Local Processes of Homophone Acquisition

Deniz Beser
Department of Computer Science
University of Pennsylvania
dbeser@seas.upenn.edu

Spencer Caplan
Department of Linguistics
University of Pennsylvania
spcaplan@sas.upenn.edu

Abstract

The Naïve Generalization Model (NGM) (Caplan, 2018) explains word learning phenomena as grounded in the local, dynamical process of category formation. A range of experimental evidence (Xu and Tenenbaum, 2007; Spencer et al., 2011; Lewis and Frank, 2018) supports the NGM over prior models of word learning such as Bayesian inference (Xu and Tenenbaum, 2007). Despite such progress, a number of theoretical phenomena remain unaddressed by previous accounts. In this paper, we present a novel extension to NGM which offers a strong fit to and explanation of experimental data on homophone acquisition (Dautriche et al., 2016).

1 Introduction

While the mutual-exclusivity constraint in word learning (Markman and Wachtel, 1988) demonstrates that concepts correspond to only a single word form, in the case of homophony, a single pronunciation can map to multiple distinct words each with their own meaning. Thus when encountering a novel label like ‘bat’ (which might refer to both a type of mammal as well as a tool), the learner must correctly determine whether this label corresponds to a pair of distinct homophones rather than a single word with a very broad meaning. Dautriche and Chemla (2016) and Dautriche et al. (2016) used a variant of the Xu and Tenenbaum (2007) word learning paradigm to demonstrate that participants are more likely to posit homophony - as opposed to a single broad generalization - as the largest semantic distance between attested exemplars increases. I.e. training items sampled from a bimodal distribution are more likely to be acquired as homophones compared to learning over a unimodal distribution (Figure 2). This effect is consistent over both artificially generated and naturally occuring stimuli, suggesting that it is driven by the category formation process of word learning rather than a top-down effect of prior world knowledge (Dautriche and Chemla, 2016).

2 Model

2.1 The Naïve Generalization Model

The Naïve Generalization Model (NGM) (Caplan, 2018) offers an explanation of word learning phenomena grounded in category formation (Smith and Medin, 1981). The model explains the mechanism by which hearing novel words invites a learner to create a new category from component ‘features’. Such ‘features’ can simply be thought of as properties that hold for some item or any component piece of a semantic representation. While any two properties will be equally true of an object, in the sense that they are formal operators, it should be clear intuitively that some properties are more salient than others in context.

Consider the number 73. It is probably easier to determine that 73 is odd than it is to determine that 73 is a prime; its not that its prime-ness is less valid than its being odd, rather it is simply a matter of salience (i.e. how noticeable it is to an average person quickly). The NGM captures this intuition by encoding mental representations of words as a vector of features whose magnitudes are proportional to salience sampled from learning experience.

2.2 The Extension

During training, initial exposure to a novel phonetic label triggers a category formation process; the model generates a vector representation of the training item. At all subsequent instances of learning, the learner evaluates whether newly exemplars are a member of the the hypothesized category. Encountering a labeled exemplar which is judged inconsistent with the extant meaning hy-

Proceedings of the Society for Computation in Linguistics (SCiL) 2019, pages 326-329.
New York City, New York, January 3-6, 2019

pothesis triggers the generation of a competitor meaning hypothesis. In the original NGM, each phonetic label was a unique identifier of a word, which mapped to a single category, thus a only one competitor hypothesis could 'win'. In the present extension, however, an edit distance is computed between hypotheses for any learning trials which result in multiple possible meanings: If the representations are sufficiently close with respect to a parameter threshold, they are merged into a single vector, which corresponds to a broad generalization. However, hypotheses with a larger edit-distance are kept distinct, resulting in homophony (visualized in Figure 1).

2.3 Vector Representation

The vector representations are kept consistent with Caplan (2018). Training items are a randomly ordered sequence of vectors with labels. These items are designed in a hierarchical fashion, such that items contain features of items superior to them in the taxonomical tree (e.g. a dog is also an animal). While the world includes hierarchical relations between categories, a learner doesn't necessarily have direct access to that. The model, therefore, does not have access to this information of hierarchy and processes the features as a flat distribution, which ensures that the model does not trivially use this property to generate taxonomy. During learning, the model operates on the features to generate representations for labels by sampling a salience value between 0 and 1 from a Gaussian distribution with mean and deviation that are stipulated as model parameters. Then, to represent the learned meaning, the model generates a vector that contains each one of these features with the calculated salience values. This variability of salience of features is integral as it allows the system to have a non-monotonic functionality.

2.4 Vector Evaluation

When the model model encounters training items with the same label, it postulates homophony by checking whether the new exemplar and the previously learner meaning are consistent. To do this, the model evaluates the distance between the two representations, which is measured as the normalized sum of the differences between salience values of features of the representations. The model sums up the salience values of features that exist only in one of the representations, and difference of salience values of features that exist in both of the representations. Then, the sum is normalized with the maximum number of features, to the range 0-1. This process of consistency evaluation is visualized in Figure 1.

3 Evaluation

3.1 Previous Experimental Background

Experiments on adults have investigated how a single word form can map onto multiple meanings (Dautriche and Chemla, 2016; Dautriche et al., 2016). For instance, the homophone 'bat' can refer to both a baseball bat and the animal species. While in principal the ability to acquire homophones is unbounded, in practice it must be subject to important constraints. Specifically, Dautriche et al. (2016) hypothesized that people "refer to convex concepts and form lexical representations that follow this constraint, in essence showing early awareness that homophony is a possibility in natural languages." To test this, they independently provided the subjects with labeled items with two different distributions, uniform and bimodal. In the uniform distribution, the presented items are distributed evenly across the conceptual space. In the bimodal distribution, the items are unevenly distributed such that there are two clusters of items in the conceptual space. After training, the participants are asked to whether the label they just learned also applies to a series of novel items from various position in the conceptual space, i.e. if a new item also fell into the same category as the previously observed items. The findings suggest that people form conceptual spaces that match the experimental conditions and are less likely to extend a label to a new item that is not within the conceptual space.

Distribution	Semantic Gap	Outcome
Uniform	Small	Broad generalization
Bimodal	Large	Homophony

Table 1: Summary of the learner behavior

3.2 Results

Evaluation of the NGM extension was performed by replicating the stimuli and the testing procedure from Dautriche and Chemla (2016) and Dautriche et al. (2016). The model is fed in a set of word-label pairs which are either uniformly

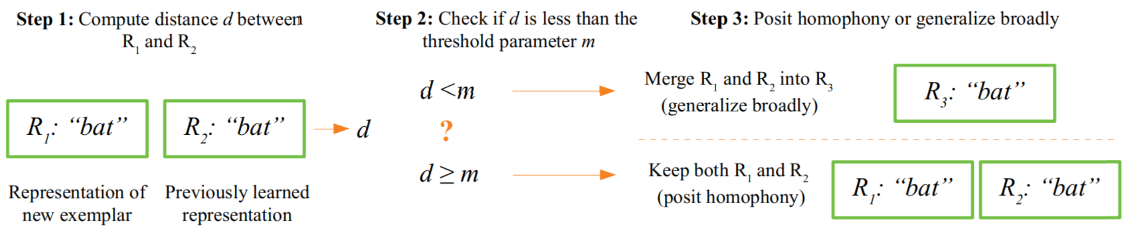

Figure 1: Flow chart of the present extension to NGM. The model compares a new exemplar and an already learned representation. If the distance between the representations is smaller than the threshold parameter, a single broad representation is generated. Otherwise, both representations are kept, indicating homophony.

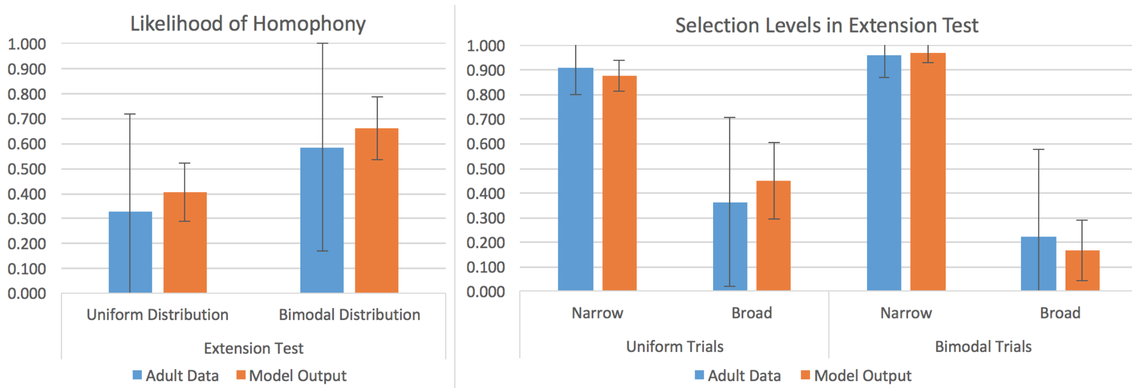

Figure 2: Comparison of adult data and model output. The model captures the increase in likelihood of positing homophony across trials; the separation of exemplars in conceptual space encourages positing homophony. Error bars show standard deviation.

or bimodally distributed. Then, the model's level of generalization is tested on novel words. We find that, matching human performance, the likelihood of positing homophony rather than a single broad meaning increases as a function of semantic distance independent of parameter value (Figure 3). Additionally, hyperparameter optimization using step-wise grid-search offers a strong quantitative fit between human empirical performance and model output (Figure 2). The average absolute difference between the model and experimental output is 0.057 and all of the model output scores are within a single standard deviation of the empirical finding.

4 Discussion

In this paper, we presented an extension to the Naïve Generalization Model, a model which explains word learning phenomena as grounded in the local, dynamical process of category formation, to a homophone acquisition. On the contrary to the Bayesian inference models, this model does not assume that hypotheses for the meanings of words exist a priori; it instead forms representation of meanings as it processes input. Based on training samples, the model creates representation that model mental states, and iteratively alters these representations with new information. On our view, learning of homophones, like word learning in general (Caplan, 2018) , is a dynamical, yet mechanistic process in which learning is driven by local computation rather than any global probability maximization.

References

Spencer Caplan. 2018. Word learning as category formation. *Proceedings of the Society for Computation in Linguistics*, 1(1):174–177.

Isabelle Dautriche and Emmanuel Chemla. 2016. What homophones say about words. *PloS one*, 11(9):e0162176.

Isabelle Dautriche, Emmanuel Chemla, and Anne Christophe. 2016. Word learning: homophony and

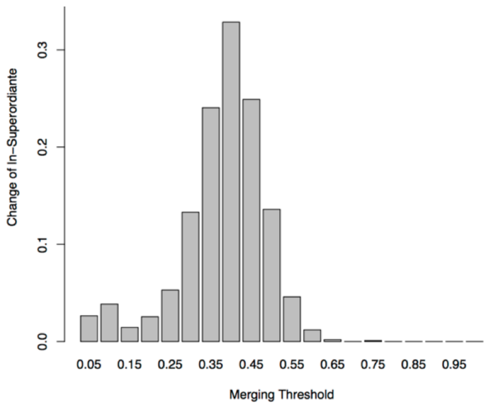

Figure 3: Parameter cutoff on x-axis, rate of broad generalization on y-axis. Independent of the parameter, the model captures the difference in selection of broadly positioned test items across uniform and bimodal distributions, supporting that the models learned conceptual space is disjoint in the bimodal trials and the likelihood of positing homophony is increased (Dautriche et al., 2016).

the distribution of learning exemplars. *Language Learning and Development*, 12(3):231–251.

Molly Lewis and Michael C Frank. 2018. Still suspicious: The suspicious coincidence effect revisited. *OpenLewis, Molly, and Michael C Frank.Still Suspicious: The Suspicious Coincidence Effect Revisited. PsyArXiv*, 9.

Ellen M Markman and Gwyn F Wachtel. 1988. Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, 20(2):121–157.

Edward E Smith and Douglas L Medin. 1981. *Categories and concepts*. Harvard University Press Cambridge, MA.

John P Spencer, Sammy Perone, Linda B Smith, and Larissa K Samuelson. 2011. Learning words in space and time: Probing the mechanisms behind the suspicious-coincidence effect. *Psychological science*, 22(8):1049–1057.

Fei Xu and Joshua B Tenenbaum. 2007. Word learning as bayesian inference. *Psychological review*, 114(2):245.