Effective Use of Linguistic Features for Sentiment Analysis of Korean*

Hayeon Jang[a] and Hyopil Shin[b]

Department of Linguistics, Seoul National University,
599 Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea
[a]hyan05@snu.ac.kr
[b]hpshin@snu.ac.kr

Abstract. In this paper, we propose a new linguistic approach for sentiment analysis of Korean. In order to overcome shortcomings of previous works confined to statistical methods, we make effective use of various linguistic features reflecting the nature of Korean such as contextual intensifiers, contextual shifters, modal affixes, and the morphological dependency chunk structures. Moreover, unlike complex statistical formulae which are hard to understand, we use simple mathematical formulae in the process of term weighting. Through experiments of news corpus, we verify an improvement on the results of sentiment analysis of Korean in comparison to the experimental results using TFIDF as popular statistical method employing word frequency. This approach, especially the chunking method, will be beneficial to sentiment analysis of other morphologically rich languages like Japanese and Turkish.

Keywords: Sentiment analysis, Linguistic feature, Morphologically rich language

1 Introduction

The Internet is now an important forum where people can express their opinions without formality and constraint. Online services such as blogs and Twitter replace private diaries, and political debates are frequently opened in the reply pages of news articles. For this reason, sentiment analysis which automatically extracts subjectivities and classifies sentiments (or polarities) about some topics in written texts has been receiving attention in the field of NLP. Moreover the techniques of sentiment analysis are applied to various applications for extracting important information on the Internet to monitor a certain brand's reputations or to make social network for peoples who have similar opinion.

Sentiment analysis of English employs various statistical and linguistic methods. In the case of Korean, however, most previous research has been confined to statistical methods which only focus either on the frequency of words or relevance of co-occurring words, due to the lack of linguistic resources properly reflecting the nature of Korean. The major drawback of a statistical-based approach is the fact that the 'real' meaning of expressions, which we feel when we read them, cannot be reflected in the analysis. Moreover, complex mathematical formulae used in statistical methods are difficult for non-specialists in statistics or mathematics to understand and, in turn, require a heavy workload.

In order to overcome such shortcomings of statistical methods, we propose a new linguistic approach for sentiment analysis of Korean. The focus of our approach is an effective utilization of various linguistic features of Korean such as contextual intensifiers, contextual shifters, modal affixes, and chunk structures. To explain briefly: contextual intensifiers strengthen the valence of opinionated terms, contextual shifters change sentimental flow of sentences or polarity of terms, modal affixes determine whether or not situation included in the sentence is

true, and chunk structures limit influential scopes of negation items. Through experiments using news corpus, we verify an improvement on the results of sentiment analysis. One of the strengths of our linguistic approach is that we can obtain advanced results by simple calculations like multiplying or dividing by two with respect to contextual and functional meanings of linguistic expressions.

This paper is mainly composed of four parts: firstly, we review previous works related to our approaches. We further explain Korean linguistic features used in sentiment analysis and present the simple term weighting method using such features. Finally, we describe our experiments and show how our linguistic approach is feasible in sentiment analysis of Korean in comparison to the experimental results using a statistical method that employs word frequency, TFIDF[1].

2 Related Works

Sentiment analysis research has been performed to distinguish the authors' polarity (sentiment orientation) on certain topics from document-level (Turney, 2002; Pang et al., 2002; Dave et al., 2003) to sentence-level (Hu and Liu, 2004; Kim and Hovy, 2004). We will focus on sentence-level sentiment classification in the assumption that the polarity of sentences in a single document can be diversified due to the inclusion of various subtopics.

In previous works of sentiment analysis of English, various linguistic features are used. One typical example is a contextual intensifier. Polanyi and Zaenen (2004) define contextual intensifiers as lexical items that weaken or strengthen the base valence of the term modified. They calculate the effects of intensifiers by adding or subtracting one point to/from the base value of a term. Banamara et al. (2007) categorize intensifying adverbs according to the degree of strength of meaning and assign a score differently. In our approach, every contextual intensifier strengthens the original polarity of opinioned terms by multiplying by two regardless of the semantic intensity.

Contextual shifters are well known for their effectiveness on sentiment analysis. Kennedy and Inkpen (2006) performs sentiment analysis of movie and product reviews by utilizing the contextual shifter information. Miyoshi and Nakagami (2007) also use this method to see the advancement of the result of sentimental analysis on electric product reviews in Japanese. In this work, we make use of the functions of each shifter to properly modify the value of the terms in the sentences and limit the number of the features, which must be observed in the analyzing process, to improve efficiency.

In addition, there are some works using structural information of the target sentiment in order to improve the results of sentiment analysis. Choi et al. (2005) and Mao and Lebanon (2006) are representative of the structured sentiment analysis approach which takes advantage of Conditional Random Fields (CRF) to determine sentiment flow. McDonald et al. (2007) also deal with sentiment analysis via the global joint-structural approach. Furthermore, as there are a lot of good parsers for English data, Meena and Prabhakr (2007) and Liu and Seneff (2009) utilize sentiment structure information by the parsers such as Berkeley Parser.

In the case of Korean, however, it is hard to find proper resources due to its nature of Korean. Korean exhibits features such as rich functional morphemes, a relatively free word-order, and frequent deletion of primary elements of sentences like the subject and object. Although much research has applied dependency grammars for reducing the complexity of sentences to match the characteristics of Korean (Kim and Lee, 2005; Nam et al., 2008), this has still caused problems which prohibited wide use. Therefore we suggest a new

[1] TFIDF(Term Frequency-Inverse Document Frequency): For a term i in document j

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents

morphological chunking method that binds semantically related concatenations of morphemes. This helps to define boundaries of semantic scopes of opinionated terms and is faster, simpler and more efficient on sentiment analysis than a general full parser.

3 Linguistic Features for Sentiment Analysis of Korean

Our approach is basically dictionary-based in that it determines the value of terms by matching with lexical items that have lexical valence in the polarity dictionary. The polarity dictionary contains 5,249 Korean lexical items that have POS tags of morphological analysis and are classified as positive or negative according to the meaning of the lexical items. Linguistic features introduced in this section are used in the following process of term weighting. The formulas used in term weighting are simple calculations such as multiplying by two to strengthen the valence of terms or dividing by two to weaken the valence of terms.

All lexical items included in linguistic features are chosen in the 21st Century Sejong Electronic Dictionary[2] according to semantic classes and meanings of the terms.

3.1 Contextual Intensifiers

Contextual shifters such as *too* in *too difficult* and *very* in *be very pleased* act to strengthen the base polarity valence of the term. In our approach, a total of 83 lexical items which include 81 adverbs (such as 강력히 *kanglyekhi*[3] 'strongly', 이토록 *itholok* 'so', and 훨씬 *hwelssin* 'much') and two nouns (사상 *sasang* 'all-time' and 제일 *ceyil* 'the most') play a role as contextual intensifiers. In our approach, we calculate the effect of contextual shifter by doubling the base value of a term.

(1) 자료/nc 가/jc 너무/a 부족/ncs 하/xpa 다/ef[4]
calyo *ka* *nemwu* *pwucok* *ha* *ta*
material Nominative too insufficiency Adjective-Deriving Declarative
'Material is too insufficient.'

In example (1), the stative noun 부족 *pwucok* 'insufficiency' is judged as negative by matching with the polarity dictionary, and so receives the negative value −1. After every term in this sentence get its own value, the contextual intensifier 너무 *nemwu* 'too' decreases the base value of 부족 *pwucok* to −2 in order to emphasize negative valence of 부족 *pwucok* in this sentence.

(2) 그/npp 의/jcm 사상/nc 은/jx 제법/a 명확/ncs 하/xpa 다/ef
ku *uy* *sasang* *un* *ceypep* *myenghwak-ha* *ta*
he Possessive thought Topic-Contrast pretty clear Declarative
'His thought is pretty clear.'

In example (2), the adjective 명확하- *myenghwakha-* 'clear' has the intensified positive value +2 according to the positive label of the polarity dictionary and the intensifier adverb 제법 *ceypep* 'pretty'.

[2] 21st century Sejong Project is one of the Korean information policies run by the Ministry of Culture and Tourism of Korea. The project was named after King Sejong the Great who invented Hangeul. (http://www.sejong.or.kr/)

[3] In this paper, all Korean examples are transliterated via the Yale Romanization system.

[4] POS tags of morphological analysis: a(adverb), ad(demonstrative adverb), ecs(subordinative conjunctive ending), ecx(auxiliary conjunctive ending), ef(final ending), efp(prefinal ending), i(interjection), jc(case particle), jcm(adnominal case particle), jcp(predicative case particle), jx(auxiliary particle), md(demonstrative adnoun), nc(common noun), nca(active common noun), ncs(stative common noun), npp(personal pronoun), pa(adjective), pv(verb), px(auxiliary verb), xpa(adjective-derived suffix), xpv(verb-derived suffix)

3.2 Contextual Shifters

In this paper, the term 'contextual shifter' covers both the negation shifter and the flow shifter: the former refers to the term which can change semantic orientation of other terms from positive to negative and vise versa, the latter the term which can control sentiment flow in sentences; for example, in English *not*, *nobody* (negation shifters), *however*, *but* (flow shifters). Contextual shifters in Korean consist of 13 negation shifters (adverbs such as 안 *an* 'not', 못 *mos* 'cannot' and auxiliary verbs such as 않 *anh* 'not', 말 *mal* 'stop', 없 *eps* 'no') and 23 flow shifters (sentence-conjunctive adverbs such as 그러나 *kulena,* 하지만 *haciman* 그래도 *kulayto* 'but, nevertheless, though', subordinative conjuctive suffixes –ㅂ니다만 *pnitaman,* -는데 *nuntey* and conjunctive suffixes such as –어도 *eto*).

(3) 흠/nc 없/pa 는/exm 사람/nc
hum *eps* *nun* *salam*
fault no Adnominal person
'The person who has no fault'

Since negation shifters play the role of shifting the polarity of the sentiment terms in our approach, we multiply them by -1. In example (3), the base value of the common noun 흠 *hum* 'fault' is the negative value −1. The negation shifter 없 *eps* 'no' affects the value of 흠 *hum* and then, the value of 흠 *hum* is changed to positive +1. Therefore the total value of example (3) properly reflects the positive meaning of the phrase 'The person who has no fault'.

In the case of flow shifters, we limit the number of features to the terms after the shifter appears. We believed it more important to understand an author's empathetic point, rather than to catch the full sentiment flow in the sentences. Also, such emphasized contents mostly exist after the flow shifters. Therefore we utilize this characteristic to reduce the work load and to prevent confusions caused by other minor sentiment terms.

(4) 참/i 유치하/pa ㄴ/exm 내용/nc 이/jcp ㄴ데/ecs 많/pa ㄴ/exm 사람/nc 이/jc 공감/ncs 하/xpa 었/efp 다/ef
cham *yuchiha* *n* *nayyong* *i* *ntey* *manh* *n* *salam* *i* *kongkam* *ha* *ess* *ta*
very immature Adnominal content Predicative though many Adnominal people Nominative sympathy Verb-Deriving Past Declarative
'Though the content was very immature, many people felt sympathy.'

The sentence of example (4) has two conflicting polarity items; the negative adjective 유치하- *yuchiha-* 'immature' and the positive stative noun 공감 *kongkam* 'sympathy'. The total value of the sentence calculated regardless of the flow shifter ㄴ데 *ntey* 'though' is negative −1 because the contextual intensifier 참 *cham* 'very' increases the value of 유치하- *yuchiha-* to the emphasized negative value −2. If we use the function of the flow shifter ㄴ데 *ntey*, however, the number of target morphemes is restricted to eight after the flow shifter and then, the total value of the sentence becomes positive +1. As the most important point that must be made in this sentence is the fact that many people feel sympathetic to the content, the latter result can be judged as the right result.

3.3 Modal affixes

Language makes a distinction between events or situations which are asserted to have happened or are happening, i.e. realis events, and those which *might*, *could*, *should*, *ought to*, or *possibly*

occur or are going to occur, i.e. irrealis events[5]. Korean as the agglutinative language has various functional affixes and the combinations of them set up a crucial meaning of contexts such as possibility, necessity and evidentiality. For this reason, we utilize various modal affixes of Korean to distinguish realis events from irrealis events and to weaken the weight of irrealis events in computing an evaluation of the author's attitude by dividing by two.

Conjectural. Conjectural mood is used in Korean to express some suppositions about the future, present or past. It can be translated to English words such as 'probably', 'perhaps' etc. Since we cannot be assured that opinionated words in the scope of conjectural mood are the actual sentiments of the author, the conjectural lexical items decrease the value of opinioned terms in their scopes in the process of our term weighting. A total of 18 Korean lexical items (13 final suffixes such as -르텐데 *-ltheyneey* 'would' and -르걸 *-lkel* 'probably', four pre-final suffixes such as -겠 *-keyss* 'wish', one adnominal suffixes –르- *-l-* 'might') play a role as Conjectural.

(5) 내/npp 딸/nc 이/jc 공부/nc 잘하/pv 면/ecs 좋/pa 겠/efp 다/ef
nay ttal i kongpwu calha meyn coh keyss ta
my daughter Nominative study do-well if good Conjectural Declarative
'I wish my daughter had studied well'

In example (5), the opinionated term 좋 *coh* 'good' is in the influential scope of the Conjectural affix -겠 *-keyss* 'wish'. Since the positive sentiment of 좋 *coh* is not in fact an actual thing but rather the author's hope, the value of 좋 *coh* is weakened to `+0.5` from `+1`.

Imperative. Imperative mood denotes the speaker's degree of requirement of conformity to the proposition expressed by an utterance, especially in commands. In a view from sentiment analysis, Imperative also expresses the author's wish like the above Conjectural. In our approach a total of 13 final suffixes play a role as Imperative. -라 *-la* and -어다오 *-etao* are typical examples.

(6) 경제/nc 를/jc 살리/pv 어라/ef
kyengcey lul salli ela
economy Objective revive Imperative
'Revive the economy!'

The verb 살리- *salli-* 'revive' has the base positive value `+1`; however, since it is in the scope of the Imperative suffix –어라 *–ela,* the valence of 살리- *salli-* is weakened to half. The situation that the economy is revived is just the desire included in the imperative sentence, not the actual state. The reason why such a value is not excluded in the process of calculation is that we do not want to completely ignore the author's desire, a component of his opinion.

Interrogative. An Interrogative mood connotes how much certainty or evidence a speaker has on the proposition expressed by an utterance and presents questions to elicit information concerning the topic of an utterance from the addressee. A total of 16 final suffixes such as -르까 *–lkka* and -느가요 *-nkayo* are used in this work to control the weight of terms in the scope of Interrogative.

(7) 왜/a 저리/ad 흥분/ncs 하/xpa 르까/ef
way celi hungpwun ha lka
why like-that excitement Adjective-Derived Interrogative
'Why are they excited like that?'

[5] Polanyi and Zaenen (2004)

The stative noun 흥분 *hungpwun* ‘excitement’ in example (7) has the base positive value +1 and the value for the word is strengthened to +2 by the effect of the intensifier adverb 저리 *celi* ‘like that’. Up to this point, the polarity of the sentence is highly positive. Since the Interrogative suffix 르까 *lka* ‘Interrogative’ is included in this sentence, however, the value of 흥분 *hungpwun* is reduced to half, +1. In opinionated sentences, Interrogative suffixes mainly play a role as the key to understanding the sentences as the opposition or doubt of authors to other’s opinions or sentiments. So there is great potential that the terms having sentiment polarity are not the attitude of the author in the sentences including Interrogative suffixes. However, there is still potential that the terms in the influential range of interrogative suffixes express the opinion of the author as in 난 왜 이 영화가 재미있을까? *nan way i yenghwa-ka caymiiss-ulkka* I why this movie-Agent interesting-Interrogative ‘Why am I interested in this movie? (in comparison to others who are not)’. For this reason, our work does not totally exclude the value of opinionated terms in interrogative mood during the calculation.

Quotative. Quotative means that opinionated terms which are in the same phrase express another person’s opinions. In our approach, a total of five final suffixes such as –라고 *-lako*, -다고 *-tako*, -ㄴ다는군 *–ntanunkwun* are used as Quotative to weaken the value of terms which are the attitude of other people, not the author. Such suffixes can be translated as ‘he/she said that~’ in English.

(8) 그녀/npp 는/jx 그/npp 의/jcm 대처/nc 가/jc 미봉책/nc 이/jcp 라고/ecs 비판/nca 하/xpv 었/efp 다/ef
kunye nun ku uy dayche ka mipongchayk i lako piphan ha ess ta
she Topic-Contrast he Possessive treatment Nominative temporary-expedient Predicative Quotative criticism Verb-Derived Past Declarative
‘She criticized his treatment as a temporary expedient.’

The negative term 미봉책 *mipongchayk* ‘temporary expedient’ gets the base value -1 by matching with the polarity dictionary and the value is changed to -0.5 affected by the Quotative ending –라고 *–lako*.

In our approach, we equally reduce the values of terms in the scope of the above modal affixes to half regardless of semantic functional distinction of each type of affixes. It is possible that such a simple term weighting method leads to wrong classification of the author’s attitude in the text. Nevertheless, we attempt to verify that if linguistic features are utilized effectively, the results of sentiment analysis can be highly improved without applying the multidimensional complex calculation.

3.4 Morphological Dependency Chunking

In our approach, instead of doing complete syntactic parsing we use a chunking method based on the dependency relation of morpheme sequences.

Korean is a head-final language: in terms of dependency grammar, governors are always located after their dependents. We reflect upon this characteristic to form a relation if a certain morpheme acts as the governor of the previous morpheme. Chunks are formed until an unrelated morpheme appears. The terms in a single chunk exert their own semantic influence to each other and control the values. After determining the values of every morpheme in each chunk, this process is replicated at a higher level and finally the ultimate values of every term in the sentence are determined. For example, in the sentence of (9), the noun 문제 *mwuncey* ‘problem’ plays a role as the governor of the definite article 그 *ku* ‘the’ and the following particle 는 *nun* is the governor of the noun 해결 *haykyel* ‘resolution’. Since the forth morpheme

해결 *haykyel* and the previous 는 *nun* do not have a dependency relation, three morphemes 그/md, 문제/nc, and 는/jx form the first chunk like example (10).

(9) 그/md 문제/nc 는/jx 해결/nca 되/xpv 지/ecx 않/px 았/efp 다/ef
ku mwuncey nun haykyel toy ci anh ass ta
the problem Topic-Contrast resolution Verb-Derived Conjunctive not Past Declarative
'The problem wasn't solved.'

(10) {{{{그/md}+문제/nc}+는/jx}+{{{{{<해결/nca+되/xpv>/pv}+지/ecx}+않/px}+았/efp}+다/ef}}

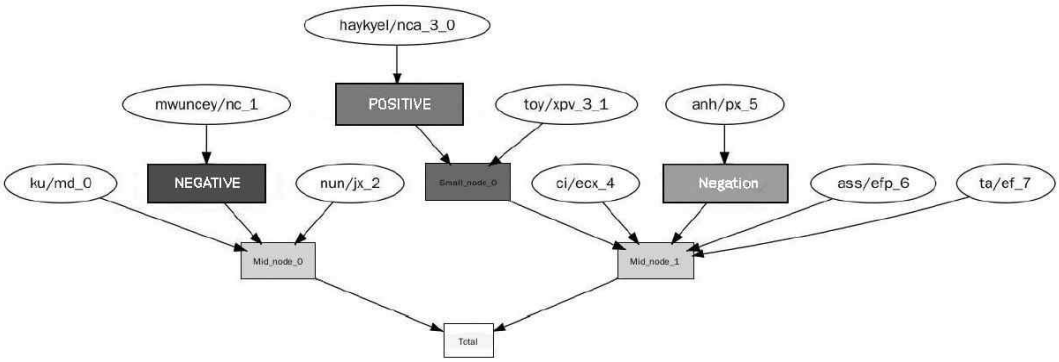

Figure 1: The Chunk Structure of the Sentence of (9) and the graphical representation of (10).

Our morphological dependency chunking method helps effective sentiment analysis by providing the structural information which properly limits the semantic influential scope of functional terms such as negation shifters. The active noun 해결 *haykyel* 'resolution' in example (9) has the base positive value `+1`. In the simple classification which does not use information of chunking structure, a [-2, +2] window determines the influential scope of functional items. Therefore, the negation shifter 않 *anh* 'not' cannot change the value of 해결 *haykyel* at -3 position and the value of 해결 *haykyel* incorrectly maintains as positive. Through our chunking method, the sentence is analyzed as in Figure 1. In this structure, the active noun 해결 *haykyel* and the verb-derived suffix 되 *toy* make the complete passive verb 해결되- *haykyeltoy*- 'be solved'. This verb is treated as one single item in a chunk and is included in the same chunk with the negation shifter 않 *anh*. Consequently, the influential scope of the negation shifter can cover the noun 해결 *haykyel* and then, the value of 해결 *haykyel* becomes `-1` by multiplying by -1.

4 Experiments

4.1 Corpora

We collected 79,390 news articles from the web site of the daily newspaper, The Hankyoreh[6] in the period of January 1, 2009 to April 7, 2010 (total 146.6MB). Since the news data includes both objective and subjective sentences, we categorize the news corpus into three groups by the following characteristics related to subjectivity: 71,612 general news articles, 3,743 opinionated news articles having subjective subtopics such as 'Yuna Kim (a Korean figure skater), terrorism, etc.' and 3,432 editorial articles including columns and contributions. The

[6] http://www.hani.co.kr/

collection of sample sentences consists of 1,225 general news sentences, 1,185 subtopic news sentences and 2,592 sentences of editorial articles by randomly extracting 100 articles from each data group.

News articles have no marks representing subjectivity or polarity of sentences compared to the grading systems found in movie review texts. Therefore, in our work, two native Korean annotators manually attached polarity labels to each sentence. Sentences are classified as subjective when they contain opinions pertaining to a certain object. Even if the opinion is not expressed on the surface using direct sentiment terms, the sentences are classified as subjective when the annotator can feel the subjectivity through the tone of voice. Only when the sentences are classified as subjective, the polarity tags are attached. The agreement rate of the two annotators in the manual annotation of polarity is 71%.

4.2 Results

Figure 2 and Table 1 show the results of a 10-fold cross variation experiment on the sentiment analysis of each of the three groups of news articles using SVMlight[7], a powerful tool for binary classification. In Table 1, numbers in bold face are the best results in each dataset.

First of all, all of our proposed linguistic methods obtain higher results than TFIDF, except in the cases of F-measure score of the subtopic news article corpus which do not use a chunking method. Accuracy values are increased from at least about 2% to 28% and F-measure values are improved from about 1% to 21%. This shows that by utilizing language-specific features which reflect Korean linguistic characteristics well, even without making use of complex mathematical measuring techniques, we could obtain better results than statistical methods in sentiment analysis. The most subjective dataset, editorial article corpus, shows the greatest improvements in both accuracy and F-measure values. This means that various linguistic features used in this work properly grasp the sentimental meanings included in subjective texts. On the other hand, the subtopic news article corpus shows the smallest improvements. This is due to the characteristics of news that aims to provide facts, not particular opinions. In order to maintain objectivity, news articles dealing with subjective topics tend to express opinions related to the topics in a roundabout way instead of using direct opinionated terms. For this reason, our dictionary-based term weighting system has difficulty in finding the implicit meanings.

Table 1: The Results of Sentiment Analysis of Korean News Corpus.

Data	Method			Accuracy (%)	F-measure[8] (%)
Editorial articles	Statistical	TFIDF		45.555	36.499
	Our linguistic approach	No chunking	No shifter	71.874	56.011
			Yes shifter	70.193	55.462
		Yes chunking	No shifter	**72.743**	**57.774**
			Yes shifter	71.684	57.189
Subtopic News articles	Statistical	TFIDF		49.753	65.911
	Our linguistic approach	No chunking	No shifter	52.915	65.881
			Yes shifter	51.574	64.948
		Yes chunking	No shifter	55.346	67.147
			Yes shifter	**55.818**	**67.347**
News articles	Statistical	TFIDF		42.416	57.074
	Our linguistic approach	No chunking	No shifter	47.589	59.455
			Yes shifter	49.149	60.115
		Yes chunking	No shifter	50.484	59.898
			Yes shifter	**50.598**	**60.656**

[7] http://svmlight.joachims.org/

[8] F-measure = 2*Precision*Recall/(Precision+Recall)

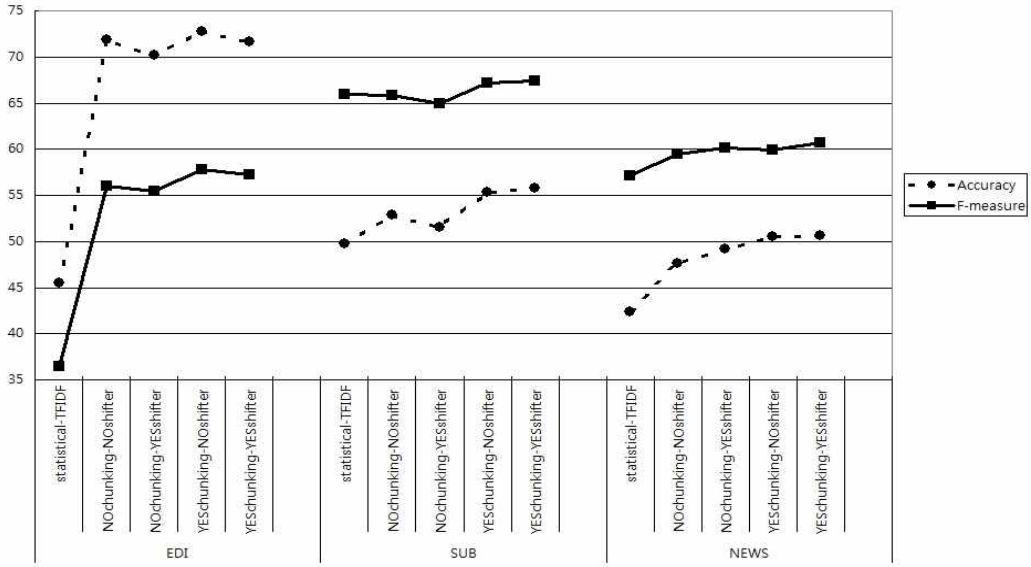

Figure 2: The Results of Sentiment Analysis of Korean News Corpus.

Secondly, we obtain higher values of sentiment classification by using chunking structures consistently in all of our data. This implies that the restriction on the semantic scope of functional terms related to sentiment polarity using structural information of Korean helps to classify sentiments of sentences correctly. Our morphological dependency chunking method (see the parts labeled 'Yes chunking' in Table 1 and Figure 2), in comparison to the experimental results which do not use the chunking method, improves accuracy values from about 1% to 4% and F-measure values from about 0.4% to 2%.

Finally, the less subjective the data, the more improved results we obtain when we use the function of contextual shifters. The effects of using contextual shifters in the process of sentiment analysis are summarized as follows; 1) in the editorial article corpus which is the most subjective dataset, utilizations of contextual shifters lower the efficiency of sentiment analysis, 2) in the subtopic news article corpus which is less subjective than the editorial articles, the results are improved only when the chunking method is used together, and 3) using contextual shifters advances the results of sentiment analysis regardless of utilizing the chunking method in the news article corpus which is the least subjective. This shows that the methods reducing target features by flow shifters and modifying values of polarity terms by negation shifters have merits in sentiment analysis of data that contain both objective and subjective contents, because contextual shifters help to focus on more important information in confusing data.

5 Discussion and Future Work

In this paper, we verified that the effective use of linguistic features can improve the results of sentiment analysis just by simple measurements. In the process of term weighting, our approach utilizes various Korean lexical items that reflect the nature of Korean well, such as contextual intensifiers, contextual shifters (negation shifters and flow shifters), modal affixes, and morphological dependency chunk structures. The proposed chunking method using dependency relations of morpheme sequences is particularly expected to aid the sentiment analysis of other agglutinative languages such as Turkish and Japanese.

Future work includes the study of elaborate measuring methods using linguistic-specific features of Korean, a morphologically rich language, more accurately and effectively. For example, more finely grained classification of Korean modal affixes by semantic functions can

help to assign different weighting points and then, it makes us catch more precise polarity valences of phrases and whole sentences.

In addition, we have plans to utilize Korean ontology in order to disambiguate the proper sense of polysemous words according to the context.

References

Benamara, F., C. Cesarano, A. Picariello, D. Reforgiato, and V. S. Subrahmanian. 2007. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. *In Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*.

Choi, Y., C. Cardie, E. Riloff, and S. Patwardhan. 2005. Identifying sources of opinions with conditional random fields and extraction patterns. *In Proceedings of the HLT/EMNLP*.

Dave, K., S. Lawrence, and D. M. Pennock. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *In Proceedings of the WWW-2003*.

Hu, M. and B. Liu. 2004. Mining and summarizingcustomer reviews. *In Proceedings of KDD*.

Jang, Hayeon and Hyopil, Shin. 2010. Language-Specific Sentiment Analysis in Morphologically Rich Languages. *In Proceedings of COLING 2010*.

Kennedy, A. and Inkpen, D. 2006. Sentiment Classification of Movie and Product Reviews Using Contextual Valence Shifters. *Computational Intelligence*, 22(2); 110–125.

Kim, Mi-Yong and Jong-Hyeok Lee. 2005. Syntactic Analysis based on Subject-Clause Segmentation. *In Proceedings of KCC 2005*, 32(9), pp.936-947. In Korean.

Kim, S. M., and E. Hovy. 2004. Determining the sentiment of opinions. *In Proceedings of the COLING 2004*.

Liu, Jingjing, and Stephanie Seneff. 2009. Review sentiment scoring via a parse-and-paraphrase paradigm. *In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 1(1).

Mao, Y., and G. Lebanon. 2006. Isotonic conditional random fields and local sentiment flow. *In Proceedings of the NIPS*.

McDonald, R., K. Hannan, T. Neylon, M. Wells, and J. Reynar. 2007. Structured Models for Fine-to-Coarse Sentiment Analysis. *In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp.432-439.

Meena, Arun and T. V. Prabhakar. 2007. Sentence Level Sentiment Analysis in the Presence of Conjuncts Using Linguistic Analysis. *Lecture Notes in Computer Science*, 573-580.

Miyoshi Tetsuya and Nakagami Yu. 2007. Sentiment classification of customer reviews on electric products. *In Proceedings of IEEE International Conference on Systems Man and Cybernetics*, pp.2028-2033.

Nam, Sang-Hyub, Seung-Hoon Na, Yeha Lee, Yong-Hun Lee, Jungi Kim, and Jong-Hyeok Lee. 2008. Semi-Supeervised Learning for Sentiment Phrase Extraction by Combining Generative Model and Discriminative Model. *In Proceedings of the KCC(Korea Computer Congress) 2008*, 35(1):268-273. in Korean.

Pang, Bo, Lillian Lee, Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. *In Proceedings of the ACL-2002 conference on Empirical methods in natural language processing*, 10.

Polanyi, Livia, and Annie Zaenen. 2004. Contextual valence shifters. *In Proceedings of the AAAI Symposium on Exploring Attitude and Affect in Text: Theories and Applications*.

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