

A polaritásváltás problémája a szentimentelemzés szempontjából

Szabó Martina Katalin

SZTE BTK Nyelvtudományi Doktori Iskola
szabomartinakatalin@gmail.com

Kivonat: A dolgozatban a polaritásváltás problémáját vizsgálom a szentimentelemzés nyelvtechnológiai feladata szempontjából. A dolgozat hiánypótló, hiszen – bár az elméleti nyelvészetben számos szerző vizsgálja a polaritásváltás jelenségét – a nyelvtechnológiai alkalmazások, így a szentimentelemzés nem fordít kellő figyelmet rá, valamint az elméleti nyelvészet eredményeit sem aknázza ki. A polaritásváltás jelensége a szentimentelemzés egy kulcsfontosságú problémájának tekinthető. A szentimentelemzést célzó kutatásokra ugyanis általánosságban jellemző, hogy a feladatot az ún. szentimentkifejezések szótári formába rendezett listáival igyekeznek – legalább részben – megoldani. Belátható ugyanakkor, hogy egy hatékony működésre képes szentimentelemző rendszer nem képzelhető el anélkül, hogy a polaritásváltási jelenséget ne tudná kezelni.

1 Bevezetés

A dolgozatban a polaritásváltás problémáját vizsgálom a szentimentelemzés nyelvtechnológiai feladata szempontjából.

A szentimentelemzés a számítógépes nyelvészet egy részfeladata, amely arra irányul, hogy az értékelést, az értékelő tartalmakat megtalálja a szövegekben, meghatározza ezeknek az értékeknek a polaritását, azaz a pozitív vagy negatív voltát, illetve esetlegesen a polaritás intenzitásának a mértékét, valamint megállapítsa azok tárgyát (az angol nyelvű terminológia alapján: targetjét), tehát azt, hogy az értékelés mire irányul. Mindezt a természetesnyelv-feldolgozás kínálta automatikus eszközökkel igyekszik megvalósítani (vö. Szabó–Vincze 2015). (A polaritás fogalma a nyelvtechnológiai alkalmazásokban tehát nem azonos azzal a polaritásfogalommal, amelyet az elméleti nyelvészeti kutatás alkalmaz.)

A szentimentelemzés növekvő nemzetközi népszerűsége (vö. Прохоров–Керимов 2012; Витехновский 2013: 34) ellenére a magyar nyelvű szövegek szentimentelemzésével csekély számú dolgozat foglalkozik (vö. Liu 2012; Berend–Farkas 2008; Miháltz 2010, 2013; Hangya et al. 2015). Jelentős probléma azonban, hogy ezek a dolgozatok sem tárgyalják a feladat elméleti vonatkozású problémáit, illetve nem nyújtanak kellő információt az e problémák vonatkozásában alkalmazott megoldásokról. Az elméleti kérdések tárgyalása azonban nélkülözhetetlen egy olyan elemző rendszer létrehozásához, amely képes a hatékony szentimentelemzésre.

A szentimentelemzésben már-már vezérelvnek tekinthető az a megállapítás, hogy az értékelő szemantikai tartalmak legfontosabb indikátorai az ún. szentimentki fejezések, vagyis azok a nyelvi elemek, amelyek jelentésük részeként pozitív vagy negatív értékelő tartalmat hordoznak, a domén-, illetve kontextuális sajátságoktól függetlenül. Ezzel összefüggésben bevett gyakorlatnak számít, hogy a szentimentelemzés feladatát e szentimentki fejezések szótárba rendezett listáival igyekeznek – legalább részben – megoldani.

A szótáralapú automatikus elemzés egyik részproblémája az ún. polaritásváltás, amelynek során:

a) a lexikai szinten negatív tartalommal rendelkező kifejezések pozitív irányú polaritásváltozáson esnek át (pl. *durva sportkupé*; *brutális alaplap*),

b) a lexikai szinten negatív vagy pozitív jelentéstartalommal rendelkező kifejezések lexikális jelentésüket elveszítve egy másik szentimentki fejezés polaritásának a fokozóivá (intenzifikálóivá) válnak (pl. *borzasztóan jó*; *jó bonyolult*) (vö. Tolcsvai Nagy 1988; Laczkó 2007; Andor 2011; Kugler 2014; Szabó 2015).

Bár az elméleti nyelvészetben számos szerző vizsgálja a polaritásváltás jelenségét, a nyelvtechnológiai alkalmazások, így a szentimentelemzés nem fordít figyelmet a jelenségre, valamint az elméleti nyelvészet eredményeit sem aknázza ki.

A jelen dolgozat célja, hogy a polaritásváltás jelenségét – különös tekintettel az intenzifikáló funkciójú elemek polaritásváltására – az automatikus szentimentelemzés feladatkorre szempontjából, valós nyelvi adatok alapján beható vizsgálat tárgyává tegye. Arra a kérdésre keresem a választ, hogy vajon detektálhatók-e olyan lényegi szintaktikai, szemantikai, doménfüggő, frekvenciabeli vagy kollokációs sajátságok, amelyek hathatós segítséget nyújthatnának a vizsgált kifejezések aktuális polaritásának automatikus felismerésében.

2 A polaritásváltás

2.1 A jelenség általános bemutatása

Több szerző (Tolcsvai Nagy 1988; Laczkó 2007; Andor 2011; Kugler 2014) is felhívja a figyelmet egy az értékelést kifejező elemek kapcsán megfigyelhető terjedőben levő jelenségre, miszerint „a negatív jelentéstartalmú és használatú lexikai egységek pozitív irányú jelentésváltozását vagy jelentésbővülését, jelentésük kiterjesztését figyelhetjük meg” (Andor 2011: 33). Andor (2011: 33) megemlíti, hogy a változás az ellentétes polaritás irányába is lehetséges, azaz pozitívból negatívba, ez a változat azonban jóval ritkább előfordulású.

Magam a polaritásváltás jelenségét a következőképpen definiálom: polaritásváltásról akkor beszélünk, ha egy lexikai szinten pozitív vagy negatív polaritással rendelkező elem az aktuális kontextusban elveszíti prior polaritását úgy, hogy az neutralizálódik vagy az ellenkezőjére változik. Ez utóbbi változás csupán a negatív polaritástól a pozitív polaritás irányába történhet. (Fontos hangsúlyozni, hogy az iróniát nem tekintem a polaritásváltás körébe tartozó jelenségnek.)

Vizsgálati eredményeim (vö. Szabó 2015) alapján szemantikai szempontból a polaritásváltásnak alapvetően két típusát különböztethetjük meg. Az egyik esetben a lexikai szinten negatív jelentéstartalommal rendelkező kifejezés pozitív irányú polari-

tásváltozáson esik át (pl. *brutális alaplap*) (vö. Andor 2011). A másik esetben a lexikai szinten negatív vagy pozitív jelentéstartalmú kifejezés lexikális jelentését elvezítve egy másik kifejezés polaritásának a fokozójává (intenzifikálójává) válik (pl. *borzasztóan jó; jó bonyolult*) (vö. Tolcsvai Nagy 1988; Laczkó 2007; Székely 2007; Kugler 2014). Ami a szerkezeti sajátságokat illeti, ugyancsak két nagy szerkeztípust különböztethetünk meg. Az egyik az, amelyben a polaritást váltott kifejezés a szentimentkifejezés szerepét tölti be, azaz önmagában fejezi ki az értékelő tartalmat. Ebben az esetben a vizsgált kifejezés állhat mind állítmányi [1a], mind jelzői funkcióban [1b] (vö. Andor 2011: 37).

- [1] a) (...) *tök durva volt, amit a hegedűs művelt a hangversenyen.*
 [2] b) *Nagyon durva sportkupét mutatott be a Volvo.*

A másik típusban a polaritást váltott elem egy szentimentkifejezés szemantikai tartalmát fokozza, másképpen, intenzifikálja azt. Ebben az esetben a vizsgált elem módosíthat határozót [2a] és jelzőt [2b] egyaránt, és nem csupán negatív, de pozitív prior polaritással rendelkező elem is állhat ebben a funkcióban [2c].

- [3] a) *A mocskosul jól kinéző CGI-vérnek például búcsút inthetünk.*
 [4] b) (...) *a miskolci Jameson Cinefesten baromi jó filmeket lehet nézni tökingyen.*
 [5] c) *Beyonce rajongó vagy? Biztos jó hülye lehetsz!*

A második típus esetében nem ritka az egybeírt változat sem [3].

- [6] (...) *megállás nélkül váltják egymást a fergeteges szövegek és a durvajó akciójelenetek.*

Ami a szemantikai motiváltság kérdését illeti, amellet érvelek, hogy a polaritásváltás hatására részleges vagy teljes deszemantizálódás következik be, s ezzel együtt a pragmatikai motiváltságra helyeződik a hangsúly. Abban az esetben, amikor a polaritást váltott elem a szentimentkifejezés szerepét tölti be, a deszemantizálódás teljes mértékű. Vessük össze az [4a] alatti példát az [1b] alatti példával, amelyet itt [4b] alatt megismétlek!

- [7] a) (...) *belehalt a durva támadásban szerzett sérüléseibe.*
 [8] b) (...) *Nagyon durva sportkupét mutatott be a Volvo.*

Azokban az esetekben azonban, ahol a polaritást váltott elem intenzifikálói funkciót tölt be, a szemantikai motiváltság és motivátlanság között egy fokozati skála tételezhető fel (vö. Szabó 2015). Tekintsük az [5] alatti példákat!

- [9] a) *pokoli meleg; pokoli hideg*
 b) *pokoli rossz; pokoli jó*

Az [5a] alatti példák intenzifikált elemei Ahn és munkatársai (2012: 5–6) alapján relatív polaritású elemek. A szerzők (Ahn et al. 2012: 6) szerint relatív polaritásról akkor beszélünk, ha a polaritás az aktuális domén, illetve target függvényében áll. Amennyiben azonban egy elem abszolút polaritással rendelkezik, úgy polaritása nem függ az adott doméntól, amelyben szerepel, illetve targettől, amit minősít. Belátható, hogy a példákban a *meleg* és a *hideg* melléknevek eltérő polaritással rendelkezhetnek attól függően, hogy az aktuális target szempontjából ezt a sajátságot pozitívnak vagy negatívnak tekinthetjük-e, azaz polaritásuk relatív. A *pokol* szó önmagában a nagyon meleg helyellyel asszociálódik. Ahhoz tehát, hogy ellentétes szemantikai tartalmú elemekkel is kolloká-

lódhasson, bizonyos mértékű deszemantizálódásra van szükség. Amellett érvelek ugyanakkor, hogy a *pokoli* deszemantizáltsága nem teljes mértékű a kifejezésben, ugyanis a szókapcsolat a negatív értékelő tartalmát csupán a *pokol* negatív konnotációjából nyerheti, a modifikált elem relatív polaritása okán (vö. Szabó 2015).

Teljes deszemantizáltságról véleményem szerint csupán az [5b] alatti példák esetében beszélhetünk, tehát ott, ahol a modifikált elem abszolút polaritással rendelkezik. Ezekben a konstrukciókban ugyanis a polaritás az alaptag által eleve meghatározott, így a polaritást váltó kifejezés betöltheti a „puszta” intenzifikáló szerepét.

2.2 A polaritásváltás a szentimentelemzés szempontjából

A polaritásváltás jelensége a szentimentelemzés egy kulcsfontosságú problémájának tekinthető. A szentimentelemzést célzó kutatásokra ugyanis általánosságban jellemző, hogy a feladatot az ún. szentimentkifejezések szótári formába rendezett listáival igyekeznek – legalább részben – megoldani (vö. Vázquez – Bel Rafecas 2012; Liu 2012: 12–13). A szentimentelemzésben már-már vezérelvnek tekinthető az a megállapítás, hogy az értékelő szemantikai tartalmak legfontosabb indikátorai e puszta szentimentkifejezések, vagyis azok a legkisebb nyelvi egységek, amelyek értékelő tartalmat hordoznak. E meggondolás alapján a szentimentszótárak a legtöbb elemzésben kimagasló jelentőséggel bírnak, mintegy az elemzés alapköveinek tekinthetők (vö. Szabó 2014, 2015).

A szentimentszótárak létrehozásának elméleti alapvetése az, hogy a polarizált kifejezéseknek legalább egy csoportja rendre pozitív vagy negatív értékelő szemantikai tartalmat hordoz, tehát a kontextuális sajátságoktól függetlenül rendelkezik az adott polaritással mint jelentéskomponenssel (vö. Moghaddam–Popowich 2010: 3; Baroni–Vegnaduzzo 2004; Kamps et al. 2004; Vegnaduzzo 2004).

Kérdéses azonban, hogy a polaritásváltás jelensége komoly elméleti dilemmát vet fel, hiszen feltétlen meggondolást sürget az a kérdés, hogy hogyan egyeztethető össze a polaritás szemantikai komponensként való tételezése a polaritásváltás lehetőségével. (A probléma teoretikus vonatkozásaival e dolgozatban részletesen nem foglalkozom, arról bővebben l. Szabó 2015).

A nyelvtechnológia szempontjából nyilvánvaló, hogy egy hatékony működésre képes szentimentelemző rendszer nem képzelhető el anélkül, hogy a polaritásváltási jelenséget kezelni ne tudná. Olyan támogató megoldásokra van tehát szükség, amely képessé teheti az elemzőt a polarizált kifejezések kontextusfüggő kezelésére.

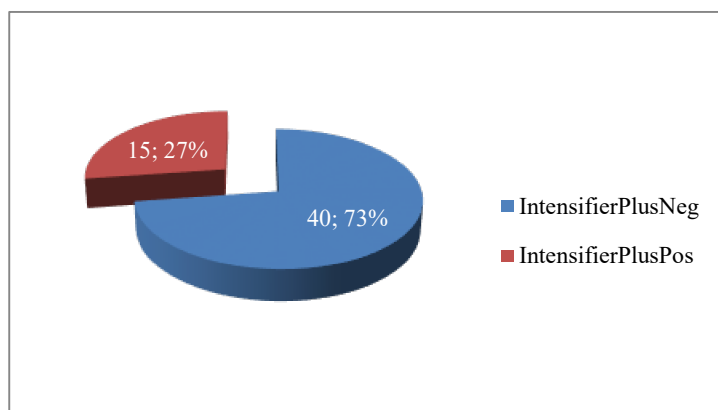
A dolgozat további részében, korpuszadatok alapján azt próbálom meg feltérképezni, hogy van-e olyan szintaktikai, szemantikai, doménfüggő, frekvenciabeli vagy kollokációs sajátság, amely hathatós segítséget nyújthat az aktuális polaritás automatikus felismerésében.

3 A polaritásváltás jelensége a korpuszadatok tükrében

Korpuszvizsgálataim túlnyomó többségét a PrecoSent korpuszon, valamint a Precorpuson végeztem.

A PrecoSent egy manuálisan annotált szentimentkorpusz, amelyet a magyar nyelvű szövegek szentimentelemzésének feladatához hoztunk létre, elemzési, fejlesztési és tesztelési céllal (vö. Szabó–Vincze 2015). A PrecoSent a jelenlegi egyetlen olyan szentimentre annotált korpusz, amelyben a polaritás a szentimentkifejezések szintjén van annotálva. Ennek, valamint a további annotációk a sajátosságoknak köszönhetően az eszköz a polaritásváltás korpuszalapú vizsgálatához is jelentős lehetőséget nyújt.

Tekintettel arra, hogy a korpuszannotáció lehetővé tette az intenzifikáló elemek statisztikai adatainak egyszerű lekérdezését, kutatásom első lépéseként ezeket a kifejezéseket vizsgáltam meg. Arra a kérdésre kerestem a választ, hogy vajon milyen gyakorisággal fordulnak elő a negatív prior értékű intenzifikáló elemek negatív, és milyen gyakorisággal pozitív polaritású kifejezésben. A vizsgálat eredményeit az 1. ábra mutatja be.



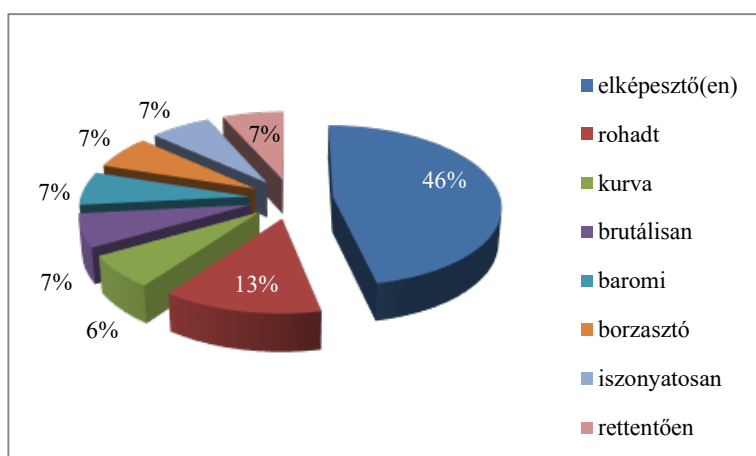
1. ábra. A negatív prior értékű intenzifikálók gyakorisági megoszlása a negatív és a pozitív kontextusok között a PrecoSent korpuszban

Amint az az ábrán látható, a korpusz negatív prior polaritású intenzifikálói csupán az esetek 73%-ban fordulnak elő negatív polaritású kifejezésben; 27%-uk pozitív polaritású kifejezésben szerepel. Ez utóbbiak esetében tehát a vizsgált kifejezés neutralizálódva pusztán intenzifikáló funkcióját tölti be (részletesebben l. fentebb, 2.1). Az arányok azt mutatják, hogy a polaritásváltás az intenzifikálók esetében igen gyakori jelenség, és amennyiben azokat nem tudjuk helyesen kezelni, úgy az elemzés eredményében jelentős torzulást okozhatnak.

Megvizsgáltam, hogy melyek azok a negatív prior értékű intenzifikáló elemek, amelyek pozitív polaritású kifejezésben szerepelnek, illetve hogy ezeknek az elemeknek milyen a gyakorisági megoszlása egymás között. Az eredményeket a 2. ábra mutatja be.

Ahogy a 2. ábra mutatja, a pozitív környezetben szereplő negatív prior értékű intenzifikálók között az *elképesztő(en)* a leggyakoribb, a *rohadt* a második leggyako-

ribb előfordulása, a további elemek között pedig az előfordulási gyakoriság körülbelül egyenlő mértékben oszlik meg. Összevettem a kapott adatokat azokkal a negatív prior értékű intenzifikálókkal, amelyek negatív környezetben szerepelnek, és azt találtam, hogy semmilyen lényeges eltérés nem mutatkozik a két használati módban. Azok az elemek tehát, amelyek pozitív környezetben előfordulnak, rendre megtalálhatóak negatív környezetben is, továbbá a pozitív környezetben leggyakoribb *elképesztő(en)* és *rohadt* a negatív környezetben is a leggyakoribb előfordulású elemek között szerepel. Azt látjuk tehát, hogy megközelítőleg ugyanazok az elemek fordulnak elő a negatív és a pozitív környezetekben, és megközelítőleg ugyanazzal a frekvenciális sajátossággal.



2. ábra. A negatív prior értékű intenzifikálók gyakorisági megoszlása a pozitív polarítású kifejezésekben a PrecoSent korpuszban

A korpuszadatok vizsgálata alapján úgy vélem, hogy az aktuális polaritás felismerésében a módosított elem polarítására hatékonyan lehetne támaszkodni. Korábban (l. fentebb, 2.1) részletesen tárgyaltam a relatív és az abszolút polaritás közötti különbséget. Néhány példa alapján amellet érveltem, hogy a polaritást váltó kifejezés teljes deszemantizáltságáról akkor beszélhetünk, ha a modifikált elem abszolút pozitív vagy negatív polaritással rendelkezik, így mellettük a polaritást váltó kifejezés „puszta” intenzifikálóként funkcionálhat. Mivel a korpuszadatok összhangban állnak a fenti hipotézissel, úgy vélem, az abszolút polarítású módosított elem esetében az alábbi szemantikai kompozíciós szabályok sikerrel volnának alkalmazhatók:

[10] I. Neg + Poz \rightarrow IntPoz (pl. *borzasztóan szép* \rightarrow 'nagyon szép' [+])

[11] II. Poz + Neg \rightarrow IntNeg (pl. *jó hülye* \rightarrow 'nagyon hülye' [- \rightarrow])

A fentebbiekkel azonos eredményt produkálnak az olyan szerkezetek is, amelyekben a két elem prior polarítása azonos egymással. Kezelésük az alábbi kompozíciós szabályokkal lehetséges:

[12] III. Poz + Poz \rightarrow IntPoz (*varázslatosan szép* \rightarrow 'nagyon szép' [+])

[13] IV. Neg + Neg \rightarrow IntNeg (*pokoli hülye* \rightarrow 'nagyon hülye' [- \rightarrow])

Vegyük észre, hogy az I. és a III., valamint a II. és a IV. alatt bemutatott szabályok a szentimentelemzés szempontjából azonos eredményt produkálnak.

A relatív polaritású modifikált elemek esetében a szerkezet polaritása nem ennyire egyértelmű. A relatív polaritás ugyanis – amint arról korábban már részletesebben is szóltam (l. fentebb, 2.1) – az aktuális domén, illetve target függvényében áll. A korpuszadatok alapján e konstrukciók esetében két szentimenttartalom is kifejeződésre juthat: Egyrészt lehetséges, hogy – ahogyan amellet korábban a *pokoli meleg* és a *pokoli hideg* kifejezések kapcsán érveltem (l. fentebb, 2.1) – az intenzifikáló nem teljesen deszemantizálódott elem, mivel a szókapcsolat a negatív értékelő tartalmát annak negatív konnotációjából nyeri [6]. Másrészt az is lehetséges, hogy a szerkezet szentimentértéke nem negatív, sőt akár pozitív is lehet, és a negatív prior értékű intenzifikálóval a beszélő pusztán hatáskeltési céllal él [7]. Tekintsük az alábbi, valós nyelvi példákat!

- [14] a) a *marha kemény futómű miatt „rázós” használni a magyar utakon*
 b) *őrült sok szenvedés árán rájön arra, miért és kiért érdemes élni*
- [15] a) *Lehetetlen helyzetből őrült nagy passz*
 b) *Így vághat ketté iszonyú gyorsan marha sok citromot*

A relatív polaritású alaptaggal alkotott szerkezetek kezelése tehát egyszerű szemantikai kompozíciós szabályok segítségével nem tűnik megoldhatónak, helyes interpretálásukhoz a tágabb kontextus figyelembe vételére van szükség.

A fentebbihez hasonlóan problematikus azoknak a szerkezeteknek a kezelése is, ahol a polaritást váltó elem nem intenzifikáló, hanem szentimentkifejezés funkcióját tölti be, azaz önmagában értékelő tartalmat hordoz. A korábban [1] alatt hozott példákat itt [8] alatt megismétlem:

- [16] a) (...) *tök durva volt, amit a hegedűs művelt a hangversenyen.*
 [17] b) *Nagyon durva sportkupét mutatott be a Volvo.*

E kifejezések aktuális polaritását semmilyen szemantikai kompozíciós szabállyal nem tudjuk lefedni.

Azt feltételeztem, hogy mind a relatív polaritású alaptaggal alkotott szerkezetek, mind a polaritást váltó szentimentkifejezések esetében segítséget nyújthatnának használatuk doménfüggő sajátosságai. Valószínűsítettem ugyanis, hogy mivel a vizsgált kifejezések váltott polaritás esetében stiláris többlettel rendelkeznek, azok frekvenciája eltérő lehet az egyes doménekből. Annak céljából, hogy megvizsgáljam, helytálló-e a hipotézis, elvégeztem egy vizsgálatot a *brutális* szó előfordulási gyakoriságát illetően a Precorpus nyolc doménből álló korpuszán (Szabó 2014). Az eredményt a 3. ábra prezentálja.

Azt látjuk tehát, hogy a vizsgált elem a politikai tárgyú szövegekben, valamint a bulvár és a kultúra doménben mutat legnagyobb előfordulást. (A kultúra alkorpusz szöveganyagának jelentős része filmes témájú internetes oldalakról származik, valószínűleg ezzel magyarázható a kifejezés gyakorisága.) A legkevesebb előfordulás a technológia és a tudomány alkorpuszokat jellemzi.

Megvizsgáltam, hogy milyen frekvenciával rendelkezik a *brutális* szó pozitív polaritású elemként ezekben a doménekből. Az elemzés tapasztalatait a 4. ábra mutatja be.

A politikai tárgyú, valamint a tudományos szövegekben tehát alapvetően nem jellemző a *brutális* szó váltott polaritású használata, így a kifejezés prior polaritása mu-

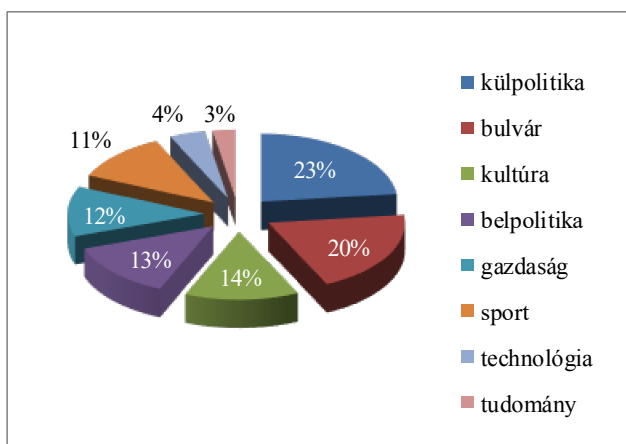
tatkozik meg [9]. Ugyanakkor, például a kultúra és a gazdaság alkorpuszban az előfordulások közel 1/3-a esetében a szó nem a prior szentimentértékében szerepel [10].

[18] *Őrízetben a brutális kettős gyilkosság gyanúsítottja*

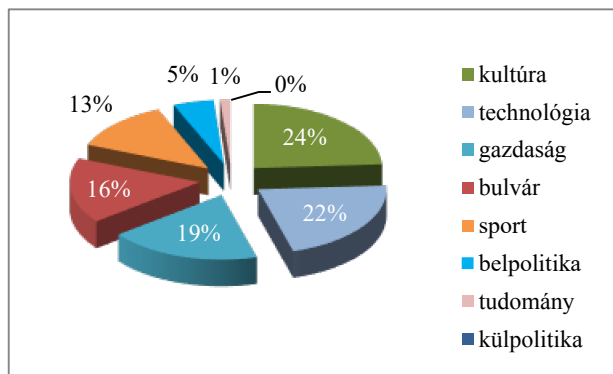
[19] *Brutális GDP-adatra izmosodik a forint*

Különösen figyelemre méltóak a kifejezésnek a technológia alkorpuszban mutatott megoszlásai arányai. Itt ugyanis alig találkozni prior polaritású előfordulással [11].

[20] *hamarosan a 370 lóerős R Evo is színre lép, brutális literteljesítménnyel*

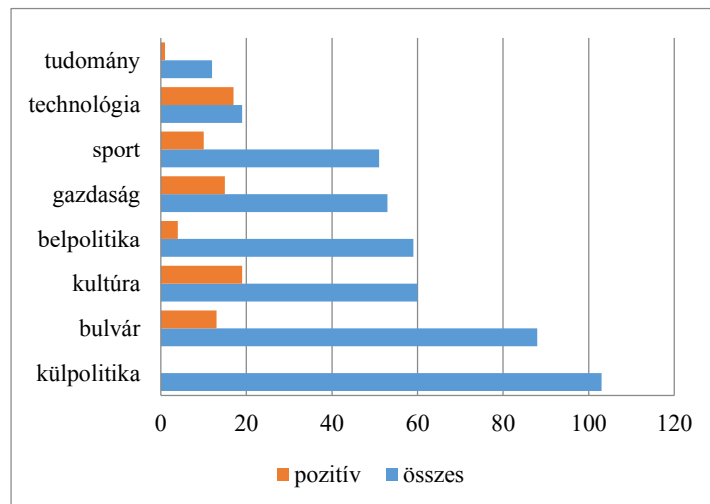


3. ábra. A *brutális* szó összes előfordulásának az aránya a Precorpusban



4. ábra. A *brutális* szó előfordulási aránya pozitív polaritású elemként a Precorpusban

A vizsgálat eredményei véleményem szerint arra mutatnak, hogy a polaritást váltott kifejezések olyan stiláris többletértékkel rendelkeznek, amelyet bizonyos domének nem, vagy csupán elvétve engednek meg, míg más domének hatáskeltési céllal gyakorta aknáznak ki.



5. ábra. A *brutális* szó összes, valamint pozitív polarítású előfordulási aránya a Precorpusban

4 Összegzés

A dolgozatban a polaritásváltás problémáját vizsgáltam meg behatóbban a szentimentelemzés nyelvtechnológiai feladata szempontjából.

A polaritásváltás mind nyelvtechnológiai, mind elméleti nyelvészeti szempontból figyelemre méltó jelenség, ugyanis alapjaiban kérdőjelezi meg annak a megközelítési módnak a helytállóságát, amely a nyelvi értékelést lexikális szinten, szemantikai komponensként tételezi (vö. Szabó 2015). A dolgozat hiánypótló, hiszen – bár az elméleti nyelvészetben számos szerző vizsgálja a polaritásváltást – a nyelvtechnológiai alkalmazások, így a szentimentelemzés a problémára nem fordít kellő figyelmet, és az elméleti nyelvészeti kutatások eredményeit sem aknázza ki.

A dolgozat elején, valós nyelvi adatok alapján áttekintettem a polaritásváltási jelenség szemantikai és szintaktikai sajátosságait. Ezt követően korpuszvizsgálatok segítségével meg próbáltam feltérképezni, hogy van-e olyan szintaktikai, szemantikai, doménfüggő, frekvenciabeli vagy kollokációs sajátosság, amely hathatós segítséget nyújthatna az aktuális polaritás automatikus felismerésében.

Vizsgálati eredményeim arra mutattak, hogy a polaritásváltás az intenzifikálók esetében igen gyakori jelenség, és amennyiben azokat nem tudjuk helyesen kezelni, a szentimentelemzés eredményében jelentős torzulást okozhatnak. Ugyanakkor – összevetve a negatív prior értékű intenzifikálók negatív és pozitív környezeti előfordulásait – azt találtam, hogy azok az elemek, amelyek pozitív környezetben előfordulnak, rendre megtalálhatók negatív környezetben is, továbbá frekvenciájuk tekintetében is hasonlóan viselkednek.

Ezt követően amellett érveltem, hogy az intenzifikálók aktuális polarításának a felismerésében a modifikált alaptag polarítására lehetne támaszkodni. Úgy vélem, abszolút polarítású alaptag esetében bizonyos szemantikai kompozíciós szabályok sikerrel volnának alkalmazhatók.

Ugyanakkor relatív polaritású alaptag esetében a fentebbi megoldást nem találtam kielégítőnek, tekintettel azok domén-, illetve targetfüggési sajátságára.

Mind a relatív polaritású alaptaggal alkotott szerkezetek, mind a szentimentkifejezés funkcióját betöltő polaritást váltó elemek esetében amellett érveltem, hogy kezelésükben doménfüggő sajátságaik nyújthatnának segítséget. Azt valószínűsítettem ugyanis, hogy mivel a vizsgált kifejezések váltott polaritás esetében stílári többlettel rendelkeznek, azok frekvenciája eltérő az egyes doménekben. Annak céljából, hogy a hipotézis helytállóságát megvizsgáljam, megnéztem a *brutális* szó előfordulási gyakoriságát nyolc különböző doménben, majd a kapott eredményeket összevettem a szó pozitív polaritású előfordulási arányával. Azt találtam, hogy a politikai tárgyú és a tudományos szövegekre alapvetően nem jellemző a *brutális* szó váltott polaritású használata. Ugyanakkor a kultúra és a gazdaság alkorpuszok szövegeiben, különösen a technológia doménben kifejezetten gyakori előfordulású.

A vizsgálat eredményei véleményem szerint arra mutatnak, hogy a polaritásváltás kezelésében szemantikai kompozíciós szabályok segíthetnének, és ahol azok alkalmazására nincs mód, ott a doménfüggő sajátságokat lehetne kiaknázni.

Irodalom

- Ahn, A., Laporte É., Nam J. 2012. Semantic Polarity of Adjectival Predicates in Online Reviews. In: Young-se, K., Jong-yuri Y., Jong-seon H., Jiun-shiung W., Seong-ha R., Kyoung-ae K., Dong-ho Ch., Kee-ho K., Hye-kyung K. (szerk.) *Seoul International Conference on Linguistics (SICOL '10)* (Seoul, South Korea, 2010. június 23–25). Seoul: Hankookmunhwasa. Elérhető: <http://arxiv.org/ftp/arxiv/papers/1211/1211.4161.pdf>. Letöltve: 2015. július 13.
- Andor, J. 2011. De durva ez a téma! – Megfigyelések a melléknévi polaritásváltásról. *Hungarológiai Évkönyv*, 12(1): 33–42.
- Baroni, M., Vegnaduzzo, S. 2004. Identifying Subjective Adjectives through Web-based Mutual Information. In: Buchberger, E. (szerk.) *Proceedings of KONVENS* (Bécs, 2004. szeptember 14–17). Bécs: GAI. 17–24.
- 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.
- 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, 2015. január 14–15.). Szeged: Szegedi Tudományegyetem. 227–234.
- Kamps, J., Marx M., Mokken R., Rijke M. 2004. Using WordNet to Measure Semantic Orientations of Adjectives. In: *Proceedings of the fourth international conference on Language Resources and Evaluation* (Lisszabon, 2004. május 26–28.). Párizs: ELRA. 1115–1118.
- 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. 2012. *Sentiment Analysis and Opinion Mining*. Kézirat. Elérhető: <http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>. Letöltve: 2015. 06. 11.
- Miháltz, M. 2010. OpinHu: online szövegek többnyelvű véleményelemzése. In: Tanács, A., Vincze, V. (szerk.) *VII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY 2010)* (Szeged, 2010. december 2–3.). 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, 2013. január 7–8.). Szeged: Szegedi Tudományegyetem. 343–345.

- Moghaddam, S., Popowich, F. 2010. Opinion polarity identification through adjectives. *CoRR abs/1011.4623*. Elérhető: <http://arxiv.org/ftp/arxiv/papers/1011/1011.4623.pdf>. Letöltve: 2015. július 13.
- Szabó, M. K. 2014. Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai. „*Nyelv, kultúra, társadalom*” című alkalmazott nyelvészeti konferencia (Budapest, 2014. november 3–4.). Publikálásra benyújtva.
- Szabó, M. K. 2015. A nyelvi értékelés mibenléte a számítógépes értékelélemzés (szentimentelemzés) szempontjából. *Nyelvészdoktoranduszok 18. Országos Konferenciája (LingDokKonf)* (Pécs, 2014. november 20–21.). Publikálásra benyújtva.
- Székely, G. 2007. *Egy sajátos nyelvi jelenség, a fokozás*. Budapest: Tinta Könyvkiadó.
- Tolcsvai Nagy, G. 1988. A mai magyar nyelv normarendszerének egy jelentős változásáról az „ifjúsági nyelv” kapcsán. *Magyar Nyelvőr*, 112(4): 398–406.
- Vázquez, S., Bel Rafecas, N. 2012. A Classification of Adjectives for Polarity Lexicons Enhancement. In: Calzolari, N., Choukri K., Declerck T., Doğan M. U., Maegaard B., Mariani J., Moreno A., Odijk J., Piperidis S. (szerk.) *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 12)* (Isztambul, 2012. május 21–27.). Párizs: ELRA. 3557–3561.
- Vegnaduzzo, S. 2004. Acquisition of subjective adjectives with limited resources. In: *AAAI Spring Symposium Technical Report: Exploring Affect and Attitude in Text: Theories and Applications* (Palo Alto, 2004. március 22–24.). Menlo Park: AAAI Press.
- Витехновский, В. И. 2013. Автоматизация определения тональности текста. *Вестник магистратуры*, 5: 34–36.
- Прохоров, А., Керимов, А. 2012. Сентимент-анализ и продвижение в социальных медиа. *КомпьютерПресс*, 7: 98–105.

Források

- PrecoSent = 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)* (Szeged, 2015. január 14–15). Szeged: Szegedi Tudományegyetem. 219–226.
- Precorpus = Szabó, M. K. 2014. Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai. „*Nyelv, kultúra, társadalom*” című alkalmazott nyelvészeti konferencia (Budapest, 2014. november 3–4.). Publikálásra benyújtva.