Slang Generation as Categorization

Zhewei Sun (zheweisun@cs.toronto.edu)
Department of Computer Science
University of Toronto

Richard Zemel (zemel@cs.toronto.edu)
Department of Computer Science
University of Toronto
Vector Institute

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

Abstract

Slang is a common device for expressivity in natural language. While slang has been studied extensively as a social phenomenon, its cognitive bases are not well understood. We formulate the processes of slang generation as a categorization problem. We explore a set of cognitive models of categorization that recommend slang words based on intended referents of the speaker beyond the existing senses of words. We test these models against a large repertoire of slang sense definitions from the Online Slang Dictionary and show that the categorization models predict slang word choices substantially better than chance, without explicit consideration of external social factors. We also show that words similar in existing senses tend to extend to similar novel slang senses, reflecting a process of parallel semantic change. Our work helps to ground theories of slang in cognitive models of categorization and provides the potential for machine processing of informal natural language.

Keywords: informal language; slang; generative model; categorization; language and cognition

Introduction

Slangs—a representative form of informal language—are ubiquitous in natural language, making up approximately 52% of words in all English books written in the past two centuries (Michel et al., 2011). Slang is a common device for enhancing expressivity in human language, allowing us to express a multitude of ideas beyond the standard lexicon. Slang also adds stylistic richness to language, often allowing the identification of social groups (Millhauser, 1952). Although slangs are prevalent and accountable for language expressivity, the cognitive processes that give rise to slangs are not well understood.

Previous work has characterized slang as a social phenomenon. For instance, Labov (1972, 2006) studied how informal language emerges as a result of differing ethnicity and social-economic status. More recent work has also suggested how slang might be influenced by multiple social factors including ethnicity (Blodgett, Green, & O'Connor, 2016), gender (Bamman, Eisenstein, & Schnoebelen, 2014), and geography (Eisenstein, O'Connor, Smith, & Xing, 2010). Although it is undeniable that slang is a social phenomenon, recent work on social media analysis has suggested that slangs are more likely to catch on if they are also linguistically appropriate (Stewart & Eisenstein, 2018). We extend these work by exploring the bases of slang from a cognitive perspective, complementary to the social factors that could influence slang formation.

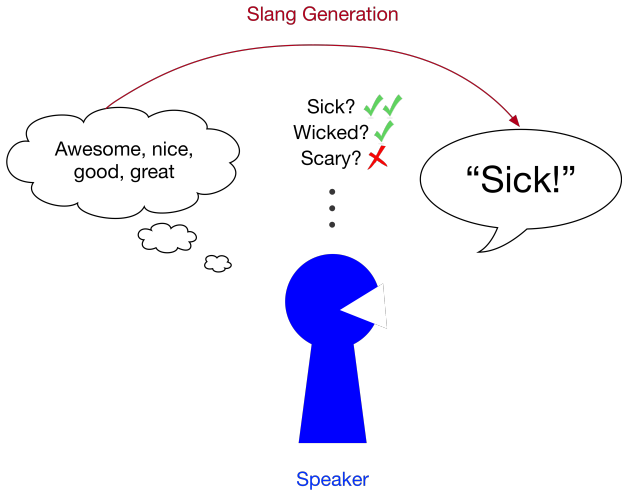

Figure 1: Illustration of the slang generation problem.

Recent work in cognitive science has explored related topics in the context of non-literal language, particularly the comprehension of metaphors (Kao, Wu, Bergen, & Goodman, 2014; Kao, Bergen, & Goodman, 2014). While slangs can often emerge from metaphorical relations, there exist many cases suggesting otherwise. For example, the slang word *sick* has the existing sense "ill" while its slang sense refers to "awesomeness". In this case, the link between the slang and existing senses are not metaphorical, but instead accounts to a polarity shift in sentiment from the existing sense.

Here we consider the general problem of slang generation by asking what cognitive processes can give rise to slang word choices for novel senses. Specifically, given a new intended slang referent one wishes to convey, how does the speaker choose an appropriate word for expressing that sense? Figure 1 illustrates this problem of *slang generation*.

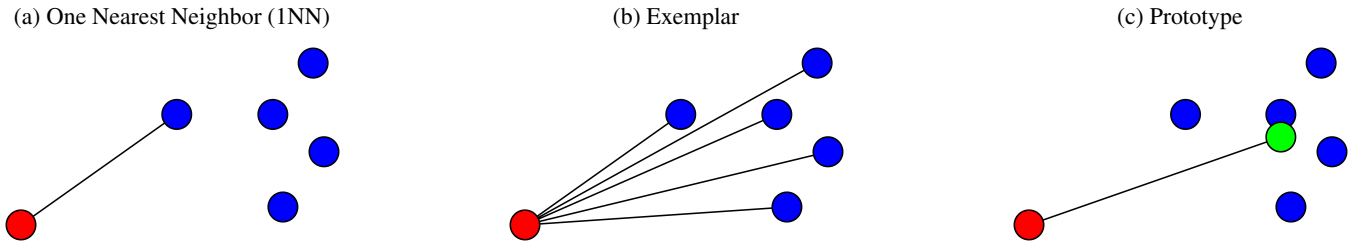

Figure 2: Illustration of categorization models for slang generation. Red (bottom-left) dot denotes novel slang sense. Blue dots denote existing senses of a candidate word. Green dot denotes prototype (or mean) of the existing senses.

Given a slang sense such as "awesome/nice", we wish to predict the word choice made by the speaker among possible alternative candidate words. In the illustrated case, the target word *sick* might be chosen if its existing senses relate to the novel slang sense, and words similar to the target word *sick* such as *wicked* might also have a good chance of being chosen. We formalize these intuitive notions of slang generation in terms of lexical choice via categorization, where we consider each candidate word as a category of existing word sense definitions. For this study, we focus on the problem of slang generation from words that are part of the existing lexicon, so we do not consider out-of-vocabulary or novel word forms for slang (e.g., Kulkarni & Wang, 2018).

We explore slang generation based on two key ideas from recent work on lexical semantic change, particularly historical word sense extension: 1) Words that bear closely related senses to a novel sense are likely to be extended to express that novel sense, a process known as semantic chaining (Lakoff, 1987; Malt, Sloman, Gennari, Shi, & Wang, 1999; Ramiro, Srinivasan, Malt, & Xu, 2018); 2) Words that begin with similar senses tend to extend to similar novel senses, a process also known as the law of parallel semantic change (Lehrer, 1985; Xu & Kemp, 2015). We formalize these ideas along with classic proposals of categorization in a simple computational framework and test them against a large online dictionary of slang.

To preview our findings, we show that cognitive models of categorization predict slang word choices substantially better than chance, and these models can be enriched by a mechanism of collaborative filtering that accounts for parallel semantic change.

Computational formulation

Models of categorization

We formulate slang generation as a categorization problem. Given a set of candidate words as categories $\{w_1, w_2, \cdots, w_N\}$ with sets of existing senses as exemplars $\{\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_N\}$ associated with those words, we wish to find the word w_s that is most appropriate for expressing a novel slang sense s, where we represent word senses by embedding their dictionary definitions into a high-dimensional vector space (see details in the next section). For a given slang sense s, a categorization model specifies a distribution over the space of candidate words based on similarities between s and existing senses of the candidate word w_j in $\mathcal{E}_j$.

We recommend a slang word choice based on the probability distribution $p(w_j|s)$ via Bayes' rule:

$$p(w_j|s) \propto p(s|w_j)p(w_j) \tag{1}$$

Here $p(s|w_j)$ is the likelihood of the novel slang sense s given the word w_j or equivalently the collective set of its existing senses E_j, and $p(w_j)$ is the prior on the candidate word. Because we constrained our analyses to words with slang senses, we used a uniform prior on the set of candidate words. We thus estimate $p(w_j|s)$ using the maximum likelihood formulation:

$$p(w_j|s) \propto p(s|w_j) = p(s|E_j) \tag{2}$$

We specify the likelihood by considering similarity relations between existing senses of the word w_j in E_j and the slang sense s. Given a set of existing senses $E_j = \{e_1, e_2, \cdots e_M\}$, we compute its similarity with the slang sense s by considering how individual exemplars in E_j are similar to s:

$$p(s|E_j) = f(s, E_j) = f(\{sim(s, e_i); e_i \in E_j\}) \tag{3}$$

We consider the specific forms of the similarity function based on three existing models of categorization: One Nearest Neighbor (1NN), Exemplar, and Prototype. We illustrate these models in Figure 2.

One Nearest Neighbor (1NN) model. Motivated by work on semantic chaining (Ramiro et al., 2018), this model predicts that a novel word sense is attached to an existing sense of a word that is closest in semantic space. We test this hypothesis in slang generation by postulating that a novel slang sense would be attached to the most similar existing sense of a word:

$$f(s, E_j) = \max_{e_i \in E_j} sim(s, e_i) \tag{4}$$

Exemplar model. Motivated by the exemplar theory (Nosofsky, 1986), this model evaluates similarities between the novel sense s and all existing senses of a word. Here we postulate that slang choice depends on the aggregated similarities of existing senses of a word to the slang sense:

$$f(s,E_j) = \sum_{e_i \in E_j} sim(s,e_i) \quad (5)$$

Prototype model. Motivated by the prototype theory (Rosch, 1975), this model predicts that category membership is established by similarity between the slang sense and a representative or prototypical existing sense:

$$f(s,E_j) = sim(s,E_j^{prototype}) \quad (6)$$

Because we do not have an accurate estimate of sense frequencies, we consider the simple version of this model where the prototypical sense is taken as the average of the existing senses, i.e., by assuming senses are equally frequent:

$$E_j^{prototype} = \frac{1}{M} \sum_{e_i \in E_j} e_i \quad (7)$$

Where M is the set size of E_j.

Similarity. To estimate individual similarities between s and e_i, we consider vector-based embeddings that transform word sense definitions into a high-dimensional vector space. We then compute the similarity as follows:

$$sim(s,e_i) = \exp(-\frac{d(s,e_i)^2}{h_s}) \quad (8)$$

Here $d(s,e_i)$ is the Euclidean distance between the vector representations of senses and h_s is a parameter controlling the degree of sense specificity that we fit to data.

Collaborative filtering

We consider an enriched version of the categorization models by taking into account parallel semantic change, cast as a variant form of collaborative filtering (Goldberg, Nichols, Oki, & Terry, 1992) that is commonly used in recommendation systems. The rationale is that words similar in existing senses may extend to label similar novel slang senses. For example, *massive* and *stellar* both refer to *large* in their existing senses, but both of them can refer to *impressiveness* in the slang context. We capture parallel semantic change by considering the influence of neighboring words to candidate words w_j's by nested likelihoods:

$$p(w_j|s) \propto \sum_{w' \in \mathcal{L}(w_j)} p(w_j,w'|s) = \sum_{w' \in \mathcal{L}(w_j)} p(w_j|w')p(w'|s) \quad (9)$$

Here $\mathcal{L}(w_j)$ indicates a small neighborhood around the word w_j in word embedding space. We estimate $p(w_j|w')$ by computing similarity between w_j and its neighboring words:

$$p(w_j|w') \propto sim(w_j,w') = \exp(-\frac{d(w_j,w')^2}{h_w}) \quad (10)$$

For the word itself, $sim(w_j,w_j) = 1$. h_w is a free parameter that controls the strength of influence from the neighbors. This nested model estimates $p(w'|s)$ using the same likelihood functions described in the previous section. The resulting collaborative filtering model effectively provides a weighted average of the likelihoods corresponding to words in the neighborhood $\mathcal{L}(w_j)$.

Materials and methods

We collected lexical data from the freely available Online Slang Dictionary (OSD; `http://onlineslangdictionary.com`) and WordNet (Miller, 1998) for novel slang and existing word sense definitions respectively. In OSD, we considered all available slang word forms with at least one available example usage. We removed words that do not exist in WordNet and extracted all word-definition pairs from the remaining words, resulting in 4,805 slang definitions from 2,357 distinct slang words. We also extracted existing definitions from WordNet by first querying the slang word and then extracting definition sentences from all retrieved *synsets*, resulting in 11,780 existing definitions. On Average, each candidate word in our dataset has 2.00 slang definitions (SD: 1.74) and 5.54 existing definitions (SD: 6.82).

We excluded acronyms because they do not extend to new senses. We removed all slang definitions containing the word 'acronym' and words that have fully capitalized spellings. Finally, we excluded slang definitions that are already part of WordNet by performing two pre-processing steps: 1) Remove a slang definition if one of the corresponding existing definitions in WordNet has at least 50% overlap in the set of content words. 2) Remove WordNet definitions that contain the token 'slang' and remove slang words that no longer have corresponding WordNet definitions. We performed a manual sanity check on 100 randomly sampled slang definitions and only 6 of them have close definitions in WordNet. After preprocessing, there are $N = 4{,}256$ slang definitions from $V = 2{,}128$ slang words. We used these words as the vocabulary for candidate slang words. We partitioned the data of sense definitions by randomly splitting into a 90% training set and a 10% test set for model evaluation.

To represent the sense definitions in a vector space, we used distributed word embeddings from fastText (Bojanowski, Grave, Joulin, & Mikolov, 2017) pretrained with subword information on 600 billion tokens from Common Crawl (`http://commoncrawl.org`). To obtain a fixed dimensional representation for the definition sentence, we take the average word embedding of all content words within the definition sentence (Landauer, Laham, & Rehder, 1997). The average pooling scheme has been shown to be a competitive sentence encoder in machine learning literature (Wieting & Kiela, 2019) and has consistently achieved better results in our experiments compared to pre-trained deep sentence encoders. We apply the same encoding method to both existing and slang definitions with no distinction. We estimated the free model parameters (h_s, h_w) using L-BFGS-B (Byrd, Lu, Nocedal, & Zhu, 1995), a quasi-newton method for bound constrained optimization, to minimize negative

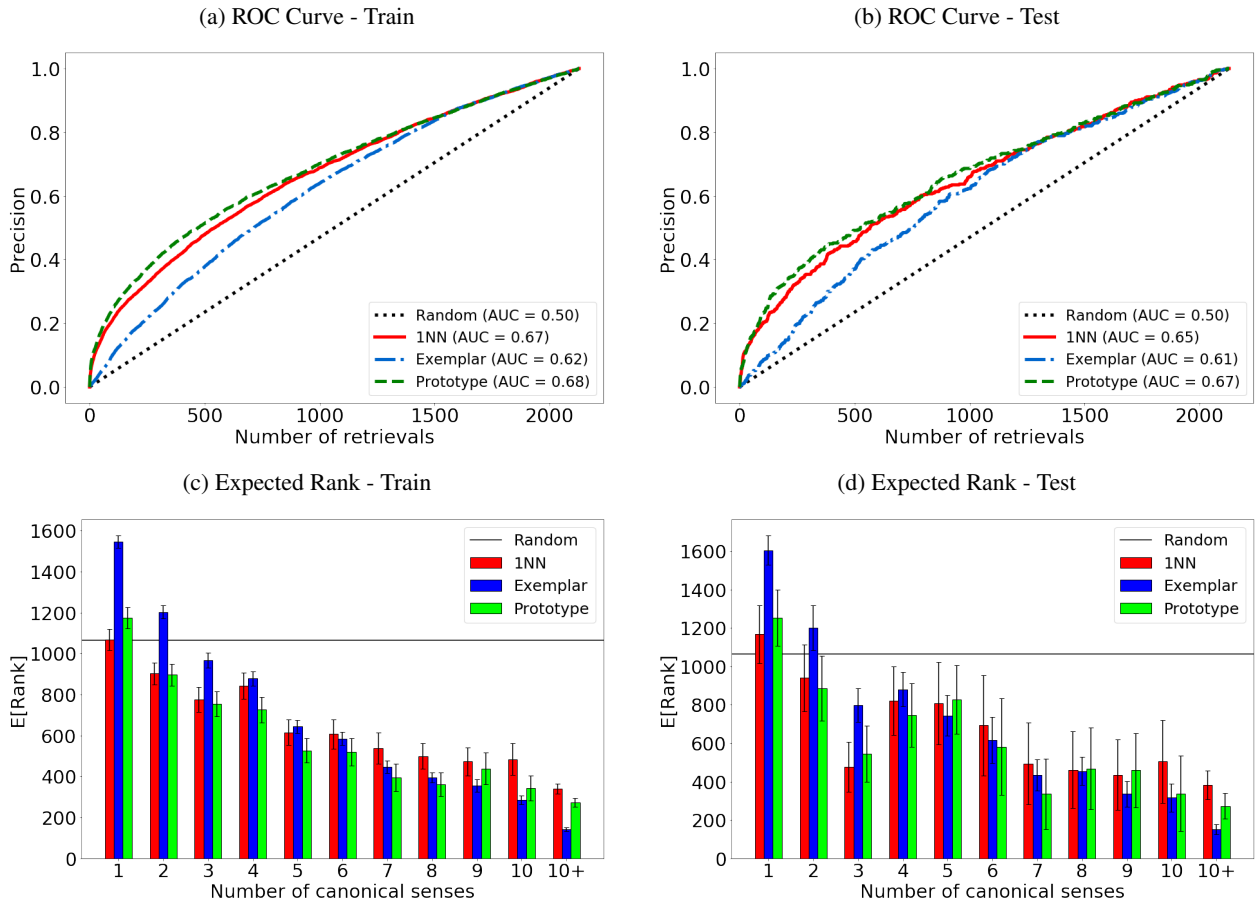

Figure 3: Top row: ROC-type curve for rank retrieval. Bottom Row: Expected Rank with respect to the number of existing senses. Ranks are computed amongst all candidate words. Whiskers denote 95% confidence intervals.

log-likelihood of the posterior:

$$\min(-\log \mathcal{L}) = \min(-\sum_{s} \log p(w_s|s)) \tag{11}$$

Here w_s is the ground truth word corresponding to the slang sense s. We estimate the free parameters on the training set while keeping them fixed in testing. For all analyzed models, we set the initial h values to 1 with bounds $[10^{-2}, 10^{2}]$. For the collaborative filtering models, both free parameters were jointly optimized.

Results

We evaluate our approach by first examining prediction of slang word choices from the three categorization models: 1NN, Exemplar, and Prototype. We then examine how collaborative filtering influences these basic categorization models on the same predictive task.

Evaluation of models of categorization

We assessed our models by ranking all candidate words according to the posterior distribution $p(w_j|s)$ from the categorization models that we described. For each slang sense definition s in the dataset, we assigned a rank to all candidate words in the vocabulary for a given model.

We first present receiver-operater curves (ROC) of model accuracy: How probable is each model to predict the correct target slang word in the first n guesses? We computed the standard Area-Under-Curve (AUC) statistics to compare cumulative precision of the models. The top row of Figure 3 shows both the ROC curves and AUC statistics of the three categorization models. All three models perform substantially better than chance. In particular, 1NN and Prototype perform better than exemplar on average in both training and testing data, which suggests that slangs are unlikely to be generated based on aggregate similarities between the existing senses and the slang sense.

Differing from previous findings on historical word sense extension where the 1NN model outperforms Prototype (Ramiro et al., 2018), we observed no substantial difference between the two models in predicting slang choices. We also considered a k-nearest-neighbor extension of the 1NN model, but we did not find any improvement in performance. We observed little difference between training and testing performances from all models, which suggests that the models did not overfit to free parameters.

For the same set of models, we also computed the expected rank of the ground-truth target words over all slang defini-

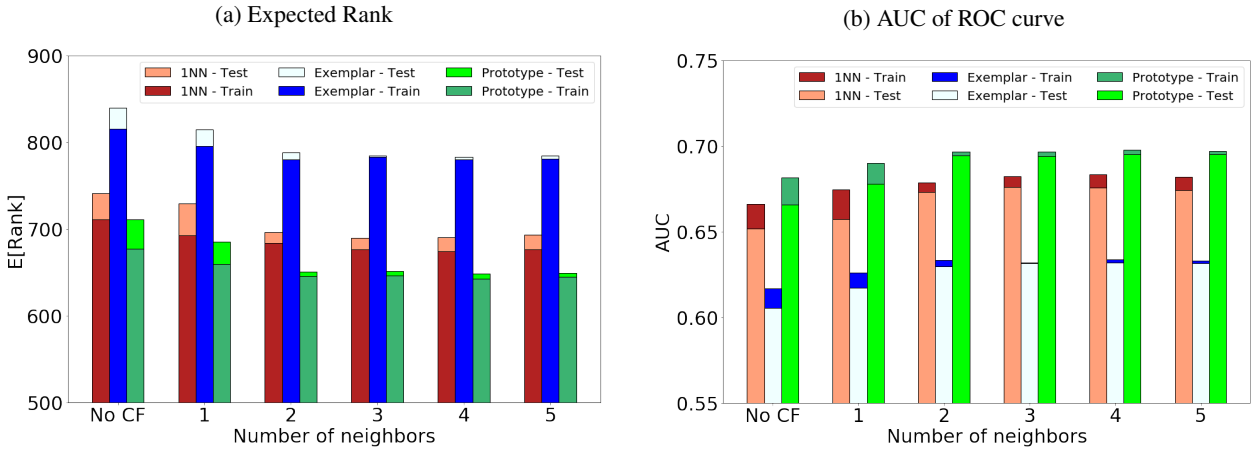

Figure 4: Summary statistics of collaborative filtering models. a): Expected Rank, b): Area-Under-Curve of ROC curves (AUC)

Table 1: Expected ranks from the categorization models.

Model	E[Rank] - Train	E[Rank] - Test
Random	1064.0	1064.0
1NN	710.89	741.29
Exemplar	815.54	839.71
Prototype	677.44	711.30

tions, based on both training and testing data. A lower expected rank indicates better predictive power. Table 1 summarizes the results. We observed similar findings with results based on AUC: All three models perform better than chance, while 1NN and Prototype models both perform better than the Exemplar model. Although these models perform above chance, the predicted expected ranks are quite high. Although some predicted words differ from the ground truth word, they may still be valid candidates for slang given sufficient social popularity. How to improve and better evaluate these model predictions will be topics of future research.

The bottom row of Figure 3 visualize the expected ranks via binning the slang definitions by degree of polysemy of their respective ground-truth candidate words w_s. We observed that all three categorization models generally perform better on more polysemous words. In particular, all three models perform better than chance when the target word has at least three existing senses. This behavior is the most prominent on the Exemplar model. Although the Exemplar model performs worse than the other two models on average, it tends to perform better on highly polysemous words. However, the Exemplar model has a natural tendency to favor those words by construction because it computes a sum of similarities instead of averaging. Both 1NN and Prototype also perform better as the number of existing senses increases. With more existing senses, it is more likely for one of them to have a close match with the slang sense, thus the improvement on 1NN. The prototypical senses would also become more accurate due to a larger sample for estimation. Compared to 1NN, the Prototype model performs slightly worse when the target word has few senses, but it outperforms 1NN as the degree of polysemy increases.

In sum, these results show that slang word choices are predictable without considering external social factors and provide evidence that simple models of categorization can capture non-arbitrariness in the generative processes of slang.

We provide examples of model success and failure in Table 2. In the *wicked* example, our models captured polarity shift in slang generation, indicated by low expected ranks from all models. The second example shows how our model can have limited predictability when the slang and existing senses are cognitively distant. In both examples, the Exemplar model consistently gave low ranks to candidate words *broken*, *play*, and *cut* because they are some of the most polysemous words in our vocabulary with more than 50 existing senses each.

Evaluation of collaborative filtering

We next examined the influence of collaborative filtering on each of the three categorization models. For each model, we considered variants of these models with up to five neighboring words.

Figure 4 summarizes the results. All collaboratively filtered models achieve better AUC and expected rank on both the training set and testing set compared to their respective basic categorization models. The improvement is most prominent on the test set, lowering expected rank by more than 50 and improving AUC by over two percent for all three models. In particular, collaborative filtering improved model prediction most substantially when two closest neighboring words were considered. Consideration of more neighbors did not improve model prediction further, suggesting that information about slang word choice is sufficiently encapsulated in a small set of neighboring words.

Table 3 illustrates collaborative filtering with two examples. In both cases, the basic categorization models perform poorly because existing senses of the ground-truth words do not have strong similarity with the slang senses. The neighboring words however, contain senses that are more rele-

Ground truth target word [w]:	*wicked*
Slang sense in OSD [s]:	impressive.
Corresponding WordNet senses [E]:	(1) morally bad in principle or practice; (2) having committed unrighteous acts; (3) intensely or extremely bad or unpleasant in degree or quality; (4) naughtily or annoyingly playful; (5) highly offensive; arousing aversion or disgust.
Model expected rankings [E(Rank)]:	(1NN): 93/2128; (Exemplar): 369/2128; (Prototype): 33/2128
Top ranked words:	(1NN): *bonzer, spot, point, tall, grand*; (Exemplar): *broken, play, cut, point, heavy*; (Prototype): *bonzer, good, tall, grand, hot*
Ground truth target word [w]:	*breezy*
Slang sense in OSD [s]:	an unimportant girlfriend or girlfriend on the side.
Corresponding WordNet senses [E]:	(1) fresh and animated; (2) abounding in or exposed to the wind or breezes.
Model expected rankings [E(Rank)]:	(1NN): 1977/2128; (Exemplar): 1829/2128; (Prototype): 1762/2128
Top ranked words:	(1NN): *man, buddy, pal, beard, associate*; (Exemplar): *broken, play, cut, run, line*; (Prototype): *front, mate, face, joker, associate*

Table 2: Examples of model success and failure.

Ground truth target word [w]:	*icky*
Slang sense in OSD [s]:	gross, unappealing.
Corresponding WordNet senses [E]:	(1) very bad; (2) soft and sticky.
5 neighboring words used in collaborative filtering [$\mathcal{L}(w)$]:	*yucky, nasty, stinky, freaky, dirty*
Ground truth target word [w]:	*scary*
Slang sense in OSD [s]:	ugly, weird.
Corresponding WordNet senses [E]:	provoking fear terror.
5 neighboring words used in collaborative filtering [$\mathcal{L}(w)$]:	*freaky, crazy, nightmare, awesome, stupid*

Table 3: Examples that illustrate how collaborative filtering helps predicting slang word choice.

vant to the probe slang sense, hence informing the model better about the ground-truth words. We also observed that the neighboring words used in collaborative filtering have strong semantic correlations, which explains the diminishing effect in performance when introducing additional neighboring words.

Conclusion

We have presented slang generation as a categorization problem. Our formulation relies on few free parameters and sheds light on the cognitive processes that give rise to slang word choice. Although the full slang generation processes are beyond the models we have explored, our framework was able to capture substantial predictability without explicitly modeling external social variables. Furthermore, we incorporated parallel semantic change in slang generation using collaborative filtering and found that it improves slang prediction beyond the basic categorization models. Future work should explore richer semantic representations of slang and extend the current framework to novel slang word forms.

Acknowledgments

We thank members of the Language, Cognition, and Computation (LCC) Group at the University of Toronto for their thoughtful feedback, particularly Suzanne Stevenson, Barend Beekhuizen, and Renato Ferreira Pinto Junior. This research is supported by an NSERC DG grant and a Connaught New Researcher Award to YX.

References

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, *18*, 135–160.

Blodgett, S. L., Green, L., & O'Connor, B. (2016). Demographic dialectal variation in social media: A case study of african-american english. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 1119–1130). Association for Computational Linguistics.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Byrd, R., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*, 1190-1208.

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2010). A latent variable model for geographic lexical variation. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 1277–1287). Association for Computational Linguistics.

Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Commun. ACM*, *35*, 61–70.

Kao, J. T., Bergen, L., & Goodman, N. D. (2014). Formalizing the pragmatics of metaphor understanding. In *Proceedings of the 36th annual conference of the cognitive science society* (pp. 719–724). Cognitive Science Society.

Kao, J. T., Wu, J. Y., Bergen, L., & Goodman, N. D. (2014). Nonliteral understanding of number words. *Proceedings of the National Academy of Sciences*, *111*, 12002–12007.

Kulkarni, V., & Wang, W. Y. (2018). Simple models for word

formation in slang. In *Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 1424–1434). ACL.

Labov, W. (1972). *Language in the inner city: Studies in the black english vernacular*. University of Pennsylvania Press.

Labov, W. (2006). *The social stratification of english in new york city*. Cambridge University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: what categories reveal about the mind*. University of Chicago Press.

Landauer, T., Laham, D., & Rehder, R. (1997). How well can passage meaning be derived without using word order? a comparison of latent semantic analysis and humans. In *Proceedings of the 19th annual conference of the cognitive science society* (pp. 412–417). Cognitive Science Society.

Lehrer, A. (1985). The influence of semantic fields on semantic change. *Historical Semantics: Historical Word Formation*, *29*, 283–296.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language*, *40*, 230–262.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*, 176–182.

Miller, G. (1998). *Wordnet: An electronic lexical database*. MIT press.

Millhauser, M. (1952). The case against slang. *The English Journal*, *41*, 306–309.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Ramiro, C., Srinivasan, M., Malt, B. C., & Xu, Y. (2018). Algorithms in the historical emergence of word senses. *Proceedings of the National Academy of Sciences*, *115*, 2323–2328.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192–233.

Stewart, I., & Eisenstein, J. (2018). Making "fetch" happen: The influence of social and linguistic context on nonstandard word growth and decline. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 4360–4370). Association for Computational Linguistics.

Wieting, J., & Kiela, D. (2019). No training required: Exploring random encoders for sentence classification. In *International conference on learning representations.*

Xu, Y., & Kemp, C. (2015). A computational evaluation of two laws of semantic change. In *Proceedings of the 37th annual conference of the cognitive science society* (pp. 2703–2708). Cognitive Science Society.