

SUMMARY OF PHD THESIS

MARTINA KATALIN SZABÓ



Problems of Sentiment Analysis in Hungarian Texts, with Particular Regard to Loss of Sentiment Value and Value Changing

Supervisors:

Károly Bibok, PhD

Veronika Vincze, PhD

University of Szeged

Doctoral School in Linguistics

Ph.D. Programme in Hungarian Linguistics

June 2018

Introduction

Broad and narrow scope of the research

In this thesis, one of the cardinal difficulties of automatic sentiment analysis, the sentiment shift is examined from a theoretical and computational linguistic point of view. Sentiment analysis is a subdivision of computational linguistics and it belongs to content analysis tasks. The goal of sentiment analysis is to automatically detect information of texts carrying positive or negative evaluative meaning toward a given target. The task is presented with the help of the sample below:

- (1) a. Sentiment analysis is a very interesting task.

In the sample (1) the target is the expression *Sentiment analysis* which is evaluated as a very interesting task. I consider the word *interesting* as an element having inherently positive evaluative meaning at lexical level. As a result, the target of the examined utterance is evaluated positively and this value is intensified with the help of the element *very*. In example (1b) the scheme of the analysis is provided:

- (1) b. [_{targ}Sentiment analysis] is a [_{szen} [_{int}Very] interesting] task.

There are a lot of theoretical and applied linguistic challenges concerning sentiment analysis, however, there are only few papers that aim to examine them thoroughly. One of the cardinal problems from the application point of view is the so-called *sentiment shift* that can lead to unsatisfactory results especially when the automatic analysis is carried out with the help of dictionary-based method. In the case of *sentiment shift* the lexical sentiment value of a given word differs from the sentiment value of the whole utterance including the given word (cf. Szabó et al. 2016a, 2017a).

In this thesis, two types of sentiment shifts are examined with particular regard to *loss of sentiment value* and *value changing*. The *loss of sentiment value* is a language phenomenon when an element carrying negative evaluative semantic content fully or partially loses its meaning when functioning as an intensifier of a sentiment word (2a). In the case of *value changing* a sentiment word inherently carrying negative evaluative semantic content

expresses not negative evaluative meaning but a positive opinion or surprise in the actual context (2b) (cf. Szabó 2015a: 632, 2015c: 52–54). Elements examined are emphasized in bold font in the examples:

- (2) a. **crazy** big, **damn** good
b. the concert was **brutal**

Both of the phenomena are interesting from a theoretical and applied linguistic, as well as language technological point of view. The *loss of sentiment value* and *value changing* have not been analysed exhaustively and the automatic handle of these language phenomena is still far from sufficient (cf. Andor 2011, Kugler 2014, Szabó 2015a: 52, Szabó et al. 2017a: 252, Drávucz et al. 2017: 236). Lexicon-based automatic sentiment analysers produce false results due to their negative semantic content at lexical level.

The research gap

Most of the international and Hungarian papers (e.g. Berend–Farkas 2008, Loughran–McDonald 2011, Liu 2012, Ruppenhofer–Rehbein 2012, Miháltz 2010, 2013, Hangya et al. 2015, 2017) examine the problems of sentiment analysis from the point of view of language technology, disregarding their applied and theoretical linguistic aspects. I do not know of any paper that aims to examine these challenges thoroughly. Researchers focus mainly on only one part of the task, for instance negation (cf. Na et al. 2004, Kennedy–Inkpen 2006, Wilson et al. 2009). Only few papers examine these problems (cf. Tolcsvai Nagy 1988, Partington 1993, Wierzbicka 2002, Jing-Schmidt: 2007, Laczkó 2007, Andor 2011, Kugler 2014, Dragut–Fellbaum 2014) and most of them discuss the phenomenon of the loss of sentiment value exclusively. Moreover, there is merely one paper (Dragut–Fellbaum 2014) concerning the loss of sentiment value from a language technological point of view, and none that would discuss value changing.

As for Hungarian language – aside from a short mentioning (Hangya et al. 2015: 212) – Szabó (2015a, 2015c, 2016, 2017) and Szabó et al. (2017a, 2017b) examine the phenomena exclusively.

The main goals and novelties of the PhD research

- One of the main goals of the PhD research was to do a thorough and multifarious analysis taking into consideration theoretical, applied linguistic and language technological aspects.
- The particular examination of the two special types of sentiment shift, the loss of sentiment value and the value changing from the point of view of theoretical linguistics, psychology and semantic motivation, was also one of the main purposes of this work.
- I considered it important to do the analyses based on a substantial amount of language data in order to reveal the quantitative and qualitative features of the phenomena.
- One of the focal points of the research was to elaborate the semantic representation of elements losing their sentiment value when functioning as intensifier. During this work, I use a theoretical framework of lexical pragmatics and spreading activation theory of semantic processing in the mental lexicon. I wished to create an automatic sentiment analyser based on the results of the examination to provide a more efficient way of processing Hungarian texts compared to the lexicon-based method. It can also be used to handle the phenomena of loss of sentiment value and value changing to some extent.

The structure of the thesis and the main results of the research

1. The challenge of defining *evaluation in language*

One of the primary tasks of automatic sentiment analysis is to define the phenomenon of evaluation in language, in other words, to consider what evaluative semantic content means. In spite of the importance of this step, neither the international, nor the Hungarian specialized literature on computational linguistics puts adequate emphasis on it.

In this part of the thesis the phenomenon of evaluation in language is thoroughly examined from a theoretical as well as computational linguistic point of view (cf. Szabó 2016). One of the main goals of the analysis was to determine whether statements of theoretical linguistics concerning the abovementioned question can provide an appropriate basis for language technological applications. I established that most of the authors consider the evaluation as a semantic component in the semantic structure of a word (cf. Berényi 1988, Péter 1991b, Szilágyi 1996) and connect the phenomenon to the notion of modality (cf. Kiefer 1986, Berényi 1988, Murvai 2001, Pete 2002, Péter 2008). Besides this, authors relate it to the phenomenon of speaker's subjectivity (cf. Kiefer, 1986, Berényi 1988, Szilágyi 1996) and only one paper (Péter 1991b) makes mention of objective type of evaluation in language. It is also worth noticing that in international specialized literature the term *sentiment analysis* is frequently used as the equivalent of the term *subjectivity analysis* (e.g. Wilson et al. 2005, Liu 2010: 627). Based on the multiphasic examination of the phenomenon of evaluation in language I concluded that the statements of theoretical linguistics concerning this language phenomenon cannot provide an appropriate basis for applied linguistic researches and language technological applications. In addition, a more significant scrutiny of the pragmatic features of texts is required in order to achieve a more appropriate automatic sentiment content analysis.

Results of this analysis were applied during the construction of the sentiment dictionary (see 5).

2. Analysis of the phenomena of loss of sentiment value and value changing

As the first step of my work, I clarified the terms related to the subject of evaluation in language and some gaps in terminology had been filled.

I established the followings about the statements of theoretical linguistics concerning the loss of sentiment value and value changing: authors do not describe the phenomena thoroughly and the statements are not based on the inspection of a proper amount of language data. What is more, descriptions in specialized literature are often vague and not adequate (cf. Tolcsvai Nagy 1988, Nemesi 1998, Székely 2007, Jing-Schmidt 2007, Balogh 2009, Andor 2011, Kugler 2014).

As for the psychological motivation of the existence of loss of sentiment value and value changing, I presented three phenomena (*negativity bias*, *positivity bias* or *the Pollyanna principle* and *congruence*) that can be linked to the phenomena in question (cf. Szabó 2017). Besides this, I examined the phenomena from the point of view of semantic prosody as well. I aspired to point out the dynamic interference of these tendencies. I also systematised the elements of the Hungarian language based on the system that Jing-Schmidt (2007) applied on the English language. This means that I classified the examined words based on their emotive semantic content (cf. six basic emotions, Ekman–Friesen 1969). I established that Hungarian words that can lose or change their sentiment value are semantically motivated by fear, anger or disgust. No element motivated by sadness was revealed during the analysis, which may be the result of negativity bias (cf. Öhman et al. 2001: 392–394). It is also worth noticing that authors examining these phenomena in the Hungarian language mention exclusively fear in this regard (cf. Nemesi 1998, Kugler 2014).

In the next step, the phenomena were analysed thoroughly from a semantic motivational point of view with the help of corpus and other language data (MNSZ2, Oravec et al. 2014) (cf. Szabó 2017).

As for the phenomenon of loss of sentiment value I established the followings: negative emotive intensifiers behave uniformly from the point of view of semantic motivation only when the word modified by the given intensifier carries inherent (positive or negative) evaluative meaning. In this case only the semantic content of intensity plays a role in the interpretation or – in case of interpersonal function of the expression – the emotive semantic component does as well. When the intensifiers examined modify words that do not have inherent (positive or negative) evaluative meaning, intensifiers have different semantic

contents: 1) only the semantic content of intensity comes to affect, 2) the negative semantic component is also active, or 3) the emotive semantic component plays an important role as well (in case of interpersonal function of the expression).

As far as value changing is concerned, I established that the speaker can express two types of information with it: positive evaluation or surprise on the communicated content of the utterance.

I also pointed out that all of the examined negative emotive language elements can function as intensifiers, but much fewer are able to change their value and function as a sentiment word to express positive evaluation or surprise. I have not found any element that would be able to change its sentiment value, but not able to function as an intensifier. I linked this phenomenon to the emotive semantic contents of the examined language elements.

In the next sub-chapters of the thesis, the phenomenon of loss of value (and partially value changing) was examined in the framework of lexical pragmatics and spreading activation theory of semantic processing in the mental lexicon. The goal of this research was to give an appropriate dictionary representation of the examined intensifiers that can also be applicable from a computational linguistic point of view.

During the research I used the theoretical framework of lexical pragmatics and I applied the principle of underspecified semantic representation relying on the representational solution of Bibok (2014, 2016). With the help of language data and based on the results of my previous examination concerning the problem of semantic motivation, I established the followings: The semantic component of intensity of the examined intensifiers is active in each case, therefore this is considered to be the core meaning of the word. In some of the contexts, this component contributes to the meaning of the utterance exclusively (cf. Szabó 2016: 168), while in other cases other semantic components build onto the core meaning during interpretation. Accordingly, the latter are only optional. The schematised lexical representation is presented below:

[intensive] ([affective]) ([negative]),

where the round parentheses denote optionality.

I assumed a separate lexical item for each examined element functioning as an intensifier. However, I argue that the spreading activation theory of semantic processing in the mental lexicon provides an opportunity to assume only one lexical item in case of the different usages, and then to model the interpretation mechanism with the help of spreading

activation. This approach also helps handling those kind of problems (for instance the question of processing time) that the applied lexical pragmatic framework cannot appropriately account for (cf. Wilensky 1987: 20, Gibbs 1994, Nemesi 2007, Forgács et al. 2012). Namely, this model enables us to assume that the choice of the adequate lexical representation and the semantic components are driven by the spreading activation in the mental lexicon (cf. Caramazza (1999), quoted by Kovács (2013: 187), Huszár (2005: 63–64)). Taking into consideration the advantages of the model I examined the phenomenon of loss of value in this framework as well, and I represented the interpretation mechanism of these elements with the help of the spreading activation theory.

The results of the analysis were applied during the construction of the NegEmotív corpus (see 4.), for the word association examination (see 4) and the creation of the sentiment analyser tool (see 5.).

3. Automatic sentiment analysis

In this part of the thesis I presented the goal and the importance of this computational linguistic task. I also gave a detailed description of the scientific and technological challenges that encourage the international and Hungarian researches concerning sentiment analysis. The main problems, methods and tools were presented related to the results of international research works. Papers on the automatic processing of Hungarian texts were reviewed exhaustively. The phenomenon of sentiment shift was discussed in a separate sub-chapter.

This part of the thesis served as a theoretical foundation for the database-construction tasks (see 4) and the creation of the sentiment analyser tool (see 5).

4. Examination of sentiment shift with the help of databases

The examination of sentiment shift was carried out with the help of three research tools that were created for the purpose of this research specifically.

In the first part of this chapter I gave a detailed description of the important role of different databases (corpora and association databases) in linguistic researches and developments, and I also discussed their types, the methods and tools of creating databases, and the quality assurance solutions. Then I presented the three main tools of the given research: 1) the sentiment corpus annotated manually on the level of fragments and targets

(Dívány corpus, cf. Szabó–Vincze 2015, Szabó et al. 2016a, 2016b), 2) the corpus annotated manually on the phenomena of loss of value and value changing (NegEmotív corpus, cf. Szabó et al. 2017a, 2017b) 3) and a word association database that contains stimulus words (the words used in association tests to evoke a response) that are able to lose or change their sentiment value, and the response words (associations) (cf. Szabó et al. 2017d). In order to create comparable data, stimulus words were not just negative emotive elements but other kinds of words (for instance, positive or neutral) as well. An important outcome of the research is that all three niche databases were first created in the framework of the given PhD research.

I presented the purpose and the goal of the creation of databases, the principles, the course and the tools of the work, and the applied quality assurance solutions. Concerning the corpora, I demonstrated the features of the processed texts and, in connection with this, I motivated the choice of text genres. Finally, I presented the basic and detailed statistic data of each databases together with examples, as well as the most relevant quantitative and qualitative results of the analysis from the point of view of the given thesis. I examined the different types of sentiment shift thoroughly based on the Dívány corpus. The NegEmotív corpus enabled me to analyse the phenomena of loss of value and value changing specifically, and I examined the results from a gender point of view as well. The conclusions of the analysis correlate with the results of the previous examination on semantic motivation, and contradict some of the statements of some papers scrutinizing the phenomena (cf. e.g. Jing-Schmidt 2007, Balogh 2009). Data of the word association database were analysed from the point of view of distribution of parts of speech, sentiment and emotive semantic contents, associative overlap, as well as patterns of gender differences. I would like to emphasize that one of the results is the fact that although the examined negative emotive elements call for response words with negative semantic content most of the time, in some cases patterns denoting lexicalisation process are observable as well.

Based on the results of these analyses the automatic sentiment analyser tool was created (l. 5), and the results were also compared to the problem of lexical representation (see 2).

5. Development of a sentiment analyser

In this chapter of the thesis a sentiment analyser tool is presented that renders a sentiment score to a given sentence using different types of dictionaries and combination rules. Rules

were created based on the experiences of the analysis and dictionaries were also composed during the research (cf. Szabó 2014, 2015b). In order to provide an opportunity to measure the efficiency of the tool a sentiment corpus annotated manually at sentence level was created.

I compared the efficiency of the analyser with the manual annotation, and the efficiency of the dictionary-based sentiment analysis method, and I concluded that the application of combination rules enables a much better quality of processing than the simple dictionary-based method. What is more, the tool can manipulate some types of loss of value and value changing adequately.

As the final step of the work, a user interface for the analyser tool was also created which enables us to test the efficiency of the software on texts from different domains.

References:

- Andor József 2011. De durva ez a téma! – Megfigyelések a melléknévi polaritásváltásról. [What a sick topic! – Observations on polarity changing of adjectives.] In *Hungarológiai Évkönyv* **12**. 33–42.
- Balogh, P. 2009. *Gender-markerek a nyelvben*. [Gender markers in language.] (<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. [Terminology and subjectivity]. *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? [How many levels of processing are in lexical accessing?] In Bánréti Z. ed. *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. [Measuring the efficiency of sentiment and emotion dictionaries on emotion and sentiment corpora.] In Vincze V. ed. *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 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. [Entity oriented opinion mining from online news.] In Tanács A. – Varga V. – Vincze V. eds. *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*. [From thought to word.] 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. [On the notion of modality.] *Nyelvtudományi közlemények* **88**(1–2). 1–38.
- Kovács, J.L. 2013. *Fogalmi rendszerek és lexikai hálózatok a mentális lexikonban*. [Concept systems and lexical networks in the mental lexicon.] Budapest, Tinta.

- Kugler N. 2014. A nyelvi polaritás kifejezésének egy mintázata, avagy milyen a félelmetesen jó? [A pattern of expression of language polarity, or what does *frighteningly good* mean?] *Magyar Nyelvőr* **138**(2). 129–139.
- Laczkó M. 2007. Napjaink tizenéveseinek beszéde szóhasználati jellemzők alapján. [Speech of present-day teenagers based on attributes of word usage.] *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. [OpinHu: multilingual opinion mining on online texts.] In Tanács A.–Vincze V. eds. *VII. Magyar Számítógépes Nyelvészeti Konferencia* (MSZNY 2010). Szeged, Szegedi Tudományegyetem. 14–23.
- Miháltz M. 2013. OpinHuBank: szabadon hozzáférhető annotált korpusz magyar nyelvű véleményelemzéshez. [OpinHuBank: an open accessible annotated corpus for opinion mining in Hungarian.] In Tanács A.–Vincze V. (eds.) *IX. Magyar Számítógépes Nyelvészeti Konferencia* (MSZNY 2013). Szeged, Szegedi Tudományegyetem. 343–345.
- Murvai O. 2001. Szemantikai és pragmatikai kiegészítések a modalitás vizsgálatához. [Semantic and pragmatic complements to the examination of modality.] 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). Würzburg, Germany, Ergon Verlag. 49–54.
- Nemesi A.L. 1998. Társalgási divatkifejezések a benyomáskeltés szolgálatában. [Conversational trendy expressions for the purpose of conveying an impression.] *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. [Hierarchy of the three main types of language modality in the Hungarian language.] *Magyar nyelv* **98**(2). 173–191.
- Péter M. 1991b. A nyelvi érzelemkifejezés eszközei és módjai. [Tools and ways of expressing emotions in language.] Budapest, Tankönyv.
- Péter M. 2008. A magyar aspektusról – más aspektusból. [On aspect in the Hungarian language – from a different aspect.] *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*. [Experiences of creating a Hungarian sentiment lexicon.] 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. [Possibilities of handling polarity changing and shift in sentiment analysis.] *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. [Experiences and problems of creating a Hungarian sentiment lexicon.] In Gecső T.–Sárdi Cs. eds. *Nyelv, kultúra, társadalom. Segédkönyvek a nyelvészet tanulmányozásához* 177. 278–285.
- Szabó M.K. 2015c. A polaritásváltás problémája a szentimentelemzés szempontjából. [The problem of polarity shift from the point of view of sentiment analysis.] In Váradi T. ed. *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. [The challenge of defining evaluation in language.] In Gécseg Zs. ed. *LingDok 15. Nyelvésztdoktoranduszok 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. [Examination of the phenomena loss of value and value changing from psychological and semantic motivational points of view.] *MANYE 26. konferenciakötete*. Megjelenés előtt.
- Szabó M.K.–Vincze V. 2015. Egy magyar nyelvű szentimentkorpusz létrehozásának tapasztalatai. [Experiences of ceating a Hungarian sentiment corpus.] In Tanács A.–Varga V.–Vincze V. eds. *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. [Aspect-level annotation and sentiment shifters in a Hungarian sentiment corpus.] In Tanács A.–Varga V.–Vincze V. eds. *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. [Examination of sentiment shift based on a corpus annotated at semantic-pragmatic level.] In Vincze V. ed. *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. [Examination of negative emotive elements from a gender point of view.] In Ludányi Zs. ed. *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. [Word association database of negative emotive elements.] *MANYE* 26. konferenciakötete. Megjelenés előtt.
- Székely G. 2007. Egy sajátos nyelvi jelenség, a fokozás. [A special language phenomenon, the intensification.] *Segédkönyvek a nyelvészet tanulmányozásához* 66. Budapest, Tinta.
- Szilágyi N.S. 1996. *Hogyan teremtsünk világot?* [How to create a world?] Kolozsvár, Erdélyi Tankönyvtanács.
- Tolcsvai N.G. 1988. A mai magyar nyelv normarendszerének egy jelentős változásáról az „ifjúsági nyelv” kapcsán. [On a notable change in the norm system in connection to the language of the youth.] *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.

Publications related to the thesis:

- 1) Szabó M.K. 2014. *Egy magyar nyelvű szentimentlexikon létrehozásának tapasztalatai.* [Experiences of creating a Hungarian sentiment lexicon.] 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. [Possibilities of handling polarity changing and shift in sentiment analysis.] *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. [Experiences and problems of creating a Hungarian sentiment lexicon.] In Gecső T.–Sárdi Cs. eds. *Nyelv, kultúra, társadalom. Segédkönyvek a nyelvészet tanulmányozásához 177.* 278–285.
- 4) Szabó M.K. 2015c. A polaritásváltás problémája a szentimentelemzés szempontjából. [The problem of polarity changing from the point of view of sentiment analysis.] In Váradí T. ed. *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 értékelélemzés (szentimentelemzés) szempontjából. [The challenge of defining evaluation in language.] In Gécseg Zs. ed. *LingDok 15. Nyelvészdoktoranduszok 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. [Examination of the phenomena loss of value and value changing from psychological and semantic motivational points of view.] *MANYE 26. konferenciakötete.* Megjelenés előtt.
- 7) Szabó M.K.–Morvay G. 2015. Emócióelemzés magyar nyelvű szövegeken. [Emotion analysis of Hungarian texts.] In Gecső T.–Sárdi Cs. *Nyelv, kultúra, társadalom.* 286–292.
- 8) Szabó M.K.–Vincze V. 2015. Egy magyar nyelvű szentimentkorpusz létrehozásának tapasztalatai. [Experiences of creating a Hungarian sentiment corpus.] In Tanács A.–Varga V.–Vincze V. eds. *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 értékelélemzési feladatában. [Questions of handling negators during sentiment analytical tasks of Hungarian texts.] In Reményi A.Á.–Sárdi Cs.–Tóth Zs. eds. *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. [Experiences of creating a spontaneous Hungarian speech corpus.] *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. [Aspect-level annotation and sentiment shifters in a Hungarian sentiment corpus.] In Tanács A.–Varga V.–Vincze V. eds. *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*. [Theoretical linguistic and language technological problems of emotion analysis of Hungarian texts.] 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. [Examination of sentiment shift based on a corpus annotated at semantic-pragmatic level.] In Vincze V. ed. *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. [Examination of negative emotive elements from a gender point of view.] In Ludányi Zs. ed. *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. [Examination of negative emotive intensifier elements from the point of view of transtation from English to Russian and Russian to English.] In Ludányi Zs. ed. *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. [Word association database of negative emotive elements.] *MANYE 26. konferenciakötete*. Megjelenés előtt.
- 19) Drávucz F.–Szabó M.K.–Vincze V. 2017. Szentiment- és emóciószótárak eredményességének mérése emóció- és szentimentkorpuszokon. [Measuring the

efficiency of sentiment and emotion dictionaries on sentiment and emotion corpora.] In Vincze V. ed. *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 emóciósztárakkal. [Analysis of narrative structures in a corpus of Russian texts (NarRu) with internally developed dictionaries of emotion and sentiment expressions.] In Vincze V. ed. *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. [Quality assurance of typing and annotation process of HuTongue, the Hungarian spontaneous speech corpus.] In Vincze V. ed. *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. [The text network analysis and visualization of Rákosi-era reports]. In Vincze V. ed. *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.