Improving Human Needs Categorization of Events with Semantic Classification

Haibo Ding[*], **Ellen Riloff**[†], **Zhe Feng**[*]

[*]Bosch Research and Technology Center, Sunnyvale, CA 94085, USA
[†]School of Computing, University of Utah, Salt Lake City, UT 84112, USA
{haibo.ding,zhe.feng2}@us.bosch.com, riloff@cs.utah.edu

Abstract

Human Needs categories have been used to characterize the reason why an affective event is positive or negative. For example, *"I got the flu"* and *"I got fired"* are both negative (undesirable) events, but getting the flu is a Health problem while getting fired is a Financial problem. Previous work created learning models to assign events to Human Needs categories based on their words and contexts. In this paper, we introduce an intermediate step that assigns words to relevant semantic concepts. We create lightly supervised models that learn to label words with respect to 10 semantic concepts associated with Human Needs categories, and incorporate these labels as features for event categorization. Our results show that recognizing relevant semantic concepts improves both the recall and precision of Human Needs categorization for events.

1 Introduction

Affective events have a positive or negative impact on the people who experience the event. For example, being hired for a job is typically a beneficial (positive) event, but being fired is usually a detrimental (negative) event. Recognizing affective events is critical to understand people's motivations, goals, desires, and empathy in narrative stories and conversations. Previous research has proposed several methods to recognize affective events and their polarity (e.g., (Deng et al., 2013; Vu et al., 2014; Reed et al., 2017; Ding and Riloff, 2016)). To achieve a deeper level of understanding, Ding and Riloff (2018a) further classified affective events into categories associated with theories of Human Needs (Maslow et al., 1970; Max-Neef et al., 1991) in psychology: *Physiological*, *Health*, *Leisure*, *Social*, *Finance*, and *Cognition*, to characterize the reason for the event's affective polarity. For example, breaking your arm is a negative event because it violates a need to maintain one's *Health*, but fighting with your spouse is negative because it violates a need for good *Social* relations with friends and family.

Human Needs categories naturally align with several broad conceptual classes, and we hypothesized that learning to recognize relevant semantic concepts would lead to more effective Human Needs categorization. For example, the *Physiological* need corresponds to basic functions such as breathing, sleeping, eating, and drinking. Learning to recognize FOOD/DRINK concepts should help identify events that belong to this category. Broadly, semantic concepts should help in two ways. First, semantic features are more general than words, which can suffer from sparsity. Second, given semantic features, a classifier can directly learn interactions between them, which should be more robust than interactions between individual words.

In this paper, we present lightly supervised classifiers that label words with respect to 10 semantic concepts associated with Human Needs categories: EMOTION, ENTERTAINMENT, EQUIPMENT, FOOD/DRINK, INTERPERSONAL, MEDICAL, MENTAL-PROCESS, MONEY/JOB, PEOPLE, and OTHER. Seed words for each semantic class are used as supervision, and pre-trained embedding vectors are used as word features. We explore three classification models: logistic regression, instance-based learning, and prototypical neural networks (Snell et al., 2017). Finally, the semantic class predictions are used as features for Human Needs event categorization, improving both recall and precision for this task.

2 Related Work

Previous work in NLP on affective events has primarily focused on identifying the affective polarity of events in narrative fables (Goyal et al.,

2013), tweets (Li et al., 2014), news (Deng et al., 2013; Deng and Wiebe, 2014), and personal blogs (Ding and Riloff, 2016; Reed et al., 2017; Ding and Riloff, 2018b). Recently, Ding et al. (2018) further characterized affective events in terms of human needs categories: *Physiological*, *Health*, *Leisure*, *Social*, *Financial*, *Cognition*, and *Freedom*, which can explain why an event is positive or negative. Subsequently, Ding and Riloff (2018a) designed supervised learning models and a semi-supervised co-training model to assign affective events to Human Needs categories.

Our work exploits recent advances in distributed word representations and semantic embeddings (e.g., (Mikolov et al., 2013; Pennington et al., 2014; Peters et al., 2018; Devlin et al., 2018)), and is related to domain adaptation of word embeddings (e.g., (Sarma et al., 2018)). But we aim to recognize a specific set of semantic concepts, rather than a specialized domain with domain-specific texts. The LIWC dictionary (Pennebaker et al., 2007) contains word lists for semantic categories that are similar to our targeted semantic concepts, but its coverage is insufficient. Our work is also related to semantic lexicon induction (Thelen and Riloff, 2002; McIntosh and Curran, 2009; Qadir and Riloff, 2012; De Benedictis et al., 2013; Gupta and Manning, 2015), contextual semantic tagging (Huang and Riloff, 2010), and fine-grained entity recognition (Fleischman and Hovy, 2002; Ling and Weld, 2012). Our goal, however, is not to generate a dictionary, or assign semantic meanings in sentence contexts. Instead, our classifier learns to recognize words that belong to a small set of relevant semantic classes based only on their pre-trained embedding vectors.

3 Lightly Supervised Semantic Classification

We hypothesized that Human Needs categorization would benefit from recognizing words that belong to the 10 semantic concepts listed in Table 1, based on our analysis of the Human Needs definitions presented in (Ding et al., 2018). In this section, we present lightly supervised models that learn to assign words to these classes.

3.1 Seeding

The input to our classifiers is a small set of seed words for each targeted semantic class. The affective events data set that we will use for our study

ENTERTAINMENT: play game movie story trip party birthday song music video
INTERPERSONAL: meet visit kiss share lie relationship hug marry agree admit
MENTAL-PROCESS: know remember read guess dream forget understand explain study memory
MEDICAL: die hurt pain kill sick blood hospital dead drug surgery
MONEY/JOB: job pay money deal sell business price sale purchase dollar
EQUIPMENT: car phone computer bike camera chair boat machine desk laptop
FOOD/DRINK: eat water food dinner drink lunch breakfast cake meal chocolate
PEOPLE: people friend guy girl man kid mom someone everyone family
EMOTION: good love nice fun bad happy best better smile enjoy laugh hate kind beautiful wrong amazing awesome funny crazy worry
OTHER: be have do get go time say make think take day come look thing tell start way try last year

Table 1: Semantic Concepts and Seed Words

(Section 5) was generated from the ICWSM 2009 and 2011 blog corpora (Burton et al., 2009, 2011), so we selected seed words from these corpora as well. We used the following procedure to identify commonly used words for each category: we sorted all word lemmas by frequency and selected the k top-ranked words belonging to each semantic concept. We set k=10 for all classes, except we set k=20 for EMOTION and OTHER because they are extremely large categories.[1] Table 1 shows the seeds selected for each semantic class.

3.2 Classification

We created three classification models: logistic regression, instance-based learning, and prototypical neural networks. For all three classifiers, we used the Word2Vec 300D pre-trained embeddings (Mikolov et al., 2013) as features. The seed words served as training examples, along with 500 randomly selected unlabeled words as additional seeds for the OTHER category, since it needs to represent a large and diverse "None-of-the-Above" class.

The first model is a one-vs.-rest logistic regression classifier, built using the scikit-learn toolkit (Pedregosa et al., 2011) with default parameters.

The second model uses instance-based classification. This method first creates a prototype representation for each semantic class as the mean of the word embeddings of its seeds. Given a new

---

[1]The k values were chosen arbitrarily without experimentation, so tuning these values could potentially further improve performance.

word, a probability distribution is computed over the semantic classes as the softmax of the negative Euclidean distance to each class prototype. The class with the highest probability is chosen.

The third model uses prototypical neural networks (Snell et al., 2017), which have performed well on "few-shot" learning tasks with limited labeled training data, because of its simple inductive bias. We created a single layer feed-forward network with ReLU activations as the embedding function f. To learn parameters for f, we use the same training algorithm as Snell et al. (2017) except that we train on all semantic classes in each training episode, and both the support set and query set consist of 5 randomly selected examples per class. During training, we use the following parameters: the dimension of the embedding representation layer is 32, the learning rate is .01, and the weight decay is .0001. We train the model for 20 epochs with 100 episodes for each epoch.

To predict the class label for a new word, the process is the same as the instance-based model, except that the learned embedding is used. First, we create a prototype embedding c_k for each semantic class k using Equation 1, where S_k contains all the labeled seed words for class k.

$$c_k = \frac{1}{|S_k|} \sum_{x_i \in S_k} f(x_i) \qquad (1)$$

Given a new word, a probability distribution over the classes is computed as the softmax of the negative Euclidean distance d to each prototype, as shown in Equation 2.

$$p(y = k|x) = \frac{\exp(-d(f(x), c_k))}{\sum_{k'} \exp(-d(f(x), c_{k'}))} \qquad (2)$$

4 Human Needs Categorization

Our goal is to explore whether semantic classification of terms can improve Human Needs categorization of affective events. Toward this end, we used the Human Needs categorization framework described in Ding and Riloff (2018a) which is a *co-training model* that iteratively trains two models with different views of the data: (1) an *event expression classifier* that uses the words in an event expression as input, and (2) an *event context classifier* that uses the sentence contexts that mention an event as input. An event expression is represented as a tuple consisting of 4 components: (Agent, Predicate, Theme, PP). The event

expression classifier is a logistic regression model that takes the embedding of an event expression as input, which is computed as the average over the embeddings of its individual words. The architecture and models are the same, but in this paper we aim to improve the event expression classifier by incorporating semantic classification.

Given an event expression, we extract two types of semantic features from the head words of its 4 components. For each of the 4 head words, we create 10 real-valued features representing the confidence scores produced by the classifier for each of the 10 semantic classes. In addition, we create 10 binary features (one per semantic class) indicating whether *any* of the head words belongs to each class, based on the classifier's predicted labels. Consequently, for each event expression the semantic classifier generates 50 semantic features.

5 Evaluation

We conducted two sets of experiments to evaluate the impact of our semantic classifiers. First, we show the results of adding semantic features to the event expression classifier for Human Needs categorization. Second, we evaluate the impact of the enhanced event expression classifier in the full co-training model. We used the evaluation data set created by Ding and Riloff (2018a), which contains 542 affective events with manually assigned Human Needs labels. To ensure a fair comparison, we used the same evaluation settings: we perform 3-fold cross-validation on the evaluation data and report the average Precision, Recall, and F1 scores over the folds.

5.1 Human Needs Categorization Results

Table 2 shows the results of experiments with the event expression classifier. The first row, Embed (D&R 2018a), shows the performance of Ding & Riloff's original event expression classifier, which is a logistic regression model with event expression embeddings as features. The next three rows show the performance of a logistic regression model that uses our semantic features instead. We show results for semantic features produced by each of our three models: instance-based classification (Sem:InstBased), logistic regression (Sem:LR), and the prototypical neural network (Sem:ProtoNets). The LR and ProtoNets models

Features	Precision	Recall	F1
Embed (D&R 2018a)	64.2	51.7	54.8
Sem:InstBased	53.7	42.4	45.1
Sem:LR	72.3	46.8	52.0
Sem:ProtoNets	63.1	48.8	52.4
Embed+Sem:InstBased	64.2	55.2	58.1
Embed+Sem:LR	**68.7**	57.3	60.8
Embed+Sem:ProtoNets	67.5	**58.7**	**61.9**

Table 2: Results for Human Needs Categorization with Event Expression Classifiers

achieve an F1 score[2] $\geq$ .52, which is not far below the performance of the D&R model that uses embedding features.

The last three rows of Table 2 show results for the event expression classifier with features for *both* the event expression embedding and the 50 semantic features produced by one of our three semantic classifiers. All of these models outperform the original D&R model. The best model, ProtoNets, substantially improves Human Needs categorization from 54.8% → 61.9%.

Human Need Category	D&R 2018a			Our Results		
	Pr	R	F1	Pr	R	F1
Physio	81	68	74	86.7	68.3	**76.3**
Health	68	50	57	71.6	59.3	**64.8**
Leisure	69	63	66	74.8	64.0	**68.8**
Social	68	79	73	73.9	82.4	**77.9**
Finance	67	44	52	60.5	47.8	**52.9**
Cognition	92	46	58	83.8	54.2	**65.5**
Emotion	64	74	69	65.5	74.2	**69.6**
None	48	52	50	49.0	53.3	**51.1**
AVG	69.7	59.5	62.4	70.7	62.9	**65.9**

Table 3: Results for Human Needs Categorization with Co-Training Models

In the next set of experiments, we evaluated the impact of the new event expression classifier in the co-training model for Human Needs categorization. Table 3 shows the results for the original co-training model reported in our previous work (Ding and Riloff, 2018a) alongside the results for our enhanced co-training model, which is identical except that we replaced the original event expression classifier with our Embed+Sem:ProtoNets model. We used the same experimental settings, running our co-training model for 20 iterations and reporting the best results[3]. Table 3 shows the Precision (Pr), Recall (R), and F1 scores for each

---

[2]Note that this result is the average F1 score over the cross-validation folds, not the F1 score of the average precision and average recall.

[3]Our co-training method achieved the best result after 17 iterations.

Human Needs category and the macro-averaged (AVG) scores. The enhanced co-training model improves performance on every Human Needs category, increasing the average F1 score from 62.4% to 65.9%.
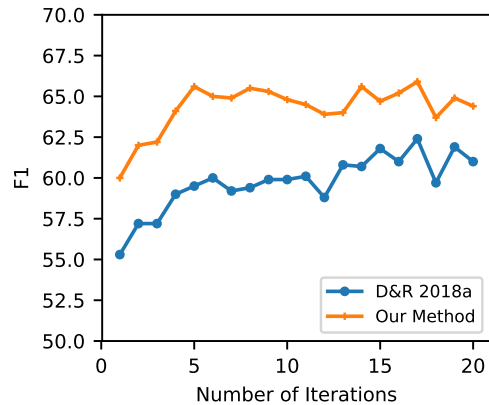

Figure 1: Performance of Co-training Models with Semantic Class Features at Different Iterations

Figure 1 shows the performance of the original co-training method D&R 2018a and our new method with semantic features learned by Embed+Sem:ProtoNets model after each iteration. This result shows that enhancing the event expression classifier with semantic features helps the co-training model improve more rapidly and achieve better performance than the original model for each iteration.

5.2 Analysis of Semantic Classification

We also informally evaluated the quality of the semantic labels assigned by the semantic classifier, to better understand its strengths and weaknesses. One of the authors assigned each word[4] in the evaluation data to one of the 10 semantic classes. Then we compared these human labels to the predicted labels from the semantic classifier.

Table 4 shows the performance for each semantic class. Overall, the classifier achieved a macro-averaged F1 score of 68.8%. Performance across the semantic classes varies, with several classes achieving high precision and high or moderate recall (OTHER, PEOPLE, ENTERTAINMENT, EMOTION, INTERPERSONAL, MENTAL-PROCESS), a few achieving high recall but low to moderate precision (FOOD/DRINK, MONEY/JOB), and a few with moderate recall and precision (EQUIPMENT, MEDICAL). Overall, these results demonstrate

---

[4]Except not pronouns.

Semantic Classes	Precision	Recall	F1
OTHER	79.5	91.0	84.9
PEOPLE	94.2	75.4	83.8
FOOD/DRINK	57.1	92.3	70.6
ENTERTAINMENT	89.1	54.7	67.8
EQUIPMENT	62.5	66.7	64.5
EMOTION	80.0	51.2	62.5
INTERPERSONAL	84.6	47.8	61.1
MENTAL-PROCESS	77.3	48.6	59.6
MEDICAL	64.5	51.3	57.1
MONEY/JOB	37.9	73.3	50.0
AVG	72.7	65.2	68.8

Table 4: Semantic Classification Results

that the semantic classifier produced fairly good predictions for most categories given only light supervision. One could almost certainly further improve these scores with more seed examples or by incorporating readily available external resources for categories such as EMOTION and MEDICAL, which would likely yield further gains for Human Needs categorization. More generally, our lightly supervised approach for training a semantic classifier demonstrates that one can rapidly create a classifier for a specific set of semantic concepts that are important for an application domain.

$\langle$I, gain, pleasure$_{\text{EMOTION}}$, $\rangle$
$\langle$I, be, busy, with my homework$_{\text{MENTAL-PROCESS}}$ $\rangle$
$\langle$we, finish, tour$_{\text{ENTERTAINMENT}}$, $\rangle$
$\langle$our pizza$_{\text{FOOD/DRINK}}$, arrive, , $\rangle$
$\langle$my eye, hurt$_{\text{MEDICAL}}$, bad$_{\text{EMOTION}}$, , $\rangle$
$\langle$I, try to cooperate$_{\text{INTERPERSONAL}}$, , $\rangle$
$\langle$I, dance$_{\text{ENTERTAINMENT}}$, , with my mom$_{\text{PEOPLE}}$ $\rangle$
$\langle$I, dance$_{\text{ENTERTAINMENT}}$, , with my friend$_{\text{PEOPLE}}$ $\rangle$
$\langle$I, not function, , at work$_{\text{MONEY/JOB}}$ $\rangle$
$\langle$day, be, magical$_{\text{EMOTION}}$, $\rangle$
$\langle$people$_{\text{PEOPLE}}$, buy$_{\text{MONEY/JOB}}$, home, $\rangle$
$\langle$we, grow, hungry$_{\text{FOOD/DRINK}}$, $\rangle$
$\langle$I, forget$_{\text{MENTAL-PROCESS}}$, paper, $\rangle$
$\langle$I, not learn$_{\text{MENTAL-PROCESS}}$, something, $\rangle$
$\langle$I, buy$_{\text{MONEY/JOB}}$, film$_{\text{ENTERTAINMENT}}$, $\rangle$
$\langle$house phone$_{\text{EQUIPMENT}}$, not work$_{\text{MONEY/JOB}}$, , $\rangle$
$\langle$epidural$_{\text{MEDICAL}}$, start to work$_{\text{MONEY/JOB}}$, , $\rangle$

Table 5: Examples of Affective Events with Automatically Predicted Semantic Classes

Table 5 presents some examples of affective events and their semantic classes that are assigned by the Prototypical Networks Model. All the unlabeled words in the table were assigned to OTHER class[5]. Besides the words

---

[5]In our experiments, we did not apply the semantic clas-

(e.g., "pleasure$_{\text{EMOTION}}$", "pizza$_{\text{FOOD/DRINK}}$", "cooperate$_{\text{INTERPERSONAL}}$") that were classified correctly, some words in events also received incorrect semantic category labels. For example, the "work" in "house phone not work" and "epidural start to work" was incorrectly classified to MONEY/JOB, which suggests that it may be beneficial to further improve the performance of semantic categorization of words in events by considering the contextual meaning of polysemous words. In addition, our set of semantic classes was selected based on our intuition about what concepts are most relevant to the human needs categories, but it might be worthwhile for future work to more thoroughly explore a large set of semantic concepts.

6 Conclusions

We proposed to improve human needs categorization of affective events by adding semantic features that classify terms into related semantic concepts. We designed lightly supervised models that learn to classify words with respect to semantic concepts using only their pre-trained word embedding vectors and seed words as training data. We then showed that representing semantic concepts improves both the precision and recall for Human Needs categorization of events.

Acknowledgements

References

Kevin Burton, Akshay Java, and Ian Soboroff. 2009. The ICWSM 2009 Spinn3r dataset. In *Third Annual Conference on Weblogs and Social Media*.

Kevin Burton, Niels Kasch, and Ian Soboroff. 2011. The ICWSM 2011 Spinn3r dataset. In *Proceedings of the Annual Conference on Weblogs and Social Media*.

---

sifier to prepositions and pronouns, but directly assign first person pronouns to OTHER, and assign second and third person pronouns to PEOPLE.

Flavio De Benedictis, Stefano Faralli, and Roberto Navigli. 2013. Glossboot: Bootstrapping multilingual domain glossaries from the web. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*.

Lingjia Deng, Yoonjung Choi, and Janyce Wiebe. 2013. Benefactive/malefactive event and writer attitude annotation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*.

Lingjia Deng and Janyce Wiebe. 2014. Sentiment propagation via implicature constraints. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Haibo Ding, Tianyu Jiang, and Ellen Riloff. 2018. Why is an event affective? classifying affective events based on human needs. In *the AAAI-18 Workshop on Affective Content Analysis*.

Haibo Ding and Ellen Riloff. 2016. Acquiring knowledge of affective events from blogs using label propagation. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Haibo Ding and Ellen Riloff. 2018a. Human needs categorization of affective events using labeled and unlabeled data. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2018)*.

Haibo Ding and Ellen Riloff. 2018b. Weakly supervised induction of affective events by optimizing semantic consistency. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Michael Fleischman and Eduard Hovy. 2002. Fine grained classification of named entities. In *Proceedings of the 19th International Conference on Computational Linguistics*.

A. Goyal, E. Riloff, and H. Daumé III. 2013. A Computational Model for Plot Units. *Computational Intelligence*, 29(3):466–488.

Sonal Gupta and Christopher D Manning. 2015. Distributed representations of words to guide bootstrapped entity classifiers. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Ruihong Huang and Ellen Riloff. 2010. Inducing domain-specific semantic class taggers from (almost) nothing. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*.

Jiwei Li, Alan Ritter, Claire Cardie, and Eduard Hovy. 2014. Major life event extraction from twitter based on congratulations/condolences speech acts. In *Proceedings of Empirical Methods in Natural Language Processing*.

Xiao Ling and Daniel S Weld. 2012. Fine-grained entity recognition. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Abraham Harold Maslow, Robert Frager, James Fadiman, Cynthia McReynolds, and Ruth Cox. 1970. *Motivation and personality*, volume 2. Harper & Row New York.

Manfred Max-Neef, Antonio Elizalde, and Martin Hopenhayn. 1991. *Human Scale Development: Conception, Application and Further Reflections*. The Apex Press.

T. McIntosh and J. Curran. 2009. Reducing semantic drift with bagging and distributional similarity. In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

James W Pennebaker, Roger J Booth, and Martha E Francis. 2007. Linguistic inquiry and word count: LIWC2007. *Austin, TX: liwc.net*.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2018)*.

Ashequl Qadir and Ellen Riloff. 2012. Ensemble-based semantic lexicon induction for semantic tagging. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*. Association for Computational Linguistics.

Lena Reed, JiaQi Wu, Shereen Oraby, Pranav Anand, and Marilyn A. Walker. 2017. Learning lexico-functional patterns for first-person affect. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*.

P. Sarma, Y. Liang, and W. Sethares. 2018. Domain adapted word embeddings for improved sentiment classification. In *Proceedings of the ACL 2018 Workshop on Deep Learning Approaches for Low-Resource NLP*.

Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*.

M. Thelen and E. Riloff. 2002. A bootstrapping method for learning semantic lexicons using extraction pattern contexts. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*.

Hoa Trong Vu, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura. 2014. Acquiring a dictionary of emotion-provoking events. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*.