



**Main**  
MPQA Home

**Corpora**  
News, debates, etc.

**Lexicons**  
Subj. clues, etc.

**Annotation**  
GATE, MPQA, gfbf

**OpinionFinder**  
Subjectivity detector

[MPOA Opinion Corpus](#)  
[Subjectivity Lexicon](#)  
[Subj. Sense Annotations](#)  
[Arguing Lexicon](#)  
[+/-Effect Lexicon](#)  
[Product Debate Data](#)  
[Political Debate Data](#)  
[goodFor/badFor Data](#)  
[OpinionFinder System](#)

**PLEASE NOTE:** our URL has recently changed from [www.cs.pitt.edu/mpqa/](http://www.cs.pitt.edu/mpqa/) to [mpqa.cs.pitt.edu](http://mpqa.cs.pitt.edu). Please update your bookmarks accordingly.

### o [MPQA Opinion Corpus](#)

The **MPQA Opinion Corpus** contains news articles from a wide variety of news sources manually annotated for opinions and other private states (i.e., beliefs, emotions, sentiments, speculations, etc.). To download the MPQA Opinion Corpus click [here](#).

For **sample documents and instructions for MPQA annotation in GATE**, click [here](#). Updated July 2011.

To learn more about the subjectivity and sentiment research that produced MPQA, please refer to the following publications:

Janyce Wiebe, Theresa Wilson, and Claire Cardie (2005). [Annotating expressions of opinions and emotions in language](#). *Language Resources and Evaluation*, volume 39, issue 2-3, pp. 165-210.

Theresa Wilson (2008). [Fine-Grained Subjectivity Analysis](#). PhD Dissertation, Intelligent Systems Program, University of Pittsburgh.

Lingjia Deng and Janyce Wiebe (2015). [MPOA 3.0: An Entity/Event-Level Sentiment Corpus](#). NAACL-HLT, 2015.

### o [Subjectivity Lexicon](#)

Made available under the terms of [GNU General Public License](#). They are distributed without any warranty.

The **Subjectivity Lexicon** (list of subjectivity clues) that is part of OpinionFinder is also available for separate [download](#). These clues were compiled from several sources (see the enclosed README). This is the version of the lexicon used in:

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann (2005). [Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis](#). *Proc. of HLT-EMNLP-2005*.

### o [Subjectivity Sense Annotations](#)

Made available under the terms of [GNU General Public License](#). They are distributed without any warranty.

The **Subjectivity Sense Annotations** used in (Wiebe and Mihalcea, 2006), (Gyamfi et al., 2009), (Akkaya et al., 2009), and (Akkaya et al., 2011) are all available for [download](#). All annotation efforts follow the annotation schema described in (Wiebe and Mihalcea 2006). Further information on the data can be found in the README of the archive you download.

Janyce Wiebe and Rada Mihalcea (2006). [Word Sense and Subjectivity](#). Joint conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics. (COLING-ACL 2006).

Yaw Gyamfi, Janyce Wiebe, Rada Mihalcea and Cem Akkaya (2009). [Integrating Knowledge for Subjectivity Sense Labeling](#). Joint Conference of the North American Chapter of the Association for Computational Linguistics and the Human Language Technologies Conference (NAACL-HLT 2009).

Cem Akkaya, Janyce Wiebe and Rada Mihalcea. (2009). [Subjectivity Word Sense Disambiguation](#). Conference on Empirical Methods on Natural Language Processing (EMNLP 2009).

Cem Akkaya, Janyce Wiebe, Alexander Conrad and Rada Mihalcea (2011). [Improving the Impact of Subjectivity Word Sense Disambiguation on Contextual Opinion Analysis](#). Conference on Computational Natural Language Learning (CoNLL 2011).

### o [Arguing Lexicon](#)

Made available under the terms of [GNU General Public License](#). They are distributed without any warranty.

The **Arguing Lexicon** is available for [download](#). The lexicon includes patterns that represent arguing. Each file (17 out of 22) represents a type (category) of arguing discussed in (Somasundaran, et al., 2007). Please refer to the README of the archive and the paper for more details.

Swapna Somasundaran, Josef Ruppenhofer and Janyce Wiebe (2007) [Detecting Arguing and Sentiment in Meetings](#), SIGdial Workshop on Discourse and Dialogue, Antwerp, Belgium, September 2007 (*SIGdial Workshop, 2007*).

### ◦ Product Debate Data

The **Product Debate Corpus** is available for [download](#). Further information on the data can be found in the README of the archive you download. This corpus was used in:

Swapna Somasundaran and Janyce Wiebe (2009). [Recognizing Stances in Online Debates](#). In Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, August 2-7, 2009, Singapore (*ACL 2009*)

### ◦ Political Debate Data

The **Political Debate Corpus** is available for [download](#). Further information on the data can be found in the README of the archive you download. This corpus was used in:

Swapna Somasundaran and Janyce Wiebe (2010). [Recognizing Stances in Ideological On-line Debates](#). In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, pages 116-124, Los Angeles, CA. Association for Computational Linguistics, 2010 (*NAACL-HLT, 2010*).

### ◦ goodFor/badFor Data

The **goodFor/badFor Corpus** investigates a kind of event which has either positive or negative affect on the event's object and how it could improve sentiment analysis. The former is defined as benefactive event (goodFor event, for short) and the later is defined as malefactive event (badFor event, for short). The goodFor/badFor data contains annotations for the benefactive event and malefactive event, their agents and objects, and the writer's attitudes towards their agents and objects.

To download the goodFor/badFor Corpus click [here](#).

For **examples and instructions for annotation in GATE**, click [here](#). Updated July 2013.

To learn more about the benefactive/malefactive event and sentiment research based on this data, a brief introduction of annotation scheme are provided [here](#). There are also annotation examples provided. You can also refer to the following publication:

Lingjia Deng, Yoonjoung Choi and Janyce Wiebe (2013). [Benefactive/Malefactive Event and Writer Attitude Annotation](#) In Annual Meeting of the Association for Computational Linguistics (*ACL-2013, short paper*).

### ◦ OpinionFinder System

**OpinionFinder** is a system that processes documents and automatically identifies subjective sentences as well as various aspects of subjectivity within sentences, including agents who are sources of opinion, direct subjective expressions and speech events, and sentiment expressions. OpinionFinder was developed by researchers at the University of Pittsburgh, Cornell University, and the University of Utah. In addition to OpinionFinder, we are also releasing the automatic annotations produced by running OpinionFinder on a subset of the Penn Treebank. To go to the OpinionFinder download page click [here](#).

### ◦ +/-Effect Lexicon

Made available under the terms of [GNU General Public License](#). They are distributed without any warranty.

The **+/-Effect Lexicon** is available for download. The lexicon is created by the work described in (Choi and Wiebe 2014). Please refer to the README of the archive and the paper for more details.

Yoonjung Choi and Janyce Wiebe (2014) [+/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference](#), Proc. of *EMNLP 2014*.

---

contact: [mpqa.project@gmail.com](mailto:mpqa.project@gmail.com)

[[nlp](#)] [[cs](#)] [[pitt](#)]