

# A beszélői szubjektívitas vizsgálata szentiment- és emóciókorpuszokon

## Absztrakt

### Bevezetés

Az interakció során megjelenő érzelmek kifejezését a pszichológia tudománya részletesen vizsgálja, azonban a nyelvi elemek emotív, illetve szubjektív tartalmának megismerésével kevesebb kutatás foglalkozik. A jelenségeket a nyelvtechnológiában részben a szentiment- és emócióelemzés vizsgálja (vö. pl. Liu 2012, Mulcrone 2012, Szabó et al. 2016a, Szabó et al. 2016b), ugyanakkor elenyésző azon kutatások száma, amelyek ezen jelenségeket, e részben rokon területeket összefüggéseiben vizsgálná. Legjobb tudomásunk szerint a magyar nyelv vonatkozásában az eddigi eredményeink úttörőnek számítanak.

Korábbi dolgozatunkban (Drávucz et al. 2017) szentimentszótárak segítségével emóciókorpust, és emóciószótárak segítségével szentimentkorpust elemeztünk. Megállapítottuk, hogy a szentimentkorpuszunkban annotált szentimenteknek csupán negyede volt jelezhető az emóciószótárainkkal. E tapasztalat alapján úgy döntöttünk, megvizsgáljuk, vajon a szentimentek, az emóciók, valamint a beszélői szubjektívitas milyen összefüggései tárhatóak fel automatikus eszközökkel.

### Módszertan

A fenti jelenségek automatikus felismerésére a természetesnyelv-feldolgozás (NLP) eszköztárából a szótárillesztés módszerét alkalmaztuk, amely például a gépi tanulás, szintaktikai elemzésen alapuló mintaillesztés mellett egyszerűbb információkinyerési módszer. A beszélő megnyilatkozásának a szubjektív elemeit vizsgáló kutatás során egy bizonytalanságjelölő elemeket tartalmazó szótárt illesztettünk olyan magyar nyelvű korpuszokra, melyeket korábban kézzel annotáltunk az emóciók, illetve a szentimentek szintjén. Vizsgálatunkhoz olyan elemcsoportra volt szükségünk, amelynek tagjai a beszélői szubjektívitas fontos indikátorai lehetnek, ezért választottuk kutatásunkhoz a bizonytalanságot kifejező elemeket.

Az emóciókorpusz szöveganyagát tévés és mozis témájú blogoldalakról származó, különböző terjedelmű és szerzőségű kritikákból, hírekből, valamint kommentekből állítottuk össze. A szentimentkorpuszunkat termékvélemény-szövegek alkotják.

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

A vizsgálat legfontosabb feltételezése az volt, hogy a nyelvi bizonytalanság szoros összefüggést mutat a beszélői szubjektívítással, ily módon az emóciófragmentumokban magasabb frekvenciával fognak szerepelni, mint a szentimentfragmentumokban.

## Eredmények, következtetések

A nyelvi bizonytalanság elemeit mind az emóció-, mind a szentimentkorpuszon illesztettük. A bizonytalanságjelölőket 7 csoportra osztva alkalmaztuk, azok szemantikai tartalma alapján (vö. Vincze 2014).

Az emóciókorpuszunk összesen 7205 tokent (központozással) tartalmaz. Ebből összesen 330 olyan kifejezést azonosítottunk az emóciókorpuszban, amely valamely emócióra annotált fragmentumon belül fordul elő, és bizonytalanság meglétére utal. Ez a teljes elemszám 4,58%-a. A legtöbb bizonytalanságjelölő elem az öröm és a bánat érzelmek kategóriában jelent meg: majdnem minden 3. bizonytalanságjelölő elemet az öröm, minden 4.-et a bánat emóciótaggel annotált fragmentumban azonosítottunk. A bizonytalanságjelölők ritkábban jelentek meg a tipikusan impulzív érzelmek (düh, undor, meglepődés) esetében.

A bizonytalanságjelölők szótáraival elemzett szentimentkorpusz 251 202 tokent (központozással) tartalmaz. A korpuszban összesen 15675 szentimentfragmentum van, ebből 8465 (54,00%) negatív és 7210 (45,99%) pozitív polaritású. Összesen 4553 olyan kifejezést azonosítottunk a szentimentkorpuszban, amely valamely szentimentre annotált fragmentumon belül fordul elő, és bizonytalanság meglétére utal. Ez a teljes elemszám 1,81%-a.

Az eredmények alapján a következő legfontosabb megállapításokat tettük: 1) Az emóciókorpuszban a tokenek számához viszonyítva jelentősen nagyobb a bizonytalanságjelölők frekvenciája. Mivel azt feltételeztük, hogy a bizonytalanságjelölők a beszélői szubjektivitással szoros összefüggést mutat, ezt az eredményt a feltételezésünk részbeni alátámasztásaként értékelhetjük. 2) A különböző típusú bizonytalanságjelölők eltérő megoszlást mutatnak a pozitív és a negatív szentimentfragmentumokban. Ezt a sajátosságot részletesen vizsgáljuk az előadásban. 3) Mind az emóció-, mind a szentimentkorpuszban a negatív polaritású fragmentumokban jóval több bizonytalanságjelölőt azonosítottunk, mint a pozitív polaritású fragmentumokban. Ez utóbbi tapasztalat véleményünk szerint azt igazolja, hogy a negatív véleményünket árnyaltabban, kevésbé direkt módon fejezzük ki.

## Irodalom

Drávucz, F., Szabó, M. K., Vincze V. Szentiment- és emóciószótárak eredményességének mérése emóció- és szentimentkorpuszokon. Benyújtva: *XIII. Magyar Számítógépes Nyelvészeti Konferencia*. Szeged, 2017. január 26-27.

Liu, B. 2012. *Sentiment Analysis and Opinion Mining*. Draft. (<http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>)

Mulcrone, K. 2012. Detecting Emotion in Text. Elhangzott: UMM CSci Senior Seminar Conference. Amerikai Egyesült Államok, University of Minnesota, Morris. 2012. ápr. 28. ([https://wiki.umn.edu/pub/UmmCSciSeniorSeminar/Spring2012Talks/KaitlynM ulcrone.pdf](https://wiki.umn.edu/pub/UmmCSciSeniorSeminar/Spring2012Talks/KaitlynM%20ulcrone.pdf))

Szabó M.K., Vincze V., Simkó K., Varga V., Hangya V. 2016a. 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. Portoroz, European Language Resources Association (ELRA). 2873-2878.

Szabó M.K., Vincze V., Morvay G. 2016b. Magyar nyelvű szövegek emócióelemzésének elméleti nyelvészeti és nyelvtechnológiai problémái. Megjelenés előtt

Vincze, V. 2014. [Uncertainty Detection in Hungarian Texts](#). In: *Proceedings of COLING 2014*, Dublin. 1844-1853.

