SLIDE – a Sentiment Lexicon of Common Idioms

Charles Jochim, Francesca Bonin, Roy Bar-Haim, Noam Slonim
IBM Research
{charlesj,fbonin}@ie.ibm.com, {roybar,noams}@il.ibm.com

Abstract

Idiomatic expressions are problematic for most sentiment analysis approaches, which rely on words as the basic linguistic unit. Compositional solutions for phrase sentiment are not able to handle idioms correctly because their sentiment is not derived from the sentiment of the individual words. Previous work has explored the importance of idioms for sentiment analysis, but has not addressed the breadth of idiomatic expressions in English. In this paper we present an approach for collecting sentiment annotation of idiomatic multiword expressions using crowdsourcing. We collect 10 annotations for each idiom and the aggregated label is shown to have good agreement with expert annotations. We describe the resulting publicly available lexicon and how it captures sentiment strength and ambiguity. The *Sentiment Lexicon of IDiomatic Expressions (SLIDE)* is much larger than previous idiom lexicons. The lexicon includes 5,000 frequently occurring idioms, as estimated from a large English corpus. The idioms were selected from Wiktionary, and over 40% of them were labeled as sentiment-bearing.

Keywords: Idiom, Lexicon, Sentiment Analysis

1. Introduction

Multiword expressions (MWE) are a key challenge in Natural Language Processing (Sag et al., 2002). Among MWEs, *idioms* are often defined as non-compositional multiword expressions, the meaning of which cannot be deduced from the literal meaning of constituent words (Nunberg et al., 1994).

Sentiment analysis systems typically consider words as the basic sentiment units. Word sentiments are either learned from the training data or looked up in a sentiment lexicon. Text sentiment is then derived by means of aggregation over word sentiments, often with some treatment of compositional phenomena such as valence shifters (Polanyi and Zaenen, 2004) and mixed sentiment (Kiritchenko and Mohammad, 2016). Other approaches are based on bottom-up sentiment composition, starting at the word level and computing the sentiment of each phrase based on the semantics and sentiment of its daughter phrases, according to the syntactic structure of the sentence (Moilanen and Pulman, 2007; Nakagawa et al., 2010; Socher et al., 2013).

Due to their non-compositionality, idioms are often not handled correctly by current sentiment analysis systems (Balahur et al., 2010). Word-level sentiment analysis would miss the positive sentiment in *two thumbs up*, and on the other hand, we might incorrectly assign positive sentiment to *as <u>well</u> as*, because of the positive sentiment of *well*. Similarly, we would like to know that it is not good if something *bites the dust*, while we would be happy to hear that our handing of idioms was *dead on*. Ignoring idioms overlooks an important signal of the sentiment of the text, as figurative and idiomatic language often directs sentence polarity (Rentoumi et al., 2012). For the above reasons, idioms have begun to receive some attention in recent sentiment analysis literature (Williams et al., 2015; Liu et al., 2017). Yet, robust treatment of idioms is hindered by their limited coverage in current sentiment datasets and lexicons.

In this work we introduce *SLIDE (Sentiment Lexicon of IDiomatic Expressions)*, a new resource for sentiment analysis, created via crowdsourcing. Our lexicon is an order of magnitude larger than previous idiom sentiment lexicons and focuses specifically on the most frequently used idioms.[1] In creating this resource, we are somewhat agnostic to the exact definition of *idiom*. We are more generally interested in sentiment analysis that can handle MWEs. In this paper, we have initially focused on idioms because they are the most problematic for sentiment analysis, being strictly non-compositional.

In the rest of the paper, we first describe the crowdsourcing-based idiom annotation process and its quality assessment (Section 2). We then provide a description of the resulting lexicon (Section 3). Section 4 describes an auxiliary annotation step aimed at identifying frequent idioms that have non-idiomatic meaning and neutral sentiment in most contexts. Section 5 covers the previous work related to idioms, sentiment lexicons, and sentiment analysis. Finally, we conclude and discuss future work in Section 6.

2. Lexicon Creation

2.1. Idiom Selection

We start with the list of all idioms available in Wiktionary,[2] which resulted in 8,772 idioms. To narrow down the list of idioms we send for annotation, we take the following steps: 1) We remove some special Wiktionary links that start with `Appendix:` or `Citation:` (only 22 in total); 2) We remove all unigrams, which we do not consider to be idioms here; however, hyphenated words are not removed; 3) Finally, we remove all idioms comprised only of stopwords (e.g., *about to, and so on*). This results in a list of 8,637 idiomatic expressions.

The resulting list contains many idioms that are infrequently used or are less popular in current usage. To ease the annotation effort, we have chosen to select frequently used idioms, which would be the most useful for sentiment analysis tools. To determine the most frequent idioms we

[1] SLIDE is available at `http://www.research.ibm.com/haifa/dept/vst/debating_data.shtml`

[2] `https://en.wiktionary.org/w/index.php?title=Category:English_idioms`

use a proprietary corpus of news articles and other publications coming from thousands of sources. The corpus includes about 200 million documents. We count the idiom's frequency in this corpus and then select the most frequent 5,000 for annotation.

Many idioms have personal pronouns or possessive adjectives and they are listed as idioms using *one*, *anyone*, or *someone*, and *one's*, *anyone's*, or *someone's*. For example we have the idioms *keep someone posted* and *on one's feet* in our lexicon. To count the frequency of these idioms we expand the generic pronoun with the various pronouns for person, number, and gender that we may encounter in the corpus. Specifically, in the case of possessive adjectives, we replace occurrences of *one's*, *anyone's*, or *someone's* with *my, your, his, her, our*, and *their*. For the remaining pronouns (*one*, *anyone*, and *someone*), if they follow a verb they are treated as object pronouns and replaced with *me, you, him, her, us*, and *them*, otherwise they are treated as subject pronouns and replaced with *I, you, he, she, we*, and *they*. The frequencies for each of the pronoun-substituted idioms are summed for the frequency of the original idiom. After the frequency is collected for all idioms, we select the top 5,000 most frequent idioms to be manually annotated for sentiment. This threshold covers the idioms that occur in at least 1,266 documents in our corpus.

2.2. Idiom Annotation

We use the CrowdFlower platform for crowdsourcing annotation.[3] CrowdFlower is an online labor market where workers are paid to complete tasks uploaded to the platform. The system requires the submission of a task via a web interface.

We first ran several exploratory pilots for the annotation task to refine the guidelines and settings for the annotation. Following CrowdFlower conventions, the annotation guidelines are succinct. Annotators are provided with the following overview of the task:

> *In this job, you will be shown expressions and their Wiktionary definitions and you should indicate if the expression has a positive, negative, or neutral sentiment in most contexts in which it is found. You should consider the most common meaning of the expression, which may or may not be idiomatic. Please look at the definitions provided to help guide your decision.*

And we suggest the following steps for performing the annotation:

1. Read the expression and definitions and imagine different contexts in which it can be found.

2. Decide if the expression tends to contribute positively, negatively, or neutrally to the sentiment of these contexts.

3. Select positive, negative, or neutral accordingly (if the content is inappropriate or vulgar, select only the Inappropriate/Vulgar box).

Some expressions can have several meanings, which may be more or less idiomatic. For example *rip up* is defined as (i) "to destroy by ripping" (less idiomatic), and also (ii) "to move quickly or violently" (more idiomatic). Our instructions ask the annotator to label the most commonly found sentiment of the expression, whether it be idiomatic or not. Annotating the most frequent sentiment of the expression is due to practical considerations, as we assume that the users of our lexicon will not perform sense disambiguation, which is a very challenging task for idioms. A number of idioms still have similar sentiment across their different meanings; so even in cases where the expression's meaning may be ambiguous, the sentiment is not. For example, the more literal and more idiomatic senses of *fall apart* are used to express negative sentiment.

Crowdsourced annotation has been shown to be an effective and reliable source of labeled data when a sufficiently large number of non-expert annotators are employed (Snow et al., 2008; Nowak and Rüger, 2010).[4] We collected 10 annotations for each idiom for greater confidence in our annotation. The CrowdFlower platform includes mechanisms to ensure consistent annotation by including random test questions with known labels to be sure the annotator understands the task and is annotating in good faith. If an annotator falls below a predefined threshold of accuracy on test questions, then the annotator is removed from the task and their annotations are discarded. This further preserves the quality of the annotation. CrowdFlower has three levels of annotation expertise and our annotation task used Level 2. We selected annotators from the United States, Canada, United Kingdom, Ireland, and Australia in an effort to have as native fluency as possible.

We found that even native speakers may be less familiar with some idioms. In pilot annotation we included links to Wiktionary definitions for when annotators might be in doubt, but we did not present the definition to the annotators. CrowdFlower annotators were likely reticent to look up the definition and even as native speakers could easily have overlooked some expression's meaning when many idioms have multiple meanings or senses.[5] We resolved this by embedding the Wiktionary page directly in the CrowdFlower annotation (see Figure 1). This encourages the annotators to check the definition and makes the resulting annotation more consistent.

To confirm the quality of our annotation we measure Fleiss' kappa (Fleiss, 1971) between an expert annotator and the most voted label from the crowdsourced annotation (Table 1). We took a random sample of 400 idioms from the 5,000 annotated and had them annotated by an in-house expert annotator. The kappa score for this sample is 0.55. We expect some disagreement in the annotation because there is inherent subjectivity in sentiment assessment, in addition to the ambiguity of some of the idioms. When presented out of context, the assessment of an idiom's sentiment sometimes depends on the annotator's own biases and beliefs.

[3] https://www.crowdflower.com/

[4] Snow et al. (2008) suggest at least four annotators to reach expert level annotation; Nowak and Rüger (2010) found a majority vote of nine non-expert annotations matched expert annotation.

[5] 36% of the idioms we collect from Wiktionary have more than one meaning.

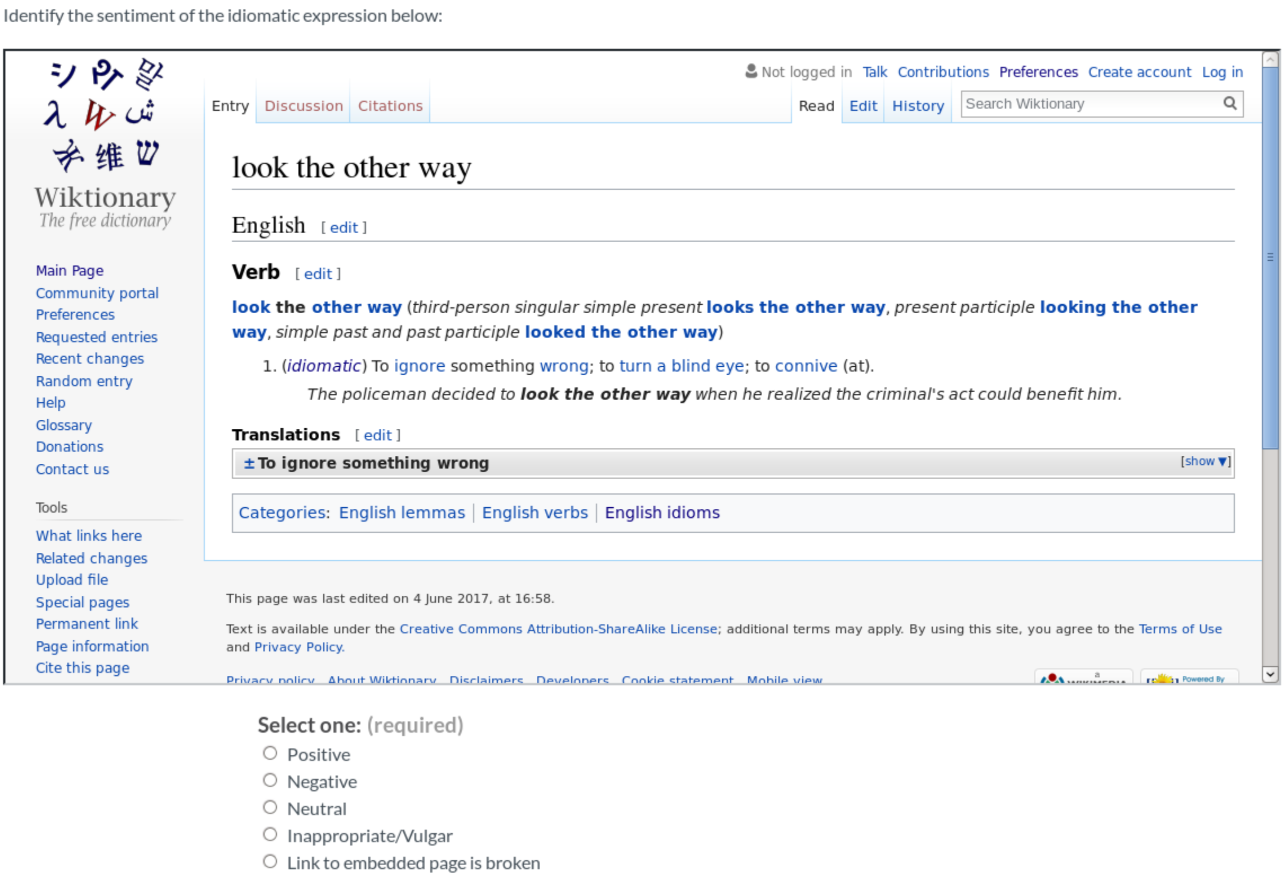

Figure 1: Annotation example of Wiktionary idioms in CrowdFlower.

	All	≥ 60%	≥ 80%
Kappa	0.55	0.61	0.74
Sample size	400	356	244

Table 1: Fleiss' kappa agreement between expert and crowd. Latter columns refer to percent agreement in crowd.

	Label	≥ 20%	≥ 60%	≥ 80%
Positive	946	1717	745	426
Negative	1108	1819	917	517
Neutral	2945	4252	2842	2021
Inapprop.	1	29	1	0

Table 2: Lexicon statistics. Percentage columns refer to distribution of total annotation, e.g., 1717 idioms are *at least* 20% positive.

For example, the idiom *live on the edge* was assessed as positive by two annotators, as negative by three annotators, and as neutral by five annotators.

Users of our lexicon may also choose to filter out lower-argreement idioms. As expected, higher agreement among crowd annotators leads to higher agreement with the expert annotator. If we only consider idioms where at least 60% of the crowd agreed (90% of the lexicon), then on this sample of 356 we have Fleiss' kappa of 0.61. Likewise with 80% crowd agreement (60% of the lexicon) we have kappa of 0.74 over a sample of 244 idioms.

3. Resource Description

The result of our annotation task is a lexicon of 5,000 idioms with at least 10 annotations. The lexicon we release includes the percentage of *positive, negative, neutral*, and *inappropriate* annotations so that future users can decide the degree of polarity that they would like to include, e.g., only the most positive and negative idioms or also idioms with weaker sentiment.

The lexicon includes a sentiment label along with the distribution of sentiment annotations. Our labels are assigned by taking the label with the greatest number of votes from the crowdsourced annotation. In the case of ties between *positive* (or *negative*) and *neutral*, the label is *positive* (resp. *negative*). In the rare cases of ties between *positive* and *negative*, we use the *neutral* label. The resulting lexicon has 946 *positive* idioms, 1,108 *negative*, 2,945 *neutral*, and 1 *inappropriate* (see Table 2). Table 2 includes additional columns that capture the lexicon's makeup. By looking at all idioms with more than 20% *positive* or *negative* annotation, we get a sense of the number of idioms with weak sentiment, i.e., over 2/3 have at least weak polarity. One may also consider using a smaller lexicon with stronger sentiment. Using 60% or 80% agreement from CrowdFlower we reduce the number of idioms with weak or ambiguous sentiment (see the kappa values in Table 1), which results in smaller *positive* and *negative* lexicons.

As stated in the guidelines, we want to capture the sentiment of the expression's common usage. However, we are also interested in the ambiguity that is reflected in the answer distribution. Williams et al. (2015) include an option for directly labeling an idiom as ambiguous (which is ultimately combined with *neutral*). Our approach allows us to handle ambiguous idioms by considering their most frequent sentiment, while the approach of Williams et al. only assigns sentiment to unambiguous idioms. We opted for only options of *positive, negative*, and *neutral*.[6] Our hope was that the distribution of annotator answers would capture the ambiguity of the idioms and the resulting label distribution could represent that ambiguity. For example, *make the cut* is labeled 80% positive and 20% neutral. It is defined in Wiktionary as "to succeed at something or meet a requirement; to be chosen out of a field of candidates or possibilities," which is largely positive, however the ambiguity might come from usage where *making the cut* is not excelling but sufficient. This contrasts *at the ready*[7] which is largely neutral with very weak positive sentiment (80% neutral vs 20% positive). These are only two examples and there are others with ambiguity that is more debatable. However, we think this highlights the advantage of having *positive, negative*, and *neutral* annotation so that we may have more insights into the ambiguity, e.g., mixed sentiment with equal percentages for *positive* and *negative*, or weak sentiment with a low percentage of *positive* (resp. *negative*) but for the most part *neutral*. In this respect our lexicon is similar to SentiWordNet (Esuli and Sebastiani, 2006; Baccianella et al., 2010), which has scores for positivity, negativity, and objectivity.

In Table 3 we have listed some examples from our lexicon that are potentially problematic when handling words alone. The list is broken down into frequency bins, and for each bin, the idioms' frequency is above the bin's threshold, but lower than the previous bin's threshold. We can see that, while the sentiment of the 500 most frequent idioms is obviously important for sentiment analysis, there is also important idiom sentiment in the long tail of less frequent idioms. This motivated our choice in drawing idioms from a much larger pool of idioms than previous work.

Positive	Negative	Neutral
document frequency > 160,900 (10% of lexicon)		
make a difference	behind bars	as well as
bounce back	under fire	on the table
on one's feet	in the red	keep an eye on
document frequency > 11,430 (50% of lexicon)		
in shape	red flag	on the clock
over the moon	in the hole	outside the box
breath of fresh air	wide of the mark	change of heart
document frequency > 1,266 (entire lexicon)		
on cloud nine	up a tree	dead ringer
have a ball	on thin ice	scratch the surface
bury the hatchet	booby prize	birthday suit

Table 3: Frequent idioms by label. For each bin, the idioms' frequency is above the bin's threshold, but lower than the previous bin's threshold.

4. Context-Based Post-Filtering

As previously mentioned, some phrases in our lexicon have both idiomatic and non-idiomatic interpretations, which may differ in polarity. The guidelines ask the annotators to choose the sentiment most commonly associated with the phrase, considering both idiomatic and non-idiomatic meanings. However, showing the annotators the idiomatic definitions from Wiktionary created bias towards these idiomatic senses. For example, the phrase *make it*, which has the idiomatic meaning of achieving one's goals, was given positive sentiment in our lexicon, although in most contexts it has neutral polarity. This may introduce many errors when applying the lexicon to sentiment analysis tasks.

We observed that such situations are much more common for highly frequent phrases, for which polarity errors also have the greatest impact. We therefore apply an additional filtering step to the 300 most frequent idioms in our lexicon with *positive* or *negative* labels. For each of these expressions, we query the corpus mentioned in Section 2.1 and retrieve ten different contexts containing the expression. We give the expressions and contexts to a group of in-house annotators to determine if the expression is *positive*, *negative*, or *neutral* in a sample of actual contexts. For each of the 300 expressions there are ten contexts and we require five annotations for each context. To determine if the idiom should be filtered, we first aggregate the annotations by taking the label with the majority vote per context (if there is no majority, the label is *neutral*). Then we check the number of context labels per expression. If the majority of contexts have been labeled *positive* or *negative*, we mark the expression to keep, otherwise we mark it to filter out. This annotation marks 103 n-grams that can be filtered out, leaving 197 as idioms with sentiment in the majority of contexts in which they were found. Some examples of filtered phrases are *do in*, *play games* and *make it*. We provide this filtering as an additional layer of annotation, rather than discarding filtered phrases from the lexicon.

5. Related Work

There is a wealth of literature on sentiment analysis ((Liu, 2012)) and idioms ((Nunberg et al., 1994)). Here we focus on work related to building sentiment lexicons and the importance of handling idioms in sentiment classification.

Available sentiment lexicons do not handle idiomatic expressions and focus almost entirely on unigrams. Manually curated lexicons such as the Harvard General Inquirer (Stone et al., 1966) or MPQA (Wilson et al., 2005) have hyphenated words but no idioms or MWEs. The lexicons created by early automatic approaches (Turney and Littman, 2003; Hu and Liu, 2004) deal with words but not longer n-grams. Approaches using WordNet (Miller, 1995), like those of Esuli and Sebastiani (2006) or Blair-Goldensohn et al. (2008), will include MWEs but WordNet has low coverage of idioms in our lexicon. Other graph-based approaches using distributionally similar n-grams (Velikovich et al., 2010) can return sentiment for MWEs, but the approach is sensitive to parameter tuning and there has been

[6]The *inappropriate* option was for filtering out idioms with swear words and was rarely used or needed.

[7]Wiktionary: "ready; in a state of preparation or waiting; in position or anticipation."

no evaluation of the quality of the MWE sentiment. Recently, Williams et al. (2015) released a sentiment lexicon with 580 idioms, but the selection of idioms focused on emotional idioms, some of which are not very frequent (e.g., they showed that more than a quarter were not found in the British National Corpus). To address the lack of large-scale idiom sentiment lexicon, we manually annotated 5,000 of the most frequently used idioms, which is still feasible using crowdsourcing and avoids potential pitfalls of automatic lexicon creation.

After analysis from Balahur et al. (2010) showing the prevalence of idiom errors in sentiment classification, and some success by Xie and Wang (2014) using idioms for sentiment classification in Chinese, Williams et al. (2015) further investigated the role of idiomatic expressions in sentiment analysis. They showed that the use of sentiment annotated idiomatic expressions as features can improve the results of sentiment analysis of sentences.

Liu et al. (2017) has also recently shown that consideration of idioms can improve sentiment classification. They propose two models to address the more and less compositional idioms discussed in (Nunberg et al., 1994): one treating idioms as a fixed phrase and a second that considers morphology to account for the possible syntactic variation in the idioms. Incorporating external knowledge on idiom sentiment is likely to further improve the performance of such approaches, in particular on sentences containing idioms that did not appear in the training data.

6. Conclusion

In this paper we have motivated the need for better handling of idioms in sentiment analysis and presented a large sentiment lexicon of idiomatic expressions for this purpose. We make the final lexicon available and hope that it will be useful in improving sentiment classification and opinion mining. In future work, we plan on expanding the lexicon by utilizing other sources of idioms, as well as covering additional types of sentiment-bearing MWEs other than idioms. For example, a negative health condition such as *high blood pressure* should get negative sentiment. We then plan on using this lexicon for experiments not only in sentiment classification but other related text classification tasks that would benefit from distinguishing idioms and their sentiment, e.g., stance classification.

7. Bibliographical References

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May. European Language Resources Association (ELRA).

Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Goot, E. v. d., Halkia, M., Pouliquen, B., and Belyaeva, J. (2010). Sentiment analysis in the news. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. European Languages Resources Association (ELRA).

Blair-Goldensohn, S., Neylon, T., Hannan, K., Reis, G. A., Mcdonald, R., and Reynar, J. (2008). Building a sentiment summarizer for local service reviews. In *NLP in the Information Explosion Era*.

Esuli, A. and Sebastiani, F. (2006). SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*, pages 417–422.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.

Kiritchenko, S. and Mohammad, S. M. (2016). Sentiment composition of words with opposing polarities. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1102–1108, San Diego, California, June. Association for Computational Linguistics.

Liu, P., Qian, K., Qiu, X., and Huang, X. (2017). Idiom-aware compositional distributed semantics. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1215–1224, Copenhagen, Denmark, September. Association for Computational Linguistics.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Miller, G. A. (1995). WordNet: A lexical database for english. *Commun. ACM*, 38(11):39–41, November.

Moilanen, K. and Pulman, S. (2007). Sentiment composition. In *Proceedings of RANLP 2007*, Borovets, Bulgaria.

Nakagawa, T., Inui, K., and Kurohashi, S. (2010). Dependency tree-based sentiment classification using crfs with hidden variables. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 786–794, Los Angeles, California, June. Association for Computational Linguistics.

Nowak, S. and Rüger, S. (2010). How reliable are annotations via crowdsourcing? a study about inter-annotator agreement for multi-label image annotation. In *Proceedings of the international conference on Multimedia information retrieval - MIR '10*, page 557.

Nunberg, G., Sag, I. A., and Wasow, T. (1994). Idioms. In Stephen Everson, editor, *Language*, pages 491–538. Cambridge University Press.

Polanyi, L. and Zaenen, A. (2004). Contextual valence shifters. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*.

Rentoumi, V., Vouros, G. A., Karkaletsis, V., and Moser, A. (2012). Investigating metaphorical language in sentiment analysis: A sense-to-sentiment perspective. *ACM Trans. Speech Lang. Process.*, 9(3):6:1–6:31, November.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002). Multiword expressions: A pain in the neck for nlp. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing: Third International Conference, CICLing 2002 Mexico City, Mexico, February 17–23, 2002 Proceedings*, pages 1–15. Springer Berlin Heidelberg, Berlin, Heidelberg.

Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. Y. (2008). Cheap and fast—but is it good?: Evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 254–263, Stroudsburg, PA, USA. Association for Computational Linguistics.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A., and 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*, pages 1631–1642, Seattle, Washington, USA, October. Association for Computational Linguistics.

Stone, P. J., Dunphry, D. C., Smith, M. S., and Ogilvie, D. M. (1966). *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

Turney, P. D. and Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Trans. Inf. Syst.*, 21(4):315–346, October.

Velikovich, L., Blair-Goldensohn, S., Hannan, K., and McDonald, R. (2010). The viability of web-derived polarity lexicons. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 777–785, Los Angeles, California, June. Association for Computational Linguistics.

Williams, L., Bannister, C., Arribas-Ayllon, M., Preece, A., and Spasić, I. (2015). The role of idioms in sentiment analysis. *Expert Systems with Applications*, 42(21):7375 – 7385.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 347–354, Vancouver, British Columbia, Canada, October. Association for Computational Linguistics.

Xie, S.-x. and Wang, T. (2014). Construction of unsupervised sentiment classifier on idioms resources. *Journal of Central South University*, 21(4):1376–1384, Apr.