Prototype theory and emotion semantic change

Aotao Xu (a26xu@cs.toronto.edu)
Department of Computer Science, University of Toronto

Jennifer Stellar (jennifer.stellar@utoronto.ca)
Department of Psychology, University of Toronto

Yang Xu (yangxu@cs.toronto.edu)
Department of Computer Science, Cognitive Science Program, University of Toronto

Abstract

An elaborate repertoire of emotions is one feature that distinguishes humans from animals. Language offers a critical form of emotion expression. However, it is unclear whether the meaning of an emotion word remains stable, and what factors may underlie changes in emotion meaning. We hypothesize that emotion word meanings have changed over time and that the prototypicality of an emotion term drives this change beyond general factors such as word frequency. We develop a vector-space representation of emotion and show that this model replicates empirical findings on prototypicality judgments and basic categories of emotion. We provide evidence that more prototypical emotion words have undergone less change in meaning than peripheral emotion words over the past century, and that this trend holds within each family of emotion. Our work extends synchronic theories of emotion to its diachronic development and offers a computational characterization of emotion semantics in natural language use.

Keywords: emotion; semantic field; semantic change; prototype theory; word vector

Introduction

Emotion plays a central role in cognition and evolution (Darwin, 1872). Unique to humans, natural language enables us to communicate emotions through words such as *joy* and *anger* beyond non-verbal means (Johnson-Laird & Oatley, 1992; Jackson et al., 2019). For example, the word *awe* used to express "a feeling of fear or dread", but it now expresses "a feeling of reverential respect, mixed with wonder or fear". [1] Here we present a computational approach to characterize meaning of emotion words and identify what principles may underlie historical meaning change in the semantic field of emotion.

Prototype theory of emotion

The starting point of our inquiry is inspired by the rich psychological literature on emotion. We focus on prototype theory which postulates that 1) emotion words exhibit graded membership, with certain words of emotion judged to be more prototypical than other words (Shaver, Schwartz, Kirson, & O'connor, 1987; Rosch, 1975), and 2) the field of emotion is derived and structured from a small set of basic categories or families (Shaver et al., 1987; Johnson-Laird & Oatley, 1992).[2] Empirical work on emotion has provided evidence for this prototype view using a variety of stimuli ranging from emotion words (Storm & Storm, 1987), videos (Cowen & Keltner, 2017), and facial expressions (Russell & Bullock, 1986; Ekman, 1992). Prototype theory provides a synchronic account of the mental representation of emotion terms, but how this view extends or relates to the diachronic development of emotion words is an open problem that forms the basis of our inquiry.

Theories of semantic change

Our work also draws on an independent line of research in historical semantic change. Two generalizations made in this area appear most relevant. One generalization concerns meaning change in semantic fields or groups of words that are closely related in meaning. This line of work has shown that words within the same semantic field tend to undergo parallel change in meaning, attested in synaesthetic adjectives (Williams, 1976), animal words (Lehrer, 1985), and near-synonyms (Xu & Kemp, 2015). This view suggests unidirectionality in meaning change of a semantic field, but it does not explain how different words (within the same field) might change meaning at differential rates.

The other generalization is more directly related to prototype theory, also known as diachronic prototype semantics (Geeraerts, 1997). This view postulates that more prototypical referents of a word tend to stay prototypical, and such senses of a word are more likely to persist over time than peripheral senses. Our work is aimed at extending this theory to the level of semantic field: we explore whether prototype theory would predict rates of meaning change across emotion words (as opposed to within each emotion word).

Our hypothesis and approach

We hypothesize that emotion words considered more prototypical should tend to be more stable in meaning than peripheral emotion words. We ground the notion of prototypicality in empirical work on human judgments of representativeness of emotion words (Shaver et al., 1987; Storm & Storm, 1987; Russell & Bullock, 1986). In these studies, a word's prototypicality is typically rated by participants in terms of how good that word is perceived as an emotion word. We postulate that words considered to be more prototypical such as

[1] Entry "awe, n.1" retrieved from Oxford English Dictionary (2019) at www.oed.com/view/Entry/13911/ on January 11, 2020.

[2] Although there is no consensus on which emotions constitute the basic categories, we focus on "love", "joy", "anger", "sadness", and "fear" drawn from Shaver et al. (1987).

love and *anger* should resist meaning change for their communicative function of conveying canonical emotions, more so than peripheral emotion words such as *zest* and *optimism* (illustrated in Figure 1). Our proposal about prototypicality is necessarily confounded with factors such as word frequency (e.g., prototypical words tend to be frequently used), so we take into account these confounding variables in the evaluation of our hypothesis.

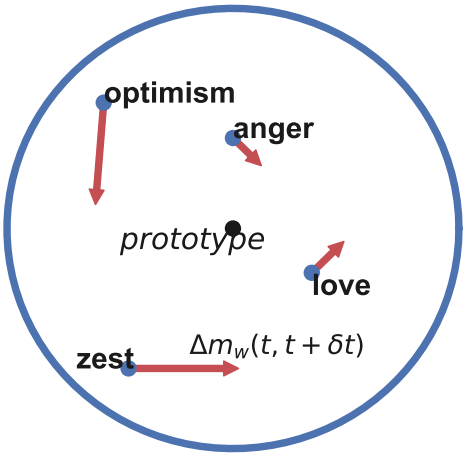

Figure 1: An illustration of our hypothesis. The center represents the prototype of the emotion semantic field. The blue circle represents the boundary of the field. Each word is an example of a member of the field. The proximity of each word to the center corresponds to its perceived prototypicality. The length of each arrow indicates the rate of semantic change, denoted by $\Delta m_w(t, t+\delta t)$, that word w undergoes over time. The direction of each arrow is for illustration only.

Our approach builds on recent computational work in diachronic word embeddings (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013; Hamilton, Leskovec, & Jurafsky, 2016). We capture meanings of emotion words using a vector-space representation trained on historical text corpora of natural language use. Although vector-space models of word meaning have been used for inducing human emotion ratings on dimensions such as valence and arousal (Buechel & Hahn, 2018) and analyzing emotion categories in documents (Calvo & Mac Kim, 2013), to our knowledge there exists no work that replicates psychological findings of emotion words with regard to their graded prototypicalities and family structures using large-scale natural semantic models.

Here we contribute a methodology for modelling emotion semantics and show how word vectors derived from independent linguistic corpora can capture both human judgments of prototypicality and human categorization of basic emotion families. We also contribute a field-level view of diachronic prototype semantics and provide evidence that prototypicality predicts stability of meaning in English emotion terms over the past century, even when factors such as word frequency are controlled for.

Computational methodology

We present a computational method to test our hypothesis using vector-space representations of meaning. We first describe a formulation of the prototypicality and family structures of emotion words in vector space. We then describe how we capture semantic change using word vectors, as well as to test theories about semantic change of emotion words.

Synchronic semantics of emotion words

We use word vectors trained on synchronic text data to model graded prototypicalities and family structures of emotion words. Concretely, we formulate the modelling of these two properties as regression and classification, respectively, and we approach these tasks using simple methods that are interpretable from a prototype-theoretical perspective.

In the following, we use E to denote an empirically determined set of emotion words and B to denote an empirically determined set of labels for basic families.

Prototypicality judgments of emotion words. We show that vector-space representations can capture human judgements of emotion prototypicality p_E. Concretely, we consider a regression task in which we use word vectors to induce prototypicality ratings, and we approach this task by constructing a prototype in vector space from a small set of seed words. We construct this vector for the emotion category v_{avg} by using the average of word vectors of emotion words with high empirical prototypicality ratings; here we use *love*, *happiness*, *anger*, *sadness*, and *fear*:

$$v_{avg} = \frac{1}{5}(v_{love} + v_{happiness} + v_{anger} + v_{sadness} + v_{fear}) \quad (1)$$

To capture prototypicality or graded membership, we approximate the prototypicality rating of a word w by computing the cosine similarity between its vector v_w and v_{avg}:

$$\hat{p_E}(w) = \frac{v_w \cdot v_{avg}}{\|v_w\|_2 \|v_{avg}\|_2} \quad (2)$$

Essentially, following prototype theory, we obtain $\hat{p_E}$ by gauging how similar the prototype v_{avg} is to a word in meaning represented by vector space.

Categorization of emotion words. We also show that it is possible to capture human categorization of emotion words in vector space. Concretely, we consider a classification task in which we use word vectors to label emotion words with empirically derived emotion families. We approach this task by constructing a prototype within each category in vector space, and use these seed words for classifying the remaining words via nearest centroid (Tibshirani, Hastie, Narasimhan, & Chu, 2002). We start with prototype vectors v_b for all categories $b \in B$:

$$v_b = \frac{1}{|E_b|} \sum_{w \in E_b} v_w \quad (3)$$

where E_b is the set of emotion words in the category b determined empirically. Because we do not have corresponding

empirical ratings, we approximate the prototypicality of an emotion word $w \in E$ with respect to category b using a formulation akin to Equation 2:

$$\hat{p_B}(w,b) = \frac{v_w \cdot v_b}{\|v_w\|_2 \|v_b\|_2} \quad (4)$$

We classify each emotion word $w \in E$ by assigning a category label $\hat{b} \in B$ such that $\hat{p_B}(w,\hat{b})$ is the highest among approximate prototypicality values over all basic categories. Essentially, following prototype theory, we assign a word to a category $\hat{b}$ if they are highly similar to the prototype $v_{\hat{b}}$ in vector space.

Diachronic semantic change of emotion words

We describe how we quantify meaning change in emotion words by using word vectors trained on diachronic text data. We then consider prototypicality p_E and other possible factors that explain rates of semantic change and evaluate our main hypothesis. We also describe evaluation of our hypothesis at the fine-grained, family level.

Quantification of semantic change. Existing methods for quantifying the degree of semantic change of a word often rely on computing the cosine distance between its word vectors trained on different historical corpora (Hamilton et al., 2016; Dubossarsky, Weinshall, & Grossman, 2017). According to this measure, a greater cosine distance implies a greater degree of semantic change of the word. However, the cosine measure is by construction dependent on frequency (Dubossarsky et al., 2017) and when vectors are trained using word2vec, rotational alignment is necessary for cosine but increases noise (Dubossarsky, Hengchen, Tahmasebi, & Schlechtweg, 2019). As a result, we use an alternate method using the Jaccard distance between sets of k-nearest neighbours in semantic space (Xu & Kemp, 2015):

$$\Delta m_w(t, t+\delta t) = 1 - \frac{|kNN(t) \cap kNN(t+\delta t)|}{|kNN(t) \cup kNN(t+\delta t)|} \quad (5)$$

where $kNN(t)$ contains the $k = 100$ closest neighbours of word w at time t, measured by cosine similarity. We take k to be 100 following Xu & Kemp (2015), but our results are robust to variation in k from 25 to 100. Compared to the cosine method, this method enables more transparent interpretations of the degree of change because we can inspect and evaluate the sets of neighbours qualitatively. We evaluate this measure qualitatively by inspecting words with the most extreme changes and their nearest neighbours.

Factors in rate of semantic change. Besides empirical prototypicality ratings p_E, there are several other potential factors that can explain the rate of semantic change in emotion words. The law of conformity suggests that frequency of a word w at the starting time t, denoted $freq(w)$, is a negative correlating factor with the rate of change (Hamilton et al., 2016); since word length, denoted $len(w)$, is related to frequency (Zipf, 1949), we probe both frequency and length alongside prototypicality. We also probe the effect of polysemy as it has been shown to affect the rate at which a word gains or loses senses (Luo & Xu, 2018); we define the degree of polysemy of a word as the number of word senses it has at t, denoted $senses(w)$. Together, we test the effect of each factor using a multiple regression model:

$$\Delta m_w(t, t+\delta t) \sim p_E(w) + freq(w) + len(w) + senses(w) \quad (6)$$

Since prototypicality and frequency may be correlated (Geeraerts, 1997; Dubossarsky et al., 2017), we further investigate the effects of prototypicality and frequency on the rate of semantic change using partial correlation.

Rate of semantic change within categories. We repeat our investigation of prototypicality and frequency at the basic level. Here we stratify our emotion words E into $|B|$ bins according to their empirically determined basic-level categorization, and compute separate partial correlations per family. Because we do not have empirical prototypicality ratings for the basic categories, we approximate the ratings by using Equation 4. Since this approximation is dependent on using historical word embeddings and thus the starting time t, we track partial correlations across time.

Data

We obtained two independent sources of data: 1) human behaviour data regarding English emotion words, and 2) historical word embeddings and related historical linguistic data regarding English words.

Behavioral data

We obtained a list of emotion words with prototypicality ratings and empirically derived basic categories from Shaver et al. (1987). The list contains 213 words, but following the original authors, our analysis focused on words that have prototypicality ratings at least 2.75 with the addition of "surprise" and exclusion of "abhorrence", "ire", "malevolence", and "titillation"; we additionally included the word "awe". This provided us with 136 emotion words. The prototypicality ratings represent how prototypical a word denotes an emotion on a scale of 1 to 4. Although views on what constitute basic emotion categories might differ, here we obtained the 5 basic categories and corresponding categorizations of emotion words from the same source (Shaver et al., 1987). The recommended labels for these categories are "love", "joy", "fear", "sadness", and "anger".

Historical data

We used word embeddings, part-of-speech tags, and frequency data provided by Hamilton et al. (2016). We used Historical Word2Vec (SGNS) embeddings and frequencies obtained from Google N-Grams eng-all. These pretrained vectors do not cover our entire list of emotion words. Because the coverage improves as the data becomes more recent, our analysis focuses on the decades between 1890 and 1990. Finally, we obtain historical word senses from the Historical Thesaurus of English (Kay, Roberts, Samuels, & Wotherspoon, 2017).

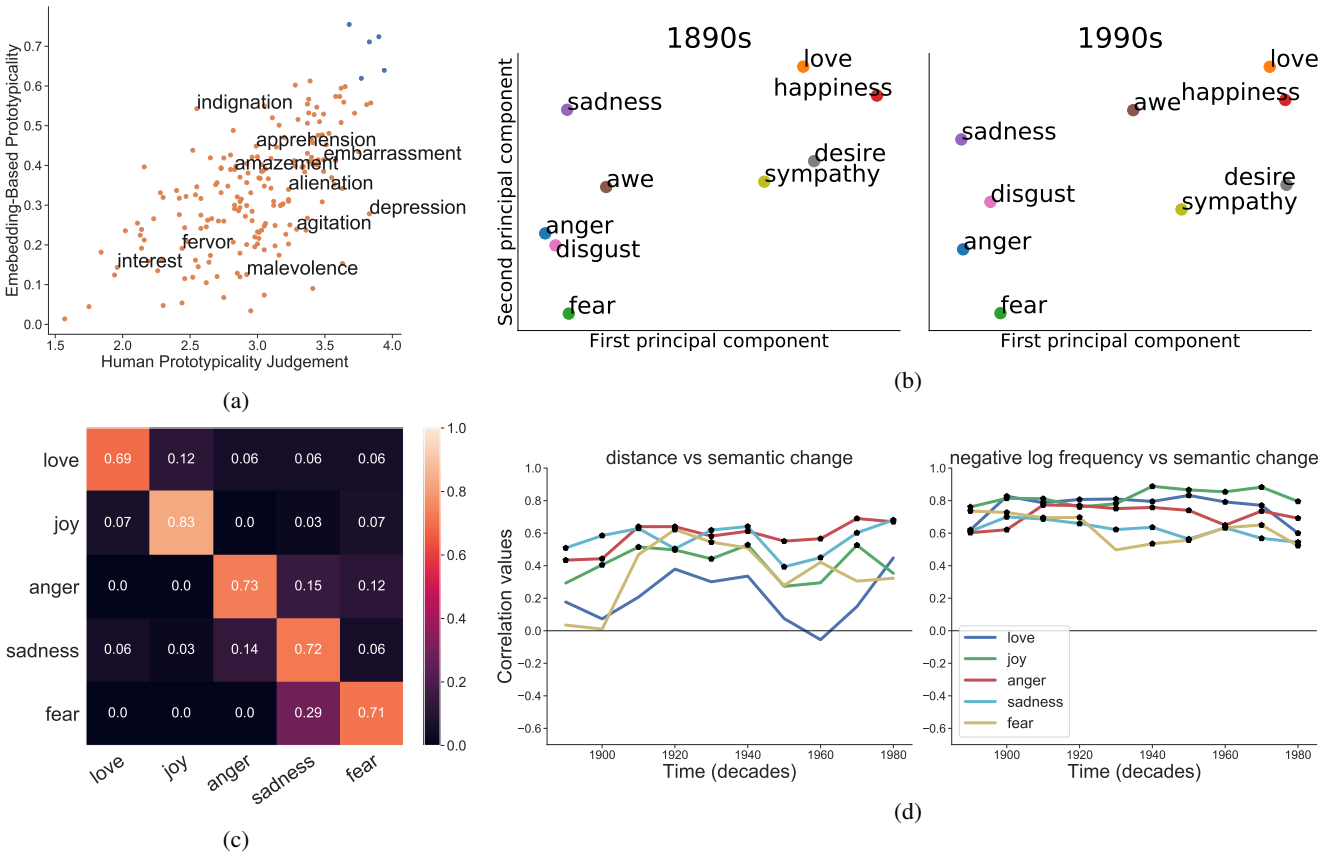

Figure 2: Summary of main results. The first row (a,b) corresponds to our investigation over all emotion words, and the second row (c,d) focuses on basic-level results; the first column (a,c) shows results for synchronic modelling, and the second column (b,d) illustrates our diachronic findings. (a) Scatter plot comparing empirical prototypicality ratings and approximated ratings; each dot corresponds to a word obtained from Table 1 of Shaver et al. (1987); blue dots indicate words used for obtaining the prototype vector. (b) An illustration of the intuition of our diachronic hypothesis. (c) Confusion matrix obtained from recreating basic-family categorizations of emotion words in vector space; vertical axis corresponds to empirical, ground-truth categorizations, and rows are normalized; horizontal axis corresponds to reconstructed categorizations. (d) Line plot comparing prototypicality and frequency predictors across time and basic families; each star indicates a significant correlation (uncorrected p-value < 0.05).

Results

We present our results in the following order: 1) the reconstruction of synchronic emotion semantics in vector space, with regards to prototypicality judgments and basic family categorizations, and 2) the evaluation of our diachronic hypothesis on the semantic change of emotion words, and an exploration of this hypothesis extended to the basic level.

Synchronic semantics of emotion words

Prototypicality judgement of emotion words. We used word vectors to induce human prototypicality judgements. Here we used the entire word list of 213 words from Shaver et al. (1987); we used word vectors trained on text data from the 1990s, close to the date of the empirical experiments. From these vectors we constructed a prototype vector defined in equation 1, and we approximated prototypicality values for all words in the list using equation 2. The Pearson correlation between empirical prototypicality ratings p_e and approximate prototypicality ratings $\hat{p}_e$ is 0.632, p-value < 0.0001. A clear positive, linear pattern can be observed in the scatter plot of words (see Figure 2a). Outliers seem to be related to broader contexts, such as society and the economy (e.g. *indignation* and *depression*). This provides some evidence that word vectors reflect human intuitions about the prototypicality of emotion words.

Categorization of emotion words. We also used word vectors to recreate human categorizations of basic emotion families, using a subset of the vectors from the previous section intersecting with Figure 1 of Shaver et al. (1987) and excluding the "surprise" family due to its small size. From these vectors we constructed prototype vectors defined in equation 3, and we approximated basic-family categorizations for all words in the list based on values obtained from equation 4. Since our method resembles a standard supervised classifica-

a. Most Changing		
Word	Nearest Neighbours in 1890s	Nearest Neighbours in 1990s
zest	relish, enjoyment, sprightliness	juice, teaspoons, vinegar
infatuation	priestcraft, devastations, misanthrope	inhomogeneity, palates, pleurisy
sentimentality	cant, sentimentalism, rusticity	polyphony, sterne, mandel
optimism	pessimism, aptness, sentimentalism	pessimism, insecurity, enthusiasm
exhilaration	mountebank, festivity, tulip	joy, sadness, excitement
b. Least Changing		
Word	Nearest Neighbours in 1890s	Nearest Neighbours in 1990s
pity	compassion, love, sympathy	compassion, shame, sadness
grief	sorrow, anguish, joy	sorrow, sadness, anguish
misery	wretchedness, miseries, degradation	sorrow, bitterness, anguish
disgust	horror, aversion, indignation	sadness, annoyance, amazement
surprise	astonishment, amazement, dismay	astonishment, amazement, dismay

Table 1: Top 5 most changing and least changing words as well as their 3 nearest neighbours in the flanking decades.

tion task, we used leave-one-out cross validation to evaluate our approach (Molinaro, Simon, & Pfeiffer, 2005). The overall cross-validated accuracy is 0.744. Details are summarized in Figure 2c: we observe emotion words tend to be correctly categorized over all 5 families; error cases for "love" tend to occur in the positive-valence "joy" category; similarly, the bottom-right block of the confusion matrix also shows that errors tend to occur among the negative-valence categories "anger", "sadness", and "fear". This provides some evidence that word vectors reflect human intuitions about categorization of emotion words with respect to basic families.

Diachronic semantic change of emotion words

Factors in rate of semantic change. We tested our hypothesis at the superordinate level. We first conducted multiple regression on semantic change using the model defined by equation 6. The adjusted r^2 is 0.541, p-value < 0.0001, $n = 123$. The coefficient and p-value for each variable are -0.0566, p-value $= 0.011$ for prototypicality, -0.0553, p-value < 0.001 for log frequency, 0.0013, p-value $= 0.623$ for length, and 0.0065, p-value $= 0.001$ for number of senses. Since both prototypicality and frequency are statistically significant but also correlated, we used partial correlation to measure the strength of correlation between one of these predictors and semantic change while controlling for the other predictor. Controlled for log frequency, the partial correlation between prototypicality and semantic change is -0.233, p-value $= 0.0096$; controlled for prototypicality, the partial correlation between log frequency and semantic change is -0.665, p-value < 0.0001. While frequency is dominant, prototypicality is a competitive factor in explaining the semantic change of emotion words.

We provide an intuitive demonstration of this result in Figure 2b using principal component analysis; all axes were produced by taking the first two principal components of the vectors of emotion words from 1890, and the location of the plotted words were obtained by projecting word vectors from respective decades to these axes. Consider a somewhat prototypical emotion word, *disgust*, and a less prototypical word, *awe*, and note that they have similar log frequencies (-7.024, -6.918 respectively). In 1890, both *disgust* and *awe* are in the neighbourhood of negative-valence words (e.g. *sadness* and *fear*). However, in 1990, while *disgust* still remains among negative-valence words, *awe* becomes much closer to positive words (e.g. *love* and *happiness*).

We evaluated the measure of semantic change defined by equation 5 qualitatively by inspecting nearest neighbours retrieved using cosine similarity. Overall we observe that the qualitative changes in nearest neighbours of a word are intuitively related to the word's quantitative rate of semantic change: for example, in Table 1, we can observe *zest*, which used to primarily convey joy but later became primarily associated with food, is among the most changing emotion words; similarly, we can observe words like *surprise* barely changed.

Rate of semantic change within categories. We also tested our hypothesis at the basic level. We obtained partial correlations for every decade between 1890 and 1990 (see Figure 2d). We can observe that frequency is still a strong predictor of semantic change for all basic categories. On the other hand, we can observe that prototypicality is a strong predictor for the "anger" and "sadness" categories; it is somewhat strong for the "joy" category. However, prototypicality is not a consistently strong predictor for the "fear" category and it is weak for the "love" category; "fear" and "love" are the smallest categories (17 and 16; compare with anger 26, joy 30, sadness 30). This offers some support for our hypothesis at the basic level.

Table 2 offers a snapshot of the ranking of emotion words by rate of change at the basic level. We observe these ranks tend to reflect our results: for example, we can observe that short, common words like "love" and "joy" changed less than long, infrequent words like "alienation" and "isolation".

Family Name	Most Changing	Least Changing
love	infatuation, fondness sentimentality	affection, desire, love
joy	zest, optimism, exhilaration	happiness, joy, pride
anger	aggravation, ferocity, exasperation	disgust, anger, envy
sadness	dejection, alienation, isolation	pity, grief, misery
fear	hysteria, worry, nervousness	horror, fear, terror

Table 2: Top 5 most changing and least changing words per emotion family in the flanking decades.

Conclusion

The importance of emotions in human cognition and the unique human ability to express emotions through language signify any underlying historical changes in the meanings of emotion words. We proposed a hypothesis that explains these changes by drawing from prototype theory and linguistics, and we presented a computational approach to evaluate it. We made two main findings. First, we leveraged existing vector-space representations of word meaning and demonstrated that this representation reflects human psychology of emotion categories. Second, we used these representations to show that prototypical emotion words tend to be more stable in meaning, even when frequency is controlled for. Future work should explore if these findings generalize beyond English and to other semantic fields.

Acknowledgments

We thank members of the Language, Cognition, and Computation Group at UofT and Department of Linguistics at Yale for helpful suggestions. AX is supported by a UofT Entrance Scholarship. YX is funded through an NSERC Discovery Grant, a SSHRC Insight Grant, and a Connaught New Researcher Award.

References

Buechel, S., & Hahn, U. (2018). Word emotion induction for multiple languages as a deep multi-task learning problem. In *Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: Human language technologies, volume 1 (long papers)* (pp. 1907–1918).

Calvo, R. A., & Mac Kim, S. (2013). Emotions in text: dimensional and categorical models. *Computational Intelligence*, *29*(3), 527–543.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, *114*(38), E7900–E7909.

Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray.

Dubossarsky, H., Hengchen, S., Tahmasebi, N., & Schlechtweg, D. (2019). Time-out: Temporal referencing for robust modeling of lexical semantic change. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 457–470). Florence, Italy: Association for Computational Linguistics.

Dubossarsky, H., Weinshall, D., & Grossman, E. (2017). Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 1136–1145).

Ekman, P. (1992). Are there basic emotions? *Psychological review*, *99*(3), 550–553.

Geeraerts, D. (1997). *Diachronic prototype semantics: A contribution to historical lexicology*. Oxford University Press.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1489–1501).

Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., ... Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, *366*(6472), 1517–1522.

Johnson-Laird, P. N., & Oatley, K. (1992). Basic emotions, rationality, and folk theory. *Cognition & Emotion*, *6*(3-4), 201–223.

Kay, C., Roberts, J., Samuels, M., & Wotherspoon, I. (2017). *The historical thesaurus of english, version 4.21*. Glasgow, UK: University of Glasgow. Retrieved from `http://historicalthesaurus.arts.gla.ac.uk/`

Lehrer, A. (1985). The influence of semantic fields on semantic change. *Historical semantics. Historical word-formation. Berlin: Mouton*, 283–296.

Luo, Y., & Xu, Y. (2018). Stability in the temporal dynamics of word meanings. In *Cogsci.*

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Molinaro, A. M., Simon, R., & Pfeiffer, R. M. (2005). Prediction error estimation: a comparison of resampling methods. *Bioinformatics*, *21*(15), 3301–3307.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of experimental psychology: General*, *104*(3), 192.

Russell, J. A., & Bullock, M. (1986). Fuzzy concepts and the perception of emotion in facial expressions. *Social Cognition*, *4*(3), 309–341.

Shaver, P., Schwartz, J., Kirson, D., & O'connor, C. (1987). Emotion knowledge: further exploration of a prototype approach. *Journal of personality and social psychology*,

52(6), 1061.
Storm, C., & Storm, T. (1987). A taxonomic study of the vocabulary of emotions. *Journal of personality and social psychology*, *53*(4), 805.
Tibshirani, R., Hastie, T., Narasimhan, B., & Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Sciences*, *99*(10), 6567–6572.
Williams, J. M. (1976). Synaesthetic adjectives: A possible law of semantic change. *Language*, 461–478.
Xu, Y., & Kemp, C. (2015). A computational evaluation of two laws of semantic change. In *Cogsci.*
Zipf, G. K. (1949). *Human behavior and the principle of least effort.* addison-wesley press.