

Chapter 6

Opinion Analysis Corpora Across Languages



Yohei Seki

Abstract At *NTCIR-6*, 7, and 8, we included a new multilingual opinion analysis task (*MOAT*) that involved Japanese, English, and Chinese newspapers. This was the first task that compared the performance of sentiment retrieval strategies with common subtasks across languages. In this paper, we introduce the research question posed by *NTCIR MOAT* and present what has been achieved to date. We then describe the types of tasks and research that have involved our test collection both previously and in current research. Finally, we summarize our contributions and discuss future research directions.

6.1 Introduction

Sentiment analysis (sometimes called “opinion mining”) is a research topic that has been actively discussed and developed for some 20 years, particularly in the fields of natural language processing (NLP) and information retrieval (IR) (Pang and Lee 2008). In this paper, we introduce the multilingual opinion analysis task (*MOAT*) (Seki et al. 2010, 2008, 2007), which was included in *NTCIR-6*, 7, and 8 (2006–2010). We then discuss the role and novelty of the task in sentiment analysis research.

Sentiment analysis research began in 2002 (Pang et al. 2002; Turney 2002; Wiebe et al. 2002). Various frameworks for classifying documents in terms of positivity or negativity that use either supervised learning (Pang et al. 2002) or unsupervised learning (Turney 2002) have been proposed. In parallel, many researchers started to build opinion corpora based on newspaper articles (Wiebe et al. 2002) for multi-perspective question answering (*MPQA*). Other early research work was published at the *AAAI 2004 Spring Symposium: Exploring Attitude and Affect in Text: Theories and Applications* (Shanahan et al. 2006).

Y. Seki (✉)

University of Tsukuba, 1-2 Kasuga, Tsukuba, Ibaraki 305-8550, Japan
e-mail: yohei@slis.tsukuba.ac.jp

© The Author(s) 2021

T. Sakai et al. (eds.), *Evaluating Information Retrieval and Access Tasks*,
The Information Retrieval Series 43,
https://doi.org/10.1007/978-981-15-5554-1_6

At the Text Retrieval Conference (TREC) in 2006, a new “Blog Track” was introduced, and was continued until 2010.¹ The original organizers released the TREC Blogs06 Collection (Macdonald and Ounis 2006), for which there have been 100,649 blog posts (excluding duplicate documents) and over 3.2 million permalinks. This dataset was used for the opinion finding (blog post) retrieval task in the *TREC 2006 Blog Track* and for the polarity opinion finding (blog post) retrieval task in the *TREC 2007 Blog Track*. In addition, the *MPQA* opinion corpus from the University of Pittsburgh (Wiebe et al. 2005), which defines a framework for opinion annotation using multiple assessors, has been released.

Building on this previous work, we introduced our opinion analysis task at *NTCIR-6* in 2006. The novel aspects of the *NTCIR MOAT* task can be summarized as follows:

1. We have released an opinion annotation corpus for evaluation workshops. The annotation units include opinionatedness, topic relevance, polarity, opinion holder (from *NTCIR-6*), and opinion target (from *NTCIR-7*).
2. We have provided a multilingual opinion corpus that includes material in English, Chinese, and Japanese.
3. The topic set in the evaluation corpus is shared across languages.

In Sect. 6.2, we give details of the *NTCIR MOAT* design to clarify its novel features and suggest an opinion corpus annotation strategy for evaluation workshops. In Sect. 6.3, we explain the evolution of opinion analysis research since the introduction of *MOAT*. Finally, in Sect. 6.4, we conclude our remarks and discuss future research directions.

6.2 NTCIR MOAT

6.2.1 Overview

NTCIR MOAT was held at *NTCIR-6* (Seki et al. 2007), *NTCIR-7* (Seki et al. 2008), and *NTCIR-8* (Seki et al. 2010). The task definition evolved through the three sessions, as shown in Table 6.1.

The goal of the task is to form a bridge between element technologies such as opinion/polarity sentence classification or opinion holder/target phrase recognition to an application such as (opinion) IR or question answering. The target languages include English, Chinese (both Traditional and Simplified), and Japanese, and the topic set for IR or question answering is shared across languages. We have prepared a document set relevant to the topics retrieved from newspaper articles published in each target language, and have evaluated the system using these document sets annotated with multiple assessors.

¹<http://trec.nist.gov/data/blog.html>.

Table 6.1 *MOAT* progress during *NTCIR-6*, *7*, & *8*

	<i>NTCIR-6</i>	<i>NTCIR-7</i>	<i>NTCIR-8</i>
Target	English, Japanese, Traditional Chinese		
Language	–	+Simplified Chinese	
Subtasks	Opinionated, Relevance, Polarity, Holder	+Target	+Cross-lingual
Annotation Unit	Sentence	Opinion Clause	
Focused Application	Information Retrieval	Q&A (<i>ACLIA</i> ^a)	Opinion Q&A
Target Corpora	Mainichi, Yomiuri, CIRB, Xinhua English, Hong Kong Standard, etc.	+Xinhua Chinese	+NYT, UDN
(Period)	1998–2001		2002–2005

^a<http://research.nii.ac.jp/ntcir/permission/ntcir-7/perm-ja-ACLIA.html>

6.2.2 Research Questions at *NTCIR MOAT*

Many researchers have focused on a resourceless approach to sentiment analysis (Elming et al. 2014; Le et al. 2016). Blitzer et al. (2007) proposed a domain adaptation approach for sentiment classification. Wan (2009) addressed the Chinese sentiment classification problem by using English sentiment corpora on the Internet. This type of research can be categorized as a semi-supervised approach to opinion/sentiment analysis that aims to solve the resource problem by using small labeled and large unlabeled datasets. We recognize that addressing language resource problems in sentiment analysis for nonnative languages is an important research area. Alternatively, applications such as the *Europe Media Monitor (EMM) News Explorer*² provide an excellent service by including viewpoints from different countries. We also understand that providing these varied opinions from different countries offers opportunities for better worldwide communications. *NTCIR MOAT* is the first task to provide opportunities for nonnative researchers to develop a sentiment analysis system for low-resource languages and to bridge cultures by clarifying opinion differences across different languages.

²<http://emm.newsbrief.eu/NewsBrief/clusteredition/en/latest.html>.

6.2.3 Subtasks

With the broad range of information sources available on the web and in social media, there has been increased interest by both commercial and governmental parties in trying to analyze and monitor the flow of prevailing attitudes from anonymous users automatically. As a result, the research community has given much attention to automatic identification and processing of the following.

- Sentences in which an opinion is expressed (Wiebe et al. 2004),
- The polarity of the expression (Wilson et al. 2005),
- The opinion holders of the expression (Choi et al. 2005),
- The opinion targets of the expression (Ruppenhofer et al. 2008), and
- Opinion question and answering (Stoyanov et al. 2005), (Dang 2008).³

With these factors in mind, we defined the subtasks in *NTCIR MOAT* as follows.

1. Opinionated sentences

The judgment of opinionated sentences is a binary decision for all sentences.

2. Relevant sentences

Each set contains documents that are found to be relevant to an opinion question, such as that shown in Fig. 6.1. For those participating in the relevance subtask evaluation, each opinionated sentence should be judged as either relevant (Y) or non-relevant (N) to the opinion questions. In *NTCIR-8 MOAT*, only opinionated sentences were annotated for relevance.

3. Opinion polarities

The polarity is determined for each opinion clause. In addition, the polarity is to be determined with respect to the topic description if the sentence is relevant to the topic, and based on the attitude of the opinion if the sentence is not relevant to the topic. The possible polarity values are positive (POS), negative (NEG), or neutral (NEU).

4. Opinion holders

The opinion holders are annotated in terms of opinion clauses that express an opinion. However, the opinion holder for an opinion clause can occur anywhere in the document. The assessors performed a kind of co-reference resolution by marking the opinion holder with the opinion clause if the opinion holder makes an anaphoric reference noting the antecedent of the anaphora. Each opinion clause must have at least one opinion holder.

5. Opinion targets

The opinion targets were annotated in a similar manner to the opinion holders. Each opinion clause must have at least one opinion target.

6. Cross-lingual opinion Q&A

The cross-lingual subtask is defined as the opinion Q&A task. Together with the questions in English, the answer opinions should be extracted in different languages. To keep it simple, the extraction unit is defined as a sentence. The

³<https://tac.nist.gov/2008/qa/index.html>.

```

<TOPIC>
<NUM>N03</NUM>
<TITLE>Bali Island Terrorist Bombing</TITLE>
<QUESTION>What reasons behind the 2002 Bali bombings were discussed?</QUESTION>
<POLARITY>Neutral</POLARITY>
<OPTYPE>Reason controversy</OPTYPE>
<CONC>Bali Island Terrorist Bombing</CONC>
<PERIOD>2002-1010</PERIOD>
</TOPIC>

```

Fig. 6.1 Example: opinion question fields at *NTCIR-8 MOAT*

answer set is defined as the combination of the annotations for the conventional subtasks, with opinionatedness, polarity, and answeredness being matched with the definition in the question description.

6.2.4 Opinion Corpus Annotation Requirements

Opinion corpus annotation for multiple domains (as in news topics) usually requires expert linguistic knowledge because crowdsourcing annotation (such as the Amazon Mechanical Turk) does not fit the *NTCIR MOAT* annotation framework. We conducted our evaluation using agreed (intersection) annotations from multiple expert assessors. To check the stability of this evaluation strategy, we compared the evaluation results for agreed (intersection) annotation and selective (union) annotation to arrive at a gold standard for using *NTCIR-8 MOAT* submission data.

For the English cases in Table 6.2 (the κ coefficient between assessor annotations was 0.73) and the Traditional Chinese cases in Table 6.3 (κ coefficient 0.46), the rank of the participants' systems is different. Although the rank differences for the English cases were within statistical significance, among the Traditional Chinese cases, the precision-oriented systems (*CTL* and *WIA*) tended to be ranked higher for cases of agreed (intersection) annotation, and recall-oriented systems (*KLELAB-1* and *NTU*) tended to be ranked lower. For the Simplified Chinese cases in Table 6.4 (κ coefficient 0.97) and the Japanese cases in Table 6.5 (κ coefficient 0.72), there was no rank difference for the participants' systems despite the different strategies because of either high κ agreement (Simplified Chinese) or a low number of participants (Japanese). From these observations, we concluded that the κ coefficient between assessor annotations should exceed 0.7 for stable evaluation. We also found that strong opinion definition and online annotation tools were helpful, but using expert linguistic annotators remained necessary to achieve high κ agreement.

Table 6.2 Evaluation strategy analysis using *NTCIR-8 MOAT* English raw submission dataEnglish (F1-score)/ $\kappa = 0.73$

Rank on agreed	Significance								Rank on non-agreed
UNINE-1	A								UNINE-1
NECLC-bsf	A	B							NECLC-bs1
NECLC-bs0	A	B	C						NECLC-bsf
NECLC-bs1	A	B	C	D					NECLC-bs0
UNINE-2		B	C	D					UNINE-2
KLELAB-3		B	C	D	E				NTU-2
KAISTIRNLP-2			C	D	E				KLELAB-2
KLELAB-2			C	D	E				KLELAB-3
KAISTIRNLP-1			C	D	E				NTU-1
NTU-2				D	E	F			KLELAB-1
KLELAB-1				D	E	F			KAISTIRNLP-2
NTU-1				D	E	F			KAISTIRNLP-1
OPAL-2					E	F	G		OPAL-1
OPAL-3						F	G		OPAL-2
OPAL-1						F	G		OPAL-3
PolyU-1							G		SICS-1
SICS-1							G	H	PolyU-1
PolyU-2								H	PolyU-2

6.2.5 Cross-Lingual Topic Analysis

We ranked topics by averaging their F1-scores, the harmonic mean of precision and recall, obtained from all *NTCIR-8 MOAT* raw submissions in the opinionated judgment subtask. The best three (easy) topics and worst three (difficult) topics and the opinion percentage in the source documents are shown in Table 6.6.

From these results, we found that the topic difficulty is strongly related to each language. We also found that, with many opinions in the source, the topics tended to be easier. Exceptions to this rule included the opinion question for topic *N16*: “What reasons have been given for the anti-Japanese demonstrations that took place in April, 2005 in Peking and Shanghai in China?” We surmise that this was caused by the systems’ difficulty in judging quite sensitive opinions expressed in newspaper articles in each language.

Table 6.3 Evaluation strategy analysis using *NTCIR-8 MOAT* traditional Chinese raw submission data

Traditional Chinese (F1-score)/ $\kappa = 0.46$

Rank on agreed	Significance				Rank on non-agreed
CityUHK-2	A				CityUHK-1
CTL-1	A				CityUHK-3
CityUHK-1	A				KLELAB-1
CityUHK-3	A				NTU-1
WIA-1	A				NTU-2
WIA-2	A				CityUHK-2
KLELAB-3		B			cyut-1
KLELAB-1		B			KLELAB-3
NTU-2		B			WIA-1
NTU-1		B			WIA-2
cyut-1		B			cyut-2
cyut-2		B	C		CTL-1
UNINE-1			C	D	UNINE-1
cyut-3				D	cyut-3

Table 6.4 Evaluation strategy analysis using *NTCIR-8 MOAT* simplified Chinese raw submission data

Simplified Chinese (F1-score)/ $\kappa = 0.97$

Rank on agreed	Significance				Rank on non-agreed
PKUTM-2	A				PKUTM-2
PKUTM-1	A	B			PKUTM-1
BUPT-2	A	B			BUPT-2
CTL-1		B			CTL-1
PKUTM-3		B	C		PKUTM-3
BUPT-1		B	C		BUPT-1
WIA-1			C	D	WIA-1
WIA-2			C	D	WIA-2
NECLC-bsf				D	NECLC-bsf
NECLC-bs0				D	NECLC-bs0
NECLC-bs1				D	NECLC-bs1
PolyU-1				E	PolyU-1

Table 6.5 Evaluation strategy analysis using *NTCIR-8 MOAT* Japanese raw submission data

Japanese (F1-score)/ $\kappa = 0.72$

Rank on agreed	Significance				Rank on non-agreed
TUT-1	A				TUT-1
TUT-3	A	B			TUT-3
IISR-3		B	C		IISR-3
TUT-2		B	C		TUT-2
IISR-1		B	C		IISR-1
IISR-2			C		IISR-2
UNINE-1				D	UNINE-1

Table 6.6 Cross-lingual topic analysis using *NTCIR-8 MOAT* raw submission data

	English		Traditional Chinese		Simplified Chinese		Japanese	
	Topic	Opinion % in doc set	Topic	Opinion % in doc set	Topic	Opinion % in doc set	Topic	Opinion % in doc set
Easy topics	N27	25.4	N14	56.5	N18	24.5	N41	34.6
	N39	21.2	N05	55.6	N20	20.7	N11	35.3
	N14	21.3	N27	57.6	N06	22.7	N13	28.1
Difficult topics	N18	7.6	N16	19.4	N07	9.5	N24	35.3
	N13	8.9	N13	15.0	N41	14.9	N18	37.7
	N06	10.0	N20	18.8	N16	20.6	N32	27.0
Average	Avg.	16.7	Avg.	32.1	Avg.	18.6	Avg.	33.9

6.3 Opinion Analysis Research Since *MOAT*

6.3.1 Research Using the *NTCIR MOAT* Test Collection

Some researchers have used the *NTCIR MOAT* test collection and presented their work at top-rated conferences, particularly those focused on cross-lingual sentiment analysis. Two representative examples are as follows.

1. Joint Bilingual Sentiment Classification

Lu et al. (2011) hypothesized that aligned sentences between languages should be similar in opinion polarity and strongness. They proposed a method for improving the polarity classification performance that used the *MPQA* opinion corpus and the *NTCIR MOAT* corpus as labeled corpora, and aligned news corpora in Chinese and English as unlabeled corpora. They extended their work by using a cross-

lingual mixture model (Meng et al. 2012) to improve performance when learning polarity clues from unlabeled corpora.

2. Cross-lingual Sentiment Lexicon Learning

Gao et al. (2015) proposed a method for generating low-resource language sentiment lexicons using available English sentiment lexicons. They created Chinese sentiment lexicons using a bilingual word graph label propagation approach. They evaluated Chinese sentiment classification at the sentence level by using the *NTCIR MOAT* corpus and found increased effectiveness of sentiment classification when using their generated sentiment lexicon to generate features.

6.3.2 *Opinion Corpus in News*

Several opinion corpora involving news have been developed after *NTCIR MOAT* was published. In this subsection, we introduce the *SemEval-2007 Task 14: Affective Corpus* (Strapparava and Mihalcea 2007) and the *sentiment-annotated quotation set* (Balahur and Steinberger 2009; Balahur et al. 2010).

In the *SemEval-2007 Affective Corpus*, six emotion labels and two polarity labels have been annotated to headlines collected from 1,250 news websites and newspaper articles. The *sentiment-annotated quotation set* contains a set of 1,590 English language quotations (reported speech), manually annotated by two independent sets of annotators for sentiment (positive, negative, or objective/neutral) expressed toward the entities mentioned inside the quotation. Web crawling for news articles employed the *EMM* (Steinberger et al. 2009)⁴ developed by the *European Commission Joint Research Centre*.

The *NTCIR MOAT* corpus, however, remains in use as a large cross-lingual news opinion corpus targeted at Chinese, Japanese, and English.

6.3.3 *Current Opinion Analysis Research: The Social Media Corpus and Deep NLP*

After *NTCIR MOAT* was published, Twitter⁵ and other microblog media came into widespread use by many users. The NLP/IR researchers also focused on tweet sentiment analysis (Martinez-Camara et al. 2013). To improve sentiment classification in Twitter, specific clues were found to be useful because a tweet is much shorter than a news article, including tweet context (Jiang et al. 2011), emoticons and hashtags (Purver and Battersby 2012), lengthened words (Brody and Diakopoulos 2011), and emoji (Felbo et al. 2017).

⁴<http://emm.newsbrief.eu/overview.html>.

⁵<http://twitter.com>.

On the other hand, deep NLP research such as *Stanford Sentiment Treebank* (Socher et al. 2013)⁶ has become mainstream from a technological point of view. In this research, the learning model builds up a representation of whole sentences based on the sentence structure. An opinion corpus called the *Stanford Sentiment Treebank* has been developed to estimate compositionality in the sentiment detection task. It includes the fine-grained sentiment labels “very negative”, “negative”, “neutral”, “positive”, and “very positive” for 215,154 phrases in trees parsed with the Stanford Parser from 11,855 sentences extracted from movie reviews (Pang and Lee 2005).

In *SemEval 2018* (Mohammad et al. 2018), an opinion corpus has been created from 10,983 English, 4,381 Arabic, and 7,094 Spanish tweets, and used to evaluate the systems. Several tasks are defined that provide annotations for the mental state of the tweeter, including (1) the intensities of the four basic emotions (anger, fear, joy, and sadness), (2) the intensity of sentiment/valence (very negative, moderately negative, slightly negative, neutral or mixed, slightly positive, moderately positive, and very positive), and (3) multi-label emotion classification across 12 emotions (anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, trust, and neutral). The corpus used *best-worst scaling* (Louviere et al. 2015), a comparative annotation method in which assessors were asked what was the best (highest in terms of the property) and worst (lowest in terms of the property), given n items (typically $n = 4$). Real-valued scores for the association between the items and the property were determined based on the number of times an item was chosen as the best and the worst. The median number of assessors for each tweet was seven. The inter-annotator agreements (*Fleiss's κ*) for the multi-label emotion classification were 0.21, 0.29, and 0.28 for the 12 classes, and 0.40, 0.48, and 0.45 for the four basic emotions in English, Arabic, and Spanish. Most of the participants employed SVM/SVR, LSTMs, and Bi-LSTMS as machine learning algorithms, and also took word embedding, affect lexicon features, and word n-grams as features.

Although the document genres being focused on and the annotation properties have changed over time, cross-lingual opinion corpora remain important in current research.

6.4 Conclusion

In this paper, we have discussed the contributions made by our development of *NTCIR MOAT*. We created a cross-lingual opinion corpus using the news document genre, following which, several researchers have conducted cross-lingual opinion research using our test collections. Although sentiment classification accuracy is improved by using a cross-lingual corpus, research investigating linguistic opinion properties characterized by languages rooted in different cultures and opinion retrieval strategies preferable for different language characteristics remain to be undertaken.

⁶<https://nlp.stanford.edu/sentiment/>.

- Jawahar G, Sagot B, Seddah D (2019) What does BERT learn about the structure of language? In: Proceedings of the 57th annual meeting of the association for computational linguistics, Florence, Italy, pp 3651–3657. <https://doi.org/10.18653/v1/P19-1356>
- Jiang L, Yu M, Zhou M, Liu X, Zhao T (2011) Target-dependent twitter sentiment classification. In: Proceedings of the 49th annual meeting of the association for computational linguistics (ACL 2011), Portland, Oregon, pp 151–160
- Le TA, Moeljadi D, Miura Y, Ohkuma T (2016) Sentiment analysis for low resource languages: a study on informal Indonesian tweets. In: Proceedings of the 12th workshop on Asian language resources (ALR), The COLING 2016 Organizing Committee, Osaka, Japan, pp 123–131. <https://www.aclweb.org/anthology/W16-5415>
- Louviere JJ, Flynn TN, Marley AAJ (2015) Best-worst scaling: theory, methods and applications. Cambridge University Press, Cambridge
- Lu B, Tan C, Cardie C, Tsou BK (2011) Joint bilingual sentiment classification with unlabeled parallel corpora. In: Proceedings of the 49th annual meeting of the association for computational linguistics (ACL 2011), Portland, Oregon, pp 320–330
- Macdonald C, Ounis I (2006) The TREC Blogs06 collection: creating and analysing a blog test collection. Technical report TR-2006-224, Department of Computing Science, University of Glasgow
- Martinez-Camara E, Martin-Valdivia MT, Urena-Lopez LA, Montejo-Raez AR (2013) Sentiment analysis in twitter. *Nat Lang Eng* 11(2):1–28
- Meng X, Wei F, Liu X, Zhou M, Xu G, Wang H (2012) Cross-lingual mixture model for sentiment classification. In: Proceedings of the 50th annual meeting of the association for computational linguistics (ACL 2012), Jeju, Republic of Korea, pp 572–581
- Mohammad SM, Bravo-Marquez F, Salameh M, Kiritchenko S (2018) SemEval-2018 task 1: affect in tweets. In: Proceedings of the 12th international workshop on semantic evaluation (SemEval-2018), New Orleans, Louisiana, pp 1–17. <https://doi.org/10.18653/v1/S18-1001>
- Pang B, Lee L (2005) Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL 2005), Ann Arbor, Michigan, pp 115–124
- Pang B, Lee L (2008) Opinion mining and sentiment analysis. Now Publishers Inc, Boston
- Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the 2002 conference on empirical methods in natural language processing (EMNLP 2011), pp 79–86
- Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L (2018a) Deep contextualized word representations. In: Proceedings of the 2018 conference of the North American Chapter of the association for computational linguistics: human language technologies, volume 1 (long papers), New Orleans, Louisiana, pp 2227–2237. <https://doi.org/10.18653/v1/N18-1202>
- Peters M, Neumann M, Zettlemoyer L, tau Yih W (2018b) Dissecting contextual word embeddings: architecture and representation. In: Proceedings of the 2018 conference on empirical methods in natural language processing (EMNLP 2018), Brussels, Belgium, pp 1499–1509
- Purver M, Battersby S (2012) Experimenting with distant supervision for emotion classification. In: Proceedings of the 13th European chapter of the association for computational linguistics (EACL 2012), Avignon, France, pp 482–491
- Ruppenhofer J, Somasundaran S, Wiebe J (2008) Finding the sources and targets of subjective expressions. In: Proceedings of the 6th international language resources and evaluation (LREC'08), Marrakech, Morocco
- Seki Y, Evans DK, Ku LW, Chen HH, Kando N, Lin CY (2007) Overview of opinion analysis pilot task at NTCIR-6. In: Proceedings of the 6th NTCIR workshop meeting, NII, Japan, pp 265–278
- Seki Y, Evans DK, Ku LW, Sun L, Chen HH, Kando N (2008) Overview of multilingual opinion analysis task at NTCIR-7. In: Proceedings of the 7th NTCIR workshop meeting, NII, Japan, pp 185–203

- Seki Y, Ku LW, Sun L, Chen HH, Kando N (2010) Overview of multilingual opinion analysis task at NTCIR-8 - a step toward cross lingual opinion analysis. In: Proceedings of the 8th NTCIR workshop meeting, NII, Japan, pp 209–220
- Shanahan JG, Qu Y, Wiebe JM (2006) Computing attitude and affect in text: theory and applications. Springer, Berlin
- Socher R, Perelygin A, Wu J, Manning JCCD, Ng A, Potts C (2013) Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the (2013) conference on empirical methods in natural language processing (EMNLP 2013). Association for Computational Linguistics, Seattle, Washington, USA, pp 1631–1642
- Steinberger R, Pouliquen B, van der Goo E. (2009) An introduction to the Europe media monitor. In: Proceedings of ACM SIGIR 2009 workshop: information access in a multilingual world, Boston, MA, USA
- Stoyanov V, Cardie C, Wiebe J (2005) Multi-perspective question answering using the OpQA corpus. In: Proceedings of the 2005 human language technology conference and conference on empirical methods in natural language processing (HLT/EMNLP 2005), Vancouver, B. C
- Strapparava C, Mihalcea R (2007) SemEval-2007 task 14: affective text. In: Proceedings of the 4th international workshop on semantic evaluations (SemEval-2007), Prague, pp 70–74
- Turney P (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, Philadelphia, USA, pp 417–424
- Wan X (2009) Co-training for cross-lingual sentiment classification. In: Proceedings of the 47th annual meeting of the association of computational linguistics (ACL), Suntec, Singapore, pp 235–243
- Wiebe J, Breck E, Buckley C, Cardie C, Davis P, Fraser B, Litman D, Pierce D, Riloff E, Wilson T (2002) NRRC summer workshop on multiple-perspective question answering final report
- Wiebe J, Wilson T, Cardie C (2005) Annotating expressions of opinions and emotions in language. *Lang Resour Eval* 39(2–3):165–210
- Wiebe JM, Wilson T, Bruce RF, Bell M, Martin M (2004) Learning subjective language. *Comput Linguist* 30(3):277–308
- Wilson T, Wiebe J, Hoffmann P (2005) Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of the 2005 human language technology conference and conference on empirical methods in natural language processing (HLT/EMNLP 2005), Vancouver, B. C
- Zhou H, Chen L, Shi F, Huang D (2015) Learning bilingual sentiment word embeddings for cross-language sentiment classification. In: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: long papers), Beijing, China, pp 430–440. <https://doi.org/10.3115/v1/P15-1042>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

