A Hungarian Sentiment Corpus Manually Annotated at Aspect Level

Martina Katalin Szabó[1,2], Veronika Vincze[3,4],
Katalin Simkó[4], Viktor Varga[4], Viktor Hangya[4]

[1]Precognox Ltd
[2]University of Szeged, Institute of Slavonic Studies, Department of Russian Philology
[3]MTA-SZTE Research Group on Artificial Intelligence
[4]University of Szeged, Department of Informatics
mszabo@precognox.com, szabo.martina@lit.u-szeged.hu; vinczev@inf.u-szeged.hu;
kata.simko@gmail.com; wyvick@gmail.com; hangyav@inf.u-szeged.hu

Abstract

In this paper we present a Hungarian sentiment corpus manually annotated at aspect level. Our corpus consists of Hungarian opinion texts written about different types of products. The main aim of creating the corpus was to produce an appropriate database providing possibilities for developing text mining software tools. The corpus is a unique Hungarian database: to the best of our knowledge, no digitized Hungarian sentiment corpus that is annotated on the level of fragments and targets has been made so far. In addition, many language elements of the corpus, relevant from the point of view of sentiment analysis, got distinct types of tags in the annotation. In this paper, on the one hand, we present the method of annotation, and we discuss the difficulties concerning text annotation process. On the other hand, we provide some quantitative and qualitative data on the corpus. We conclude with a description of the applicability of the corpus.

Keywords: Hungarian language, sentiment analysis, text mining, corpus, manual annotation

1. Introduction

In this paper we present a Hungarian sentiment corpus manually annotated at the aspect level. First, we describe the method of annotation, and discuss the difficulties concerning text annotation process. Later, we present some quantitative and qualitative data on the corpus.

Opinions have essential influence on almost all human activities and behaviors (Liu, 2012). This is not only true for individuals but also true for organizations. For this reason, in recent years, sentiment analysis (also called opinion mining) is widely studied in data mining, web mining, and text mining. Since the early 2000s, sentiment analysis has grown to be one of the most active research areas in natural language processing (Liu, 2012). At the same time, in contrast to the international research, sentiment analysis of texts written in Hungarian has so far attracted less attention (Berend and Farkas, 2008; Miháltz, 2010; Hangya et al., 2015). As for Hungarian databases, to the best of our knowledge, only one Hungarian sentiment corpus has been made so far, the OpinHuBank corpus (Miháltz, 2013). However, this corpus does not contain deep sentiment annotation (Szabó and Vincze, 2015), therefore it can be used for linguistic research and NLP purposes only in a limited way.

Based on the above, we decided to create a Hungarian sentiment corpus, annotated manually at the level of fragments and targets, as well as for different types of sentiment modifiers, which we will present in this paper.

2. Literature review

Sentiment analysis is the field of language technology that analyzes people's opinions, evaluations, appraisals and attitudes towards entities such as products, services, organizations, individuals, issues or events (Liu, 2012). There are also many names of the tasks, e.g. opinion mining, opinion extraction, sentiment mining, subjectivity analysis and so on.

Within Computational Linguistics, works dealing with the topic of sentiment analysis have already become a relevant part of the academic research in the past few decades (Berend and Farkas, 2008). The significance of sentiment analysis is basically supported by economic interests (Ahn et al., 2012; Liu, 2012). People share their opinions about products on the web (e.g. in social media), therefore the number of documents expressing opinions is constantly growing. These opinions become important resources for those who need information about products, as well as manufacturers who wish to improve their productivity (Ahn et al., 2012). Therefore, the demand for efficient automatic extraction of opinions from the web is growing day by day. However, identifying polar phrases that express opinions towards a certain target seems to be an unsolved and challenging problem so far.

With the high increase in the number of projects aiming at an effective sentiment analysis, target-dependent opinion mining is becoming a widely studied task (Hu and Liu, 2004; Dong et al., 2014; Liu, 2012; Jiang et al., 2011). Basically, two features of opinionated texts render the target-dependent analysis more difficult. On the one hand, the document- and the sentence-level sentiment classifications are based on the assumption that each document or each sentence expresses only one definite opinion on a single target (Turney, 2002; Liu, 2012). As a consequence, these methods of analysis are not applicable to documents or sentences which evaluate more than one entity (see Section 3.1.). In addition, classifying opinion texts at the entity-

level is still insufficient for applications because they do not reveal the relations between entities and their different aspects in the texts, so they are unable to handle entities and aspects adequately. For instance, a positive opinion about an entity does not mean that the author has positive opinions about all aspects of the entity. Likewise, a negative opinion on one or some aspects of a given entity does not mean that the author is negative about the whole entity. Therefore, sentiment analysis needs to decompose the opinion target into entity and its aspects. Evidently, this type of analysis requires deeper NLP capabilities, hence it introduces a suite of technological and theoretical problems as well (Liu, 2012).

Nevertheless, many of the sentiment expressions are context-dependent in the opinionated texts (Ahn et al., 2012). That is, a given polarity word may indicate different opinions depending on its domain, or even within the very same domain, depending on the given entity or aspect that the polarity expression is connected to (Ahn et al., 2012). For instance, the adjective *long* expresses a positive opinion in a sentence like *The battery life is long* but it is negative combined with distinct product features in the sentence like *It takes a long time to focus* (Ding et al., 2008; Ahn et al., 2012). It is worth noting that both sentences could occur in the same domain. Since a great amount of sentiment words are ambiguous as their polarity is concerned, it is essential to adequately identify the actual sentiment value of the given polarity expression in different contexts.

Another challenging and often discussed problem is the rule of the so called sentiment shifters in sentiment analysis (Polanyi and Zaenen, 2006; Moilanen and Pulman, 2007; Choi and Cardie, 2009; Ding et al., 2008; Feldman et al., 2010; Loughran and McDonald, 2011; Ruppenhofer, 2013). Sentiment shifters are types of expressions that are used to change the sentiment orientations, for instance, from positive to negative or vice versa (Liu, 2012). One of the classes of sentiment shifters is made up of negation words. For example, even though the sentence *I don't like this camera* contains a positive polarity word (*like*), the sentiment value of the whole sentence is negative due to the negation of the positive polarity word. At the same time, it is essential to handle the sentiment shifters with care because not all occurrences of such words can change polarity. Consider the expression *not only... but also*, where the word *not* does not change sentiment orientation of the sentence (Liu, 2012).

All the problems mentioned here emphasize the essential importance of manually annotated corpora in sentiment analysis tasks. Corpora annotated manually at the sentiment level play a key role not just in training and testing algorithms but in theoretical and applied research concerning sentiment analysis as well (Pang et al., 2002; Dave et al., 2003; Finn and Kushmerick, 2003; Salvetti et al., 2004; Aue and Gamon, 2005; Bai et al., 2005; Wang and Wang, 2007; Boiy and Moens, 2009).

3. Annotation

3.1. Principles of annotation

Liu (2012) distinguishes three levels of granularities of the existing methods of sentiment analysis. In general, senti-ment analysis can be carried out at the level of the whole document. The task at this level is to determine whether a document expresses a positive or a negative sentiment (Liu, 2012; Turney, 2002). This task is commonly known as document-level sentiment classification (Liu, 2012). This method of analysis is based on the assumption that each document expresses only one definite opinion on a single target. As a consequence, the document-level approach is not applicable to documents which evaluate more than one entity.

The so called sentence-level sentiment classification determines whether each sentence expresses a positive or negative opinion, or it is neutral from the polarity point of view. However, we should note that many sentences can imply more than one opinion on more than one target. For this, researchers also analyzed clauses (Wilson et al., 2004), but "the clause-level method is still not enough to achieve adequate results" (Liu, 2012).

On the basis of the problems mentioned above, we can conclude that only analysis carried out on the level of entities and aspects can produce satisfactory results. Aspect level sentiment analysis (also called earlier as feature-level or feature-based opinion mining (Hu and Liu, 2004)), instead of looking at documents, sentences or clauses, directly looks at the opinion itself. This type of approach tries to process each polarity item in connection with its own target (Liu, 2012). One of the underlying principles of the entity-level analysis is that a sentiment can be expressed toward a definite entity, or an aspect of this entity, too. For example, although the sentence *although the service is not that great, I still love this restaurant* expresses a positive evaluation, we cannot conclude that the sentence is entirely positive. The differentiation of entities and their aspects is an important task of sentiment analysis, with regard to the fact that in many applications, opinion targets are described by entities and/or their different aspects (Liu, 2012). Thus, the goal of this level of analysis is to produce a structured summary of opinions about entities and their aspects.

Taking into consideration the above mentioned approaches and their drawbacks and advantages, we decided to carry out the annotation of the sentiment corpus on entity and aspect level (Szabó and Vincze, 2015; Szabó et al., 2016). The corpus created is a unique Hungarian database: to the best of our knowledge, no digitized Hungarian sentiment corpus that is annotated on this level has been made so far.

3.2. Methods of Annotation

The corpus is composed of Hungarian opinion texts written about different types of products, published on the homepage http://divany.hu/. The corpus is made up of 154 opinion texts, and comprises of approximately 17 thousand sentences and 251 thousand tokens.

In the first phase of the annotation we manually processed about one-fourth of the full dataset (Szabó and Vincze, 2015). In this phase of the work we annotated: 1) the whole constructions, expressing positive or negative opinion, 2) the sentiment words, expressing positive or negative opinion at the lexeme level, 3) the targets of the sentiment words, 4) elements, modifying the prior polarity (also called sentiment shifters or semantic orientation) of the sen-

timent words (about sentiment shifters see also Section 2. above).

The whole construction that we called sentiment fragment was annotated firstly in the raw texts of the corpus. We regarded an expression as a sentiment fragment if it contained only one polarity item connected to only one target (or more than one target, coordinated to each other) (Szabó et al., 2016). As a consequence, most of the fragments of the corpus annotated are whole sentences and phrases.

As far as sentiment words are concerned, not only one-word expressions but also multiword expressions were annotated in the corpus as positive or negative sentiment words.

The targets of the sentiment expressions were annotated with the same target tag in the first phase of the annotation. Product names that functioned as a title were annotated as topics.

Concerning sentiment shifters, we distinguished three subtypes: intensifiers, negations and irreal expressions. Intensifiers are used to increase (Example 1) or decrease (Example 2) the degree to which a term is positive or negative (Kennedy and Inkpen, 2006).

(1)   nagyon jó 'very good'

(2)   kevesbé jó 'less good'

We used different tags to distinguish the two types of intensifiers at the level of annotation (IntensifierPlus and IntensifierMinus).

Negations (Example 3) are used to reverse the semantic polarity of a particular term (Kennedy and Inkpen, 2006).

(3)   nem jó 'not good'

As we have already noticed (see Section 2. above), they often change sentiment orientation, but not always.

Irreal expressions (Example 4) were handled with distinct annotation-tags on the basis of their special effect on sentiment polarity: they may change sentiment orientation (partially or completely), but this is not always the case.

(4)   jó volna 'would be good'

These types of language elements are used to modify the certainty of the semantic content of a given expression, for instance, adverbs like *valószínűleg* 'probably' and verbs like *tűnik* 'seem' or *gondol* 'think' may carry irreal meaning.

After the test annotation, we analysed the annotated data, and on the basis of the analysis we reconsidered and modified some principles and methods of the first phase of annotation. In the second phase of the work the whole database has been processed, according to the new annotation principles.

The main difference of the method implemented in the second phase of annotation was that entities and their aspects were annotated with different tags (Target 1-20) and we applied the same tag for a given target in a given document of the corpus consistently. The modification of the annotation method was based on the revelation that a more elaborated and detailed annotation scheme was essential to reveal and automatically process the relationship of entities

and aspects, as well as to identify coreference relations in our future work.

The proper handling of the relationship of entities and aspects is one of the most important and difficult task in sentiment analysis. Namely, the fact that a target of a given opinion is the entity or just an aspect of the entity plays an important role in determining the polarity of a sentiment expression. For instance, to correctly handle the sentence below it is necessary to reveal the connection between the entity *fényképezőgép* 'camera' and its aspect *ár* 'price'.

(5)   Bár az ára nem alacsony, megéri megvenni ezt a fényképezőgépet.
       'Although its price is not low, this camera is worth buying'.

As the example demonstrates, the negative opinion of the speaker is connected only to only one aspect of the given entity but he has positive evaluation of the whole product. The following example demonstrates a sample of the annotated corpus.

```
Negyedik helyezett:
<topic>
    Kolios kecskesajt
</topic>
<SentNeg>
    <target1>
        Állagra
    </target1>
    olyan, mint a
    <SentiWordNeg>
        gumi
    </SentiWordNeg>
</SentNeg>
,
<SentNeg>
    <target2>
    ízre
    </target2>
    pedig
    <SentiWordNeg>
        fanyar
    </SentiWordNeg>
</SentNeg>
.
<SentNeg>
    Nekem
    <ShiftNeg>
        nem
    </ShiftNeg>
    <SentiWordPos>
        jön be
    </SentiWordPos>
</SentNeg>
.
```

'Kolios goat cheese. Its density is rubber-like, it tastes tart. I don't like it.'

4. Results

Here we report on the most important results of corpus annotation, both from a quantitative and a qualitative point of view.

4.1. Statistical Data on the Corpus

In Table 1 we offer the main statistical data on the annotated corpus.

On the basis of the data we can conclude the following: The frequency of the positive (SentiWordPos) and the negative sentiment words (SentiWordNeg) is approximately equal in the corpus. At the same time, this statistical data does not yield that the positive (PosSentiment) and the negative sentiment fragments (NegSentiment) would show the same frequency distribution: the negative sentiment fragment is a more frequent phenomenon than the positive one. From the point of view of computational linguistics, this notable difference emphasizes the important role of the sentiment shifters in opinion mining since sentiment shifters can shift the sentiment polarity of a sentiment word (for instance, negations change the original polarity of the sentiment word).

Moreover, it is also important to point out that negations, irreals and decreasers (IntensifierMinus) are more frequent in the negative fragments than in the positive ones; e.g. ***nem jó*** 'not good'; ***kevésbé jó*** 'less good'; ***jó volna*** 'would be good'. Our results show that a negative opinion is more often expressed by positive sentiment words (e.g. ***nem jó*** 'not good') than a positive opinion with negative sentiment word (e.g. ***nem rossz*** 'not bad'). This result of the corpus analysis complies with the *Pollyanna hypothesis* (also called *positivity bias*), which asserts that "there is a universal tendency to use evaluatively positive words more frequently than evaluatively negative words in human communication" (Boucher and Osgood, 1969).

4.2. Negation

As for negation, the data also revealed that its main function in the texts is to change the polarity of the sentiment word. Also, we supposed at the beginning of the annotation that words that express negation can mostly be covered by negation words like *nem* or *sem* 'not', the negative forms of the copula like *nincs* or *sincs* 'is not' and some postpositions like *nélkül* 'without'. However, it turned out that negation is expressed by a wider variety of words and phrases than expected, for instance:

(6) *hiány* 'lack'
  elillan 'disappear'
  nélkülöz 'miss'
  bizarr lenne azt állítani 'it would be strange to say'
  helyett 'instead'
  semmi köze sincs 'it has nothing to do with'
  nulla 'zero'
  lespórol 'spare'

Altogether, there are 3516 negation words in the corpus, including 2587 adverbs, 468 verbs, 145 pronouns and 93 conjunctions.

4.3. Irreals

Irreals were mostly expressed by verbs and adverbs, about 66% of the data can be covered by these two parts-of-speech. In the case of adverbs, it is mostly the lexical content of the word that has an irreal meaning such as in *állítólag* 'allegedly', *valószínűleg* 'probably', *talán* 'perhaps'. As for verbs, their irreal content can also be determined at the lexical level: some verbs express uncertainty (*tűnik* 'seem', *hasonlít* 'resemble', *imitál* 'imitate') or subjectivity (*érez* 'feel', *gondol* 'think'). Furthermore, there are some morphological processes that can also encode uncertain or irreal meaning such as the suffix *-hat/-het*, which roughly corresponds to the English auxiliary *may* or conditional forms of the verb. These verbal forms are related to linguistic elements that refer to uncertainty (Vincze, 2014), and we intend to compare irreals and uncertain elements in a more detailed way in a future study.

4.4. Intensifiers

To get information about the proportion of the two different types of intensifiers between positive and negative fragments, we carried out a statistical analysis of these elements found in the corpus.

On the basis of the results we concluded that the two types of intensifiers together occur with the same frequency in positive (6693:2706) and negative (8053:3347) fragments, for instance, *nagyon jó* 'very good' and *nagyon rossz* 'very bad'. However, frequency of intensifiers with decreasing semantic content is not the same in the two types of sentiment fragments: they occur much more often in negative polarity fragments (8053:779) than in positive ones (6693:301). This is probably caused by the dissimilar semantic behavior of these elements: intensifiers with increasing semantic content (Example 7) cannot reverse the semantic polarity of a particular sentiment word, in contrast to intensifiers with decreasing semantic content (Example 8) (Szabó et al., 2016). See the word *jó* 'good' modified by different types of intensifiers below.

(7) nagyon jó 'very good' (the expression is still positive)

(8) kevésbé jó 'less good' (the intensifier changes the polarity from positive to negative)

In addition, it is worth mentioning that intensifiers with decreasing semantic content change positive and negative sentiment words differently. Namely, these elements invert the polarity of positive polarity items (Example 8), but it rarely changes the polarity of negative polarity items (Example 9).

(9) kevésbé rossz 'less bad' (the expression may be still negative depending on the context)

On the basis of the language phenomena mentioned above we can conclude that automatic processing of sentiment fragments containing intensifiers with decreasing semantic content need to carry out cautiously, especially when they modify positive polarity items.

tag	total frequency	frequency in positive sentiment fragments	frequency in negative sentiment fragments
PosSentiment	7200	-	-
NegSentiment	8442	-	-
SentiWordPos	8100	6247	1853
SentiWordNeg	8090	1347	6743
Topic	1371	-	-
Target	7867	3743	4124
Negation	3347	1385	1962
IntensifierPlus	5218	2538	2680
IntensifierMinus	1151	327	824
Irreal	942	273	669
OtherShifter	722	388	334
TOTAL:	**52455**	**16248**	**19189**

Table 1: The main statistical data on the annotated corpus.

lexicon	number of elements
SentiWordPos	2568
SentiWordNeg	3343
Target	2219
Negation	95
IntensifierPlus	744
IntensifierMinus	199
Irreal	195

Table 2: Statistical data of the lexicons generated from the corpus

4.5. Inter-annotator agreement rate

The corpus was annotated by two annotators with a 65.02% agreement rate. The agreement rate was by far the highest for topics, as they are mostly isolated titles in the texts; on the other hand, the agreement rate for irreals was far below average, this is probably because this category is hard to define and they vary a lot lexically and structurally.

5. Usability of the Corpus

Our corpus may be fruitfully applied in linguistic research as well as in developing and testing of sentiment analyzer software tools.

One of the most important advantages of the corpus was that the annotation made it possible for us to automatically generate dictionaries of different types of language expressions, which can be exploited in opinion mining.

In Table 2 we present the main statistical data on the generated lexicons. We expect that these lexicons will improve the results of the automatic sentiment analysis in the future.

6. Conclusions

Here we presented our Hungarian sentiment corpus manually annotated at the aspect level. From the point of view of computational linguistics, the automatic detection and handling of sentiments in texts encounter difficulties since the syntactic, semantic and pragmatic realization of opinions is far from uniform. Consequently, sentiment analysis poses challenges for computer processing.

The main aim of creating the corpus was to produce an appropriate database providing possibilities for developing software tools, automatically extracting and processing sentiments of texts. Our corpus is a unique Hungarian database: In contrast with the presented project, no previous digitized Hungarian sentiment corpus that is annotated on the level of fragments and targets has been made so far. We hope that our manually annotated database can be fruitfully applied in theoretical and computational linguistic research, connected to opinion mining from texts written in Hungarian. In addition, the corpus can be efficiently exploited in development and testing of sentiment analyzer software tools as well.

7. Bibliographical References

Ahn, A., Laporte, É., and Nam, J. (2012). Semantic polarity of adjectival predicates in online reviews. *CoRR*, abs/1211.4161.

Aue, A. and Gamon, M. (2005). Customizing sentiment classifiers to new domains: A case study. In *Proceedings of Recent Advances in Natural Language Processing (RANLP)*.

Bai, X., Padman, R., and Airoldi, E. (2005). On learning parsimonious models for extracting consumer opinions. In *System Sciences, 2005. HICSS '05. Proceedings of the 38th Annual Hawaii International Conference on*, pages 75b–75b, Jan.

Berend, G. and Farkas, R. (2008). Opinion mining in hungarian based on textual and graphical clues. In *Proceedings of the 8th conference on Simulation, modelling and optimization*, pages 408–412. World Scientific and Engineering Academy and Society (WSEAS).

Boiy, E. and Moens, M.-F. (2009). A machine learning approach to sentiment analysis in multilingual web texts. *Inf. Retr.*, 12(5):526–558, October.

Boucher, J. and Osgood, C. E. (1969). The pollyanna hypothesis. *Journal of Verbal Learning and Verbal Behavior*, 8(1):1 – 8.

Choi, Y. and Cardie, C. (2009). Adapting a polarity lexicon using integer linear programming for domain-specific

sentiment classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2 - Volume 2*, EMNLP '09, pages 590–598, Stroudsburg, PA, USA. Association for Computational Linguistics.

Dave, K., Lawrence, S., and Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th International Conference on World Wide Web*, WWW '03, pages 519–528, New York, NY, USA. ACM.

Ding, X., Liu, B., and Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, WSDM '08, pages 231–240, New York, NY, USA. ACM.

Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., and Xu, K. (2014). Adaptive recursive neural network for target-dependent twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 2: Short Papers*, pages 49–54.

Feldman, R., Rozenfeld, B., and Breakstone, M. Y. (2010). Ssa – a hybrid approach to sentiment analysis of stocks. In *The Israeli Seminar for Computational Linguistics*, ISCOL.

Finn, A. and Kushmerick, N. (2003). Learning to classify documents according to genre. In *In IJCAI-03 Workshop on Computational Approaches to Style Analysis and Synthesis*.

Hangya, V., Farkas, R., and Berend, G. (2015). Entitásorientált véleménydetekció webes híranyagokból [Entity-oriented opinion mining from web news]. In *XI. Magyar Számítógépes Nyelvészeti Konferencia*, pages 227–234.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.

Jiang, L., Yu, M., Zhou, M., Liu, X., and Zhao, T. (2011). Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT '11, pages 151–160, Stroudsburg, PA, USA. Association for Computational Linguistics.

Kennedy, A. and Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110–125.

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

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries and 10-ks. *The Journal of Finance*, 66.

Miháltz, M. (2010). OpinHu: online szövegek többnyelvű véleményelemzése. [OpinHu: multilingual opinion mining from online texts]. In *VII. Magyar Számítógépes Nyelvészeti Konferencia*, pages 14–23.

Miháltz, M. (2013). OpinHuBank: szabadon hozzáférhető annotált korpusz magyar nyelvű véleményelemzéshez [OpinHuBank: a freely available annotated corpus for Hungarian opinion mining]. In *IX. Magyar Számítógépes Nyelvészeti Konferencia*, pages 343–345.

Moilanen, K. and Pulman, S. (2007). Sentiment composition. In *Proceedings of Recent Advances in Natural Language Processing (RANLP 2007)*, pages 378–382, September 27-29.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of EMNLP*, pages 79–86.

Polanyi, L. and Zaenen, A., (2006). *Computing Attitude and Affect in Text: Theory and Applications*, chapter Contextual Valence Shifters, pages 1–10. Springer Netherlands, Dordrecht.

Ruppenhofer, J. (2013). Anchoring sentiment analysis in frame semantics. *Veredas*, pages 66–81.

Salvetti, F., Lewis, S., and Reichenbach, C. (2004). Impact of lexical filtering on overall opinion polarity identification. In *in Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*.

Szabó, M. K. and Vincze, V. (2015). Egy magyar nyelvű szentimentkorpusz létrehozásának tapasztalatai. [Lessons learnt from creating a Hungarian sentiment corpus]. In *XI. Magyar Számítógépes Nyelvészeti Konferencia*, pages 219–226.

Szabó, M. K., Vincze, V., and Hangya, V. (2016). Aspektusszintű annotáció és szentimentet módosító elemek egy magyar nyelvű szentimentkorpuszban [Aspect-level annotation and sentiment shifters in a Hungarian sentiment corpus]. In *XII. Magyar Számítógépes Nyelvészeti Konferencia*, Szeged. Szegedi Tudományegyetem.

Turney, P. D. (2002). Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424, Stroudsburg, PA, USA. Association for Computational Linguistics.

Vincze, V. (2014). Uncertainty Detection in Hungarian Texts. In *Proceedings of COLING 2014*, pages 1844–1853, Dublin.

Wang, B. and Wang, H. (2007). Bootstrapping both product properties and opinion words from chinese reviews with cross-training. In *Web Intelligence*, pages 259–262. IEEE Computer Society.

Wilson, T., Wiebe, J., and Hwa, R. (2004). Just how mad are you? finding strong and weak opinion clauses. In *Proceedings of the 19th National Conference on Artifical Intelligence*, AAAI'04, pages 761–767. AAAI Press.