Word Association Lexicons: Capturing word–emotion, word–sentiment, and word–colour associations

Contact: Saif M. Mohammad (saif.mohammad@nrc-cnrc.gc.ca)

Terms of use:

- Indicate that you agree with the terms of use here.
- The lexicons mentioned in this page are available for direct download and can be used freely for research purposes.
- The papers listed next to the lexicons provide details of the creation and use. If you use a lexicon, then please cite the associated papers.
- If interested in commercial use of any of these lexicons, send email to the contact. A one-time licensing fee may apply.
- If you use a lexicon in a product or application, then acknowledge this in the 'About' page and other relevant documentation of the application by stating the name of the resource, the authors, and NRC. For example, "This application/product/tool makes use of the <resource name>, created by <author(s)> at the National Research Council Canada." (Also, if you send us an email, we will be thrilled to know about how you have used the lexicon.)
- Rather than redistributing the data, please direct interested parties to this page.
- National Research Council Canada (NRC) disclaims any responsibility for the use of the lexicons listed here and does not provide technical support. However, the contact listed above will be happy to respond to queries and clarifications.

We will be happy to hear from you, especially if:

- you give us feedback regarding these lexicons.
- you tell us how you have (or plan to) use the lexicons.
- you are interested in having us analyze your data for sentiment, emotion, and other affectual information.
- you are interested in a collaborative research project. We also regularly hire graduate students for research internships.

Access the following resources in separate webpages dedicated to them:

- *Homepage of the NRC Word-Emotion Association Lexicon, also called EmoLex.*
- *Homepage for Sentiment Composition Lexicons.*
- *Homepage for Arabic sentiment lexicons and corpora.*
- *Homepage for Best-Worst Scaling (aka MaxDiff) software and annotations.*

Table of Word-Association Lexicons

Lexicon	Version	# of Terms	Categories	Association Scores	Method of Creation
Both Word-Emotion and Word-Sentiment Association Lexicon					
1. NRC Word-Emotion Association Lexicon (also called EmoLex). README. Explore the interactive visualization. Homepage of the Lexicon. Also available in over 40 other languages here. The sense-level annotations provided by individual annotators for the eight emotions can be downloaded by clicking here.					
	0.92 (2010)	14,182 unigrams (words)	**sentiments:** negative, positive **emotions:** anger, anticipation, disgust, fear, joy, sadness, surprise, trust	0 (not associated) or 1 (associated)	**Manual:** By crowdsourcing on Mechanical Turk. **Domain:** General
		~25,000 senses		not associated, weakly, moderately, or strongly associated	
Papers: Crowdsourcing a Word-Emotion Association Lexicon, Saif Mohammad and Peter Turney, *Computational Intelligence,* 29 (3), 436-465, 2013. Paper (pdf) BibTeX Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon, Saif Mohammad and Peter Turney, In *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text,* June 2010, LA, California. Paper (pdf) BibTeX Presentation					
Word-Emotion Association Lexicon					
1. NRC Hashtag Emotion Lexicon. The Hashtag Emotion Corpus (aka Twitter Emotion Corpus, or TEC) used to create the lexicon.					
	0.2 (2013)	16,862 unigrams (words)	**emotions:** anger, anticipation, disgust, fear, joy, sadness, surprise, trust	Real-valued score between 0 (not associated) to ∞ (maximally associated)	**Automatic:** From tweets with emotion word hashtags. **Domain:** Twitter
Papers: Using Hashtags to Capture Fine Emotion Categories from Tweets. Saif M. Mohammad, Svetlana Kiritchenko, Computational Intelligence, in press. Paper (pdf) BibTeX					

#Emotional Tweets, Saif Mohammad, In Proceedings of the First Joint Conference on Lexical and Computational Semantics (*Sem), June 2012, Montreal, Canada. Paper (pdf) BibTeX

Word-Sentiment Association Lexicons
(All lexicons below are for English terms. Arabic sentiment lexicons and corpora are available here.)

1. Sentiment Composition Lexicon of Negators, Modals, and Adverbs (SCL-NMA), aka SemEval-2016 General English Sentiment Modifiers Lexicon, created using Best-Worst Scaling (aka MaxDiff)

	1.0 (Feb. 2016)	~3200 terms	**sentiments:** negative, positive	Real-valued score between -1 (most negative) to 1 (most positive)	**Manual.** By crowdsourcing and using Best-Worst Scaling. **Domain:** General

Papers:

- The Effect of Negators, Modals, and Degree Adverbs on Sentiment Composition. Svetlana Kiritchenko and Saif M. Mohammad, In *Proceedings of the NAACL 2016 Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media (WASSA),* June 2014, San Diego, California.
  Paper (pdf) BibTeX Presentation

- Capturing Reliable Fine-Grained Sentiment Associations by Crowdsourcing and Best-Worst Scaling. Svetlana Kiritchenko and Saif M. Mohammad. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* June 2016. San Diego, CA.
  Paper (pdf) BibTeX Presentation

- Semeval-2016 Task 7: Determining Sentiment Intensity of English and Arabic Phrases. Svetlana Kiritchenko, Saif M. Mohammad, and Mohammad Salameh. In *Proceedings of the International Workshop on Semantic Evaluation (SemEval '16)*. June 2016. San Diego, California.
  Paper (pdf) BibTeX Presentation Task Website

2. SemEval-2015 English Twitter Sentiment Lexicon, created using Best-Worst Scaling (aka MaxDiff)

	1.0 (Feb. 2015)	~1500 terms	**sentiments:** negative, positive	Real-valued score between -1 (most negative) to 1 (most positive)	**Manual.** By crowdsourcing and using Best-Worst Scaling.

					Domain: Twitter

Paper:

- SemEval-2015 Task 10: Sentiment Analysis in Twitter. Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif M Mohammad, Alan Ritter, and Veselin Stoyanov. In Proceedings of the ninth international workshop on Semantic Evaluation Exercises (SemEval-2015), June 2015, Denver, Colorado.
  Paper (pdf) BibTeX

- Sentiment Analysis of Short Informal Texts. Svetlana Kiritchenko, Xiaodan Zhu and Saif Mohammad. *Journal of Artificial Intelligence Research,* volume 50, pages 723-762, August 2014.
  Paper (pdf) BibTeX

This data was used in SemEval-2015 Task 10 (Sentiment Analysis in Twitter), subtask E - Determining strength of association of Twitter terms with positive sentiment (or, degree of prior polarity). Task description, trial data, test data, and other details available here.

3. Sentiment Composition Lexicon of Opposing Polarity Phrases (SCL-OPP) aka SemEval-2016 English Twitter Mixed Polarity Lexicon, created using Best-Worst Scaling (aka MaxDiff)

	1.0 (Feb. 2016)	~1200 terms	**sentiments:** negative, positive	Real-valued score between -1 (most negative) to 1 (most positive)	**Manual.** By crowdsourcing and using Best-Worst Scaling. **Domain:** Twitter

Paper:

- Sentiment Composition of Words with Opposing Polarities. Svetlana Kiritchenko and Saif M. Mohammad. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. June 2016. San Diego, CA.
  Paper (pdf) BibTeX Poster

- *Happy Accident:* A Sentiment Composition Lexicon for Opposing Polarities Phrases. Svetlana Kiritchenko and Saif M. Mohammad. In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference*, May 2016, Portorož (Slovenia).
  Paper (pdf) BibTeX Poster

- Semeval-2016 Task 7: Determining Sentiment Intensity of English and Arabic Phrases. Svetlana Kiritchenko, Saif M. Mohammad, and Mohammad Salameh. In *Proceedings of the International Workshop on Semantic Evaluation (SemEval '16)*. June 2016. San Diego, California.
  Paper (pdf) BibTeX Presentation Task Website

4. NRC Twitter Sentiment Lexicons (NRC Hashtag Sentiment Lexicons and Sentiment140 Lexicons)					
a. NRC Hashtag Sentiment Lexicon					
	1.0 (2013)	54,129 unigrams 316,531 bigrams 308,808 pairs	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic**: From tweets with sentiment word hashtags. **Domain:** Twitter
b. NRC Hashtag Affirmative Context Sentiment Lexicon and NRC Hashtag Negated Context Sentiment Lexicon					
	1.0 (2014)	Affirmative contexts: 36,357 unigrams Negated contexts: 7,592 unigrams Affirmative contexts: 159,479 bigrams Negated contexts: 23,875 bigrams	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic:** From tweets with sentiment word hashtags. Separate entries for affirmative and negated contexts. **Domain:** Twitter
c. Emoticon Lexicon aka Sentiment140 Lexicon (note that this is sentiment lexicon drawn from emoticons, and is not an emotion lexicon)					
	1.0 (2014)	62,468 unigrams 677,698 bigrams 480,010 pairs	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic**: From tweets with emoticons. **Domain:** Twitter
d. Sentiment140 Affirmative Context Lexicon and Sentiment140 Negated Context Lexicon					
		Affirmative contexts:			

	1.0 (2014)	45,255 unigrams Negated contexts: 9,891 unigrams Affirmative contexts: 240,076 bigrams Negated contexts: 34,093 bigrams	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic:** From tweets with sentiment word hashtags. Separate entries for affirmative and negated contexts. **Domain:** Twitter

Papers (describing the four NRC Twitter Lexicons listed above):

Sentiment Analysis of Short Informal Texts. Svetlana Kiritchenko, Xiaodan Zhu and Saif Mohammad. *Journal of Artificial Intelligence Research,* volume 50, pages 723-762, August 2014.
Paper (pdf) BibTeX

NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets, Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu, In Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013), June 2013, Atlanta, USA.
Paper (pdf) BibTeX System Description and Downloads Poster Slides

NRC-Canada-2014: Recent Improvements in Sentiment Analysis of Tweets, Xiaodan Zhu, Svetlana Kiritchenko, and Saif M. Mohammad. In Proceedings of the eigth international workshop on Semantic Evaluation Exercises (SemEval-2014), August 2014, Dublin, Ireland.
Paper (pdf) BibTeX

These lexicons were used to generate winning submissions for the sentiment analysis shared tasks of SemEval-2013 Task 2 and SemEval-2014 Task 9.

5. Yelp and Amazon Sentiment Lexicons

a. Yelp Restaurant Sentiment Lexicon
(created from the Yelp Dataset -- from the subset of entries pertaining to these restaurant-related businesses)

	1.0 (2014)	39,274 entries for unigrams (includes affirmative and negated context entries) 276,651	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic:** From customer reviews on Yelp.com. **Domain:** Restaurant

		entries for bigrams			

b. Amazon Laptop Sentiment Lexicon

	1.0 (2014)	26,577 entries for unigrams (includes affirmative and negated context entries) 155,167 entries for bigrams	**sentiments:** negative, positive	Real-valued score between -∞ (most negative) to ∞ (most positive)	**Automatic:** From customer reviews on Amazon.com. **Domain:** Laptop

Paper (describing the Yelp and Amazon Lexicons):

NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews, Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif M. Mohammad. In Proceedings of the eigth international workshop on Semantic Evaluation Exercises (SemEval-2014), August 2014, Dublin, Ireland. Paper (pdf) BibTeX

These lexicons were used to generate winning submissions for the sentiment analysis shared task of SemEval-2014 Task 4.

6. Macquarie Semantic Orientation Lexicon

	0.1 (2009)	76,400 terms	**sentiments:** negative, positive	binary distinction: negative or positive	**Automatic:** Using the structure of a thesaurus and affixes. **Domain:** General

Paper:

Generating High-Coverage Semantic Orientation Lexicons From Overtly Marked Words and a Thesaurus, Saif Mohammad, Bonnie Dorr, and Cody Dunne, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009),* August 2009, Singapore. Paper (pdf) BibTeX Presentation

Word-Colour Association Lexicon

1. NRC Word-Colour Association Lexicon

	0.92 (2011)	~14,000 words	**colours:** black, blue, brown, green, grey, orange purple, pink, red, white, yellow	0 (not associated) or 1 (associated)	**Manual:** Crowdsourcing on Mechanical Turk. **Domain:** General
		~25,000 senses		not, weakly, moderately, or strongly associated	

Papers:

Colourful Language: Measuring Word-Colour Associations, Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Cognitive Modeling and Computational Linguistics (CMCL), June 2011, Portland, OR. Paper (pdf) BibTeX Presentation

Even the Abstract have Colour: Consensus in Word-Colour Associations, Saif Mohammad, In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, June 2011, Portland, OR. Paper (pdf) BibTeX Poster