Concreteness and Subjectivity as Dimensions of Lexical Meaning

Felix Hill
Computer Laboratory
Cambridge University
felix.hill@cl.cam.ac.uk

Anna Korhonen
Computer Laboratory
Cambridge University
anna.korhonen@cl.cam.ac.uk

Abstract

We quantify the lexical subjectivity of adjectives using a corpus-based method, and show for the first time that it correlates with noun concreteness in large corpora. These cognitive dimensions together influence how word meanings combine, and we exploit this fact to achieve performance improvements on the semantic classification of adjective-noun pairs.

1 Introduction

Concreteness, the degree to which language has a perceptible physical referent, and *subjectivity*, the extent to which linguistic meaning depends on the perspective of the speaker, are well established cognitive and linguistic notions. Recent results suggest that they could also be useful knowledge for natural language processing systems that aim to extract and represent the meaning of language.

Insight into concreteness can help systems to classify adjective-noun pairs according to their semantics. In the non-literal expressions *kill the process* or *black comedy*, a verb or adjective that occurs with a concrete argument in literal phrases takes an abstract argument. Turney et al. (2011) present a supervised model that exploits this effect to correctly classify 79% of adjective-noun pairs as having literal or non-literal meaning.

Subjectivity analysis has already proved highly applicable to a range of NLP applications, including sentiment analysis, information extraction and text categorization (Pang and Lee, 2004; Riloff and Wiebe, 2003). For such applications, subjectivity is analyzed at the phrasal or document level. However, recent work has highlighted the application of subjectivity analysis to lexical semantics, for instance to the tasks of disambiguating words according to their usage or sense (Wiebe and Mihalcea, 2006; Banea et al., 2014).

The importance of concreteness to NLP systems is likely to grow with the emergence of multi-modal semantic models (Bruni et al., 2012; Roller and Schulte im Walde, 2013). Such models, which learn representations from both linguistic and perceptual input, outperform text-only models on a range of evaluations. However, while multi-modal models acquire richer representations of concrete concepts, their ability to represent abstract concepts can be weaker than text-only models (Hill et al., 2013). A principled treatment of concreteness is thus likely to be important if the multi-modal approach is to prove effective on a wider range of concepts. In a similar vein, interest in subjectivity analysis is set to grow with interest in extracting sentiment and opinion from the web and social media (Benson et al., 2011). Moreover, given that humans seem to exploit both concreteness (Paivio, 1990) and subjectivity (Canestrelli et al., 2013) clues when processing language, it is likely that the same clues should benefit computational models aiming to replicate human-level performance in this area.

In this paper, we show how concreteness and subjectivity can be applied together to produce performance improvements on two classification problems: distinguishing literal and non-literal adjective-noun pairs (Turney et al., 2011), and classifying the modification type exhibited by such pairs (Boleda et al., 2012). We describe an unsupervised corpus-based method to quantify adjective subjectivity, and show that it effectively predicts the labels of a hand-coded subjectivity lexicon. Further, we show for the first time that adjective subjectivity correlates with noun concreteness in large corpora. In addition, we analyse the effect of noun concreteness and adjective subjectivity on meaning combination, illustrating how the interaction of these dimensions enables the accurate classification of adjective-noun pairs according to their semantics. We conclude by dis-

Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 725–731,
Baltimore, Maryland, USA, June 23-25 2014.

cussing other potential applications of concreteness and subjectivity to NLP.

2 Dimensions of meaning

Concreteness A large and growing body of empirical evidence indicates clear differences between concrete concepts, such as *donut* or *hotdog* and abstract concepts, such as *guilt* or *obesity*. Concrete words are more easily learned, remembered and processed than abstract words (Paivio, 1991), while differences in brain activity (Binder et al., 2005) and cognitive representation (Hill et al., 2013) have also been observed. In linguistic conctructions, concreteness appears to influence compound and phrasal semantics (Traugott, 1985; Bowdle and Gentner, 2005; Turney et al., 2011). Together with the practical applications outlined in Section 1, these facts indicate the potential value of concreteness for models aiming to replicate human performance in language processing tasks.

While automatic methods have been proposed for the quantification of lexical concreteness, they each rely on dictionaries or similar hand-coded resources (Kwong, 2008; Turney et al., 2011). We instead extract scores from a recently released dataset of lexical concepts rated on a 1-5 scale for concreteness by 20 annotators in a crowdsourcing experiment (Brysbaert et al., 2013).[1]

Subjectivity Subjectivity is the degree to which language is interpretable independently of the speaker's perspective (Langacker, 2002). For example, the utterance *he sits across the table* is more subjective than *he sits opposite Sam* as its truth depends on the speaker's position. Language may also be subjective because it conveys evaluations or opinions (Mihalcea et al., 2007).

Computational applications of subjectivity, including sentiment analysis and information extraction, have focused largely on phrase or document meaning.[2] In contrast, here we present six corpus-based features designed to quantify the *lexical* subjectivity of adjectives. The features *Comparability* and *Modifiability* are identified as predictors of subjectivity by Wiebe (2000). The remainder are motivated by corpus studies and/or observations from the theoretical literature.[3]

Adverbiability: Quirk et al. (1985) theorizes that subjective adjectives tend to develop derived adverbial forms, whereas more objective adjectives do not. We thus define adverbiability as the frequency of derived adverbial forms relative to the frequency of their base form, e.g.

$$\frac{\sum hotly}{\sum hot + \sum hotly}$$

Comparability: Wiebe (2000) oberve that *gradable* are more likely to be subjective. Following Wiebe, we note that the existence of comparative forms for an adjective are indicative of gradability. We thus define comparability as the frequency of comparative or superlative forms relative to the frequency of the base form, e.g.

$$\frac{\sum hotter + \sum hottest}{\sum hot + \sum hotter + \sum hottest}$$

LeftTendency: Adamson (2000) proposes that more subjective adjectives typically occur furthest from the noun in multiple-modifier strings such as (<u>*hot*</u> *crossed buns*). We consequently extract the LeftTendency of our adjectives, defined as the frequency of occurrence as the leftmost of two adjectives as a proportion of the overall frequency of occurrence in multiple-modifier strings.

Modifiability: Another characteristic of gradable adjectives noted by Wiebe (2000) is that they admit degree modifiers (<u>*very/quite*</u> *delicious*). We therefore extract the relative frequency of occurrence with one of a hand-coded list of English degree modifiers.

Predicativity: Bolinger (1967) proposed that subjective adjectives occur in predicative constructions (*the cake is sweet*), rather than attributive constructions (*the German capital*) more frequently than objective adjectives. We therefore extract the relative frequency of occurrence in such constructions.

Non-nominality: Many adjectives also function as nouns (<u>*sweet*</u> *cake* vs. (*boiled* <u>*sweet*</u>). Unlike nouns, many adjectives are inherently subjective, and the number of adjectives in texts correlates with human judgements of their subjectivity (Hatzivassiloglou and Wiebe, 2000). We therefore extract the frequency with which concepts are tagged as adjectives relative to as nouns, on the

[1] Available at http://crr.ugent.be/archives/1330.
[2] See e.g. (Wiebe and Riloff, 2011).
[3] Several of the features here were applied by Hill (2012), to the task of ordering multiple-modifier strings.

assumption that ‘pure’ adjectives are on average more subjective than nominal-style adjectives.

Concreteness meets Subjectivity Demonstrable commonalities in how different people perceive the physical world suggest that concrete language may be more objective than abstract language (Langacker, 1997). Intuitively, adjectives ascribing physical properties (*wooden shed*) are more objective than those conveying abstract traits (*suspicious man*). Indeed, in certain cases the original, apparently objective, senses of polysemous adjectives are not modifiable (*very wooden shed*?), while their more abstract sense extensions are (*very wooden personality*).

Motivated by these observations, in the following sections we test two hypotheses. (1) Subjective / objective adjectives are more likely to modify abstract / concrete nouns respectively. (2) Subjectivity and concreteness can predict aspects of how adjective and noun concepts combine.

3 Analysis

In addressing (1), we extracted the 2,000 highest-frequency nouns from the Brysbaert et al. (2013) concreteness dataset. We denote by $CONC(n)$ the mean concreteness rating for noun n. For the 24,908 adjectives that occur in some adjective-noun pair with one of these nouns in the British National Corpus (BNC) (Leech et al., 1994), we extracted subjectivity features from the Google Books Corpus (Goldberg and Orwant, 2013). Since each of the six features takes values on $[0, 1]$, we define the overall subjectivity of an adjective a with feature vector $\mathbf{s}^a = [s_1^a \dots s_6^a]$ as

$$SUBJ(a) = \sum_{i=1}^{6} s_i^a.$$

To verify the quality of our subjectivity features, we measured their performance as predictors in a logistic regression classifying the 3,250 adjectives labelled as subjective or not in the Wilson et al. (2005) lexicon.[4] The combination of all features produced an overall classifiction accuracy of 79%. The performance of individual features as predictors in isolation is shown in Figure 1 (top).

We first tested the relationship between concreteness and subjectivity with a correlation analysis over noun concepts. For each noun n we defined its *subjectivity profile* as the mean of the subjectivity vectors of its modifying adjectives

$$SUBJpfl(n) = \frac{1}{|A^n|} \sum_{a \in A^n} \mathbf{s}^a$$

where the bag A^n contains an adjective a for each occurrence of the pair (a, n) in the BNC. As hypothesized, $CONC(n)$ was a significant predictor of the magnitude of the subjectivity profile (Pearson $r = -0.421, p < 0.01$). This effect is illustrated in Figure 1 (bottom).

To explore the relationship between concreteness, subjectivity and meaning, we plotted the 20,000 highest frequency (a, n) pairs in the BNC in the *CONC-SUBJ* semantic space (Figure 2, top). In addition, to examine the effect of concreteness alone on adjective-noun semantics, we

[4]Available at http://mpqa.cs.pitt.edu/

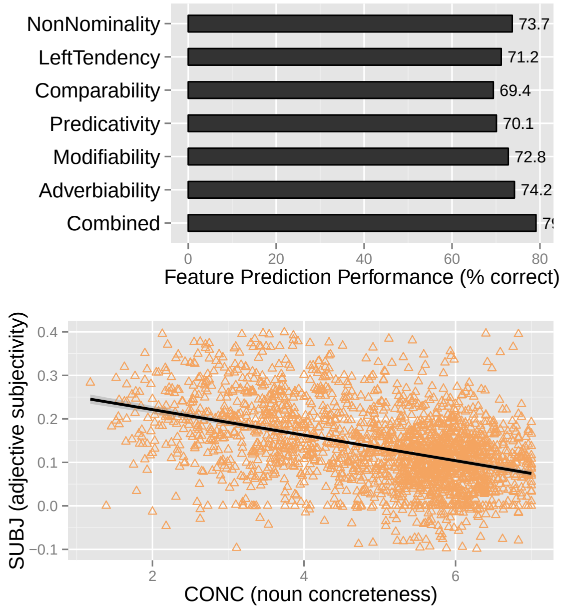

Figure 1: Top: Performance of features in predicting subjectivity labels from the Wilson et al. (2005) lexicon. Bottom: Concreteness-subjectivity correlation in adj-noun pairs.

a	$SUBJ(a)$	n	$CONC(n)$
flashy	1.98	*umbrella*	5
honest	1.63	*flask*	5
good	1.59	*automobile*	5
Siberian	6.9×10^-4	*virtue*	1.49
interglacial	6.3×10^-4	*pride*	1.46
Soviet	1.9×10^-4	*hope*	1.18

Table 1: The most and least subjective adjectives and most and least concrete nouns in our data.

(a,n)	Δ	(a,n)	Δ
white hope	4.61	*mature attitude*	4.05
fresh hope	4.34	*injured pride*	4.03
male pride	4.28	*black mood*	3.99
wild hope	4.06	*white spirit*	3.93

Table 2: The eight pairs with highest $\Delta = ExpCONC(a) - CONC(n)$ in our data.

defined a new adjective feature

$$ExpCONC(a) = \frac{1}{|N^a|} \sum_{n \in N^a} CONC(n)$$

where the bag N^a contains noun n for each occurrence of the pair (a,n) in the BNC. Figure 2 (bottom) illustrates the the *CONC-ExpCONC* space.

In both spaces, the extremities reflect particular meaning combination types. Pairs in the bottom-left region of the *CONC-SUBJ* space (objective adjectives with abstract nouns, such as *green politics*) seem to exhibit a non-literal, or at least non prototypical modification type. In contrast, for pairs in the objective+concrete corner, the adjectives appear to perform a classifying or categorizing function (*baptist minister*).

In the *CONC-ExpCONC* space, on the diagonal, where noun-concreteness is 'as expected', pairings appear to combine literally. Away from the diagonal, meaning composition is less predictable. In the top-left, where *ExpCONC* is greater than *CONC*, the combinations are almost all non-literal, as shown in Table 2.

In this section we have outlined a set of corpus features that, taken together, enable effective approximation of adjective subjectivity. The results of our analyses also demonstrate a clear interaction between subjectivity and concreteness scores for nouns attributed by human raters. Specifically, objective adjectives are more likely to modify concrete nouns and subjective adjectives are more likely to modify abstract nouns. Qualitative investigations further suggest the interaction between these dimensions to be useful in the semantic characterization of adjective-noun pairs, a proposition we test formally in the next section.

4 Evaluation

We evaluate the potential of our adjective subjectivity features, together with noun concreteness, to predict adjective-noun semantics, based on two existing classification tasks.

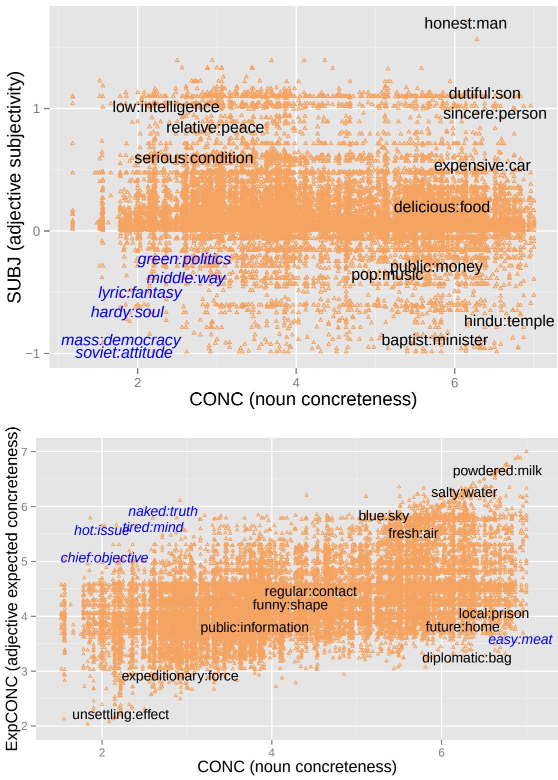

Figure 2: Adjective-noun pairs in two semantic spaces. Selected pairs are labelled for illustration, italics indicate non-literal meaning combinations.

4.1 Non-literal Composition Task

To evaluate their model of lexical concreteness, Turney et al. (2011) developed a list of 100 common adjective-noun pairs classified as either *denotative* (used literally) or *connotative* (non-literal) by five annotators. Using an identical supervised learning procedure to Turney et al. (logistic regression with 10-fold cross-validation), we test whether our lexical representations based on subjectivity and concreteness convey sufficient information to perform the same classification.

4.2 Modification-type Classification

Boleda et al. (2012) introduce a set of 370 adjective-noun pairs grouped into modification types by human judges. Because a *red car* is both a car and red, the pair is classed as *intersective*, whereas *dark humour*, which is not literally dark, is classed as *subsective*. To create a distinct but analogous binary categorization problem to the composition task, we filtered out pairs not unanimously allocated to either class. We again aim to classify the remaining 211 intersective and 93 subsective pairs with a logistic regression.

Feature type	Composition	Modification
Baseline	55.0	69.4
Concreteness	83.0	72.7
Subjectivity	64.0	70.4
Combined	**85.0**	**75.0**
Turney et al.	79.0	-

Table 3: Prediction accuracy (%) of models with different features on the two tasks. The baseline method allocates all test pairs to the majority class.

4.3 Results

Models were trained with concreteness features (*CONC* and *ExpCONC*), subjectivity features (*SUBJ* and *SUBJpfl*) and the combination of both types (*Combined*). The performance of each model is presented in Table 3, along with a baseline score reflecting the strategy of allocating all pairs to the largest class.

On the non-literal composition task, the concreteness (83.0) and combined (85.0) models outperform that of Turney et al. (79.0). The concreteness model performance represents further confirmation of the link between concreteness and composition. The improvement of this model over Turney et al. (2011) is perhaps to be expected, since our model exploits the wide scope of the new Brysbaert et al. (2013) crowdsourced data whereas Turney et al. infer concreteness scores from a smaller training set. Notably, our combined model improves on the concreteness-only model, confirming that the interaction of concreteness and subjectivity provides additional information pertinent to meaning composition.

The modification-type task has no performance benchmark since Boleda et al. (2012) do not use their data for classification. Although all models improve on the majority-class baseline, the combined model was again most effective. Additive improvement over the baseline in each case was lower than for the composition task, which may reflect the greater subtlety inherent in the subsective/intersective classification. Indeed, interannotator agreement for this goldstandard (Cohen's $\kappa = 0.87$) was lower than for the composition task (0.95), implying a less cognitively salient distinction.

5 Conclusion

We have shown that objective adjectives are most likely to modify concrete nouns, and that non-literal combinations can emerge when this principle is violated (Section 3). Indeed, the occurrence of an adjective with a more abstract noun than those it typically modifies is a strikingly consistent indicator of metaphoricity (Table 2). In addition, we showed that both concreteness and subjectivity improve the automatic classification of adjective-noun pairs according to compositionality or modification type (Section 4). Importantly, a classifier with both subjectivity and concreteness features outperforms concreteness-only classifiers, including those proposed in previous work.

The results underline the relevance of both subjectivity and concreteness to lexical and phrasal semantics, and their application to language processing tasks. We hypothesize that concreteness and subjectivity are fundamental to human language processing because language is precisely the conveyance of information about the world from one party to another. In decoding this signal, it is clearly informative to understand to what extent the information refers directly to the world, and also to what extent it reports a fact versus an opinion. We believe these dimensions will ultimately prove essential for computational systems aiming to replicate human performance in interpreting language. As the results suggest, they may be particularly important for capturing the intricacies of semantic composition and thus extending representations beyond the lexeme.

Of course, two dimensions alone are not sufficient to reflect all of the subtleties of adjective and noun semantics. For instance, our model classifies *white spirit*, a transparent cleaning product, as non-literal, since the lexical concreteness score does not allow for strong noun polysemy. Further, it makes no allowance for wider sentential context, which can be an important clue to modification type in such cases.

We aim to address these limitations in future work by integrating subjectivity and concreteness with conventially acquired semantic representations, and, ultimately, models that integrate input corresponding to the perceptual modalities.

6 Acknowledgements

The authors are supported by St John's College, Cambridge and The Royal Society.

References

Sylvia Adamson. 2000. A lovely little example. In Olga Fischer, Annette Rosenbach, and Deiter Stein, editors, *Pathways of change: Grammaticalization in English*. John Benjamins.

Carmen Banea, Rada Mihalcea, and Janyce Wiebe. 2014. Sense-level subjectivity in a multilingual setting. *Computer Speech & Language*, 28(1):7–19.

Edward Benson, Aria Haghighi, and Regina Barzilay. 2011. Event discovery in social media feeds. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 389–398. Association for Computational Linguistics.

Jeffrey R Binder, Chris F Westbury, Kristen A McKiernan, Edward T Possing, and David A Medler. 2005. Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, 17(6):905–917.

Gemma Boleda, Eva Maria Vecchi, Miquel Cornudella, and Louise McNally. 2012. First-order vs. higher-order modification in distributional semantics. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1223–1233. Association for Computational Linguistics.

Dwight Bolinger. 1967. Adjectives in english: attribution and predication. *Lingua*, 18:1–34.

Brian F Bowdle and Dedre Gentner. 2005. The career of metaphor. *Psychological review*, 112(1):193.

Elia Bruni, Gemma Boleda, Marco Baroni, and Nam-Khanh Tran. 2012. Distributional semantics in technicolor. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 136–145. Association for Computational Linguistics.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2013. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, pages 1–8.

Anneloes R Canestrelli, Willem M Mak, and Ted JM Sanders. 2013. Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes*, 28(9):1394–1413.

Yoav Goldberg and Jon Orwant. 2013. A dataset of syntactic-ngrams over time from a very large corpus of english books. In *Second Joint Conference on Lexical and Computational Semantics, Association for Computational Linguistics*, pages 241–247. Association for Computational Linguistics.

Vasileios Hatzivassiloglou and Janyce M Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th conference on Computational linguistics-Volume 1*, pages 299–305. Association for Computational Linguistics.

Felix Hill, Douwe Kiela, and Anna Korhonen. 2013. Concreteness and corpora: A theoretical and practical analysis. *ACL 2013 Workshop on Cognitive Modelling and Computational Linguistics, CMCL 2013*, page 75.

Felix Hill. 2012. Beauty before age? Applying subjectivity to automatic english adjective ordering. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop*, pages 11–16. Association for Computational Linguistics.

Oi Yee Kwong. 2008. A preliminary study on the impact of lexical concreteness on word sense disambiguation. In *PACLIC*, pages 235–244.

Ronald W Langacker. 1997. Consciousness, construal and subjectivity. *Language structure, discourse and the access to consciousness. Advances in Consciousness Research. John Benjamins*, pages 49–57.

Ronald W Langacker. 2002. Deixis and subjectivity. In Frank Brisard, editor, *Grounding: The epistemic footing of deixis and reference*, pages 1–28. De Gruyter.

Geoffrey Leech, Roger Garside, and Michael Bryant. 1994. Claws4: the tagging of the british national corpus. In *Proceedings of ACL*, pages 622–628. Association for Computational Linguistics.

Rada Mihalcea, Carmen Banea, and Janyce Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. In *Annual Meeting of the Association for Computational Linguistics*, volume 45, page 976.

Allan Paivio. 1990. *Mental Representations: A Dual Coding Approach*. Oxford University Press.

Allan Paivio. 1991. Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 45(3):255.

Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*, page 271. Association for Computational Linguistics.

Randolph Quirk, David Crystal, and Pearson Education. 1985. *A Comprehensive Grammar of the English Language*, volume 397. Cambridge Univ Press.

Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in* *natural language processing*, pages 105–112. Association for Computational Linguistics.

Stephen Roller and Sabine Schulte im Walde. 2013. A multimodal LDA model integrating textual, cognitive and visual modalities. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1146–1157, Seattle, Washington, USA. Association for Computational Linguistics.

Elizabeth C Traugott. 1985. On regularity in semantic change. *Journal of literary semantics*, 14(3):155–173.

Peter D Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and metaphorical sense identification through concrete and abstract context. In *Proceedings of the 2011 Conference on the Empirical Methods in Natural Language Processing*, pages 680–690.

Janyce Wiebe and Rada Mihalcea. 2006. Word sense and subjectivity. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 1065–1072. Association for Computational Linguistics.

Janyce Wiebe and Ellen Riloff. 2011. Finding mutual benefit between subjectivity analysis and information extraction. *Affective Computing, IEEE Transactions on*, 2(4):175–191.

Janyce Wiebe. 2000. Learning subjective adjectives from corpora. In *AAAI/IAAI*, pages 735–740.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics.