How nouns surface as verbs: Inference and generation in word class conversion

Lei Yu (jadeleiyu@cs.toronto.edu)
Department of Computer Science
University of Toronto

Lana El Sanyoura (lana.elsanyoura@mail.utoronto.ca) **Yang Xu (yangxu@cs.toronto.edu)**
Department of Computer Science, Cognitive Science Program
University of Toronto

Abstract

Word class conversion refers to the extended use of a word from one grammatical class to another without overt morphological marking. Noun-to-verb conversion, or denominalization, is one form of word class conversion studied extensively in the literature. Previous work has suggested that novel denominal verb usages are comprehensible if the listener can compute the intended meaning based on shared knowledge with the speaker. However, no existing work has explored the computational mechanism under this proposal. We propose a frame-semantic generative model, *Noun2Verb*, that supports the inference and generation of novel denominal verb usages via semi-supervised learning. We evaluate this framework in a dataset of denominal verbs drawn from adults and children against a state-of-the-art model from natural language processing. Our results show that *Noun2Verb* aligns better with human interpretation and bridges the gap between machines and humans in lexical innovation.

Keywords: word class conversion; denominal verb; frame semantics; lexical innovation; generative model

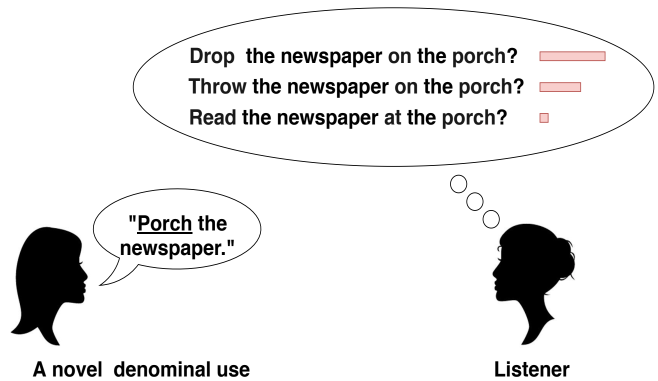


Figure 1: An example of noun-to-verb word class conversion. Given a novel denominal usage of the noun *porch*, the listener infers the meaning of that utterance.

The problem of word class conversion

Word class conversion refers to the extended use of a word from one grammatical class to another without overt morphological marking (Baeskow, 2006). Noun-to-verb conversion, or denominalization, is one common form of word class conversion. For instance, consider the denominal verb *to google*, which refers to using the Google engine to research the topic at hand. The phrase *to google* therefore extends the noun *Google* into a verb. Innovative denominal verb usages like this are not only attested in adult speech but also in child language. For example, data from MacWhinney (2014) show denominal usages such as “bee my cereal” (meaning “add honey in my cereal”) in young children, and similar lexical innovations have been reported in child speakers of English, German and French (Clark, 1982). Although denominal verbs are an extensively studied linguistic subject (Jespersen, 2013; Clark & Clark, 1979), the computational mechanism that supports the inference and generation for novel denominal verb usages is an under-explored area that we pursue in this work.

Clark and Clark (1979) proposed that “the innovative denominal verb convention” is established if the intended meaning of a denominal verb can be computed from context. They suggested that the successful comprehension of a (novel) denominal verb usage relies on the fact that the speaker denotes the kind of state, event, or process that she believes the listener can readily and uniquely compute on the basis of their mutual knowledge. They illustrated this idea with the example “the boy porched the newspaper” (see also Figure 1). Upon hearing this sentence, the listener is expected to identify the scenario of a boy delivering the newspaper onto a porch, based on the shared knowledge about entities in the utterance: the boy, the porch, and the newspaper. Recent studies offered evidence that children may also take this convention as a guide for the context in which lexical innovations should be used (Lippeveld, 2013). Similarly, computational work in historical word class conversion suggested that verbs derived from nouns tend to be more semantically specific than their parent nouns (Kisselew, Rimell, Palmer, & Padó, 2016).

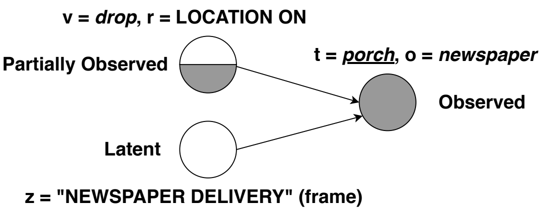
The empirical literature has provided important clues as to “when” noun-to-verb conversion occurs in communication (Clark & Clark, 1979), but not so much “how” the meaning of a denominal verb can be correctly inferred or what forms of semantic representations and mechanisms are required for noun-to-verb conversions. One piece of evidence for the lack of such a formal account is that despite recent progress in computational models for natural language understanding, even the state-of-the-art neural language models often failed to differentiate between the novel usage of words from their conventional meaning (Iacobacci, Pilehvar, & Navigli, 2016). Here we explore the potential of developing generative machine algorithms that support human-like lexical innovation.

©2020 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

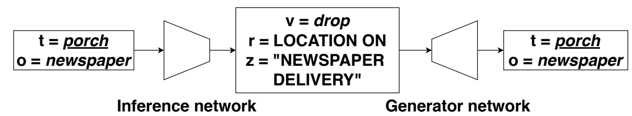
We illustrate our approach in Figure 1. Here the listener infers the intended meaning of a novel denominal usage, operationalized as a distribution over possible interpretations of the utterance from the speaker. We consider listener interpretations that can be decomposed into a paraphrase verb, e.g., *drop*, and a prepositional phrase that describes the semantic relation between the verb and the noun context, e.g., location *on*. One goal of our proposed framework is to learn how to automatically infer the meaning of a novel denominal verb usage by rephrasing it with a canonical verb and an appropriate prepositional phrase—this is the *inference* problem. We also consider the inverse problem of *generation*, where the aim is to automatically extend a canonical noun (e.g., *porch*) to novel instances of denominal usage given the intended meaning, made up of a clue verb and a prepositional phrase.

Our framework is inspired partly by the Generative Lexicon theory (Pustejovsky, 1991), where it was proposed that meanings of a noun can be represented by a set of relations referred to as *qualia* that encode information relating to hidden events and activities associated with the word. The approach we take will support choosing the most probable role(s) under a linguistic context over all possible qualia roles. Importantly our work is also rooted in the earlier theory of frame semantics (Fillmore, 1968; Gropen, Pinker, Hollander, & Goldberg, 1991). As Barsalou (1992) suggested, frame-like representational components provide a powerful productive apparatus for generating highly flexible concepts within a field.

We contribute a semi-supervised framework that learns latent semantic frames in verb denomination which we show to improve both understanding and production of novel denominal usages. The inference of semantic relations between converted verbs and parent nouns also bears resemblance to the line of NLP research on noun-compound classification, where computational models are trained to determine the semantic relation that holds between the constituents of a noun-compound (Shwartz & Dagan, 2018). However, few studies have focused on automatic inference for relations in noun-verb compounds. Our work goes beyond these classification tasks and considers more cognitively grounded generation problems for innovative lexical usage.



(a) Graphical illustration of N2V



(b) Illustrative diagram of the reconstruction game.

Figure 2: (a) A graphical illustration of the proposed *Noun2Verb* (*N2V*) model. The observed variable refers to a set of denominal verb usages (e.g., "porch the newspaper"). The partially observed variable refers to the fact that only a proportion of the denominal cases would have human-annotated paraphrases (e.g., "drop the newspaper on the porch"), whereas a greater proportion would have no annotation which the model needs to learn. The latent variable captures implicit frames associated with denominal usages (e.g., "NEWSPAPER DELIVERY") which the model marginalizes out. (b) An illustration of the semi-supervised learning paradigm based on the reconstruction game, equivalent to a variational Bayes autoencoder. The networks are provided with only partially annotated denominal usage data, but they learn to interpret the unannotated data by encoding the variables v, r, and z and minimizing the difference between outputs from the generator network (reconstruction) and inputs to the inference network.

Computational formulation

We first formalize noun-to-verb conversion as a probabilistic generative model based on a listener (or inferencer) and a generator. We then describe how this framework can be effectively learned through a reconstruction game between the inferencer and the generator via semi-supervised learning.

Variable definition. The listener hears a *query utterance* $u_q = \{t, o\}$ consisting of a denominal verb (which we call *target word*) t and its object o (e.g., "porch the newspaper") from the generator. The listener then tries to interpret this query utterance by inferring the following semantic roles: 1) a *paraphrase verb* v representing the novel meaning of converted noun t (e.g., *drop* for *porch*); 2) a prepositional *semantic relation* r between v and t (e.g., "location on" between *drop* and *porch*); 3) a latent *frame variable* z specifying the topic, or semantic frame of the scenario conveyed by the query utterance. Inference of the query utterance will take the form of a *paraphrase utterance* $u_p = \{v, r, t, o\}$ that construes the meaning of the denominal usage. We consider 8 semantic relational types drawn mainly from work by Clark and Clark (1979). As we shall see, the frame variable z is essential to our generative framework, because it allows the generator to produce novel denominal verbs across different scenarios.

We present the *Noun2Verb* (abbreviated as *N2V*) model to simulate both the inference and generation of denominal verbs. We consider the five semantic roles described above as random variables that fall into three groups: observed variables $\{t, o\}$ of denominal usage, partially observed variables $\{v, r\}$ (as we only consider a small proportion of denominal usages with human annotations; most denominal usages we consider do not have annotations), and latent variable $\{z\}$, as illustrated in Figure 2a. Inspired by frame semantics, our model handles cases where the main verbs associated with the frames are *null-instantiated* (Ruppenhofer & Michaelis,

2014) – they are omitted and left for the listener to infer.

Generation. In the generation process, the generator produces a denominal verb usage by sampling a frame z and a verb-relation pair $\{v, r\}$ from a prior distribution, and then drawing the sample target word and its object from the following conditional probability:

$$p_{\text{gen}}(u_{\text{q}}) = \sum_{z,u_{\text{p}}} p_{\text{prior}}(u_{\text{p}}) p_{\text{prior}}(z) p_{\text{gen}}(u_{\text{q}}|u_{\text{p}}, z) \quad (1)$$

where we set $p_{\text{prior}}(u_{\text{p}})$ to be uniform for both verb and relation, and $p_{\text{prior}}(z) = \mathcal{N}(0,1)$ as a standardized normal. The sampled relation-verb pairs are expressed as vectors via pre-trained word2vec embeddings that become the input of a multi-layer perceptron (MLP), which independently draws variables $u_{\text{q}} = (t, o)$ from the conditional distribution:

$$p_{\text{gen}}(u_{\text{q}}|u_{\text{p}}, z) = \pi_o(\sigma(f_o(u_{\text{p}}))) * \pi_t(\sigma(f_t(u_{\text{p}}))) \quad (2)$$

where $\pi_i(\cdot)$ are categorical (or multinomial) distributions, $\sigma(\cdot)$ denotes the softmax function, and $f_i(\cdot)$ stands for non-linear functions parametrzied by MLPs.

Inference. In the inference process, the listener performs probabilistic inference over a set of candidate paraphrase verbs v and semantic relations r, given a query utterance $u_{\text{q}} = \{t, o\}$. We model this as conditional probability $p(v, r, z|t, o)$ where the latent variable z is then marginalized out:

$$p_{\text{inf}}(u_{\text{p}}|u_{\text{q}}) = \sum_{z} p_{\text{inf}}(u_{\text{p}}, z|u_{\text{q}}) \quad (3)$$

Similar to generation, the listener network consisting of three MLPs $g_i(\cdot)$ computes the factorized distribution over the candidate verbs, relations and frames. The only difference is that the posterior of z is modeled by a normal distribution, mean and variance of which are computed by MLPs:

$$\begin{aligned} & p_{\text{inf}}(u_{\text{p}}, z|u_{\text{q}}) \\ & = \pi_v(\sigma(f_v(u_{\text{q}}))) * \pi_r(\sigma(f_r(u_{\text{q}}))) * \mathcal{N}(f_{z_1}(u_{\text{q}}), f_{z_2}(u_{\text{q}})) \end{aligned} \quad (4)$$

where $\pi_i(\cdot)$, $\sigma(\cdot)$, $f_i(\cdot)$ are defined the same way as those in the generation model.

Inference and generation via semi-supervised learning. The parameters of the two neural networks described form a large hypothesis space for model selection. To find optimal configurations, it is possible to train two groups of MLPs by maximizing log-likelihood independently in the generation and inference tasks (Equations 2 and 4). However, this schedule treats the networks separately from denominal and paraphrased utterances produced by human annotators, hence it is unable to capture the interaction between the inferencer and the generator. We consider an alternative approach through a "reconstruction game": first, we provide the listener with a denominal utterance, and force it to "think out loud" the paraphrase by sampling a verb-relation pair, which serves as the input to generator network. The generator then tries to recover the denominal utterance by sampling a set of semantic components u_{o}. The distance between the original utterance and the reconstructed utterance is taken as the loss function. Since both the inferencer and the generator contribute to the outcome of the game, when applying the standard back-propagation algorithm, the two neural networks can be trained simultaneously. The overall learning process consists of reconstructing query utterances, in conjunction with learning on human-paraphrased examples alternately—a paradigm known as *semi-supervised learning* from machine learning.

Let $X_u = \{u_{\text{q}}^{(i)}\}_{i=1}^{m}$ denote the set of query utterances, and $X_s = \{u_{\text{q}}^{(j)}, u_{\text{p}}^{(j)}\}_{j=1}^{n}$ for those with paraphrase utterances from human annotations, we minimize the following loss function J:

$$J = \mathcal{L} + \mathcal{U} \quad (5)$$

where the supervised loss $\mathcal{L}$ in Equation (5) refers to:

$$\mathcal{L} = \sum_{u_{\text{p}}^{(j)}, u_{\text{q}}^{(j)} \in X_s} \log p_{\text{gen}}(u_{\text{q}}) + \log p_{\text{inf}}(u_{\text{p}}|u_{\text{q}}) \quad (6)$$

and the unsupervised loss $\mathcal{U}$ in Equation (5) refers to:

$$\mathcal{U} = \sum_{u_{\text{q}}^{(i)} \in X_u} \mathbb{E}_{\text{inf}}[\log p_{\text{gen}}(u_{\text{q}}) - \log p_{\text{inf}}(u_{\text{p}}|u_{\text{q}})] \quad (7)$$

$\mathbb{E}_{\text{inf}}[\cdot]$ in Equation 7 refers to sampling u_{p} from $p_{\text{inf}}(u_{\text{p}}|u_{\text{q}})$ to compute the entire loss multiple times and take the expected loss. Therefore, minimizing the loss $\mathcal{J}$ Equation 5 is equivalent to minimizing both supervised loss $\mathcal{L}$ and unsupervised loss $\mathcal{U}$. We illustrate the reconstruction game and the two networks in Figure 2b. This learning paradigm is equivalent to *variational Bayes* autoencoder in deep learning (Kingma & Welling, 2013) used widely to generate complex data.

Data

We collected data from four sources: 1) denominal utterances from adults; 2) denominal utterances from children; 3) augmented denominal utterances from a web corpus, and 4) paraphrased utterances, or human annotations, via crowdsourcing.

Meta dataset of denominal verb usages. We extracted pairs of denominal usages and paraphrases from 1) a list of query utterances produced by adults from Clark and Clark (1979), and 2) a similar set from child speech reported in Clark (2004). For these cases, each denominal utterance has already been annotated in terms of the semantic relation by the authors, but no paraphrased verbs are available. To obtain ground-truth paraphrases, we searched for the top-3 verbs that co-occur most frequently with every query utterance on the large, comprehensive iWeb 2015 corpus (`https://corpus.byu.edu/iweb/`). We performed searches through the Sketch Engine online corpus tool (`https://www.sketchengine.eu`) via regular-expression queries. We obtained 786 query-paraphrase pairs

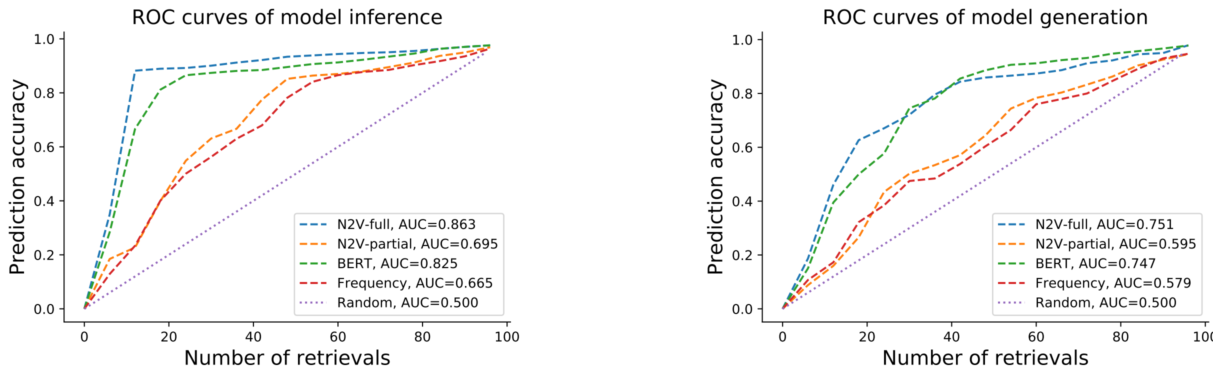


Figure 3: Receiver operating characteristic (ROC) curves that summarize model performance in inference and generation. "N2V-full" is the proposed Noun2Verb word class conversion model; "N2V-partial" is the same model yet without the latent frame variable z; "BERT" is a state-of-the-art natural language processing model; "Frequency" and "Random" refer to frequency-based and random baseline models. "AUC" refers to area under the curve: a better model has a higher AUC.

from adult data (denoted by U_{adult}), and 32 examples from children (denoted by U_{children}).

Augmented data from WordNet. While a small proportion (about 5%) of denominal verb usages in our data has paraphrases, a greater proportion lacks such information. Furthermore, we expect our model to be able to interpret novel utterances by generalizing over learned examples: if the model is told that "send the resume via email" is the right interpretation of "email the resume", then on hearing a similar utterance like "mail the package", it should generalize and understand that utterance also has something to do with the transportation frame. We therefore obtained a set of new query utterances by replacing target noun of each query example described in the previous section with a semantically related noun. We took the taxonomy from WordNet and extract all synonyms of each target from the same synset as substitutes. This yielded 1,129 queries (denoted by U_{aug}) excluding targets without synonyms.

Ground-truth human annotations of paraphrases. We also collected human paraphrases for a small set of query utterances via Amazon Mechanical Turks (AMT) crowdsourcing marketplace for model evaluation. We chose 132 query utterances (denoted by U_{eval}) – 100 of which came from U_{adult} and 32 of which from U_{children} – and collected their human annotations. Each query has a paraphrased utterance with the top-3 paraphrase verbs collected from the iWeb corpora, and the participants were asked to choose, among the three candidates, all verbs that serve as good paraphrases for the target word in the query utterance. If none of them is appropriate, then the participants must provide a good alternative paraphrase verb by themselves. For each query utterance, about 15 to 20 responses were collected. We found that for over 85% cases the participants considered at least one candidate verb as a good paraphrase, suggesting that our iWeb-based approach to bootstrap paraphrases is reasonably accurate.

Experiments and results

Model training. We trained both the inferencer and generator networks based on data from U_{adult}, U_{children} and U_{aug}, and we evaluated the model on novel instances of denominal usage in U_{eval} (discussed later). We used Pyro (Bonawitz, Denison, Griffiths, & Gopnik, 2014), a probabilistic programming language, to implement our generative model. Optimizing the objective in Equation 5 was performed by stochastic gradient descent via Adam optimizer (Kingma & Ba, 2014), with learning rate $\alpha = 0.001$ and $\{\beta_1, \beta_2\} = \{0.9, 0.999\}$.

We compared our model against a strong alternative, state-of-the-art language model named BERT (Devlin, Chang, Lee, & Toutanova, 2018). BERT is a neural language model pre-trained on a large set of text-based sequential prediction tasks, and it has yielded human-level performance over 11 natural language understanding problems. To adapt BERT to denominal verbs, we fine-tuned the BERT model (with 12 hidden layers and dimensionality of 768) following standard supervised learning paradigm: the listener takes embedded $\{u_{\text{q}}\}$ as input, and outputs categorical probabilities for each element in u_{p}. The generator model computes the reversed probability $p(u_{\text{q}}|u_{\text{p}})$. Our *Noun2Verb* model was trained on both paraphrased utterances U_{adult}, U_{children} and non-paraphrased utterances U_{aug} through the reconstruction game described. The BERT model, which cannot learn denominal verbs without fully labelled paraphrases, was trained only on U_{adult} and U_{children}. To demonstrate the effectiveness of the augmented data, we also trained the *Noun2Verb* model without the reconstruction game: both inferencer and generator learned separately on paraphrased utterances U_{adult}, U_{children}. We denote this model as *N2V*-partial, to distinguish it from the full model *N2V*-full. We also considered a frequency-based baseline and a random model to verify if the proposed model performs above chance. Each model takes word vectors encoded by 100-dimensional GloVe word embeddings as initial input, and we re-trained GloVe embeddings over a cropped corpus with all denominal verbs in our data set screened out, thus

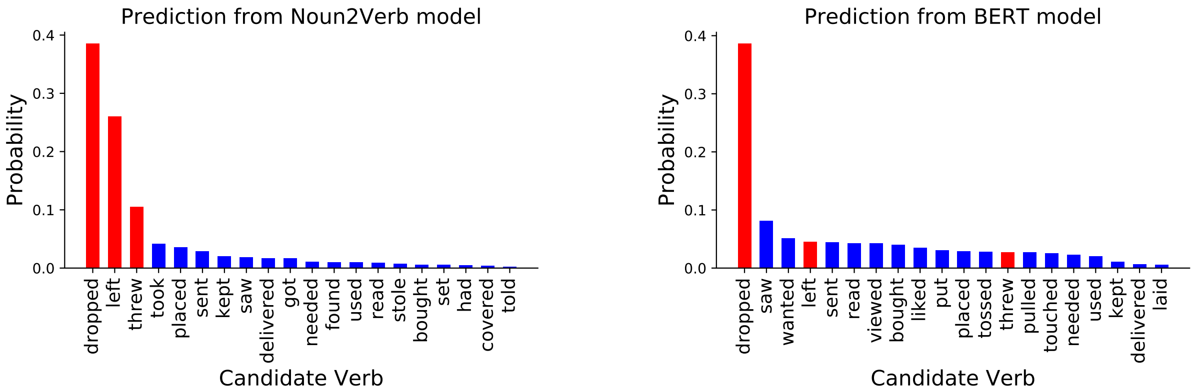


Figure 4: Example model inferences for the novel denominal usage "porched the newspaper". The horizontal axis shows a set of candidate verb paraphrases inferred from each model for the denominal verb *porched*. The bars represent model posterior probabilities over the paraphrases, with the top 3 human-annotated choices shown in red: (1) *dropped*, (2) *left*, and (3) *threw*.

preventing N2V models from exposure to denominal usages prior to the learning stage.

Model evaluation. We evaluated the models in terms of their ability to interpret and produce novel denominal verbs on U_{eval} set of 132 query-paraphrase test cases. For each query in U_{eval}, we ranked all ground-truth paraphrased verbs (from human annotation) by computing the posterior probability $p_{\text{inf}}(v|u_{\text{o}})$. We also ranked the posterior probabilities for all semantic relations r.

We first summarize the model performances in inference and generation via receiver operating characteristics curves in Figure 3. This measure examines whether each model can predict the correct paraphrase verb/target in the top k guesses. We found that all non-baseline models achieved good accuracy in predicting semantic relational types (lowest accuracy = 96%), so we focused on inference over t and v. We computed the area-under-the-curve (AUC) statistics to compare the cumulative precision of the models, summarized also in Figure 3. All models performed substantially better than chance on both tasks. In particular, *N2V*-full offers the best AUC scores for both the inference and generation tasks, outperforming BERT which in turn outperforms *N2V*-partial. These findings show that the unsupervised reconstruction procedure has helped in the inference and generation of novel noun-to-verb conversions. It is worth noting that BERT has already gained a large repertoire of knowledge from supervised pre-training, and our results indicate that unsupervised learning has helped *Noun2Verb* to make better generalizations over the limited annotated data.

To gain insights into the best two models, we visualize the posterior distributions $p_{\text{inf}}(v|u_{\text{q}} = \{t,o\})$ under each model. Figure 4 shows the posterior inferences from *Noun2Verb* and BERT based on the query utterance "the boy porched the newspaper" (top 20 candidate verbs whose probabilities are above zero are shown). We found that the *Noun2Verb* model assigned the highest posterior masses on the three ground-truth human-annotated verb paraphrases *dropped*, *left*, and *threw*. In contrast, BERT only assigned the highest posterior mass for *drop* and minimal masses on the other two alternative paraphrases. Moreover, BERT chose two non-sensical paraphrases *saw* and *wanted* as the second and third most likely candidates, most possibly because these are commonly associated words in the pre-training of BERT. This caveat of BERT not being able to explain the full distribution of paraphrases and only locking onto a single best solution has appeared to be a general issue, since we observed the same phenomenon in many other test cases.

To further verify our intuition of this issue, we considered a second measure for assessing how faithfully a model can infer the distribution of verb paraphrases from human annotations. We used Kullback-Leibler (KL) divergence, $KL(p_{\text{inf}}(v|u_{\text{q}} = \{t,o\})|p_{\text{human}}(v|u_{\text{q}} = \{t,o\}))$, to quantify the discrepancy between model posterior and human-generated distributions over the verb space, where $p_{\text{human}}(v|u_{\text{q}} = \{t,o\}))$ denotes the empirical distribution of paraphrases collected from AMT workers. Figure 5 shows that the average KL divergence on all 132 test cases is the smallest in the *Noun2Verb* model, suggesting that this model best represents the uncertainty in paraphrase choices from humans. This set of results shows that our proposed model offers better flexibility in interpreting novel instances of denominal usage and predicts human data better than the state-of-the-art language model.

Model interpretation. We provide example inferences and generated instances of novel denominal verbs. Table 1 shows example inferences for 5 query utterances. The top three rows correspond to cases where *Noun2Verb* made reasonable inference of the denominal meaning, which is manifested in the low ranks (lower is better) that the model has assigned to the ground-truth paraphrases. However, the bottom two rows suggest that *Noun2Verb* can fail in the inference task. In particular, the model assigned poor ranks to ground-truth paraphrases for the query "mine the gold". By checking the training data we found that there is no utterance describing a similar or related scenario, so our model failed to under-

Table 1: Example inferences of novel denominal usages made by the Noun2Verb model.

Query	Semantic relation	Human paraphrases with model-predicted ranks in ()	Top verb paraphrases inferred from the model
carpet the floor	locatum on	put(1), place(3), lay(11)	put, drop, cover
paper my hands	instrument	cut(1), hurt(2)	cut, hurt, wound
fox the police	agent	deceive(4), baffle(2), fool(3)	cheat, baffle, fool
mine the gold	location out	dig(327), extract(609), get(25)	put, bury, find
bee the cereal	locatum on	add(54)	get, find, eat

Table 2: Examples of novel denominal usages generated by the Noun2Verb model.

Clue verb	Semantic relation	Ground-truth phrases	Novel denominal usages sampled from the model posterior
remove	locatum out	shell the peanuts, fin the fish, skin the rabbit	stem the flowers
hit	instrument	stick my sister, rock the police	bottle the head, rope the back
repeat	agent	parrot my words	chimpanzee my gestures

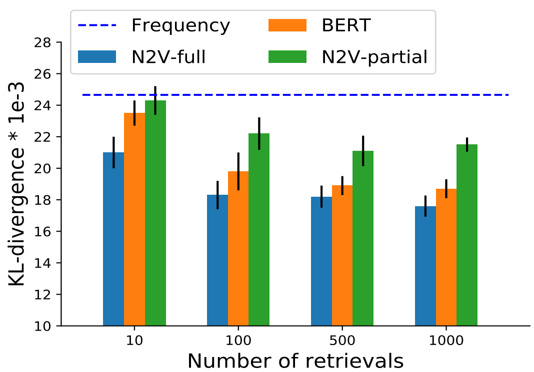


Figure 5: Summary of model performance in inference. Each bar represents the average Kullback-Leibler (KL) divergence between the empirical distribution of human-annotated paraphrases (with standard error) and the model-inferred posterior across 132 test cases. "N2V-full", "N2V-partial", "BERT", and "Frequency" correspond to the proposed Noun2Verb model, a partial version without the z variable, BERT language model, and frequency-based baseline. A lower value in KL indicates better performance. The KL divergence for random baseline (not shown) is the worst: $44.03 \times 1e^{-3}$.

stand the denominal meaning. In the last example, our model also failed to provide a reasonable paraphrase for *bee* in "bee the cereal" which presumably refers to "put honey produced by bees into the cereal", which exemplifies a rare or remotely extended word use by children. Table 2 shows examples of model generation. Our model is able to generate both conventional cases of denominal verb usage such as "stem the flowers" (although such cases did not appear in model training) and novel cases such as "chimpanzee my gestures".

Conclusion

The extended usage of a word across grammatical classes is a fundamental form of lexical innovation. We present a framework that builds on structured semantic representations informed by frame semantics and learns to interpret and generate denominal verb usages with a small amount of annotated data. We show the potential of this framework in how it outperforms a state-of-the-art language model that finds difficulty in offering flexible inference beyond idiosyncratic solutions.

Acknowledgments

We thank Bai Li and Suzanne Stevenson for constructive comments on the draft. We thank John Xu, Emmy Liu, and Zhewei Sun for helping with the experimental design. We are also thankful to members of the Computational Linguistics group at the University of Toronto for comments on an early version of this work. This research is supported by an NSERC Discovery Grant, a SSHRC Insight Grant, and a Connaught New Researcher Award to YX.

References

Baeskow, H. (2006). Reflections on noun-to-verb conversion in english. *Zeitschrift für Sprachwissenschaft*, *25*(2), 205–237.

Barsalou, L. W. (1992). Frames, concepts, and conceptual fields. *Frames, fields, and contrasts: New essays in semantic and lexical organization*, 21–74.

Bonawitz, E., Denison, S., Griffiths, T. L., & Gopnik, A. (2014). Probabilistic models, learning algorithms, and response variability: sampling in cognitive development. *Trends in cognitive sciences*, *18*(10), 497–500.

Clark, E. (1982). The young word maker: a case study of innovation in the child's lexicon. *Language Acquisition: The State of the Art*, 390–425.

Clark, E., & Clark, H. (1979, 12). When nouns surface as verbs. *Language*, *55*. doi: 10.2307/412745

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Fillmore, C. (1968). The case for case. *Universals in Linguistic Theory*, 1–89.

Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991). Affectedness and direct objects: The role of lexical semantics in the acquisition of verb argument structure. *Cognition*, *41*(1-3), 153–195.

Iacobacci, I., Pilehvar, M. T., & Navigli, R. (2016). Embeddings for word sense disambiguation: An evaluation study. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 897–907).

Jespersen, O. (2013). *A modern english grammar on historical principles: Volume 7. syntax*. Routledge.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kisselew, M., Rimell, L., Palmer, A., & Padó, S. (2016). Predicting the direction of derivation in english conversion. In *Proceedings of the 14th sigmorphon workshop on computational research in phonetics, phonology, and morphology* (pp. 93–98).

Lippeveld, M. (2013). *The acquisition of class extension rules for flexible noun-verb pairs*. Unpublished doctoral dissertation, McGill University Libraries.

MacWhinney, B. (2014). *The childes project: Tools for analyzing talk, volume ii: The database*. Psychology Press.

Pustejovsky, J. (1991). The generative lexicon. *Computational linguistics*, *17*(4), 409–441.

Ruppenhofer, J., & Michaelis, L. (2014, 01). Frames and the interpretation of omitted arguments in english. In (p. 57-86). doi: 10.1075/pbns.244.04rup

Shwartz, V., & Dagan, I. (2018). Paraphrase to explicate: Revealing implicit noun-compound relations. *arXiv preprint arXiv:1805.02442*.