Annotation Scheme for Constructing Sentiment Corpus in Korean

Hyopil Shin, Munhyong Kim, Yu-Mi Jo, Hayeon Jang, and Andrew Cattle

[a]Department of Linguistics, Seoul National University
1 Gwanak-no Gwanak-gu, Seoul 151-741
{hpshin, likerainsun, jmocy84, hyan05, acattle}@snu.ac.kr

Abstract

This paper describes the first year of work constructing the Korean Sentiment Corpus, focusing on the theoretical background such as the annotation scheme. Our aim is to provide a solid theoretical background for the corpus which reflects the characteristics of the Korean language and includes approximately 8,050 sentences taken from news articles. The corpus annotation scheme, based on the MPQA, is described along with the results of inter-annotator agreement tests with a view to improving the annotation scheme.

1   Introduction

There has been much research on the automatic identification and extraction of sentiments and opinions in text. Researchers have been working on these issues by focusing mainly on subjectivity and sentiment classification either at the document or sentence level. Classifying editorials or movie reviews as positive or negative are examples of a document classification tasks while classifying individual sentences as subjective or objective would be an example of a sentence-level task (Wiebe et al., 2005).

Along with these lines of research, a need for corpora annotated with rich information about opinions and emotions has also emerged. This would allow for the development of statistical and machine learning approaches for various practical NLP applications. As such a resource, the Multiperspective Question Answering (MPQA) Opinion Corpus, developed by Wiebe (2002), Wiebe et al. (2005), and Wilson et al. (2008), plays an important role in sentiment and opinion analysis. It contains the manual annotation of a 10,000 sentence-corpus of articles from the world press. Since this corpus provides a fine-grained annotation scheme, it is widely used as a source for training data in machine learning approaches and serves as the gold standard in sentiment classification tests.

We started constructing a cross-language sentiment corpus, called the Korean Sentiment Corpus. We received two years of support in this project by the Korean Research Foundation (KRF) for two years. We aim to provide both a solid theoretical background for the Corpus, reflecting the characteristics of the Korean language, as well as fine-grained annotations for the 8,050 sentence-corpus of news articles. The total number of annotated sentences is less than that of the MPQA, but since our annotation is morpheme-based due to the agglutinative nature of Korean, the number of annotation units is much greater. We have also adopted the basic annotation scheme of the MPQA for comparative research purposes.

This paper describes the first year of work constructing the Korean Sentiment Corpus, focusing on the theoretical background such as the annotation scheme. Inter-annotator agreement tests were performed to improve annotation quality. The remainder of this paper is organized as follows. Section 2 gives a brief overview of the MPQA corpus as a starting point. Section 3 elaborates on the annotation scheme for the Korean sentiment corpus, providing examples of annotations with attributes. Section 4 shows observations on the inter-annotator agreements. Section 5 presents future work and conclusions.

2   The MPQA Corpus

As a fundamental resource for sentiment corpus construction in Korean, this work takes advantage of the Multiperspective Question Answering (MPQA) Opinion Corpus which began with the conceptual structure for private states in Wiebe (2002) and developed manual annotation instructions. The MPQA Corpus version 1.0 was released in 2003, and now version 2.0 is available with more detailed attitude annotations. In this section we briefly review the annotation scheme and structures of the corpus with a view to providing a theoretical background.

2.1   Private States

According to Quirk et al. (1985), a private state refers to mental and emotional states such as the opinions, beliefs, and intentions of a writer. Wiebe et al. (2005) focused on identifying private state expressions in contexts and presented numerous examples annotated with schemes that cover a broad range of linguistic expressions and phenomena.

Private states and speech events are the core of the MPQA corpus. Private states cover *opinions*, *beliefs*, *thoughts*, *feelings*, *emotions*, *goals evaluations*, and *judgments* (Wiebe et al. 2005). Private state frames cover expressive subjective element frames, which are used to represent expressive subjective elements, as well as direct subjective element frames, which are used to represent subjective speech events. In order to distinguish opinion-oriented material from fact, objective speech event frames are also defined in terms of speech events. Private state frames have the following attributes directly excerpted from Wiebe et al. (2005)

Direct subjective frame:
- text anchor: a pointer to the span of text that represents the speech event or explicit mention of a private state
- source: the person or entity that is expressing the private state, possibly the writer
- target: what the speech event or private state is about
- properties
    - intensity: the intensity of the private state (low, medium, high, or extreme)
    - expression intensity: the contribution of the speech event or private state expression itself to the overall intensity of the private state (neutral, low, medium, high, or extreme)
    - insubstantial: true, if the private state is not substantial in the discourse
    - attitude type: represents the polarity of the private state. The possible values are positive, negative, other, or none

Expressive subjective element frame:
- text anchor
- source
- properties
    - intensity
    - attitude type

2.2   Objective Speech Event

Objective speech event in the MPQA is used to distinguish opinion-oriented material from material presented as factual and has the following frames.

Objective speech event frame:
- text anchor
- source
- target

Unlike the MPQA, we do not distinguish direct subjective frames from expressive subjective elements. Rather, those two frames are merged into SEED subjective expressions in our approach.

2.3   Nested Sources

In sentiment analysis, it is very useful to recognize the person whose opinion or emotion is being expressed. Thus source is introduced in the MPQA. The source of a speech event is implicitly the speaker or the writer while the source of a private state is the experiencer. However, there are situations where speech events and private states are assessed by more than one source. In this case, an additional explicit source was introduced. This source generally corresponded to the subject of the embedded predicate. This is a so-called nested

source, as adopted by Wiebe et al. (2005), Wilson (2008), and Sauri (2008). Nested sources include other people's speech events and private states as well as speaker's. Please look the following examples adopted from Wiebe et al. (2005: 9):

   (1) a. Sue said, "The election was fair."
       b. Sue thinks that the election was fair.
       c. Sue is afraid to go outside.

In the above sentences, Sue is the source of speech event (1a) and of private states (1b, 1c). However, we do not know what Sue says, thinks, or feels directly. We only know Sue's speech event according to the writer. In the MPQA Corpus, such a nested source would be represented as <*writer*, *Sue*>. Private states can be directed toward the private states of others. Consider Wiebe et al. (2005)'s example:

   (2) "The U.S. fears a spill-over," said Xirao-Nima.

In (2), it is not *the U.S.* that directly states its fear. Rather, according to the writer, the *Xirao-Nima* states that the U.S. fears a spill-over. Thus the nested source of the fear can be expressed as <*writer, Xirao-Nima, U.S*>.

3   Outline of Annotation Scheme for Korean Sentiment Corpus

Our work essentially follows the idea of the MPQA, but we have also modified the existing MPQA attributes as well as introduced new attributes to address the characteristics of Korean.

The annotation scheme starts with distinguishing a SEED from a whole sentence in terms of subjectivity. In a SEED, each individual unit expresses a private state. By contrast, the subjectivity of the whole sentence is about whether we feel the sentence is objectively true or not in terms of the speech event. Even though a sentence bears many subjective expressions in it, the sentence can carry objective facts. Thus our annotation principle separates basic subjective expressions from subjectivity of a whole sentence. That is, unlike the MPQA, we explicitly annotate subjectivity or objectivity of the sentence. This principle can be illustrated as follows.
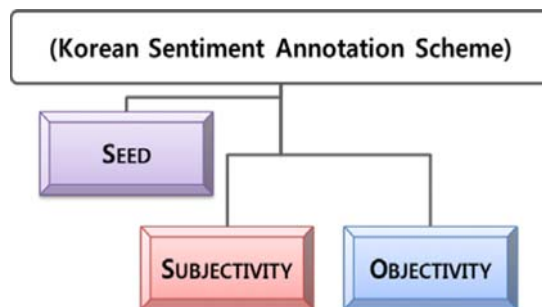

Figure 1. Korean Sentiment Annotation Scheme

As a basic annotation unit, we chose a morpheme rather than a word. Korean is an agglutinative language and many meaning-bearing particles and sentence endings can carry private states, therefore we need to be able to pinpoint these precise segments as a basic unit. Although such morpheme-based annotation helps to produce a fine-grained corpus, the trade-off is that it also requires a great deal of time and effort spent on annotating.

3.1   SEED

The elements of SEED are as follows:
- anchor: morpheme id(s)
- id: tag id
- expressive type: direct-explicit, direct-speech, direct-action, indirect, writing-device
- subjectivity type: emotion-pos, emotion-neg, emotion-neutral, emotion-complex, judgment-pos, judgment-neg, judgment-neutral, agreement-pos, agreement-neg, agreement-neutral, argument-pos, argument-neg, argument-neutral, intention-pos, intention-neg, speculation-pos, speculation-neg, others
- nested-source: w-sources
- target: target id(s)
- polarity: positive, negative, neutral, complex
- intensity: low, medium, high
- insubstantial: TRUE, FALSE

According to Wiebe et al. (2005: 4) private states are states of *experiencers* holding *attitudes*, optionally toward *targets*. For example, in the sentence *John hates Mary*, the experiencer is *John*, the attitude is *hate*, and the target is *Mary*. Thus, in order to annotate subjective expressions, all three attributes of the private state should be properly represented. In the MPQA, the following

183

three main types of private states expressions were included: explicit mentions of private states, speech events expressing private states, and expressive subjective elements. Expressive subjective elements, speech events, and attitudes of a private state in the MPQA, roughly correspond to SEED, expressive type, and subjectivity type in our scheme.

3.1.1 Expressive Types

Express types specify either speech events (acts) that express private states (or other subjective elements) or non-speech events. These fit into five subtypes: *direct-explicit*, *direct-speech*, *direct-action*, *indirect*, and *writing-device*. While the former three types are related to speech events and usually originate from subject-predicate relations, *indirect* and *writing-device* are used for a writer to show his/her own subjectivity through non-predicate expressions. These include using a nominal as an argument, adverbials, conjunctive endings, or some particles in Korean. *Indirect* and *writing device* are common in that subjectivity is not carried through speech event. In the case of *indirect*, the source of the expression is not clear compared to *direct* or *writing device*. The following is some examples of each expression type.

- explicit: cikyepta 'boring' inkita 'be popular'
- direct speech: cwucanghata 'insist,' pinanhata 'blame,'
- direct action: elkwulsayki pyenhata 'turn pale,' hwanhohata, 'acclaim'
- indirect: isanghan salam 'strange people,' huylluynghi 'greatly'
- writing-device: -man 'only,' isanghakeyto 'strangely'

3.1.2 Subjectivity Types

The attribute subjectivity type is used to classify subjective expressions according to their sources' attitudes; lexically determined as the core meaning of subjective expression. It consists of the following subtypes: *emotion*, *judgment, agreement, argument, intention*, and *speculation*. These types can be further combined with other polarity attributes such as *positive, negative, neutral* and *complex* according to their semantic orientations which may lead to complex attributes such as *emotion-positive*, *emotion-negative*, and so on.

Generally, a *complex* attribute is due to a combination of positive and negative words, such as in the Chinese character expression '幸不幸,' 'happiness and unhappiness'. The MPQA does not provide this kind of detailed classification. Considering our previous sentiment research, we think that classifying subjectivity into more refined types provides the benefits not just when determining whether a document is subjective but also when determining what kind of attitude the document contains. The subjectivity types are exemplified as follows:

Type	Values	Examples
emotion	emotion-positive	*kipputa* 'glad,' *miso-lul cista* 'make a smile'
	emotion-negative	*mwusepta* 'afraid,' *kothongsulepta* 'feel pain'
	emotion-neutral	*kamtong-i epsta* 'not touching'
	emotion-complex	*hayngpwulhayng* 'happiness and unhappiness'
judgment	judgment-positive	*yongkamhata* 'be brave,' *cangcem* 'merit'
	judgment-negative	*napputa* 'bad,' *kepcayngi* 'a coward'
	judgment-neutral	*aymayhata* 'vague,' *cal molukessta* 'don't know well'
agreement	agreement-positive	*tonguyhata* 'agree,' *yongnaphata* 'accept'
	agreement-negative	*pantayhata* 'do not agree,' *kikak* 'rejection'
	agreement-neutral	*kikwenhata* 'give up,' *cwunglip* 'be in the middle'
argument	argument-positive	*cungmyenghata* 'verify,' *seltukhata* 'persuade'
	argument-negative	*panpakhata* 'refute,' *kecisita* 'not true'
	argument-neutral	*cham kecis-ul kwupwunhal swu epsta* 'can't know if it is true or not

Intention	intention-positive	*uytohata* 'intend,' *kyelsimhata* 'make one's mind'
	intention-negative	*~hal maum-i epsta* '~not willing to,' *wuyenhi* 'accidentally'
speculation	speculation-positive	*chwuchukhata* 'speculate,' *somang* 'wish'
	speculation-negative	*epsta* 'there is not'

Table 1. Subjectivity types

3.1.3 Targets

Attribute targets are used to specify objects or themes to which the subjective expressions are directed. In many cases targets can be clearly specified but in some cases pinpointing source and target is not that simple. The following is a complicated example of target which requires an embedded clause as target.

(3) Mary-nun ku-wa   hamkkey issnun
    Mary-subj he-with together be-adnom
    kes-i   koylowessta
    that-sub   feel uncomfortable-past
"That he was with Mary made her feel uncomfortable

The target of *koylowessta* 'be hard' is not *ku* 'he' but an embedded clause which has a meaning of 'the fact that he was with Mary'. Next, due to the possibility of double subjects in Korean, some expressions can have more than two targets.

(4) Sakwa-ka   pwumcil-i   cohta.
    apple-subj   quality-subj good
    "The apple has a good quality"

3.1.4 Nested Sources

Since source information is crucial to sentiment analysis, the MPQA elaborates on sources and nested sources in annotations. As described in 2.3, nested sources include other people's speech events or private states as well as those of the speaker or writer. Table 2 shows some examples of nested sources. Here, underlining means a subjective expression and bold face means a nested source.

Following the MPQA, we specify nested sources from left to right. That is, *<w-Tom-Mary>* means that writer states Mary's speech event through Tom's eye. *<w,>* and *<w-implicit>* represent generic sources and implicitly specified sources, respectively. In (f), we can guess the source of 'be popular' from the context. Meanwhile, general population is the source of the belief 'good' in (e).

3.1.5 Polarity, Intensity, and Insubstantial

The attribute polarity describes whether the (nested) source has an positive or negative subjectivity toward the target. An example of a positive value would be *coh-(ta)* 'good/well' while an example of a negative value would be *nappu-(ta)* 'bad'. In addition, there are two more values: neutral and complex. The value of attribute intensity depends on how intensely subjectivity is expressed. For example, (*i chayk-un*) **kucekuleh-ta** '(this book is) so-so' shows a neutral intensity while (*i chayk-un*) **ssuleki-ta** '(this book is) trash' shows a highly intense negative subjectivity. Similarly, intensity modifiers, e.g. *maywu* 'very,' *sangtanghi* 'considerably,' or *nemwu* 'too (bad),' can also affect the intensity of an expression. The attribute insubstantial specifies whether a subjective expressions carry actual or imaginary events such that a value of TRUE denotes that the event actually happened while FALSE denotes an intended event. The following illustrates a SEED annotation:

$Manh_0$-un_1 $sayongca_2$-tul_3-i_4 i_5 $ceypwum_6$-ul_7 $cohaha_8$-ko_9 iss_{10}-$ta_{11 \cdot 12}$
$Many_0$-ADNOMINAL$_1$ $user_2$-PLURAL$_3$-NOM$_4$ $this_5$ $product_6$-ACC$_7$ $like_8$-DURATIVE$_{9, 10}$-DECL$_{11 \cdot 12}$
'Many users like this product'
<SEED> anchor= "8" id= "u1" type= "direct-explicit" subjectivity-type= "emotion-pos" nested-source= "w-manhun sayongcatul" target= "5-6" polarity= "positive" intensity= "medium" insubstantial= "FALSE" </SEED>

Types	Example	Values
a. Source = writer	Kwail-un sakwa-ka <u>ceilita</u> 'fruit'-topic apple-subj best-be As for fruit, apple is best	W
b. Source=writer According to = subject Subject=writer	**Na**-to sakwa-lul <u>cohahanta</u> I –too apple-obj like I like an apple too.	w w-I
c. Source=writer According to=subject	**Tom**-un sakwa-lul <u>cohahanta</u> Tom-subj apple-obj like Tom likes an apple	w-Tom
d. Source=writer According to= A According to=B	**Tom**-un **Mary**-ka sakwa-lul <u>cohahanta</u>-ko <u>malhayssta</u> Tom-subj Mary-subj apple-obj like-comp say-past Tom said that Mary likes an apple	w-Tom-Mary
e. Source = unclear, or general population	<u>Cohun</u> kamera-nun pissata 'good' camera-sub expensive Good cameras are expensive	w-general
f. Source=not explicitly specified source in a sentence	Yocum <u>inkki- iss-nun</u> kamera-nun gf-1 ita Now popular-be-adnom camera-subj gf-1 be Now popular camera is gf-1	w-implicit

Table 2 Example of Nested Sources

3.2 Sentence Level Subjectivity

Unlike MPQA, we explicitly specify the whole sentence's subjectivity. Although each sentence consists of various numbers of subjective expressions, we feel that a sentence may be an objective fact rather than subjective. Thus we mark the subjectivity of a whole sentence on the basis of the speech event, i.e. from the writer's perspective. We believe that this can help researchers to extract relevant features for subjectivity from those sentences and to train the corpus to see what makes the sentences subjective or objective. Information on the sentence level subjectivity or objectivity differs from SEED tags as they have relatively simple structures, as follows.

- The BNF of SUBJECTIVITY
  anchor: Morpheme id(s)
  id: s1
  polarity: positive, negative, neutral, complex
  intensity: low, medium, high

The OBJECTIVITY tag consists of only the attributes *anchor* and *id*.

- The BNF of OBJECTIVITY
  anchor: Morpheme id(s)
  id: o1

Examples of SUBJECTIVITY and OBJECTIVITY tags are listed in (5). The subjectivity of objectivity of a sentence can be influenced by SEED tags, but it is not completely dependent on them. In a case of a SEED tag affecting the subjectivity of the whole sentence, usually the original source of the subjectivity indicated by the SEED tag is the writer of sentence. That is, there is no nested-source except the writer: nested-source="w". In (5c), 'was reported as a regrettable event that Yumi bought a house,' the value of nested-source "w-general" represents general population.

(5)
a. $Yumi_0$-ka_1 cip_2-ey_3 ka_4-n_5 il_6-un_7 $chamulo_8$ $yukamsulep_9$-$ta_{10\cdot11}$
$Yumi_0$-NOM$_1$ home$_2$-AT$_3$ go$_4$-ADNOMINAL$_5$ event$_6$-TOP$_7$ truly$_8$ regrettable$_9$-DECL$_{10\cdot11}$
'It is truly regrettable that Yumi went home'

\<SUBJECTIVITY\>   anchor="0-11"   id="s1" polarity="negative"   intensity="high" **\</SUBJECTIVITY\>**
\<SEED\> anchor="8-9" id="u1" type="direct-explicit"   subjectivity-type="judgment-neg" nested-source="w"   target="0-6" polarity="negative"   intensity="high" insubstantial="FALSE" **\</SEED\>**

186

Measure	Recall (A1‖A2)	Recall (A2‖A1)	F-measure	Recall (A2‖A3)	Recall (A3‖A2)	F-measure	Recall (A3‖A1)	Recall (A‖A3)	F-measure
Agreement	0.9	0.29	0.595	0.78	0.83	0.80	0.43	0.92	0.675

Table3. SEED Tag Anchor Agreement

b. $\text{Yumi}_{12}\text{-nun}_{13}$ kkoley_{14} $\text{cip}_{15}\text{-ul}_{16}$ $\text{sa}_{17}\text{-ss}_{18}\text{-ta}_{19. 20}$
$\text{Yumi}_{12}\text{-TOP}_{13}$ $\text{in.a.pathetic.state}_{14}$ $\text{home}_{15}\text{-ACC}_{16}$ $\text{buy}_{17}\text{-PAST}_{18}\text{-DECL}_{19.20}$
'Yumi was pathetic but she bought a house'

> **<SUBJECTIVITY>** anchor="12-19" id="s2" polarity="negative" intensity="high" **</SUBJECTIVITY>**
> **<SEED>** anchor="14" id="u1" type="writing-device" subjectivity-type="judgment-neg" nested-source="w" target="12" polarity="negative" intensity="high" insubstantial="FALSE" **</SEED>**

c. $\text{Yumi}_{21}\text{-ka}_{22}$ $\text{cip}_{23}\text{-ul}_{24}$ $\text{sa}_{25}\text{-n}_{26}$ $\text{il}_{27}\text{-un}_{28}$ $\text{yukamsulewu}_{29}\text{-n}_{30}$ $\text{saken}_{31}\text{-ulo}_{32}$ $\text{pokotoy}_{33}\text{-ess}_{34}\text{-ta}_{35.36}$
$\text{Yumi}_{21}\text{-NOM}_{22}$ $\text{home}_{23}\text{-ACC}_{24}$ $\text{buy}_{25}\text{-ADNOMINAL}_{26}$ $\text{event}_{27}\text{-TOP}_{28}$ $\text{regrettable}_{29}\text{-ADNOMINAL}_{30}$ $\text{event}_{31}\text{-as}_{32}$ $\text{be.reported}_{33}\text{-PAST}_{34}\text{-DECL}_{35.36}$
'It was reported as a regrettable event that Yumi bought a house'

> **<OBJECTIVITY>** anchor="21-36" id="o1" **</OBJECTIVITY>**
> **<SEED>** anchor="29" id="u1" type="indirect" subjectivity-type="judgment-neg" nested-source="w," target="31" polarity="negative" intensity="medium" insubstantial="FALSE" **</SEED>**

4 Inter-Annotator Agreement Tests

4.1 The First Agreement Test

Once we set up our preliminary annotation schemes for the Korean Sentiment Corpus, we had three different annotators (A1, A2, and A3) created sample annotations and then checked the degree of agreement amongst their annotations. After careful investigation of these pilot annotations, we continued changing and developing these schemes.

Let's briefly look at the procedure. The first agreement test focused on three main issues. The first issue was whether annotators would recognize the same subjective expressions as SEED tags. The second and the third issues were whether annotators assigned the same values to the express types and subjectivity-type attributes respectively.

Cohen's Kappa (k) is not appropriate for measuring the inter-annotator agreement for SEED tags because it is only applicable to annotators annotating the same set of expressions. Instead, our annotators annotated different expressions, thus, following Wilson (2008), we used F-measure. F-measure is a harmonic mean of recalls from annotation results. When A and B are the set of anchors annotated by annotator a and b, the recall of a with respect to b (recall (a‖b)) is as below

$$\text{Recall } (a\|b) = \frac{|A\ matching\ B|}{|A|}$$

The F-measure in turn is the mean of recall (A1‖A2) and recall (A2‖A1). The SEED tag agreement result is shown in table 3. The result shows that there is a noticeable asymmetry in the recalls between (A1‖A2) and (A2‖A1). This is because the annotator *A2* created a much larger number of SEED tags compared to *A1*. The overall F-measure was not sufficient to settle on this annotation scheme. This SEED tag agreement could not be improved much since it was a measure of what people recognize as subjective expressions. Annotators are likely to depend on their intuition about subjective expressions.

The agreements between sentence level OBJECTIVITY and SUBJECTVITY values were even worse than the previous SEED tag agreement. There was no consensus amongst annotators on when to give what values for each attribute. For these measures, we used Krippendorff's Alpha [1] (Krippendorff, 1998; 2004)

krippendorff's alpha	A1-A2	A1-A3	A2-A3
Agreement	0.408	0.730	0.132

Table 4. Expressive Type Agreement

---

[1] α = 1 indicates perfect reliability. α = 0 indicates the absence of reliability. α < 0 indicates disagreements are systematic and exceed what can be expected by chance.

krippendorff's alpha	A1-A2	A1-A3	A2-A3
Agreement	-0.343	-0.397	0.214

Table 5. Subjectivity Type Agreement

As seen in Table 4 and 5, the inter-annotator agreements for SUBJECTIVITY type were significantly different from each other. This indicated not only that all the annotators needed more training on the annotation guidelines, but also that some modification of the attributes and values was necessary.

After the first agreement test, we divided the *sentiment* value of the subjectivity type attribute into *emotion* and *judgment*, as we found the *sentiment* value category was too broad and vague to define all those expressions. Also, a *writing-device* category was added to expressive types Even though the *expressive* category seems to include many different types of subjective expressions, it was hard to make a clear boundary between expressions. Thus, we chose to mark them all as expressive, as we had previously done, except those of *writing-device* type. Beyond these two changes, many vague categories were more precisely defined and thoroughly discussed.

4.2 The Second Agreement Test

Another one-hundred sentences were annotated by the same annotators as the first agreement test. The agreement test results are stated in Table 6.

Despite some degree of disagreement for all types of tags, the overall agreement between annotators showed a marked improvement across all types except SEED tags. As mentioned, we expected that the SEED tag agreement would not increase during this second agreement test. Note how the SEED tag agreement between annotators A1 and A2 did show an increase but that this was canceled out by a decrease in agreement between the other two pairs.

On the other hand, the expressive type and subjectivity type agreements improved significantly. Despite this, we still needed to further refinement for our annotation guidelines. Due to experience gained during these evaluations,

detailed instructions about how to annotate *writing-device* type expressions were created Additionally, 'say' type expressions, which were one of the most frequently confusing cases, were discussed in more detail. Furthermore, we were able to reach a consensus on the way SEED tags and targets should be annotated.

5 Future Work and Conclusions

We have begun this project building the Korean Sentiment Corpus. The goal of this first year was to investigate theoretical foundations and to make tools for manual annotations. Regarding theoretical background, we followed the annotation scheme and the framework proposed by the MPQA corpus. The framework of the MPQA is similar to that of Appraisal Theory by Martin (2000) and White (2002). The Appraisal framework is composed of concepts including *Affect, Judgment, Appreciation, Engagement*, and *Amplification. Affect, Judgment*, and *Appreciation* represent different types of positive and negative attitudes. According to Wiebe et al. (2005) the similarity between these approaches is that they are both concerned with systematically identifying expressions of opinions and emotions in context.

Nonetheless, the MPQA corpus does not distinguish different types of private states, such as *Affect* and *Judgment*, which can provide useful information in sentiment analysis. On the other hand, the MPQA corpus distinguished different ways that private states may be expressed, i.e. *directly* or *indirectly*.

Our annotation scheme, however, not only covers many types of attitudes as in Appraisal theory but also several expressive types as in the MPQA corpus. Subjectivity types correspond to *Attitude* in Appraisal theory and Expressive types correspond to *direct subjective* or *expressive subjective elements* in the MPQA. We believe that a corpus founded on a comprehensive annotation scheme could be used by researchers as a gold standard for training and testing

Measure	Recall (A1‖A2)	Recall (A2‖A1)	F-measure	Recall (A2‖A3)	Recall (A3‖A2)	F-measure	Recall (A3‖A1)	Recall (A‖A3)	F-measure
Agreement	0.89	0.39	0.64	0.27	0.7	0.46	0.6	0.55	0.58

Table 6. SEED Tag Agreement in the Second Test

krippendorff's alpha	A1-A2	A1-A3	A2-A3
Agreement	0.62	0.50	0.37

Table 7. Expressive Type Agreement

krippendorff's alpha	A1-A2	A1-A3	A2-A3
Agreement	0.57	0.62	0.54

Table 8. Subjectivity Type Agreement

Another important aspect of our work is that, following the MPQA corpus, information on nested sources is incorporated into the annotation scheme. Specifying nested sources can help allow annotated expressions to denote their context below the sentence-level (Wiebe et al. 2005). Furthermore, analyzing nested sources along with speaker's attitudes toward subjectivity allows for a new modality or pragmatics-based methodology for further Sentiment Analysis. We will pursue this approach further after our initial annotation task has been completed.

Along with the elaboration of annotation scheme for the Korean Sentiment Corpus, we also developed annotation tools to aid manual tagging. We created a Graphical User Interface (GUI) which allowed annotators to easily search our corpus of Korean news stories by individual morphemes, by words or by article. Annotators could then select entire sentences or individual morphemes along with specify the desired annotation attributes and automatically generate the appropriate annotation.

This tool utilized the wxPython library to create the GUI while a Python core communicated with a database. This database in turn stored the corpus text, already parsed and separated into morphemes, as well as any annotations an annotator created. This allowed annotators to review and modify previously created tags.

The main goal of the annotation scheme presented in this paper was to support the development of the Korean Sentiment Corpus. We plan to complete the annotation of about 8,750 Korean sentences by April, 2013 after which the corpus will be opened to public for research purposes. We believe that researchers will be able to extract useful information from the corpus and use the data for training and testing in sentiment and opinion analysis.

References

Dan Gusfield. 1997. Algorithms on Strings, Trees and Sequences. Cambridge University Press, Cambridge, UK.

Edward Finergan. 1995. Subjectivity and Subjectification: an Introduction. In D. Stein & S. Wright (Eds.), Subjectivity and Subjectification: Linguistic Perspectives:1-15. Cambridge University Press, Cambridge.

Ellen Riloff, Janyce Wiebe, and William Phillips. 2005. Exploiting Subjectivity Classification to Improve Information Extraction. In Proceedings of the 20th National Conference on Artificial Intelligence (AAAI-2005): 1106-1111.

James Pustejovsky, Patrick Hanks, Roser Sauri, Andrew See, Robert Gaizauskas, Andrea Setzer, Beth Sundheim, Lisa Ferro, Marcia Lazo, Inderjeet Mani, and Dragomir Radev. 2003. The TimeBank corpus. In Proceedings of Corpus Linguistics 2003: 647-656.

Janyce Wiebe. 2002. Instructions for Annotating Opinions in Newspaper Articles. Department of Computer Science Technical Report TR-02-101, University of Pittsburgh.

189

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. Language Resources and Evaluations, 39(2):165-210.

Janyce Wiebe and Ellen Riloff. 2005. Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. In Proceedings of the 6th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2005): 486-497.

Krippendorff, Klaus. 1978. Reliability of Binary Attribute Data. *Biometrics*, 34 (1), 142-144.

Krippendorff, Klaus. 2004. Content Analysis: An Introduction to Its Methodology. Thousand Oaks, CA: Sage.

Lauri Karttunen. 1971. Some observations on Factivity. *Papers in Linguistics*, 47:340-358.

MPQA. 2005. Multi-Perspective Question Answering. University of Pittsburgh. http://www.cs.pitt.edu /mpqa/.

Ronald Langacker. 1985. Observations and speculations in subjectivity. In J. Haiman (Ed*.),* Iconicity in Syntax. Typological Studies in Language 6. John Benjamins, Amsterdam/Philadelphia.

Randolf Quirk, Sidney Greenbaum, Geoffry Leech, and Jan Svartvik. 1985. A Comprehensive Grammar of the English Language. New York.

Roger Sauri. 2008. A Factuality Profiler for Eventualities in Text. Ph.D Dissertation, Brandeis University.

Theresa Ann Wilson. 2008. Fine-Grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States. Ph.D Dissertation, University of Pittsburgh.

Wiebe, Janyce, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. Language Resources and Evaluation (formerly Computers and the Humanities), 39(2/3):164–210.

Wiebe, J. 2002. Instructions for annotating opinions in newspaper articles. Department of Computer Science Technical Report TR-02-101, University of Pittsburgh.