

DOKTORI ÉRTEKEZÉS TÉZISEI

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.

Előszó

A tágabb és a szűkebb problémakör megjelölése

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 (pozitív vagy negatív) értékelést fejeznek ki valamely célpont (angolul: *target*) viszonylatában. A feladat az alábbi egyszerű példával szemléltethető:

- (1) a. 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. Az elemzést tehát az alábbi módon sematizálhatjuk:

- (1) b. [_{targ} A szentimentelemzés] [_{szent} [_{int} nagyon] érdekes] feladat.

A feladat számos komoly elméleti és alkalmazott nyelvészeti kihívást rejt magában, amelyeket azonban csupán csekély számú dolgozat igyekszik behatóbban, elméleti megfontolások mellett tárgyalni. Az egyik fontos alkalmazásszintű probléma a szentimentérték módosulása, amely különösen a szótáralapú szentimentelemzés szempontjából bír kimagasló jelentőséggel. Értékmódosulás esetén ugyanis egy adott elemnek a szótárban rögzíthető értéke eltér attól az értéktől, amelyet az adott elemet magában foglaló teljes megnyilatkozás hordoz (vö. Szabó és mtsai 2016a, 2017a).

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 (2a), *értékváltó*ként pedig azokra, amelyek negatív tartalmuk ellenére bizonyos esetekben, szentimentkifejezés funkciójában pozitív értékelést vagy meglepődést is kifejezhetnek (2b) (vö. Szabó 2015a: 632, 2015c: 52–54). A példákban a vizsgált elemeket kövér szedéssel emelem ki.

- (2) a. **mocskos** nagy, **rohad**t jól
b. **brutális** volt a hétvégi koncert

A két jelenség mind elméleti, mind alkalmazott nyelvészeti, különösen nyelvtechnológiai szempontból figyelemre méltó. Az *értékváltás* és *értékvesztés* jelenségeinek kimerítő elméleti nyelvészeti leírása még nem történt meg, hatékony nyelvtechnológiai kezelésük pedig 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.

A szakirodalmi háttér

A nemzetközi és hazai publikációk túlnyomó többsége (vö. pl. Berend–Farkas 2008, Loughran–McDonald 2011, Liu 2012, Ruppenhofer–Rehbein 2012, Miháltz 2010, 2013, Hangya és mtsai 2015, 2017) nyelvtechnológiai oldalról közelít a nyelvi értékeléselemzés problémáihoz; azok alkalmazott, és különösen elméleti vonatkozásait csekély számú dolgozat igyekszik behatóbban tárgyalni. 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).

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 *értékvesztés*re fordít figyelmet. Egyetlen olyan dolgozatról (Dragut–Fellbaum 2014) van tudomásunk, amely az *értékvesztés* jelenségét nyelvtechnológiai szempontból közelíti meg. A hazai nyelvtechnológiai irodalomban az említett problémákat – egyetlen rövid megjegyzéstől

eltekintve (vö. Hangya és mtsai 2015: 212) – Szabó (2015a, 2015c, 2016, 2017) és Szabó és mtsai (2017a, 2017b) dolgozatain kívül más nem vizsgálja.

A doktori kutatómunka fő céljai és újdonságai

- Fontos célkitűzésem volt, hogy a szentimentérték módosulásának a problémáját úgy tegyem alapos és sokrétű vizsgálat tárgyává, hogy közben mind az elméleti, mind az alkalmazott nyelvészeti és nyelvtechnológiai vonatkozásokat szem előtt tartom.
- A szentimentérték módosulásának két speciális, az általam *értékváltásnak* és *értékvesztésnek* nevezett típusait különös figyelemmel kívántam tárgyalni, azok az elméleti nyelvészeti, a pszichológiai és a szemantikai motivációs sajátosságainak a részletes elemzésével.
- Fontosnak tartottam, hogy a jelenségeket nagy mennyiségű valós adat alapján, különböző adatbázisok segítségével vizsgáljam, feltárva ezzel a jelenségek eddig ismeretlen kvantitatív és kvalitatív sajátosságait.
- A vizsgálat fontos része annak a kérdésnek a körüljárása, hogy hogyan képzelhető el az értékvesztő elemek szemantikai reprezentációja, amely vizsgálathoz a lexikai pragmatika és a mentális lexikon hálózatos modelljének elméleti kereteit alkalmazom.
- A vizsgálati tapasztalatok alapján létre kívántam hozni egy automatikus szentimentelemző eszközt, amely – az értékmódosulás, köztük az értékvesztés és az értékváltás tekintetében is – hatékonyabb feldolgozási megoldást nyújt a szótáros szentimentelemzési módszerhez képest.

Az értekezés felépítése és a legfontosabb eredmények

1. A nyelvi értékelés mibenlétének kérdése

A számítógépes értékeléselemzés számára is elsődlegesnek kellene lennie, hogy megválaszolja, miben áll a nyelvi értékelő jelentés, azaz pontosan mi az elemzés tárgya. Ennek ellenére sem a nemzetközi, sem a hazai szentimentelemzési irodalom nem törekszik e kérdés vizsgálatára és megválaszolására.

E dolgozatrészben a nyelvi értékelés mibenlétének kérdését tárgyaltam részletesen, elméleti és alkalmazott nyelvészeti szempontból egyaránt (vö. Szabó 2016). Fontos célom volt, hogy megállapítsam, vajon megfelelő alapot tud-e nyújtani az elméleti irodalom a nyelvtechnológiai alkalmazások számára. Ráműtattam, 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 szerzők a beszélői szubjektivitással hozzák azt összefüggésbe (vö. Kiefer, 1986, Berényi 1988, Szilágyi 1996), 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).

A nyelvi értékelés több szempontú vizsgálata alapján megállapítottam, hogy az elméleti nyelvészeti irodalom nem nyújthat megfelelő alapot a nyelvi értékelés jelenségének sem az elméleti, sem a nyelvtechnológiai szintű kezeléséhez, és a pragmatikai sajátosságok jelentősebb figyelembevételére volna szükség egy pontosabban működő automatikus elemző rendszer létrehozásához.

Az e vizsgálatok során tett megállapításaim elsősorban a szentimentszótár összeállításakor hasznosítottam (l. 5).

2. Az értékvesztési és az értékváltási jelenségek tárgyalása

Első lépésben elvégeztem a nyelvi értékelés témaköréhez kapcsolódó fogalmak tisztázását, valamint a terminológiai hiányok pótlását.

A leíró nyelvészeti megállapításokról a következő legfontosabb megállapításokat tettem: a szerzők nem írják le a jelenségeket teljes körűen, nem nagy mennyiségű, valós

szövegből származó adat alapján vonják le következtetéseiket, 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).

A pszichológiai motivációt illetően részletesen bemutattam azt a három tendenciát (*negatív torzítás, pozitív torzítás vagy Pollyanna-jelenség és kongruencia*), amelyekkel a vizsgált jelenségeket összefüggésbe hozhatjuk (vö. Szabó 2017). Emellett megvizsgáltam azokat az ún. *szemantikai prozódia* szempontjából is. Törekedtem arra, hogy rávilágítsak a fenti tendenciák dinamikus kölcsönhatására. Az elemeket Jing-Schmidt (2007) alapján aszerint is rendszereztem, hogy azok a szemantikai tartalmuk szerint az alapemóciók (vö. Ekman–Friesen 1969) közül mely típushoz kapcsolhatóak. Megállapítottam, hogy a magyar nyelvben a vizsgált elemek a félelem, a düh és az undor érzelem által lehetnek motiváltak. A bánat érzelem által motivált elemet nem találtam, ami összefügghet a negatív torzítás működési mechanizmusával (vö. Öhman és mtsai 2001: 392–394). Ugyanitt felhívtam a figyelmet arra, hogy a magyar nyelvre vonatkozó dolgozatok az érzelmek közül kizárólag a félelmet említik (vö. Nemesi 1998, Kugler 2014).

Ezt követően az értékvesztő és az értékváltó nyelvi elemeket a szemantikai motiváció szempontjából vizsgáltam meg részletesen, korpuszból (MNSZ2, Oravecz és mtsai 2014), valamint egyéb, valós szövegekből származó példaanyag alapján (vö. Szabó 2017).

Az értékvesztésre vonatkozóan a következő legfontosabb megállapításokat tettem: A negatív fokozó elemek a szemantikai motivátlanság tekintetében csupán inherens (pozitív vagy negatív) szentimentértékű alaptagok mellett viselkednek egységesen. Ilyenkor vagy csupán az intenzitási jegyük jut szerephez, vagy – interperszonális funkció esetében – az emotív komponens is. Amikor a vizsgált elemek inherens értékkel nem rendelkező alaptagot módosítanak, nem azonos mértékben veszítik el lexikai szintű szemantikai tartalmukat, és 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.

Az értékváltó elemeket illetően megállapítottam, hogy velük 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. A jelenséget az elemek emotív szemantikai tartalmával hoztam összefüggésbe.

Az értekezés következő alfejezeteiben a lexikai pragmatika elméleti keretében, valamint a mentális lexikon hálózatos modelljének a segítségével vizsgáltam az értékvesztés (kis részben az értékváltás) jelenségét. Azt a problémát jártam körül, hogyan lehetne az értékvesztő elemeknek a megelőző fejezetekben feltárt sajátágaival egy helytálló szótári reprezentációban, ugyanakkor a számítógépes nyelvészeti alkalmazások szempontjából is hasznosítható módon elszámolni az elméleti oldalon.

A lexikai pragmatika keretét alkalmazva az alulspecifikált szemantikai reprezentáció elvét követtem, és alapvetően Bibok (2014, 2016) érvelésére és reprezentációs megoldására támaszkodtam. Példákon keresztül, a szemantikai motivációt illetően tett megállapításaim alapján felírtam az értékvesztésre képes elemek lexikai-szemantikai reprezentációját. A vizsgált elemek intenzitási jegye minden megnyilatkozásban jelen kell, hogy legyen, így ezt tekinthetjük a magjelentésnek. Bizonyos esetekben kizárólag ez a komponens vesz részt a megnyilatkozás interpretálásában (vö. Szabó 2016: 168), más esetekben a magjelentésre egyéb komponens is ráépül. Ez utóbbiak tehát nem kötelező komponensek. A lexikai reprezentációt a következőképpen adtam meg:

[intenzív] ([affektív]) ([negatív]),
ahol a kerek zárójelek opcionálitást jelölnek.

A vizsgált elemeknek tehát ebben a funkcióban önálló lexikai tételt adtam meg. Úgy véltem ugyanakkor, hogy a mentális lexikon aktivációterjedéses modellezése lehetőséget adhat arra is, hogy a vizsgált elemek különböző használataiban is csupán egyetlen lexikai tételt tegyünk fel, majd a terjedő aktivációval modellezzük az aktuális interpretációs mechanizmust. A megközelítési mód mellett másproblémák (pl. a feldolgozási idő kérdése) kezelésére is megoldást kínálhat, amelyekkel az alkalmazott lexikai pragmatikai eszköztárral nem tudunk kielégítően elszámolni (vö. pl. Wilensky 1987: 20, Gibbs 1994, Nemesi 2007, Forgács és mtsai 2012). A modell megengedi ugyanis, hogy azt feltételezzük, az aktuális nyelvhasználatban a megfelelő lexikai reprezentáció kiválasztását, majd a különböző jelentéskomponensek szerephez jutását a terjedő aktiváció irányítja (vö. Caramazza (1999), idézi Kovács (2013: 187), Huszár (2005: 63–64)). A fentebbieket megfontolva az értékvesztés jelenségét a hálózatos modell segítségével is megvizsgáltam, és a terjedő aktiváció feltételezésével reprezentáltam.

A fejezetben tett megállapításaim fontos alapot szolgáltattak a NegEmotív korpusz építéséhez (l. 4), a szóasszociációs vizsgálatomhoz (l. 4), valamint az automatikus elemző eszköz létrehozásához (l. 5).

3. A számítógépes értékeléselemzés

Tisztáztam a feladat tárgyát és fontosságát, majd feltártam, hogy melyek azok a tudományos és technológiai kihívások és hasznosítási lehetőségek, amelyek a szentimentelemzéssel kapcsolatos nemzetközi és hazai kutatásokat generálják. A legfontosabb problémákat, módszereket és eszközöket a nemzetközi kutatások bemutatásánál tárgyaltam. A magyar nyelvű szövegekkel foglalkozó dolgozatok áttekintését igyekeztem kimerítő jelleggel elvégezni. A szentimentérték-módosulási jelenséget külön alfejezetben vizsgáltam.

Az áttekintés az adatbázis-építési feladatok (l. 4), valamint az elemző eszköz elkészítésének (l. 5) az alapját képezte.

4. Az értékmodosulással kapcsolatos vizsgálatok

E vizsgálatokat az erre a célra létrehozott három vizsgálati eszköz alapján végeztem el.

A dolgozatrész elején a különböző adatbázisoknak (korpusz és szóasszociációs adatbázis) a nyelvészeti kutatásban és fejlesztésben betöltött szerepét, azok típusait, feldolgozásuk módszereit és eszközeit, valamint a minőségbiztosítási megoldásokat tárgyaltam részletesen.

Ezt követően ismerttettem a kutatás három legfontosabb vizsgálati eszközét: a fragmentum- és targetszintű annotációt tartalmazó, kézzel annotált szentimentkorpuszt (Dívány-korpusz, vö. Szabó–Vincze 2015, Szabó és mtsai 2016a, 2016b), az értékvesztés és az értékváltás jelenségei szempontjából kézzel annotált korpusz (NegEmotív korpusz, vö. Szabó és mtsai 2017a, 2017b), valamint egy szóasszociációs adatbázist (vö. Szabó és mtsai 2017d), amely az értékvesztésre – és esetlegesen az értékváltásra is képes – elemekre adott asszociációkat tartalmazza. A kutatómunka jelentős hozadéka, hogy mindezek, a vizsgálatokhoz szükséges, hiánypótló eszközök a doktori kutatómunka keretében jöttek létre.

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 tárgyaltam az azokban feldolgozott szövegtípusok

sajátságait, 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 Dívány-korpusz alapján behatóan vizsgáltam az értékmódosulás különböző sajátságait. A NegEmotív-korpusz szemantikai-pragmatikai annotációjának segítségével kimondottan az értékvesztés és az értékváltás jelenségeinek korpuszalapú vizsgálatát végeztem el és az eredményeket megvizsgáltam a férfi és a női nyelvhasználati eltérések szempontjából is. A kapott eredmények egybevágtak a korábbi szemantikai motivációs vizsgálati eredményeimmel, és több dolgozat megállapításait cáfolták (vö. pl. Jing-Schmidt 2007, Balogh 2009). A szóasszociációs adatbázis hívószóit és válaszszóit a szófaji megoszlások, a szentiment- és emóciótartalmak, az asszociatív átfedések és a férfi és a női nyelvhasználati eltérések szempontjából elemeztem. A vizsgálat eredményei közül kiemelendő, hogy bár az értékvesztésre képes elemek legtöbbször negatív szemantikai tartalmú válaszokat hívnak, bizonyos elemek esetében a lexikalizáció folyamatára utaló asszociátumokat kaptam.

A vizsgálati eredmények alapján készítettem el a szentimentelemző eszközt (l. 5), valamint visszacsatoltam a szótári reprezentációs problémához (l. 2).

5. Egy saját fejlesztésű szentimentelemző eszköz

Az eszköz szótárak és kombinációs szabályok alapján rendel összesített szentimentértéket az egyes mondatokhoz. A kombinációs szabályokat a vizsgálatok során gyűjtött vizsgálati tapasztalatok alapján hoztam létre, a különböző szótárak pedig a jelen kutatómunka során készültek (vö. Szabó 2014, 2015b). Az eszköz hatékonyságának méréséhez létrehoztam egy kézzel, mondat szinten annotált szentimentkorpuszt is.

Az eszköz hatékonyságát a humán annotációval, valamint az egyszerű szótáralapú elemzési eredményekkel összehasonlítva 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.

A tézisekhez felhasznált szakirodalom:

- Andor József 2011. De durva ez a téma! – Megfigyelések a melléknévi polaritásváltásról. In *Hungarológiai Évkönyv* **12**. 33–42.
- Balogh, P. 2009. *Gender-markerek a nyelvben* (<http://webfu.univie.ac.at/wp/565>)
- 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.
- 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.
- Caramazza, A. 1999. Hány feldolgozási szint van a lexikai hozzáférésben? In Bánréti Z. szerk. *Nyelvi struktúrák és az agy*. Budapest, Corvina. 249–283.
- 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.–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.
- Ekman, P.–Friesen, W.V. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* **1**. 49–98.
- 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.
- Gibbs, R.W. 1994. *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge, UK, Cambridge University Press.
- 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>)
- Huszár Á. 2005. *A gondolattól a szóig*. Budapest, Tinta.
- Jing-Schmidt, Z. 2007. Negativity bias in language: A cognitive-affective model of emotive intensifiers. *Cognitive Linguistics* **18**(3). 417–443.
- Kennedy, A.–Inkpen, D. 2006. Sentiment Classification of Movie Reviews using Contextual Valence Shifters. *Computational Intelligence* **22**(2). 110–125.
- Kiefer F. 1986. A modalitás fogalmáról. *Nyelvtudományi közlemények* **88**(1–2). 1–38.
- Kovács, J.L. 2013. *Fogalmi rendszerek és lexikai hálózatok a mentális lexikonban*. Budapest, Tinta.

- Kugler N. 2014. A nyelvi polaritás kifejezésének egy mintázata, avagy milyen a félelmetesen jó? *Magyar Nyelvőr* **138**(2). 129–139.
- Laczkó M. 2007. Napjaink tizenéveseinek beszéde szóhasználati jellemzők alapján. *Magyar Nyelvőr* **131**(2). 173–184.
- 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>)
- 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.
- 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.
- 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.
- 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.
- Oravecz, Cs.–Váradi, T.–Sass, B. 2014. The Hungarian Gigaword Corpus. In *Proceedings of LREC*. Reykjavik, Iceland, European Language Resources Association (ELRA). 1719–1723.
- 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.
- 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.
- Pete I. 2002. A nyelvi modalitás három fő típusának hierarchiája a magyarban. *Magyar nyelv* **98**(2). 173–191.
- 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.
- Ruppenhofer J.–Rehbein, I. 2012. *Anchoring sentiment analysis in semantic frames*. (<http://www.uni-hildesheim.de/ruppenhofer/pubs/longversion.pdf>)

- 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*. Linceum 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. Nyelvészeti doktoranduszok dolgozatai*. Szeged, Szegedi Tudományegyetem, Nyelvtudományi Doktori Iskola. 153–172.
- Szabó M.K. 2017. Az értékvesztés és az értékváltás jelenségeinek vizsgálata a pszichológiai, valamint a szemantikai motiváció szempontjából. *MANYE 26. konferenciakötete*. Megjelenés előtt.
- 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.–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.–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 Zs. szerk. *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből*. Budapest, MTA Nyelvtudományi Intézet. 161–176.
- Szabó M.K.–Ilyés V.–Nyíri Zs. 2017d. A negatív emotív elemek szóasszociációs adatbázisa. *MANYE 26. konferenciakötete*. Megjelenés előtt.
- Székely G. 2007. Egy sajátos nyelvi jelenség, a fokozás. *Segédkönyvek a nyelvészet tanulmányozásához* 66. Budapest, Tinta.
- Szilágyi N.S. 1996. *Hogyan teremtsünk világot?* Kolozsvár, Erdélyi Tankönyvtanács.

- Tolcsvai N.G. 1988. A mai magyar nyelv normarendszerének egy jelentős változásáról az „ifjúsági nyelv” kapcsán. *Magyar Nyelvőr* **112**(4). 398–406.
- Wierzbicka, A. 2002. Australian cultural scripts – *bloody* revisited. *Journal of Pragmatics* **34**(9). 1167–1209.
- Wilensky, R. 1987. *Some Problems and Proposals for Knowledge Representation*. Technical Report. EECS Department, University of California, Berkeley.
- 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.

Az értekezés témájához kapcsolódó publikációk:

- 1) 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.
- 2) 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.
- 3) Szabó M.K. 2015b. Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai és dilemmái. In Geecső T.–Sárdi Cs. szerk. *Nyelv, kultúra, társadalom. Segédkönyvek a nyelvészet tanulmányozásához 177.* 278–285.
- 4) Szabó M.K. 2015c. A polaritásváltás problémája a szentimentelemzés szempontjából. In Váradí T. szerk. *IX. Alkalmazott Nyelvészeti Doktoranduszkonferencia. Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből.* MTA Nyelvtudományi Intézet. 51–61.
- 5) Szabó M.K. 2016. A nyelvi értékelés mibenlétének kérdése a számítógépes értékeléselemzés (szentimentelemzés) szempontjából. In Gécség Zs. szerk. *LingDok 15. Nyelvészeti doktoranduszok dolgozatai.* Szeged, Szegedi Tudományegyetem, Nyelvtudományi Doktori Iskola. 153–172.
- 6) Szabó M.K. 2017. Az értékvesztés és az értékváltás jelenségeinek vizsgálata a pszichológiai, valamint a szemantikai motiváció szempontjából. *MANYE 26. konferenciakötete.* Megjelenés előtt.
- 7) Szabó M.K.–Morvay G. 2015. Emócióelemzés magyar nyelvű szövegeken. In Geecső T.–Sárdi Cs. *Nyelv, kultúra, társadalom.* 286–292.
- 8) 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.
- 9) Szabó M.K.–Vincze V. 2016. A negáló szentimentshifterek kezelési kérdései a magyar nyelvű szövegek számítógépes értékeléselemzési feladatában. In Reményi A.Á.–Sárdi Cs.–Tóth Zs. szerk. *Távlatok a mai magyar alkalmazott nyelvészetben.* Budapest, Tinta. 272–281.
- 10) Szabó M.K.–Galántai J. 2017. Egy magyar nyelvű spontán beszélt nyelvi korpusz (HuTongue) létrehozásának tapasztalatai. *MANYE 26. konferenciakötete.* Megjelenés előtt.
- 11) 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.

- 12) 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.
- 13) 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.
- 14) 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.
- 15) 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.
- 16) Szabó M.K.–Lázár B.–Nyíri Zs.–Morvay G. 2017b. A negatív emotív elemek vizsgálata a nemek közötti nyelvhasználati különbségek szempontjából. In Ludányi Zs. szerk. *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből*. Budapest, MTA Nyelvtudományi Intézet. 161–176.
- 17) Szabó M.K.–Nyíri Zs.–Lázár B. 2017c. Negatív emotív szemantikai tartalmú fokozó elemek vizsgálata az angol–orosz és orosz–angol fordítás tükrében. In Ludányi Zs. szerk. *Doktoranduszok tanulmányai az alkalmazott nyelvészet köréből*. Budapest, MTA Nyelvtudományi Intézet. 177–191.
- 18) Szabó M.K.–Ilyés V.–Nyíri Zs. 2017d. A negatív emotív elemek szóasszociációs adatbázisa. *MANYE 26. konferenciakötete*. Megjelenés előtt.
- 19) Drávucz F.–Szabó M.K.–Vincze V. 2017. Szentiment- és emóciósótárak eredményességének mérése emóció- és szentimentkorpuszokon. In Vincze V. szerk. *XIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szegedi Tudományegyetem, Informatikai Intézet. 228–239.
- 20) Nyíri Zs.–Szabó M.K.–Ilyés V. 2018. Egy orosz nyelvű korpusz (NarRu) narratívaelemzése saját fejlesztésű szentiment- és emóciósótárakkal. In Vincze V. szerk. *XIV. Magyar Számítógépes Nyelvészeti Konferencia* (MSZNY 2018). 88–99.
- 21) Gulyás A.–Galántai J.–Szabó M.K.–Szebeni Z. 2018. A HuTongue spontán beszélt nyelvi korpusz leiratozásának és annotálásának minőségbiztosítási munkálatai. In Vincze V. szerk. *XIV. Magyar Számítógépes Nyelvészeti Konferencia* (MSZNY 2018). 317–330.
- 22) Gulyás A.–Szabó M.K.–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). Szegedi Tudományegyetem, Informatikai Intézet. 109–121.

- 23) Gulyás A.–Szabó M.K.–Boros I.–Havadi G. 2018. Text network analysis and visualization of Hungarian, communist-era political reports. In *Text2Story 2018 - First Workshop on Narrative Extraction From Text*, Grenoble, France. Available: <http://ceur-ws.org/Vol-2077/>
- 24) Boros I.–Gulyás A.–Szabó M.K.–Havadi G. 2018. The advanced methods of text network analysis and visualisation. In *POLTEXT Incubator Workshop*. Budapest, Hungary.