosulási jelenséget külön alfejezetben vettem górcső alá (l. 5.2.3).

A dolgozat **negyedik nagy egységében** (6. fejezet) azokat az értékmódosulással kapcsolatos kutatásokat mutattam be, amelyeket az erre a célra létrehozott három vizsgálati eszköz alapján végeztem el.

A dolgozatrész elején a különböző adatbázisoknak a nyelvészeti kutatásban és fejlesztésben betöltött szerepét tárgyaltam részletesen (l. 6.1-6.2). A korpuszok esetében tisztáztam a korpusz fogalmát, majd ismertettem a korpuszok típusait, a feldolgozásuk módszereit és eszközeit, számos példát hozva az egyes adatbázis-típusokra (l. 6.1). Mindeközben ráirányítottam a figyelmet a korpuszoknak a kutatási és fejlesztési feladatokban betöltött fontos szerepére. Végül az annotáció minőségbiztosítási megoldásait tárgyaltam részletesen. A szóasszociációs adatbázisokat illetően ismertettem az asszociációs adatfelvételek célját és hasznát, valamint tárgyaltam az adatbázisok típusait (l. 6.2). Bemutattam az adatok elemzési megoldásait, majd a legfontosabb kutatási eredményeket, különös tekintettel a hazai vizsgálatokra.

Ezt követően ismertettem a kutatás három legfontosabb vizsgálati eszközét: a fragmentum- és targetszintű annotációt tartalmazó, kézzel annotált szentimentkorpuszt, az értékvesztés és az értékváltás jelenségei szempontjából kézzel annotált korpusz, valamint egy szóasszociációs adatbázist, amely az értékvesztésre – és esetlegesen az értékváltásra is képes – elemekre adott asszociációkat tartalmazza (l. 6.3). A disszertáció alapját képező kutatómunka jelentős hozadéka, hogy mindezek, a vizsgálatokhoz szükséges eszközök a doktori kutatómunka keretében jöttek létre. Hiánypótló magyar nyelvű kutatási és fejlesztési eszközök, amelyek számos, a magyar nyelvű szövegek szentimentelemzésével, illetve a szentimentérték módosulási jelenségével kapcsolatos további kutatási és fejlesztési tevékenységhez is fontos eszközként szolgálhatnak.

A bemutatás során tárgyaltam az adatbázisok létrehozásának az okát és célját, a munkavégzés alapelveit és pontos menetét, eszközeit, valamint a minőségbiztosítási megoldásokat. A korpuszok esetében áttekintettem az azokban feldolgozott szövegtípusok sajátosságait, és ezzel összefüggésben a szövegek megválasztását is indokoltam. Végül, mindhárom eszköz esetében, példák segítségével közöltem az alapvető és részletes statisztikai adatokat, valamint a vizsgálatoknak az értekezés szempontjából legrelevánsabb kvantitatív és kvalitatív eredményeit.

A kézzel annotált szentimentkorpusz (Dívány-korpusz) összesen 154 véleményyszöveget, 17 059 mondatot és 251 202 tokent tartalmaz (vö. Szabó–Vincze 2015,

Szabó és mtsai 2016a) (l. 6.3.1). A korpusz fragmentumszintű, részletes, az entitásokat és azok különböző aspektusait is eltérően kezelő annotációt tartalmaz. A feldolgozási megoldás többek között azért figyelemreméltó, mert lehetőséget adhat az entitás-aspektus-összefüggések gépi feltárására, segítve ezzel egy hatékonyabb automatikus szentimentelemző rendszer kidolgozását. Emellett, a targeteket érintő koreferencia-viszonyokat érintő kutatási és fejlesztési feladatokat is támogathatja.

A korpuszt az annotáció alapján részletesen megvizsgáltam, feltárva ezzel a szentimentelemzés szempontjából különösen fontos nyelvi jelenség, az értékmódosulás különböző sajátságait (l. 6.3.1.7). Bár a korpusz annotációját nem specifikusan az értékváltás és az értékvesztés vizsgálatára alakítottam ki, a két jelenség közül a különböző tagek szűrésével lehetőségem volt az utóbbit lekérdezni és elemezni a korpuszban. A korpuszelemzési eredményeket az automatikus szentimentelemző létrehozásában hasznosítottam.

Az értékvesztő és az értékváltó elemek szintjén kézzel annotált korpusz (NegEmotív korpusz) létrehozásának elsődleges célja az volt, hogy olyan vizsgálati anyagot teremtsék, amelynek segítségével az értékvesztéshez és az értékváltáshoz kapcsolódó szemantikai-pragmatikai jelenségeket a specifikus annotáció alapján, kvalitatív és kvantitatív szempontból egyaránt elemezni lehet (l. 6.3.2).

A korpuszépítéshez olyan nyelvi adatokra volt szükségünk, amelyek tartalmaztak legalább egy, értékváltásra – és esetlegesen értékvesztésre is – képes elemet. Az összesen 610 tweetből álló korpuszt egy annotáló program segítségével dolgoztuk fel úgy, hogy az annotáció segítségével a negatív emotív elemek szemantikai-pragmatikai viselkedését vizsgálni lehessen (vö. Szabó és mtsai 2017a, 2017b). A munka során összesen 383 értékvesztést vagy értékváltást jelöltünk be. A korpuszt két vizsgálatban használtam fel: Egyrészt, kvantitatív és kvalitatív szempontból elemeztem az annotált adatokat, másrészt – mivel a tweetek szerzőire vonatkozó metaadatok ezt lehetővé tették – az eredményeket megvizsgáltam a férfi és a női nyelvhasználati eltérések szempontjából is.

Az eredmények többek között megmutatták, hogy a semleges alaptagok módosítóiként a negatív emotív fokozó elemek jelentős szemantikai változatosságot mutatnak, és csupán az eseteknek valamivel több mint a felében deszemantizálódnak (l. 6.3.2.7). Mindez egybevágott a korábbi szemantikai motivációs vizsgálati eredményeimmel, és számos, a negatív fokozó elemekkel foglalkozó dolgozat megállapítását cáfolta (vö. pl. Jing-Schmidt 2007, Balogh 2009). A vizsgálat további fontos hozadéka az, hogy kvantitatív és kvalitatív hasonlóságokat, illetve eltéréseket tárt

fel a férfi és a női szerzőségű tweetek között a kutatási kérdés tekintetében, kiegészítve, illetve korrigálva ezzel a problémához kapcsolódó csekély számú irodalom megállapításait (vö. Székely 2006, Jing-Schmidt 2007, Balogh 2009) (l. 6.3.2.8).

A harmadik vizsgálati eszközként egy szóasszociációs adatbázist hoztam létre, amelyben a hívószókat bizonyos szempontok alapján válogatott értékvesztésre – és esetlegesen értékváltásra is – képes elemek, valamint az összevetés céljából néhány egyéb, fokozó funkciójú nyelvi elem adta (l. 6.3.3).

Az adatfelvételt egy online kérdőív segítségével végeztük el (vö. Szabó és mtsai 2017). Összesen 36 ingerszóval dolgoztunk, amelyek közül egyenként 4-4 elemet kaptak az adatközlők, és legfeljebb öt választ adhattak egy-egy ingerszóra. A kérdőívet összesen 313 adatközlő töltötte ki, ebből 252 nő, 61 férfi volt. Kitöltőnként a 4 ingerszóra átlagosan 16,6, összesen 5198 asszociátum érkezett.

Első lépésben az asszociátumokat a szófaji megoszlások szempontjából vizsgáltam meg (l. 6.3.3.6). Azt láttuk, hogy a válaszok szófaját alapvetően a hívószó szófaja befolyásolta. Ugyanakkor, szerényebb mértékben ugyan, de a negatív fokozók lexikalizáltsági foka is kirajzolódni látszott néhány esetben. Erre mutattak például a *borzasztó–borzasztóan* és az *átkozott–átkozottul*, valamint a *durván* hívószók adatai is. További érdekes eredmény volt, hogy amikor kizárólag az első helyen érkező válaszszókra készítettük el a statisztikát, azt láttuk, hogy amíg az első asszociátumok grammatikai sajátságai gyakran azonosak a hívószó grammatikai sajátságaival, addig a továbbiakban gyakrabban jelennek meg szemantikai vagy kollokációalapú válaszok.

Az asszociátumokat a szentiment- (vö. Szabó 2014, 2015b) és az emóciósztáraink (vö. Szabó–Morvay 2015, Szabó és mtsai 2016c) segítségével is feldolgoztuk. Az asszociátumok szentimentértékei a hívószók szemantikai tartalmával szoros összefüggést mutattak. Ily módon, még azok az elemek is, amelyek a NegEmotív korpusz elemzése alapján gyakori fokozó elemnek mutatkoznak és viszonylag magas kollokációs potenciállal rendelkeznek (pl. az *iszonyat*), negatív válaszokat eredményeztek. Hasonlóan, a NegEmotív korpusz alapján láttuk, hogy például a *durva* nagyon gyakran szerepel értékváltóként, az asszociációs vizsgálatban mégis a negatív szemantikai tartalma érvényesült a rá adott válaszokban. A hívószók szemantikai tartalmával az asszociátumok emóciótartalma ugyancsak szignifikáns összefüggést mutatott. Azt láttuk tehát, hogy kontextus hiányában rendre a hívószó negatív emotív tartalma dominált. Ez a tapasztalat véleményem szerint különösen fontos a vizsgált elemek szótári reprezentációját illetően (l. 4.6). Amint azt a lexikai pragmatika elméleti keretét alkalmazva tárgyaltam (l. 4.6.1), a

negatív emotív fokozó elemeknek önálló lexikai tételt kellene feltennünk, amelynek a [félelem/düh/undor] szemantikai komponens még opcionális elemként sem lehet a része. Ugyanakkor az asszociációs eredmény azt indokolja, hogy ezzel a komponenssel feltétlenül számolnunk kellene a jelentésreprezentációban. Ez a tapasztalat is az aktivációterjedéses hálózati modellben megmutatott reprezentáció (l. 4.6.2) létjogosultságát támogatja.

Az asszociatív átfedések vizsgálata alapján a következő három legfontosabb tendenciát láttuk kirajzolódni: Egyrészt, a legtöbb hívószó azonos grammatikai sajátságú hívószóval mutatta a legnagyobb átfedést. Másrészt, az azonos tövek gyakran egymással fedtek át a legnagyobb mértékben. Ez utóbbi tendencia képezte az első tendencia alóli kivételek legtöbb esetét. A tapasztalat egybevágott egyik kezdeti feltevésemmel, mely szerint a szemantikai tartalom és a grammatikai forma bizonyos esetekben egymás ellen ható tendenciaként is jelentkezhet. Harmadrészt, a negatív, valamint a semleges és a pozitív fokozó elemek az átfedéseket illetően erősen elkülönültek egymástól, és a negatív elemek csoportján belül az átfedések jelentősen korreláltak a hívószók emotív szemantikai tartalmával (félelem, düh vagy undor). Azt is láttuk, hogy a düh emóció elemei gyakran nem a düh emóció elemeivel fedtek át a legnagyobb mértékben, azonban erre a jelenségre ezidáig kielégítő magyarázatot nem találtam. A feltárt tapasztalatok igazolták az átfedésekre vonatkozó hipotézisemet, mely szerint azokat a hívószók grammatikai és szemantikai sajátságai fogják meghatározni.

Az asszociátumokat az értékvesztésre – és esetlegesen értékváltásra is – képes a negatív emotív elemek listájával is elemeztük. A legtöbb, a feltételnek megfelelő elem a negatív elemekre érkezett. Viszonylagosan gyakrabban adtak még hasonló választ az adatközlők a meglepetés emócióhoz köthető elemekre is. Végezetül minden tárgyalt szempont alapján azt is megvizsgáltam, mutatkozik-e valamilyen eltérés a női és a férfi adatközlők asszociátumainak sajátságai között. A mérési eredmények eltértek attól, mint amit az elemzés elején feltételeztem. Úgy véltem ugyanis, hogy bár megfigyelhetünk majd némi eltérést, nem fogunk jelentősebb különbségeket tapasztalni. Ez a hipotézis azonban sem a grammatikai, sem a szemantikai sajátságok tekintetében nem igazolódott.

Az értekezés **utolsó nagy fejezetében** (7.) bemutattam egy olyan, a jelen kutatómunka keretében létrehozott, saját fejlesztésű szentimentelemző rendszert, amely szótárak és kombinációs szabályok alapján rendel összesített szentimentértéket az egyes mondatokhoz. A munka során egy szentimentszótár, egyéb kifejezéslisták, valamint az eredményesség méréséhez szükséges tesztkorpusz is készült.

A nyelvi értékeléselemzésben a kombinációs szabályok alkalmazása nem képes megoldani az automatikus nyelvi értékeléselemzés több fontos problémáját. Ugyanakkor az adatbázisok elemzési tapasztalatai arra mutattak, hogy a kombinációs szabályok alkalmazása eredményes megoldást jelenthet bizonyos részfeladatok vonatkozásában. Így például, mint láttuk, az értékvesztés és az értékváltás kezelését illetően is sikerült egy helytállóbb megoldást kínálni az egyszerű szótáras feldolgozáshoz képest.

A fejezet elején összegeztem mindazokat a vizsgálati tapasztalatokat, amelyekre támaszkodva a kombinációs szabályokat létrehoztam, ismertettem a szabályokat, majd az elemző eszköz fontos részét képező szótárakat (l. 7.1-7.2). A szabályokat a jelen disszertációban tárgyalt kutatási tapasztalatok alapján alakítottam ki. Ezt követően bemutattam azt a kézzel, mondatszinten annotált szentimentkorpuszt, amelyet az eredményesség méréséhez készítettem (l. 7.3). Végül ismertettem az eszköz hatékonyságát a humán annotációval, valamint az egyszerű szótáralapú elemzési eredményekkel összehasonlítva (l. 7.4-7.5). A vizsgálati eredmények alapján megállapítottam, hogy a kombinációs szabályok alkalmazása az egyszerű szótáralapú elemzési megoldásnál egy jelentősen jobb minőségű feldolgozást tesz lehetővé. A módszerrel az értékvesztéshez és az értékváltáshoz kapcsolódó problémák egy része is kezelhetővé vált.

Annak céljából, hogy az elkészített automatikus elemzési eszközt további – akár más doménhez tartozó szövegeken is – tesztelni lehessen, a munka záró lépéseként létrehoztunk egy olyan programot, amely egy tetszőlegesen bevitt szövegrészt a szótárak és az elkészített elemző algoritmus segítségével feldolgoz, majd ahhoz szentimentértéket rendel (l. 7.5). Az elemző felület fontos segítséget nyújthat abban, hogy a szentimentelemzés szempontjából kardinális nyelvi jelenségeket vizsgálni, ezáltal pedig a jelenlegi szentimentelemző módszer hatékonyságát fejleszteni lehessen.

IRODALJEGYZÉK

- Ahn, A.–Laporte, É.–Nam, J. 2012. Semantic Polarity of Adjectival Predicates in Online Reviews. *CoRR abs/1211.4161* (<https://arxiv.org/ftp/arxiv/papers/1211/1211.4161.pdf>)
- Aitchison, J. 1997. *The language web*. Cambridge, Cambridge University Press.
- Aitchison, J. 2003. *Words in the Mind*. Third Edition. Malden, Oxford, Blackwell.
- 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*. Lingua Franca Csoport, Pécs. 43–59.
- Andor József 2011. De durva ez a téma! – Megfigyelések a melléknévi polaritásváltásról. In *Hungarológiai Évkönyv* **12**. 33–42.
- Antweiler, W.–Frank, M.Z. 2004. Is all that talk just noise? The information content of Internet stock message boards. *Journal of Finance* **59**. 1259–1293.
- Aue, A.–Gamon, M. 2005. *Customizing sentiment classifiers to new domains: a case study*. Technical report, Microsoft Research. (http://research.microsoft.com/~anthau/new_domain_sentiment.pdf)
- Baayen, R.H. 2007. Storage and Computation in the Mental Lexicon. In Jarema, G.–Libben, G. eds. *The Mental Lexicon*. Amsterdam, Elsevier. 81–104.
- Baccianella, S.–Esuli, A.–Sebastiani, F. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Valletta, Malta. European Language Resources Association (ELRA). 2200–2204.
- Backstrom, L.–Boldi, P.–Rosa, M.–Ugander, J.–Vigna, S. 2011. *Four degrees of separation*. (<http://www.leonidzhukov.net/hse/2017/sna/papers/websci12-fourdegrees.pdf>)
- Bai, X.–Padman, R.–Airoldi, E. 2005. On learning parsimonious models for extracting consumer opinions. In *Proceedings of HICSS-05, 38th Annual Hawaii International Conference on System Sciences*. Washington, DC, USA, IEEE Computer Society. 75–82.
- Bakró-Nagy M. 2008. A tabu születése. Előleg egy fogalomtörténeti elemzésből. In Bereczki A.–Csepregi M.–Klima L. szerk. *Ünnepi írások Havas Ferenc tiszteletére. Uralisztikai Tanulmányok* **18**. Budapest, ELTE BTK Finnugor Tanszék, Numi-Tórem Alapítvány. 69–79.
- Balaskó M. 2010. Amikor a nyelvi hálózatok metszik egymást: kibővített jelentésegységek. In Balaskó M.–Balázs G.–Kovács L. szerk. *Hálózat kutatás – Hálózatok a társadalomban és a nyelvben*. Budapest, Tinta. 98–108.
- Balogh, P. 2009. *Gender-markerek a nyelvben* (<http://webfu.univie.ac.at/wp/565>)

- Bánczerowski J. 2005. A félelem tartományába tartozó negatív érzelmek konceptualizációjáról. *Magyar Nyelvőr* **129**(2). 202–208.
- Barabási A.-L. 2003. *Behálózva*. Budapest, Magyar Könyvklub.
- Barabási A.-L. 2006. A hálózatok tudománya: a társadalomtól a webig. *Magyar Tudomány* **11**. 1298–1308.
- Baroni, M.–Vegnaduzzo, S. 2004. Identifying Subjective Adjectives through Web-based Mutual Information. In Buchberger, E. ed. *Proceedings of KONVENS*. Vienna, GAI, 17–24.
- Barrat, A.–Barthélemy, M.–Vespignani, A. 2008. *Dynamical Processes on Complex Networks*. Cambridge, Cambridge University Press.
- Barsalou, L. 1987. The instability of graded structure: implications for the nature of concepts. In *Neisser*. 101–140.
- Barsalou, L. 1992. Frames, Concepts, and Conceptual Fields. In Lehrer, A.–Kittay, E.F. eds. *Frames, Fields, and Contrasts*. Lawrence Erlbaum Associates Publishers. 21–74.
- Bauer, J. 2010. *Miért érzem azt, amit te? Ösztönös kommunikációs és a tükroneuronok titka*. Budapest, Ursus Libris.
- Baumeister, R.F.–Bratslavsky, E.–Finkenauer, C.–Vohs, K.D. 2001. Bad is stronger than good. *Review of General Psychology* **5**(4). 323–370.
- Benczédy J.–Fábián P.–Rácz E.–Velcsov M. 1976. *A mai magyar nyelv*. Budapest, Tankönyvkiadó.
- Benczes R. 2011. Putting the notion of „domain” back into metonymy. In Benczes, R.–Barcelona, A.–Ruiz de Mendoza, F. eds. *Defining Metonymy in Cognitive Linguistics: Towards a Consensus View*. Amsterdam, Philadelphia, John Benjamins.
- Berend, G.–Farkas, R. 2008. Opinion Mining in Hungarian based on textual and graphical clues. In *Proceedings of the 4th Intern. Symposium on Data Mining and Intelligent Information Processing*. Santander. 408–412.
- Berend, G.–Vincze, V. 2012. How to Evaluate Opinionated Keyphrase Extraction? In *Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. Jeju, Republic of Korea. 99–103.
- Berend G.–Nagy T.I.–Móra, Gy.–Vincze, V. 2011. Inter-domain Opinion Phrase Extraction Based on Feature Augmentation. In *Proceedings of the Student Research Workshop associated with RANLP 2011*. Hissar, Bulgaria. 41–47.
- Berényi Pálné 1988. Szaknyelv és szubjektivitás. *Magyar Nyelvőr* **112**(1). 71–80.

- Bibok, K. 2014. Lexical semantics meets pragmatics. *Argumentum* **10**. Debreceni Egyetemi Kiadó. 221–231.
- Bibok K. 2016. Encyclopedic information and pragmatic interpretation. *Intercultural Pragmatics* **13**(3). 407–437.
- Bierwisch, M. 1983. Semantische und konzeptuelle Repräsentation lexikalischer Einheiten. In Motsch W.–Ruzicka, R. Hrsg. *Untersuchungen zur Semantik*. Berlin, Akademie Verlag. 61–99.
- Blutner, R. 1995. *Ansätze zur Erzeugung und Beschränkung von Interpretationsvarianten*. Arbeitspapiere des SFB 340, Bericht Nr. 71. University of Stuttgart. 33–67.
- Blutner, R. 1998. Lexical pragmatics. *Journal of Semantics* **15**. 115–162.
- Blutner, R. 2004. Pragmatics and the Lexicon. In Horn L.R.–Ward, G. *The Handbook of Pragmatics*. Oxford, Blackwell. 488–514.
- Boiy, E.–Moens, M.-F. 2009. A machine learning approach to sentiment analysis in multilingual Web texts. *Information retrieval* **12**(5). 526–558.
- Bolinger, D. 1972. *Degree Words*. The Hague, Mouton.
- Bollen, J.–Mao, H.–Pepe, A. 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (ICWSM 2011). Barcelona, Catalonia, Spain. AAAI Press. 450–453.
- Boubel, N.–Francois, T.–Naets, H.–Cental, I. 2013. Automatic extraction of contextual valence shifters. In *Proceedings of Recent Advances in Natural Language Processing* (RANLP). Shoumen, Bulgaria, INCOMA Ltd. 98–104.
- Boucher, J.–Osgood, C. 1969. The Pollyanna hypothesis. *Journal of Verbal and Learning Behavior* **8**(1). 1–8.
- Brooke, J.–Tofiloski, M.–Taboada, M. 2009. Cross-linguistic sentiment analysis: From english to spanish. In *Proceedings of the 7th International Conference on Recent Advances in Natural Language Processing*. Borovets. 50–54.
- Brugger, W. 2005. *Filozófiai lexikon*. Budapest, Szent István Társulat.
- Cacioppo, J.T.–Berntson, G.G. 1994. Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin* **115**. 401–423.
- Caramazza, A. 1999. Hány feldolgozási szint van a lexikai hozzáférésben? In Bánréti Z. szerk. *Nyelvi struktúrák és az agy*. Budapest, Corvina. 249–283.
- Carvalho, P.C.–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.* University of Lisbon, Faculty of Sciences, LASIGE. 53–56.
- Castellanos, M.–Dayal, U.–Hsu, M.–Ghosh, R.–Dekhil, M.–Lu, Y.–Zhang, L.–Schreiman, M. 2011. LCI: a social channel analysis platform for live customer intelligence. In *Proceedings of the 2011 international conference on Management of data (SIGMOD '11)*. ACM Press, New York, USA. 1049–1058.
- Chaffin, R.–Herrmann, D.J. 1988. The nature of semantic relations: a comparison of two approaches. In Evens, M.W. ed. *Relational models of the lexicon*. Cambridge, Cambridge University Press. 289–334.
- Chen, B.–Zhu, L.–Kifer, D.–Lee, D. 2010. What is an opinion about? exploring political standpoints using opinion scoring model. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI-2010)*. 1007–1012.
- Chen, L.-S.–Liu, C.-H.–Chiu, H.-J. 2011. A neural network based approach for sentiment classification in the blogosphere. *Journal of Informetrics* 5(2). 313–322.
- Choi, Y.–Cardie, C. 2008. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 793–801.
- Cohen, J. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20. 37–46.
- Cosmides, L.–Tooby, J. 2000. Evolutionary psychology and the emotions. In Lewis, M.–Haviland-Jones, J. M. eds. *Handbook of Emotions*. 2nd edition. New York, London, The Guilford Press. 91–115.
- Cramer, P. 1968. *Word Association*. London, Academic Press.
- Croft, W.–Cruse, D.A. 2004. *Cognitive Linguistics*. Cambridge, Cambridge University Press.
- Cruse, A. 2004. *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford, Oxford University Press.
- Crystal, D. 2006. *Words, Words, Words*. Oxford, Oxford University Press.
- Cseh M. 1998. Ragtalan határozóinkról. In Sándor K. szerk. *Nyelvi változó - nyelvi változás*. Szeged, JGYF Kiadó. 31–43.
- Csendes D.–Hatvani Cs.–Alexin Z.–Csirik J.–Gyimóthy T.–Prószéky G.–Váradi T. 2003. Kézzel annotált magyar nyelvi korpusz: a Szeged Korpusz. In Alexin, Z.–Csendes, D. eds. *I. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2003)*, Szeged, Szegedi Tudományegyetem. 238–245.
- Cser J. 1939. *A magyar gyermek szókincse*. Budapest, Magyar Pedagógiai Társaság.

- Csomai L. 2012. *Szabad szóasszociációs vizsgálat 10-14 évesek körében*. Szakdolgozat. Veszprém, Pannon Egyetem.
- Dave, K.–Lawrence, S.–Pennock, D.M. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of WWW-03, 12th International Conference on the World Wide Web*. New York, ACM Press. 519–528.
- De Boot, K. 2012. The end of psycholinguistics as we know it? It's about time! In Navracsics J.–Szabó D. szerk. *Mentális folyamatok a nyelvi feldolgozásban / Mental Procedures in Language Processing*. Budapest, Tinta. 13–21.
- Derczynski, L. 2016. Complementarity, F-score, and NLP Evaluation. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. 261–266.
- Dowty, D.R.–Wall, R.E.–Stanley, P. 1981. Introduction to Montague Semantics. *Studies in Linguistics and Philosophy* 11. D. Reidel Publishing Company.
- Dragut, E.–Fellbaum, C. 2014. The Role of Adverbs in Sentiment Analysis. In *Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore*. ACL, Baltimore, MD. 38–41.
- Drávucz F.–Szabó M.K. 2017. A beszélői szubjektivitás vizsgálata szentiment- és emóciókorpuszokon. In *XI. Alkalmazott Nyelvészeti Doktoranduszkonferencia konferenciakötete*. Megjelenés előtt.
- 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*. Szegedi Tudományegyetem, Informatikai Intézet. 228–239.
- Durst P.–Szabó M.K.–Vincze V.–Zsibrita J. 2014. Using automatic morphological tools to process data from a learners' corpus of Hungarian as a foreign language. *Apples – Journal of Applied Language Studies* 8(3). 39–54.
- Dutton, D.–Aron, A.P. 1974. Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology* 23. 510–517.
- Eco, U. 1999. *Kant és a kacsacsőrű emlős*. Budapest, Európa.
- Ekman, P.–Friesen, W.V. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1. 49–98.
- Engelberg, J. 2008. *Costly information processing: Evidence from earnings announcements*. Working paper. Northwestern University.
- Esuli, A.–Sebastiani, F. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)*. 417–422.

- Feldman, R.–Rozenfeld, B.–Breakstone, M.Y. 2010. SSA – A Hybrid Approach to Sentiment Analysis of Stocks. *The Israeli Seminar for Computational Linguistics (ISCOL)*. Tel Aviv University.
- Feldman, R.–Benjamin, R.–Roy, B.H.–Moshe, F. 2011. The Stock Sonar - Sentiment analysis of stocks based on a hybrid approach. In *Proceedings of 23rd IAAI Conference on Artificial Intelligence (IAAI-2011)*. 1642–1647.
- Finn, A.–Kushmerick, N. 2003. Learning to classify documents according to genre. *Journal of the American Society for Information Science* **57**. Special issue on Computational Analysis of Style. 1506–1518.
- Fitzpatrick, T. 2006. Habits and rabbits: word associations and the L2 lexicon. In Foster-Cohen, S.–Medved Krajnovic, M.–Mihaljevic Djigunvic, J. eds. *EUROSLA Yearbook* **6**. Amsterdam, John Benjamins. 121–145.
- Fodor, J.A. 1983. *The Modularity of Mind*. Cambridge, MA, Bradford Books, MIT Press.
- Fodor, J.A.–Pylyshyn Z.W. 1988. Connectionism and cognitive architecture: a critical analysis. *Cognition* **28**. 3–71.
- Forgács B.–Bohrn I.–Baudewig J.–Hofmann M.J.–Pléh C.–Jacobs A.M. 2012. Neural correlates of combinatorial semantic processing of literal and figurative noun noun compound words. *Neuroimage* **63**. 1432–1442.
- Fóris Á. 2009. A fordítás skálafüggetlen hálómodellje. *Fordítástudomány* **XI**(1). 5–21.
- Fóris Á. 2010. A fordítás skálafüggetlen hálómodellje. In Balaskó M.–Balázs G.–Kovács L. szerk. *Hálózat kutatás. Hálózatok a társadalomban és a nyelvben. Segédkönyvek a nyelvészet tanulmányozásához* **103**. Budapest, Tinta. 122–141.
- Foster, C. A.–Witcher, B. S.–Campbell, W. K.–Green, J. D. 1998. Arousal and attraction: Evidence for automatic and controlled processes. *Journal of Personality and Social Psychology* **74**. 86–101.
- Gábor K.–Héja E.–Kuti J.–Nagy V. –Váradi T. 2008. a lexikon a nyelvtechnológiában. In Kiefer F. szerk. *Strukturális magyar nyelvtan 4. A szótár szerkezete*. Budapest, Akadémiai. 853–895.
- Gangl A. 2012. Szinonimák hálózata (újra)fordított szövegekben. In Balázs G.–Kovács L.–Szőke V. szerk. *Hálózat kutatás – Interdiszciplináris megközelítések*. Budapest, Inter–Magyar Szemiotikai Társaság–Bolyai Műhely Alapítvány–Eötvös. 27–35.
- Gibbs, R.W. 1994. *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge, UK, Cambridge University Press.
- Giora, R. 1997. Understanding figurative and literal language: the graded salience hypothesis. *Cognitive Linguistics* **8**(3). 183–206.

- Giora, R. 2003. *On our Mind: Salience, Context and Figurative Language*. New York, Oxford University Press.
- Givón, T. 1984. *Syntax: A functional-typological introduction. Vol. 1*. Philadelphia, John Benjamins.
- Glucksberg, S. 2001. *Understanding Figurative Language: From Metaphors to Idioms*. Oxford, Oxford University Press.
- Go, A.–Bhayani, R.–Huang, L. 2009. *Twitter sentiment classification using distant supervision*. Technical report, Stanford.
- Godbole, N.–Srinivasaiah, M.–Skiena, S. 2007. Large-scale sentiment analysis for news and blogs. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*. (<http://icwsml.org/papers/3--Godbole-Srinivasaiah-Skiena.pdf>)
- Gósy M. 2005. *Pszicholingvisztika*. Budapest, Osiris.
- Gósy M. 2000. Az életkor hatása a mentális lexikon működésére. *Magyar Nyelvőr* **124**(4). 410–423.
- Gósy M. 2001. Szóasszociációs műveletek az életkor függvényében. *Alkalmazott Nyelvtudomány* **1**(1). 17–30.
- Gósy M.–Kovács M. 2001. Mentális lexikon a szóasszociációk tükrében. *Magyar Nyelvőr* **125**(3). 330–354.
- Grice, H.P. 1975. Logic and conversation. In Cole, P.–Morgan, J. eds. *Studies in Syntax and Semantics III: Speech Acts*. New York, Academic. 183–198.
- Groh, G.–Hauffa, J. 2011. Characterizing Social Relations Via NLP-Based Sentiment Analysis. In Adamic, L.A.–Baeza-Yates, R.A.–Counts, S. eds. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM-2011)*. The AAAI Press. 502–505.
- Gulyás A.–Szabó M.K.–Ifj. Boros I.–Havadi G. 2018. A Rákosi-éra pártjegyzőkönyveinek feldolgozása, elemzése és vizualizációja szövegalapú kapcsolatháló-elemzési módszerekkel. In Vincze V. szerk. *XIV. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2018)*. Szeged, Szegedi Tudományegyetem, Informatikai Intézet. 109–121.
- Hangya V.–Farkas R. 2015. Doménspecifikus polaritáslexikonok automatikus előállítás magyar nyelvre. In Tanács A.–Varga V.–Vincze V. szerk. *XI. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, JATEPress. 210–218.
- Hangya, V.–Berend, G.–Farkas, R. 2013. SZTE-NLP: Sentiment Detection on Twitter Messages. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Atlanta, Georgia, USA, Association for Computational Linguistics. 549–553.

- Hangya, V.–Berend, G.–Varga, I.–Farkas, R. 2014. SZTE-NLP: Aspect Level Opinion Mining Exploiting Syntactic Cues. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Dublin, Ireland. 610–614.
- Hangya V.–Farkas R.–Berend G. 2015. Entitásorientált véleménydetekció webes híryanagokból. In Tanács A. – Varga V. – Vincze V. szerk. *XI. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2015)*. Szeged, Szegedi Tudományegyetem. 227–234.
- Hangya, V.–Szántó Zs.–Farkas R. 2017. Latent Syntactic Structure-Based Sentiment Analysis. In *2nd IEEE International Conference on Computational Intelligence and Applications*. Beijing, China. (<http://real.mtak.hu/64698/1/paper-iccai.pdf>)
- Hansen, C.H.–Hansen, R.D. 1988. Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology* **54**. 917–924.
- Hart, R.P.–Childers, J.P.–Lind, J.C. 2013. *Political Tone: How Leaders Talk and Why*. Chicago University Press.
- Hatzivassiloglou, V.–McKeown, K. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of 35th Meeting of the Association for Computational Linguistics*. Madrid. 174–181.
- Honkela, T.–Izzatdust, Z.–Lagus, K. 2012. Text Mining for Wellbeing: Selecting Stories Using Semantic and Pragmatic Features. In *Lecture Notes in Computer Science* **7553**. 467–474.
- Hu, M.–Liu, B. 2004. Mining and Summarizing Customer Reviews. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*. Seattle, Washington, USA. 168–177.
- Hu, Y.–Duan, J.–Chen, X.–Pei, B.–Lu, R. 2005. A new method for sentiment classification in text retrieval. In *Proceedings of International Joint Conference on Natural Language Processing (IJCNLP 2005)*. Berlin, Heidelberg, Springer. 1–9.
- Hogenboom, A.–Iterson, P. van–Heerschop, B.–Frasincar, F.–Kaymak, U. 2011. Determining Negation Scope and Strength in Sentiment Analysis. In *2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2011)*. IEEE. 2589–2594.
- Huszár Á. 2005. *A gondolatól a szóig*. Budapest, Tinta.
- Ikeda, D.–Takamura, H.–Ratinov, L.–A.–Okumura, M. 2008. Learning to shift the polarity of words for sentiment classification. In *Proceedings of the 3rd International Joint Conference on Natural Language Processing (IJCNLP-2008)*. 296–303.
- Imrényi 2012. A magyar mondat mint multiplex hálózat. In Balázs G.–Kovács L.–Szőke V. szerk. *Hálózat kutatás – Interdiszciplináris megközelítések*. Budapest, Inter–Magyar Szemiotikai Társaság–Bolyai Műhely Alapítvány–Eötvös. 36–42.

- Imrényi A. 2013. A magyar mondat viszonyhálózati modellje. *Nyelvtudományi Értekezések* **164**. Budapest, Akadémiai Kiadó.
- Ingerman, P.Z. 1967. Paṇini Backus Form suggested. *Communications of the ACM* **10**(3). 137.
- Israel, M. 2004. The pragmatics of polarity. In Horn, L.–Ward, G. eds. *The Handbook of Pragmatics*. Oxford, Blackwell. 701–723.
- Jiang, L.–Yu, M.–Zhou, M.–Liu, X.–Zhao, T. 2011. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Portland, Oregon. Association for Computational Linguistics. 151–160.
- Jing-Schmidt, Z. 2007. Negativity bias in language: A cognitive-affective model of emotive intensifiers. *Cognitive Linguistics* **18**(3). 417–443.
- Johnson, M. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago, University of Chicago Press.
- Kaji, N.–Kitsuregawa, M. 2007. *Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents*.
(<http://www.aclweb.org/anthology/D07-1#page=1109>)
- Kamp, H. 1975. Two theories about adjectives. In Keenan, E.L. ed. *Formal Semantics for Natural Language*. Cambridge, Cambridge University Press. 123–155.
- Kamps, J.–Marx M.–Mokken, R.–de Rijke, M. 2004. Using WordNet to Measure Semantic Orientations of Adjectives. In *Proceedings of the fourth international conference on Language Resources and Evaluation*. Lisbon, Portugal. 1115–1118.
- Kanayama, H.–Nasukawa, T. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '06)*. Sydney, Australia. 355–363.
- Károly K. 2003. Korpusznyelvészeti és fordításkutatás. *Fordítástudomány* **5**(2). 18–26.
- Katz, J.J.–Fodor, J.A. 1963. The structure of semantic theory. *Language* **39**(2). 170–210.
- Keenan, E.L. 1974. The functional principle: Generalizing the notion of Subject of. In *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*. Chicago, Illinois. 298–310.
- Kegyesné Szekeres E. 2002. Férfiak és nők: mentális lexikon és fogalmi tipizálás. In Fóris Á.–Kárpáti E.–Szűcs T. szerk. *A nyelv nevelő szerepe*. Pécs, Lingua Franca Csoport. 233–240.
- Kennedy, A.–Inkpen, D. 2006. Sentiment Classification of Movie Reviews using Contextual Valence Shifters. *Computational Intelligence* **22**(2). 110–125.

- Kertész A.–Rákosi Cs. 2008. *Adatok és plauzibilis érvelés a nyelvészetben*. Debrecen, Kossuth Egyetemi.
- Kicsi S.A. 2010. Az enantioszémia néhány főbb esete a magyarban. *Magyar Nyelv* **106**. 196–198.
- Kiefer F. 1986. A modalitás fogalmáról. *Nyelvtudományi közlemények* **88**(1–2). 1–38.
- Kiefer F. 2007. *Jelentéselmélet*. 2. kiadás. Budapest, Corvina.
- Kiefer F. 2008. A melléknevek szótári ábrázolásáról. In Kiefer F. szerk. *Strukturális magyar nyelvtan 4. A szótár szerkezete*. Budapest, Akadémiai. 805–538.
- Kim, S.-M.–Hovy, E. 2006. Identifying and analyzing judgment opinions. In *Proceedings of the conference of North American Chapter of the Association of Computational Linguistics* (HLT-NAACL '06). New York, USA. 200–207.
- Kiss, G.R. 1975. An Associative Thesaurus of English: Structural Analysis of a Large Relevance Network. In Kennedy, A.–Wilkes, A. eds. *Studies in Long Term Memory*. London, John Wiley & Sons. 103–122.
- Klaudy K. 2005. Párhuzamos korpuszok felhasználása a fordításkutatásban. In Lanstyák I.–Vanconé Kremmer I. szerk. *Nyelvészetéről változatosan*. Dunaszerdahely, Gramma. 153–185.
- Klégr, A. 2013. The limits of polysemy: Enantiosemy. *Linguistica Pragensia* **23**. 7–23.
- Kohn J. 1999. Párhuzamos szövegek számítógéppel segített elemzése a fordításoktatásban (1. rész). *Fordítástudomány* **1**(1). 67–78.
- Kouloumpis, E.–Wilson, T.–Moore, J.D. 2011. Twitter sentiment analysis: The good the bad and the OMG! In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (ICWSM 2011). Barcelona, Catalonia, Spain. AAAI Press. 538–541.
- Kovács J.L. 2008. Nyelvi hálózatok a mentális lexikonban – Agykapocskutatás. *Alkalmazott Nyelvtudomány* **VIII**(1-2). 177–196.
- Kovács J.L. 2009. Begriffsstruktur des mehrsprachigen mentalen Lexikons. In Fekete-Csizmazia, Zs.–Lang, E.–Pólay, V.–Szatmári, P. Hrsg. *Sprache – Kultur – Berührungen*. Szombathely–Wien, Savaria University Press–Praesens. 65–75.
- Kovács, J.L. 2013. *Fogalmi rendszerek és lexikai hálózatok a mentális lexikonban*. Budapest, Tinta.
- Krippendorff, K. 2004. *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, California, Sage.

- Krug, M.G. 2000. *Emerging English Modals: A Corpus-Based Study of Grammaticalization. Topics in English Linguistics* **32**. Berlin and New York, Walter de Gruyter.
- Kugler N. 2014. A nyelvi polaritás kifejezésének egy mintázata, avagy milyen a *félelmetesen* jó? *Magyar Nyelvőr* **138**(2). 129–139.
- Kugler N.–Tolcsvai Nagy G. 2000. *Nyelvi fogalmak kisszótára A-Zs*. Budapest, Korona.
- Laczkó M. 2007. Napjaink tizenéveseinek beszéde szóhasználati jellemzők alapján. *Magyar Nyelvőr* **131**(2). 173–184.
- Laczkó M. 2008. A minősítés kifejezésére használt szerkezetek a tanulók spontán beszédében. *Magyar Nyelvőr* **132**(3). 326–340.
- Lahav, R. 1989. Against compositionality: the case of adjectives. *Philosophical Studies* **55**. 111–129.
- Lahav, R. 1993. The combinatorial-connectionist debate and the pragmatics of adjectives. *Pragmatics and Cognition* **1**. 71–88.
- Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago, University of Chicago Press.
- Lakoff, G. 1990. The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics* **1**. 39–74.
- Lakoff, G.–Johnson, M. 1980. *Metaphors We Live By*. Chicago, University of Chicago Press.
- Lakoff, G.–Turner, M. 1989. *More than cool reason: A Field Guide to Poetic Metaphor*. Chicago, University of Chicago Press.
- Landis, J.R.–Koch, G.G. 1977. The measurement of observer agreement for categorical data. *Biometrics* **33**. 159–174.
- Lavallee, A. 2007. Friends Swap Twitters, and Frustration New Real-Time Messaging Services Overwhelm Some Users With Mundane Updates From Friends. *Wall Street Journal*, 2007. 03.16.
(http://www.wsj.com/public/article/SB117373145818634482-ZwdoPQ0PqPrcFMDHDZLz_P6osnI_20080315.html)
- LeDoux J. 2000. *Mózg emocjonalny. Tajemnicze podstawy życia emocjonalnego, tłum.* Poznań, A. Jankowski.
- Leech, G. 1983. *Principles of Pragmatics*. London, Longman.
- Lengyel Zs. 2010. Szóasszociációs vizsgálatok: általános elvek, célok. In Balaskó M.–Balázs G.–Kovács L. szerk. *Hálózat kutatás. Hálózatok a társadalomban és a nyelvben. Segédkönyvek a nyelvészet tanulmányozásához* **103**. Budapest, Tinta. 195–208.

- Lengyel Zs. 2008. *Magyar asszociációs normák enciklopédiája I.* Budapest, Tinta.
- Lengyel Zs. 2012. *Szóról szóra.* Budapest, Gondolat.
- Levelt, W.J.M. 1989. *Speaking. From Intention to Articulation.* Cambridge, London, MIT Press.
- Levin, B. 1993. *English verb classes and alternations: A preliminary investigation.* Chicago, University of Chicago Press.
- Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* **45**. 221–247.
- Li, S.–Lee, S.–Chen, Y.–Huang, C.–Zhou, G. 2010. Sentiment classification and polarity shifting. In *Proceedings of the 23rd International Conference on Computational Linguistics* (Coling 2010). Beijing. 635–643.
- Li, S.–Wang, Z.–Li, S.–Huang, C. 2013. Sentiment classification with polarity shifting detection. In *Proceedings of the international conference on Asian language processing* (IALP). 129–132.
- Libárdi P. 2001. A nyelvi változás tükröződése az egyén mentális lexikonában. *Alkalmazott Nyelvtudomány* **1**(1). 39–49.
- Liu, B. 2010. Sentiment Analysis and Subjectivity. In Indurkha, N.–Damerau, F.J. eds. *Handbook of Natural Language Processing*, Second Edition. Boca Raton, FL, Chapman & Hall. 627–666.
- Liu, B. 2012. *Sentiment Analysis and Opinion Mining*. Kézirat. (<http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>)
- Liu, Y.–Huang, X.–An, A.–Yu, X. 2007. ARSA: a sentiment-aware model for predicting sales performance using blogs. In *Proceedings of ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR-2007). ACM, New York, NY, USA. 607–614.
- Loughran, T.–McDonald, B. 2011. When is a liability not a liability? Textual Analysis, Dictionaries and 10-Ks. *The Journal of Finance* **66**(1). 35–66.
- Louw, B. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In: Baker, M.–Francis, G.–Tognini-Bonelli, E. eds. *Text and Technology: In Honour of John Sinclair*. Amsterdam, John Benjamins. 157–175.
- Maas, A.L.–Daly, R.E.–Pham, P.T.–Huang, D.–Ng, A.Y.–Potts, C. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (ACL-2011). 142–150.
- Malt, B.C.–Sloman, S.A.–Gennari, S.P. 2003. Universality and language specificity in object naming. *Journal of Memory and Language* **49**(1). 20–42.

- McEnery, T. 2005. Corpus Linguistics. In Mitkov, R. 2005. *The Oxford Handbook of Computational Linguistics*. Oxford, Oxford University Press. 448–463.
- McEnery, T.–Wilson, A. 1997. Teaching and language corpora (TALC). *ReCALL* 9(1). 5–14.
- Meyer, R. 1994. Probleme von Zwei-Ebenen-Semantiken. *Kognitionswissenschaft* 4. 32–46.
- Miháltz M. 2010. OpinHu: online szövegek többnyelvű véleményelemzése. In Tanács A.–Vincze V. szerk. *VII. Magyar Számítógépes Nyelvészeti Konferencia* (MSZNY 2010). Szeged, Szegedi Tudományegyetem. 14–23.
- Miháltz M. 2013. OpinHuBank: szabadon hozzáférhető annotált korpusz magyar nyelvű véleményelemzéshez. In Tanács A.–Vincze V. (szerk.) *IX. Magyar Számítógépes Nyelvészeti Konferencia* (MSZNY 2013). Szeged, Szegedi Tudományegyetem. 343–345.
- Miháltz, M.–Hatvani, Cs.–Kuti, J.–Szarvas, Gy.–Csirik, J.–Prószéky, G.–Váradi, T. 2008. Methods and Results of the Hungarian WordNet Project. In *Proceedings of the Fourth Global WordNet Conference* (GWC 2008). Szeged, Hungary. 310–320.
- Miller, G.A. 1990. WordNet: An on-line lexical database. *International Journal of Lexicography* 3(4). 235–312.
- Miller, M.–Sathi, C.–Wiesenthal, D.–Leskovec, J.–Potts, C. 2011. Sentiment flow through hyperlink networks. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (ICWSM-2011). Association for the Advancement of Artificial Intelligence.
(<https://cs.stanford.edu/people/jure/pubs/sentiflow-icwsm11.pdf>)
- Mogyoródi E. 2012. Parmenidész és a semmi. *Magyar filozófiai szemle* 56(2). 9–24.
- Moghaddam, S–Popowich, F. 2010. Opinion polarity identification through adjectives. *Computing Research Repository* (CoRR) *abs/1011.4623*.
(<http://arxiv.org/ftp/arxiv/papers/1011/1011.4623.pdf>)
- Moilanen, K.–Pulman, S. 2007. Sentiment Composition. In *Proceedings of Recent Advances in Natural Language Processing* (RANLP 2007). 378–382.
- Montague, R. 1970. Universal Grammar. *Theoria* 36. 373–398.
- Montague, R. 1974. *Formal Philosophy. Selected papers of Richard Montague*. Yale University Press.
- Muhammad, A.–Wiratunga, N.–Lothian, R.–Glassey, R. 2013. Contextual sentimentanalysis in social media using high-coverage lexicon. In *Research and Development in Intelligent Systems XXX*, Springer. 79–93.

- Murvai O. 2001. Szemantikai és pragmatikai kiegészítések a modalitás vizsgálatához. In *Officina Textologica* **5**, Debrecen. 59–69.
- Na J.–Sui, H.–Khoo, C.–Chan, S.–Zhou, Y. 2004. Effectiveness of Simple Linguistic Processing in Automatic Sentiment Classification of Product Reviews. In *Conference of the International Society for Knowledge Organization (ISKO-04)*. Wurzburg, Germany, Ergon Verlag. 49–54.
- Nádasdy Á. 2003. Modern Talking. *Magyar Narancs* 2003/27.
- Nemesi A.L. 1998. Társalgási divatkifejezések a benyomáskeltés szolgálatában. *Magyar Nyelvőr* **122**. 24–35.
- Nemesi, A.L. 2007. *The issue of figurative language in pragmatics*. Ph.D. thesis. Szeged.
- Németh T.E. 2011. A humán kommunikáció modelljei és az ember-gép kommunikáció. In Németh T. E. szerk. *Ember-gép kapcsolat. A multimodális ember-gép kommunikáció modellezésének alapjai. Segédkönyvek a nyelvészet tanulmányozásához* **133**. Budapest, Tinta. 43–61.
- Nerlich, B.–Tiddm Z.–Herman, V.–Clarke, D.D. szerk. 2003. *Polysemy: Flexible patterns of meaning in mind and language*. Berlin, Walter de Gruyter.
- Newman, M.E.J. 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **46**. 323–351.
- Nikunen, A. 2007. *Different approaches to word sense disambiguation*. Language technology and applications Essay. University of Helsinki, Department of Computer Science.
- O'Connor, B.–Balasubramanyan, R.–Routledge, B.R.–Smith, N.A. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth International AAAI Conference on Weblogs and Social Media (ICWSM 2010)*. (<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1536/1842>)
- Oláh-Nagy N. 2012. Kulturális különbségek a mentális lexikonban – szóasszociációs vizsgálat és számítógépes klaszterelemzés. In Balázs G.–Kovács L.–Szőke V. szerk. *Hálózat kutatás – Interdiszciplináris megközelítések*. Budapest, Inter–Magyar Szemiotikai Társaság–Bolyai Műhely Alapítvány–Eötvös. 50–57.
- Oravecz, Cs.–Várad, T.–Sass, B. 2014. The Hungarian Gigaword Corpus. In *Proceedings of LREC*. Reykjavik, Iceland, European Language Resources Association (ELRA). 1719–1723.
- Ortigosa, A.–Martín, J.M.–Carro, R.M. 2014. Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior* **31**. 527–541.
- Osgood, C.E.–Suci, G.J.–Tannenbaum, P.H. 1957. *The Measurement of Meaning*. Urbana, IL., University of Illinois Press.

- Öhman, A.–Lundqvist, D.–Esteves, F. 2001. The face in the crowd revisited: a threat advantage with schematic stimuli. *Journal of Personality and Social Psychology* **80**. 381–396.
- Pak, A.–Paroubek, P. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation* (LREC 2010). European Language Resource Association (ELRA). 1320–1326.
- Pang, B.–Lee, L. 2008. Opinion mining and sentiment analysis. In *Foundations and Trends in Information Retrieval* **2**(1-2). 1–135.
- Pang, B.–Lee, L.–Vaithyanathan, S. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of EMNLP-02, the Conference on Empirical Methods in Natural Language Processing*. Philadelphia, PA, Association for Computational Linguistics. 79–86.
- Pápay, K.–Szeghalmy, S.–Szekrényes, I. 2011. Hucomtech multimodal corpus annotation. *Argumentum* **7**. 330–347.
- Papp F. 1984/2006. Anyanyelv és idegen nyelv a verbális asszociációk tükrében. *Magyar Nyelv* **LXXX**(2). 202–207. Változatlan formában: In Klaudy K. szerk. *Papp Ferenc olvasókönyv*. Budapest, Tinta. 219–224.
- Paradis, C. 2001. Adjectives and boundedness. *Cognitive Linguistics* **12**. 47–64.
- Partington, A. 1993. Corpus evidence of language change: The Case of the Intensifier. In Baker, M.–Francis, G.–Tognini-Bonelli, E. *Text and Technology: In honour of John Sinclair*. Italy, University of Bologna. Amsterdam, John Benjamins. 177–192.
- Péché O. 2007. A lexikai kohézió vizsgálata a fordításkutatásban – új távlatok a korpusznyelvészeti módszernek köszönhetően. *Fordítástudomány* **IX**(1). 79–96.
- Pedregosa, F.–Varoquaux, G.–Gramfort, A. et al. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**. 2825–2830.
- Peeters, G.–Czapinski, J. 1989. Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. In Stroebe, W.–Hewstone, M. eds. *European Review of Social Psychology* **1**. Chichester, UK, Wiley. 33–60.
- Pete I. 1999. Az állító és tagadó mondatok szinonimiája a magyarban. *Magyar Nyelv* **95**(3). 305–312.
- Pete I. 2002. A nyelvi modalitás három fő típusának hierarchiája a magyarban. *Magyar nyelv* **98**(2). 173–191.
- Pete I. 2009. A fokozás mibenlétéről. *Magyar Nyelvőr* **133**(3). 295–310.

- Péter M. 1991a. Érzelmi-értékelő tartalmak a szójelentésben. *Magyar Nyelvőr* **87**. 134–144.
- Péter M. 1991b. *A nyelvi érzelmkifejezés eszközei és módjai*. Budapest, Tankönyv.
- Péter M. 2008. A magyar aspektusról – más aspektusból. *Magyar Nyelv* **CIV**(1). 1–11.
- Pinker, S. 1994. The game of the name. *The New York Times*. April 5, 1994.
- Pléh Cs.–Lukács Á.–Kas B. 2008. A szótár pszicholingvisztikája. In Kiefer F. szerk. *Strukturális magyar nyelvtan 4. A szótár szerkezete*. Budapest, Akadémiai. 789–852.
- Polanyi, L.–Zaenen, A. 2004. Contextual valence shifters. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text*. 106–111.
- Postman, L.–Keppel, G. eds. 1970. *Norms of Word Association*. New York, Academic Press.
- Poursepanj, H.–Weissbock, J.–Inkpen, D. 2013. uOttawa: System description for SemEval-2013 Task 2: Sentiment Analysis in Twitter. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Atlanta, Georgia, Association for Computational Linguistics. 380–383.
- Pontiki, M.–Galanis, D.–Papageorgiou, H.–Manandhar, S.–Androutsopoulos, I. 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Association for Computational Linguistics. 486–495.
- Powers, D.M.W. 2011. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies* **2**(1). 37–63.
- Pratto, F.–John, O.P. 1991. Automatic vigilance: the attention-grabbing power of negative social information. *Journal of Personality and Social Psychology* **61**. 380–391.
- Prohászka J. 1953. A nagyítás és a túlzás kifejezésmódjai nyelvünkben. *Magyar Nyelvőr* **77**. 196–208.
- Prószéky G.–Olaszy G.–Váradi T. 2003. Nyelvtechnológia. In Kiefer F. szerk. *A magyar nyelv kézikönyve*. Budapest, Akadémiai.
- Pustejovsky, J.–Boguraev, B. 1993. Lexical knowledge representation and natural language processing. *Artificial Intelligence* **63**. 193–223.
- Pustejovsky, J.–Boguraev, B.–Johnson, M. 1995. *A core lexical engine: The contextual determination of word sense*. Technical report. Waltham, MA, Department of Computer Science, Brandeis University.

- Pustejovsky, J.–Boguraev, B. szerk. 1996. *Lexical semantics. The problem of polysemy*. Oxford, Clarendon Press.
- Quine, W.V.O. 1960. *Word and Object*. Cambridge, Mass, MIT Press.
- Qiu, G.–Liu, B.–Bu, J.–Chen, C. 2009. *Expanding Domain Sentiment Lexicon through Double Propagation*.
(https://www.researchgate.net/profile/Bing_Liu41/publication/220813594_Expanding_Domain_Sentiment_Lexicon_through_Double_Propagation/links/5472bbf90cf24bc8ea199f8a.pdf)
- Quirk, R.–Greenbaum, S.–Leech, G.–Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. Harlow, Longman.
- Rahmati, F. 2015. Semantic shift, homonyms, synonyms and auto-antonyms. *WALLA journal* **31**(S3). 81–85.
- Rakova, M.–Pethő G.–Rákosi Cs. szerk. 2007. The cognitive basis of polysemy. New sources of evidence for theories of word meaning. *MetaLinguistica* **19**. Frankfurt am Main, Peter Lang.
- Reboul, A.–Moeschler, J. 2000. *A társalgás cselei*. Budapest, Osiris.
- Rosenzweig, M.R. 1970. International Kent-Rosanoff word association norms, emphasizing those of French male and female students and French workmen. In Postman, L.–Keppel, G. eds. *Norms of Word Association*. New York, Academic Press. 95–176.
- Rozin, P.–Fallon, A.E. 1987. A perspective on disgust. *Psychological Review* **94**(1). 23–41.
- Rozin, P.–Haidt, J.–McCauley, C.R. 2000. Disgust. In Lewis, M.–Haviland-Jones, J. M. eds. *Handbook of Emotions*. 2nd edition. New York, London, The Guilford Press. 637–653.
- Rumelhart, D.E.–McClelland, J.L.–the PDP Research Group 1986. Parallel Distributed Processing. *Explorations in the Microstructure of Cognition* **1**. Foundations. MA, USA, MIT Press/Bradford Books.
- Ruppenhofer, J.–Wiegand, M.–Brandes, J. 2014. Comparing methods for deriving intensity scores for adjectives. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* **2**. Short Papers. Gothenburg, Sweden, Association for Computational Linguistics. 117–122.
- Ruppenhofer J.–Rehbein, I. 2012. *Anchoring sentiment analysis in semantic frames*.
(<http://www.uni-hildesheim.de/ruppenhofer/pubs/longversion.pdf>)
- Sadikov E.–Parameswaran A.–Venetis P. 2009. Blogs as predictors of movie success. In *International AAAI Conference on Weblogs and Social Media (ICWSM)*. Association for the Advancement of Artificial Intelligence, Palo Alto, CA. 304–307.

- Salveti, F.–Lewis, S.–Reichenbach, C. 2004. Impact of lexical filtering on overall opinion polarity identification. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*. Stanford, CA. (<http://www-plan.cs.colorado.edu/reichenb/papers/AAAI-SSS04.pdf>)
- Sass B. 2017. Keresés korpuszban: a kibővített Magyar történeti szövegtár új keresőfelülete. In: Forgács T.–Németh M.–Sinkovics B. szerk. *A nyelvtörténeti kutatások újabb eredményei IX*. Szeged, SZTE Magyar Nyelvészeti Tanszék. 267–277.
- Schachter, S.–Singer, J. 1962. Cognitive, Social, and Physiological Determinants of Emotional State. *Psychological Review* **69**. 379–399.
- Shmelev, A. 2012. Cognitive and Communicative Sources of Enantiosemy. In *Proceedings of the 10th World Congress of the International Association for Semiotic Studies*. Universidade da Coruña. 837–844.
- Searle, J.R. 1980. The Background of Meaning. In Searle, J.R.–Kiefer, F.–Bierwisch, M. eds. *Speech Act Theory and Pragmatics Volume 10 of the series Texts and Studies in Linguistics and Philosophy* **10**. 221–232.
- Shoukry, A.–Rafea, A. 2012. Sentence-level Arabic sentiment analysis. In *Proceedings of international conference on collaboration technologies and systems*. IEEE Xplore Press. 546–550.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford, Oxford University Press.
- Sinclair, J. 1992. Trust the Text: the Implications are Daunting. In Davies, M.–Ravelli, L. eds. *Advances in Systemic Linguistics*. London, Pinter. 5–19.
- Socher, R.–Pereygin, A.–Wu, J.Y.–Chuang, J.–Manning, C.D.–Ng, A.Y.–Potts, C. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank Stanford University. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Stanford, USA. 1631–1642.
- Söderman, T. 1993. Word association of foreign language learners and native speakers: the phenomenon of a shift in response type and its relevance for lexical development. In Ringbom H. ed. *Near-native proficiency in English*. Stockholm, Åbo. 91–182.
- Spitzer, M. 1996. *Geist im Netz*. Heidelberg–Berlin, Spektrum.
- Stone, P.–Dexter, J.–Dunphy, C.–Smith, M.S.–Ogilvie, D.M. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA MIT Press.
- Stubbs, M. 2002. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford, Blackwell Publishing.
- Sweetser, E. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge, Cambridge University Press.

- Szabó M.K. 2014. *Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai*. A Magyar Tudományos Akadémia Szegedi Akadémiai Bizottsága és „A Tudomány Támogatásáért a Dél-Alföldön” Alapítvány pályázata – Oklevél-díj.
- Szabó M.K. 2015a. A polaritásváltás- és változás kezelési lehetőségei a szentimentelemzésben. *Tavaszi Szél konferencia konferenciakötete*. Líceum Kiadó, Eger és Doktoranduszok Országos Szövetsége, Budapest. 629–643.
- Szabó M.K. 2015b. Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai és dilemmái. In Gecső T.–Sárdi Cs. szerk. *Nyelv, kultúra, társadalom. Segédkönyvek a nyelvészet tanulmányozásához* 177. 278–285.
- Szabó M.K. 2015c. A polaritásváltás problémája a szentimentelemzés szempontjából. In Váradi T. szerk. *IX. Alkalmazott Nyelvészeti Doktoranduszkonferencia. Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből*. MTA Nyelvtudományi Intézet. 51–61.
- Szabó M.K. 2016. A nyelvi értékelés mibenlétének kérdése a számítógépes értékelélemzés (szentimentelemzés) szempontjából. In Gécseg Zs. szerk. *LingDok 15. Nyelvésztdoktoranduszok dolgozatai*. Szeged, Szegedi Tudományegyetem, Nyelvtudományi Doktori Iskola. 153–172.
- Szabó M.K. 2017. Az értékvesztés és az értékvtáltás jelenségeinek vizsgálata a pszichológiai, valamint a szemantikai motiváció szempontjából. *MANYE 26. konferenciakötete*. Megjelenés előtt.
- Szabó M.K.–Morvay G. 2015. Emócióelemzés magyar nyelvű szövegeken. In Gecső T.–Sárdi Cs. *Nyelv, kultúra, társadalom*. 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* (MSZNY 2015). Szegedi Tudományegyetem, Szeged. 219–226.
- Szabó M.K.–Vincze V. 2016. A negáló szentimentshifterek kezelési kérdései a magyar nyelvű szövegek számítógépes értékelélemzési feladatában. In Reményi A.Á.–Sárdi Cs.–Tóth Zs. szerk. *Távlatok a mai magyar alkalmazott nyelvészetben*. Budapest, Tinta. 272–281.
- Szabó M.K.–Galántai J. 2017. Egy magyar nyelvű spontán beszélt nyelvi korpusz (HuTongue) létrehozásának tapasztalatai. *MANYE 26. konferenciakötete*. Megjelenés előtt.
- Szabó M.K.–Vincze V.–Nagy T.I. 2012. HunOr: A Hungarian–Russian Parallel Corpus. In *Proceedings of the 8th International Conference on Language Resources and Evaluation* (LREC-2012). Istanbul, Turkey. 2453–2458.
- Szabó M.K.–Vincze V.–Hangya V. 2016a. Aspektusszintű annotáció és szentimentet módosító elemek egy magyar nyelvű szentimentkorpuszban. In Tanács A.–Varga V.–Vincze V. szerk. *XII. Magyar Számítógépes Nyelvészeti Konferencia* (MSZNY 2016). Szeged, Szegedi Tudományegyetem. 174–182.

- Szabó M.K.–Vincze V.–Simkó K.–Varga V.–Hangya V. 2016b. A Hungarian Sentiment Corpus Manually Annotated at Aspect Level. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portoroz, Szlovénia, European Language Resources Association (ELRA). 2873–2878.
- Szabó M.K.–Vincze V.–Morvay G. 2016c. *Magyar nyelvű szövegek emócióelemzésének elméleti nyelvészeti és nyelvtechnológiai problémái*. In. Távlatok a mai magyar alkalmazott nyelvészetben. Budapest, Tinta. 282–292.
- Szabó M.K.–Nyíri Zs.–Morvay G.–Lázár B. 2017a. A szentimentérték módosulásának vizsgálata szemantikai-pragmatikai szempontból annotált korpuszon. In Vincze V. szerk. *XIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szegedi Tudományegyetem, Informatikai Intézet. 251–262.
- Szabó M.K.–Lázár B.–Nyíri Zs.–Morvay G. 2017b. A negatív emotív elemek vizsgálata a nemek közötti nyelvhasználati különbségek szempontjából. In Ludányi Zsófia szerk. *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből*. Budapest, MTA Nyelvtudományi Intézet. 161–176.
- Szabó M.K.–Nyíri Zs.–Lázár B. 2017c. Negatív emotív szemantikai tartalmú fokozó elemek vizsgálata az angol–oros és orosz–angol fordítás tükrében. In Ludányi Zsófia szerk. *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből*. Budapest, MTA Nyelvtudományi Intézet. 177–191.
- Szabó M.K.–Ilyés V.–Nyíri Zs. 2017. A negatív emotív elemek szóasszociációs adatbázisa. *MANYE 26. konferenciakötete*. Megjelenés előtt.
- Szalay, L.B.–Deese, J. 1978. *Subjective Meaning and Culture: An Assessment Through Word Associations*. Hillsdale, Lawrence Erlbaum.
- 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.
- Székely G. 2006. Adalékok a női és a férfi kommunikatív kompetencia jellegzetességeinek leírásához. In Gecső T. szerk. *Nyelvi kompetencia – kommunikatív kompetencia*. KJF – Budapest, Tinta. 330–334.
- Székely G. 2007. Egy sajátos nyelvi jelenség, a fokozás. *Segédkönyvek a nyelvészet tanulmányozásához* **66**. Budapest, Tinta.
- Szerdahelyi I. 1995. *Irodalomelméleti enciklopédia*. Budapest, Eötvös József Könyvkiadó.
- Szilágyi N.S. 1996. *Hogyan teremtsünk világot?* Kolozsvár, Erdélyi Tankönyvtanács.
- Szirmai M. 2005. *Bevezetés a korpusznyelvészetbe*. Budapest, Tinta.
- Taboada, M.–Grieve, J. 2004. Analyzing appraisal automatically. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text*. Stanford. 158–161.

- Taboada, M.–Brooke, J.–Tofiloski, M.–Voll, K.–Stede, M. 2011. *Lexicon-Based Methods for Sentiment Analysis* (http://www.mitpressjournals.org/doi/pdfplus/10.1162/COLI_a_00049)
- Taylor, J.R. 1994. The two-level approach to meaning. *Linguistische Berichte* **149**. 3–28.
- Tolcsvai N.G. 1988. A mai magyar nyelv normarendszerének egy jelentős változásáról az „ifjúsági nyelv” kapcsán. *Magyar Nyelvőr* **112**(4). 398–406.
- Tolnai T. 2011. *Twitter az oktatásban (?)*. NLP Meetup - Számítógépes Nyelvészet. 2011. szeptember 22. (<http://szamitogepesnyelveszet.blogspot.hu/2011/09/twitter-az-oktatásban.html>)
- Tompa J. szerk. 1970. *A mai magyar nyelv rendszere*. Leíró nyelvtan I.-II. Budapest, Akadémiai.
- Tong, R.M. 2001. An operational system for detecting and tracking opinions in online discussions. In *Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification*. New York, NY. 1–6.
- Tukacs, T. 2015. *Túlzásba vitt szavak. A fokozó értelmű szókapcsolatok magyar–angol szótára*. Budapest, Tinta.
- Turney, P. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of 40th Meeting of the Association for Computational Linguistics*. Philadelphia, PA. 417–424.
- Turney, P.–Littman, M. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems* **21**(4). 315–346.
- Umar, A.M. 1982. *Ilm al-adillah*. Kuwait, Maktabah Dar al-Arabiya.
- Vázquez, S.–Bel, N. 2012. A Classification of Adjectives for Polarity Lexicons Enhancement. In Calzolari, N.–Choukri, K.–Declerck, T. et al. eds. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)*. Paris, ELRA. 3557–3561.
- Vechtomova, O.–Suleman, K.–Thomas J. 2014. An Information Retrieval-Based Approach to Determining Contextual Opinion Polarity of Words. In *Proceedings of the 36th European Conference on Information Retrieval (ECIR)*, Amsterdam. Springer. 553–559.
- Vegnaduzzo S. 2004. Acquisition of subjective adjectives with limited resources. In *AAAI Spring Symposium Technical Report: Exploring Affect and Attitude in Text*. (<http://www.aaai.org/Papers/Symposia/Spring/2004/SS-04-07/SS04-07-033.pdf>)
- Véronis, J. 2003. Sense tagging: does it make sense? In Wilson, A.–Rayson, P.–McEnery, T. eds. *Corpus linguistics by the lune: a festschrift for Geoffrey Leech*. Frankfurt, Peter Lang.

- Vincze V. 2011. *Semi-Compositional Noun + Verb Constructions: Theoretical Questions and Computational Linguistic Analyses*. PhD thesis. University of Szeged, August 2011.
- 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, Japan, 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*. Szeged, Szegedi Tudományegyetem, Informatikai Tanszékcsoport. 99–108.
- Vincze V. 2014b. Uncertainty Detection in Hungarian Texts. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014)*. Dublin. 1844–1853.
- Vincze V. 2016. Előadásjegyzet. (<http://www.inf.u-szeged.hu/~vinczev/oktatas.html>)
- Vincze V.–Zsibrita J.–Durst P.–Szabó M.K. 2013. HunLearner: a magyar nyelv nyelvtanulói korpusza. In Tanács A.–Vincze V. szerk. *IX. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, Szegedi Tudományegyetem. 97–105.
- Vincze, V.–Hegedűs, K.–Berend, G.–Farkas, R. 2016. Telltale Trips: Personality Traits in Travel Blogs. In *Emotion and Sentiment Analysis workshop*. Portorož, Slovenia.
- Wang, B.–Wang, H. 2007. Bootstrapping both product properties and opinion words from Chinese reviews with cross-training. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*. Washington, DC, USA, IEEE Computer Society. 259–262.
- Watts, D.J. 1999. *Small Worlds*. Princeton, Oxford, Princeton University Press.
- Watts, D.J. 2004. The “New” Science of Networks. *Annual Review of Sociology* **30**. 243–270.
- Watts, D.J.–Strogatz, S.H. 1998. Collective dynamics of ‘small-world’ networks. *Nature* **393**. 440–442.
- Weber, R.P. 1990. *Basic Content Analysis*. Thousand Oaks, California, Sage.
- Westbury, C.F.–Hollis, G. 2007. Putting Humpty Together Again: Synthetic Approaches to Nonlinear Variable Effects Underlying Lexical Access. In Jarema, G.–Libben, G eds. *The Mental Lexicon*. Amsterdam, Elsevier. 7–31.
- Wiebe, J.–Bruce, R.–O’Hara, T. 1999. Development and use of a gold standard data set for subjectivity classifications. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL-99)*. University of Maryland. 246–253.

- Wierzbicka, A. 2002. Australian cultural scripts – *bloody* revisited. *Journal of Pragmatics* **34**(9). 1167–1209.
- Wilensky, R. 1987. *Some Problems and Proposals for Knowledge Representation*. Technical Report. EECS Department, University of California, Berkeley.
- Wilensky, R. 1989. Primal content and actual content an antidote to literal meaning. *Journal of Pragmatics* **13**(2). 163–186.
- Wilson, D. 2003. Relevance and Lexical Pragmatics. *Italian Journal of Linguistics* **15**(2). 273–291.
- Wilson, D.–Sperber, D. 2004. Relevance Theory. In Ward, G.–Horn, L. eds. *Handbook of Pragmatics*. Oxford, Blackwell. 607–632.
- Wilson, T.–Wiebe, J.–Hoffmann, P. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP2005)*. (<https://people.cs.pitt.edu/~wiebe/pubs/papers/emnlp05polarity.pdf>)
- Wilson, T.–Wiebe, J.–Hoffman, P. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics* **35**(3). 399–433.
- Young, L.–Soroka, S. 2012. Affective news: The automated coding of sentiment in political texts. *Political Communication* **29**(2). 205–231.
- Zsibrita J.–Vincze V.–Farkas R. 2013. magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian. In *Proceedings of RANLP-2013*. Hissar, Bulgaria. 763–771.
- Витехновский, В.И. 2013. Автоматизация определения тональности текста. *Вестник магистратуры* **5**(20). 34–36.
- Вознесенская, М.М. 2011. Энантисемиа в русской фразеологии. In Кибрик, А.Е.–Беликов, В.И.–Богуславский, И.М.–Добров, Б.В.–Добровольский Д.О. ред. *Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог»*. Москва, РГГУ. 759–767.
- Елькина, О.М. 2009. Блог как средство асинхронной веб-коммуникации. *Университетские чтения – 2009: материалы научно-методических чтений ПГЛУ*. Пятигорск, ПГЛУ. 180–184. (http://www.pglu.ru/upload/iblock/8dd/uch_2009_v_00035.pdf)
- Копотев, М.В.–Гурин, Г.Б. 2006. Принципы синтаксической разметки Хельсинкского аннотированного корпуса русских текстов ХАНКО. In Копотев, М.В.–Гурин, Г.Б. ред. *Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции «Диалог-2006»*. Москва, РГГУ. 280–284.

Национальный корпус русского языка (НКРЯ). Институт русского языка им. В. В. Виноградова РАН, Россия. (ruscorpora.ru/).

Прохоров, А.–Керимов, А. 2012. *Сентимент-анализ и продвижение в социальных медиа*. (<http://compress.ru/Article.aspx?id=23115>)

Сабо М.К. 2014. Контрастивный анализ анафорических местоимений в венгерских и русских сложных предложениях. *Dissertationes Slavicae. Sectio linguistica XXX*. Szeged. 125–150.

Шаров С. А. 2003. Представительный корпус русского языка в контексте мирового опыта. In Шаров, С.А. ред. *Научно-техническая информация* (НТИ) 2(6). 9–17.

FORRÁSJEGYZÉK

babaszoba.hu

boldoguljgyorben.blog.hu

cotcot.hu/cikk/cimke/stoppolas

[Dívány \(divany.hu\)](http://divany.divany.hu)

edzesonline.hu/blog

esettanulmanyok.hu

facebook.hu/com

Farönk királyfi / Tibeti legendák és mesék, fordította: Róna Tas András 1958

Frányó Kriszta 2016. *Legény(álom)lak*. Publio Kiadó Kft.

gamestar.hu

gyakorikerdesek.hu

hargitanepe.eu

highlightsofhungary.hu

hirek.ma/hirek

ilajkyou.hu

index.hu

ize.hu

juniorrezervatum.hu

Krúdy Gyula 1928. *A kandúr*

Lackfi János 2014. *Egészségedre, Sándor!*

laptophardware.hu/blog

linguee.de

Loughran–McDonald Financial Sentiment Dictionary

(http://www3.nd.edu/~mcdonald/Word_Lists.html)

MTA SZTAKI online szótár (<http://szotar.sztaki.hu/angol-magyar>)

Magyar Elektronikus Könyvtár (<http://mek.oszk.hu/>)

Magyar Nemzeti Szövegtár 2 (MNSZ2) (<http://clara.nytud.hu/mnsz2-dev/>)

Magyar Történeti Szövegtár (<http://www.nytud.hu/oszt/lexi/index.html>).

Magyarország online szinonimaszótára (szinonimaszotar.hu)

minuszos.hu

mobilarena.hu

moly.hu

napidroid.hu

Népújság, 1977. január 1, 1. szám, 28. évfolyam

Neticle Technologies (<https://neticle.hu/hogyan-mukodik-a-neticle-social-listening.html#>).
nyugattolkeletig.ipolyerdo.hu/
OpinHuBank (<https://sites.google.com/site/mmihaltz/resources>)
pellagriasdiary.blogspot.hu
petofilive.hu
Poet szinonimaszótár (<https://szinonimaszotar.poet.hu/>)
port.hu
rockstation.blog.hu
sci-fi.mandiner.hu
So-CAL
(<https://github.com/DrOttenssooser/BiblicalNLPworks/tree/master/SkyDrive/NLP/CommonWorks/Data/Opion-Lexicon-English/SO-CAL>)
sorozatjunkie.hu
sorozatwiki.hu
Stanford Sentiment Treebank (<http://nlp.stanford.edu/sentiment/treebank.html>)
starity.hu
Stendhal: Vörös és fekete, fordította: Illés Endre 2013
Subjectivity Lexicon (http://mpqa.cs.pitt.edu/#subj_lexicon)
szeretlekmagyarorszag.hu
szilveszter.hu
szuloklapja.hu
tech2.hu
the6thousandmileblog.wordpress.com/
totaldamage.hu
trendalelke.hu
Twitter (twitter.com)
Twitter magyar oldala (twitter.com/?lang=hu)
Vass Tamás 1932. *Élvemaradottak útja / Az utolsó hadgyakorlat a Bácskában*
velvet.hu/celeb
WordNet (<https://wordnet.princeton.edu/>)
24.hu

FELHASZNÁLT ELEMZŐ ESZKÖZÖK

Brat online annotáló eszköz (<http://brat.nlplab.org/>)
magyarlanc elemző eszköz: Zsibrita, J.–Vincze, V.–Farkas, R. 2013. magyarlanc: A Toolkit for Morphological and Dependency Parsing of Hungarian. In *Proceedings of RANLP 2013*. 763–771. (<http://www.inf.u-szeged.hu/rgai/magyarlanc>)

MELLÉKLETEK

1. melléklet: Az értékvesztésre – és esetlegesen értékváltásra is – képes elemek listája^{138,139}

állat	éktelen	irtózat	őrült
állati	éktelenül	irtózatosan	őrülten
állatira	elképesztő	istentelen	piszkos
átkozott	elképesztően	istentelenül	piszkosul
átkozottul	embertelen	iszonyatos	piszok
baromi	embertelenül	iszonyatosan	piszokul
baromian	eszméletlen	iszonyú	pokoli
baromira	észtvesztő	iszonyúan	pokolian
baszott	észtvesztően	<i>katasztrofális</i>	pusztító
baszottul	észvesztő	<i>katasztrofálisan</i>	pusztítóan
bazi	észvesztően	kegyetlen	rém
bazira	félelmetes	kegyetlenül	rémes
beteg	félelmetesen	kibaszott	rémesen
<i>beteges</i>	förtelmes	kibaszottul	rettenetes
<i>betegesen</i>	förtelmesen	könyörtelen	rettenetesen
bődületes	geci	könyörtelenül	rettentő
bődületesen	gecire	kurva	rettentően
borzalmas	halál	kurvára	rohad
borzalmasan	haláli	marha	rohadul
borzasztó	halálian	marhára	sokkolóan
borzasztóan	halálos	megdöbbentő	szédítő
brutális	halálosan	megdöbbentően	szédítően
brutálisan	hulla	meghökkenítő	szédületes
büdös	irdatlan	meghökkenítően	szédületesen
büdösül	irdatlanul	mocskos	szörnyen
disznó	irgalmatlan	mocskosan	szörnyűséges
dög	irgalmatlanul	mocskosul	szörnyűségesen
durva	irtó	őrületes	veszett
durván	irtóra	őrületesen	veszettül

¹³⁸ A listában azokat az elemeket, amelyek nem csupán értékvesztők, de értékváltók is lehetnek, félkövérrel emeltem ki.

¹³⁹ A listát a dolgozat bírálója, Sass Bálint megjegyzései alapján a vizsgálatok elvégzése után 116 elemből állóra bővítettem. Köszönöm az értékes javaslatokat!

2. melléklet: Minta a pozitív szentimentszótárból

adományoz	angyal	barát	baró
agilis	ápol	baráti	bársonyos
agilitás	aranybánya	barátság	bátor
ajándék	aranyérmes	barátságos	bátorít
ajándékozás	aranyos	barátságosság	bátorítás
akadályozatlan	ari	bebiztosít	bátorító
akkurátus	aromás	becéz	bátorítóan
alapos	ász	becézget	bátorság
alaposság	átélés	becsszó	békés
alázatos	átgondolt	becsülendő	békésen
alázatosság	átkarolás	becsület	békesség
áld	atomzsír	becsületes	békeszerető
áldás	áttekinthető	becsületesség	béketűrés
áldott	áttekinthetőség	becsületszó	békülékeny
áldozatkész	attraktív	befogadó	beleszeret
áldozatos	átvészel	béke	bensőséges
alkalmas	autonóm	békeidők	bizakodás
alkalmasság	báj	békekötés	bizakodó
ambiciózus	bajnok	békéltető	bizalomgerjesztő

3. melléklet: Minta a negatív szentimentszótárból

abajgat	aggodalom	agyonadminisztrált	alárúg
abajgós	aggódás	agyonadósodott	alátámaszthatatlan
abajog	aggódik	agyonajnároz	alattomos
aberrált	aggódó	agyonajnározás	alattomosan
abnormális	aggszűz	agyonajnározó	alattomosság
abszurdum	agonizál	akadozik	alávág
ácsingózik	agresszió	akadozó	alávaló
ádáz	agresszív	akaratos	alávetettség
ádázul	agresszivitás	akarnok	áldozat
addiktív	agresszor	akasztófa	aljas
adócsaló	ágrólszakadt	állhatatlan	aljasság
adósság	agyalágyult	állott	aljasul
áfacsaló	agyatlan	alábecsül	alkalmatlan
aggály	agyatlanság	alábecsülik	alkalmatlankodik
aggályos	agybajos	aláígér	alkalmazhatatlan
aggályoskodik	agyhalott	alakoskodó	alkesz
aggályoskodó	agyhúgykő	alamuszi	alkotmányellenes
aggaszt	agymenés	alantas	akadékoskodó
aggasztó	agymosás	alantasság	akadozik
aggodalmasan	agyonabajgat	alaptalan	akadozó
aggodalmaskodik	agyonadóztat	alapzaj	akarnok

4. melléklet: A Dívány korpusz annotálási útmutatója

A szentimentelemzés során azokat a szövegrészeket akarjuk megtalálni és jelölni a szövegekben, amelyek valamiféle értékelést fejeznek ki egy bizonyos személy, esemény, tárgy stb. vonatkozásában. Amire vonatkoznak, azt targetnek, tehát a szentiment célpontjának fogjuk nevezni, hiszen az értékelés erre irányul.

Az (1) alatti mondat pozitív értékelést fogalmaz meg a Creator nevű táplálékkiegészítő termékkel szemben:

(1) A Creator nevű táplálékkiegészítő nagyon hatékony.

A szentimentelemzést a fenti mondat esetében a következőképpen hajtánánk végre:

(2) <PosFrag> A <target> Creator nevű táplálékkiegészítő </target> nagyon hatékony </PosFrag>.

A (2)-ben A Creator nevű táplálékkiegészítő *target*-címkét kap, mivel ez a megfogalmazott értékelés „célpontja”. A teljes értékelő kifejezés, azaz a *szentimentfragmentum* pedig kap egy *PosFrag*-címkét, mivel pozitív szentimentet fogalmaz meg.

Az annotálni kívánt korpusz termékvéleményekből áll. A termékvélemények egy termék megnevezéséből, majd azt követő rövid véleményekből állnak. A korpusz e sajátága miatt a következő annotációs kiegészítést alkalmazzuk: *Topic*-címkét kap maga a címszerű terméknév, amely minden esetben tulajdonnév! *Target1*-címkét az adott szövegben véleményezett termék. Target 1-20 címkét az adott target adott aspektusa, pl. *az íze fanyar; a színe szép; kellemes az illata* stb. Ez utóbbi címkéket egy adott véleménysszövegen belül végig következetesen alkalmazzuk, tehát ugyanazt a címkét kapja ugyanaz az aspektus.

A fragmentum az a minimális egység, amely egy véleményt fogalmaz meg egy target vonatkozásában (kivéve azokban a ritka esetekben, ha felsorolás van).

Egyetlen mondat több fragmentumot is tartalmazhat; pl.

(3) A Creator nevű táplálékkiegészítő nagyon hatékony, a vitaminok pedig tovább javítanak a teljesítményen.

Mindezzel az annotációnk azonban még nem teljes. Be kell jelölnünk ugyanis a teljes értékelő szövegrészen belül magát azt az elemet, amely önmagában, a lexika szintjén szentimentértéket hordoz. Ez az elem az (1) alatti példában a *hatékony* szó.

A szentimentértékeket a kontextus függvényében határozzuk meg.

Természetesen az (1) alatti példában a targetról megfogalmazott értékelés a teljes *nagyon hatékony* kifejezés. Azonban amikor a minimális egységet keressük, akkor kizárólag azokat az elemeket akarjuk megtalálni és bejelölni a pusztán szentimentkifejezésként, amelyről úgy véljük, hogy a szemantikai tartalmában benne foglaltatik egy értékelő jelentéskomponens. A *hatékony* szó pedig önmagában, kontextustól függetlenül magában hordozza a pozitív értékelés szemantikai komponensét. A szentimentkifejezés fogalmát tágran értelmezzük, ezért nem csupán (a klasszikusnak tekinthető) melléknévi elemet annotáljuk szentimentként, hanem azokat a főneveket, határozószókat és igéket is, amelynek az aktuális kontextusban szentimentértéket tudunk

tulajdonítani, tehát értékelést fejez ki az adott szövegben; pl. *öldökölve; agyonver; zseni*. 3) Emellett nem csupán egy szóból álló, hanem többszavas kifejezéseket is jelölhetünk szentimentként, amennyiben azok állandósult szókapcsolatok. Pl. *az egyik kutya, másik eb* kifejezést egy az egyben egy darab szentimentkifejezésnek annotálnánk a szövegben.

A *nagyon* csupán módosítja azt az értékelő jelentést, amelyet a *hatékony* hordoz. A *nagyon* egy olyan elem, amely növeli, intenzívebbé teszi az általa módosított elem szemantikai tartalmát. A pusztá szentimentkifejezések módosító elemeit *sentimentsiftereknek* nevezzük. A *sentimentsifterek* közé tartoznak tehát a *hatékony*-hoz hasonló elemek, amelyek az adott kifejezés prior értékét valamilyen mértékben és irányban módosítják. Azokat az elemeket, amelyek a *nagyon*-hoz hasonlóan viselkednek, (az angol terminológia alapján) *intenzifikálóknak* nevezzünk. Az *intenzifikálóknak* két típusát különböztetjük meg: a *növelőket* és a *csökkentőket*. Mindezt az alábbi táblázat mutatja be, példákkal:

	típusai	definíció	példák	elemek
a szentimentérték intenzifikálói	növelők	az adott kifejezés prior értékét valamilyen mértékben és irányban módosítják	<i>hihetetlenül kényelmes az ülés</i>	<i>rendkívül, rendkívüli módon, nagyon, borzasztóan, elképesztően stb.</i>
	csökkentők		<i>aránylag kényelmes az ülés</i>	<i>valamennyire, valamelyest, feliből-nagyjából, részben, kevésbé stb.</i>

Az (1) alatti példát tehát a következőképpen kellene a rendszerünk szerint annotálni:

(5) <PosFrag> A <target> Creator nevű táplálékkiegészítő </target> <IntPlus> nagyon </IntPlus> <PosSent> hatékony</PosSent> </PosFrag>.

A *sentimentsiftereknek* van azonban egy másik típusa is, amely nem csupán annyit tesz, mint az *intenzifikálók*, tehát, hogy valamelyest változtat az értéken, hanem vagy törli az értéket, vagy az ellenkezőjére változtatja azt; pl.

(6) A Creator nevű táplálékkiegészítő **nem** hatékony.

Ezeket az elemeket negálóknak nevezzük, és az alábbi táblázatban jellemezzük:

	definíció	példa	elemek
a szentimentérték negálói	az adott kifejezés prior értékét vagy az ellenkezőjére változtatják, vagy törlik	<i>nem</i> szép; szépségét <i>vesztett</i> ; <i>aligha</i> szép; szépség <i>nélkül</i> stb.	hiányában, dehogy, vesztett, nélkül, aligha, látszatra stb.

A negáló funkcióját tagadószó, névutó, névutómelléknév stb. egyaránt betöltheti.

A szentimentelemzés szempontjából az elemzett szöveg szerzőjének elkötelezettségi foka a proposíciós tartalom iránt nagy jelentőséggel bír. A problémát Kiefer (1986) alapján szemléltetjük:

- (8) a. Pisti beteg.
b. Sajnos, Pisti beteg.
c. Hallottad már, hogy Pisti beteg?
d. Nahát, Pisti beteg!
- (9) a. Pisti talán beteg.
b. Pisti biztosan beteg.
c. Pisti valószínűleg beteg.
d. Valószínű, hogy Pisti beteg. d. Pisti beteg?
e. Vajon beteg Pisti?
f. Pisti betegeskedhet.

Amíg a (8) alatti mondatok esetében a szerző elkötelezi magát a proposíciós tartalom igazsága iránt, addig az (9) alattiakban nem; azok az elkötelezettségnek csupán valamilyen fokozatára utalnak. Az erre utaló elemeket *irreálóknak* nevezzük, és külön taggel látjuk el a korpuszban.

Amennyiben az irreálás jelenségét egy szó egyetlen morféma fejezi ki, a teljes szót irreálóként annotáljuk (technikai okokból); pl. *kedves-e*; *gyárthat*.

	definíció	példák	elemek
irrealiók	részben vagy teljesen lehetetlenné teszik az adott szövegrész faktív olvasatát	- A Samsung jó televíziókat gyárt. [+] - <u>Lehet</u> , hogy a Samsung jó televíziókat gyárt. [0] - <u>Gyártson</u> jó televíziókat a Samsung! [0] - A Samsung jó televíziókat <u>gyárthat</u> . [0]	modális határozószók, modális módosítószók; modális segédigék; módosító mondatrészek; kérdőszók; óhajtószók; <i>-hat/-het</i> inflexiós toldalék, a feltételes mód <i>-na/-ne</i> és <i>-ná/-né</i> jelei stb. (vö. Kiefer 2005).

5. melléklet: A NegEmotív-korpusz annotálási útmutatója

A feladatban a szentimentérték módosulásának bizonyos típusait annotáljuk. Arról hoznak döntést az annotátorok, hogy az adott tweet adott, vizsgált eleme aktuálisan milyen szemantikai-pragmatikai szerepben van. Két altípust különböztetünk meg:

1) **értékváltás**: a vizsgált elem az aktuális kontextusban önmagában szentimentkifejezés funkciójában áll, és vele a tweet szerzője nem negatív értékelést szeretne kifejezni, pl.

- (1) a. ismét **brutális** koncertet adott az énekes
b. **durva**, ahogy ez a fickó rajzol

Ebben az esetben nem csak az értékváltó elemet, hanem az értékelés targetét is annotáljuk. A fenti példák esetében a targeteket az alábbiakban aláhúzással jelölöm:

- (2) a. ismét **brutális** koncertet adott az énekes
b. **durva**, ahogy ez a fickó rajzol

Ha a vizsgált elem negatív értékelést fejez ki, nem annotáljuk.

Megfigyelésem szerint csak olyan verzió lehetséges, ahol a lexikai szinten negatív elem pozitív értékelést fejez ki. A fordított eset nem lehetséges, tehát az, amikor a lexikai szinten pozitív elem negatív ítéletet fejez ki. Amennyiben ilyet látunk a tweetekben, akkor az valószínűleg ironia vagy a gúny, azzal ebben a kutatásban nem foglalkozunk, ezért nem annotáljuk.

2) **Értékvesztés:** egy lexikai szinten negatív elem fokozó funkcióját tölti be egy az aktuális kontextusban értékelő funkciót betöltő elem mellett, pl.

- (3) a. **brutálisan** jó volt a tegnapi esti buli
b. mostanában **borzasztó** csúnyán beszél velem

Ilyenkor bejelöljük az értékvesztő elemet, és azt a fejezést is, amelyet módosít. A fenti példák esetében a módosított elemeket az alábbiakban aláhúzással jelölöm:

- (4) a. **brutálisan** jó volt a tegnapi esti buli
b. mostanában **borzasztó** csúnyán beszél velem

Azt is annotáljuk, hogy a módosított elem (a kifejezés alaptagja) lexikai szinten pozitív (pl. *szép*), negatív (pl. *rossz*), vagy semleges-e (pl. *nagy*, *kemény*).

Továbbá azt is megállapítjuk, hogy az aktuális kontextusban a fokozó elem milyen szemantikai-pragmatikai tartalmat hordoz: deszemantizált, negatív vagy pozitív, pl.

- (4) a. **Brutál** ügyik voltunk ma is ☹️
b. Szerintem ez egyszerűen **borzasztó** hangos... Ti nem így gondoljátok?
c. **kegyetlen** helyes, egyszerűen kész vagyok tőle

Bizonytalanság esetén, mind az alaptagnál, mint a módosító elemnél lehetséges egy *egyéb*-taget alkalmazni.

6. melléklet: A szóasszociációs adatfelvételben alkalmazott kérdőív felépítése

I. Szóasszociációs kérdőív

Kedves Kitöltő!

A következő kérdőívvel egy tudományos kutatómunkához veszünk fel adatokat.

A kérdőív kitöltésével Ön Szabó Martina Katalin doktori kutatómunkáját támogatja. Az adatfelvételt az Új Nemzeti Kiválóság Program finanszírozásával a Precognox Informatikai kft. végzi.

A kérdőív névtelen és a válaszadás önkéntes. Nincsenek jó vagy rossz válaszok! A véleményére vagyunk kíváncsiak!

Köszönjük a segítségét!

II. Útmutató

A feladatsorban különböző vizuális feladatok és szavak fogják váltani egymást.

Ha szót lát, azt szeretnénk, hogy írja be azokat a szavakat, amelyek elsőként eszébe jutnak.

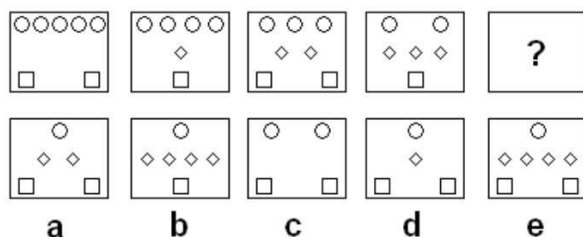
A szavak alatt 5 üres sort lát majd, kérjük az egyes asszociációit (külön-külön) ezekbe a sorokba írja be.

Amennyiben feladatot lát, kérjük, kövesse a feladatokhoz tartozó instrukciókat!

III. Hívószók és vizuális feladatok váltakozva

Két példa a feladatok közül:

1) Melyik megoldás illik a kérdőjel helyére?



2) Melyik szám illik a kérdőjel helyére?

256 -> 225 -> 196 -> 169 -> ?

☐ 124

☐ 100

☐ 144

☐ 136

☐ 108

RÖVIDÍTÉSEK JEGYZÉKE

adj	adjective	melléknév
adp	postposition	névutó
adv	adverb	határozószó
conj	coordination	mellérendelés
det	determiner	névelő
intj	interjection	indulatszó
noun	noun	főnév
num	numeral	számnév
part	particle	partikula
pron	pronoun	névmás
propn	proper noun	tulajdonnév
sconj	subordination	alárendelés
sym	symbol	szimbólum
verb	verb	ige