Modeling word overextension in a Grey Parrot

Michal Fishkin (michal.fishkin@mail.utoronto.ca)
Department of Computer Science
University of Toronto
Shereen Chang (shereen.chang@nu.edu.kz)
Department of History, Philosophy, and Religious Studies
Nazarbayev University
Yang Xu (yangxu@cs.toronto.edu)
Department of Computer Science, Cognitive Science Program
University of Toronto

Abstract

Word meaning extension refers to the process by which a single word form develops multiple related meanings. Young children exhibit the capacity to extend word meaning, and previous research shows that such word overextension relies on multimodal semantic knowledge. We explore the evolutionary trace of word meaning extension by asking whether nonhuman animals might have this shared capacity. We compare meaning extension in children with the attested cases of overextension collected from the YouTube channel of a Grey Parrot, Apollo, who has acquired some English words. Our results show that parrot overextension can be predicted by a multimodal child overextension model better than baselines, which demonstrates that Grey Parrot may be using semantic knowledge similar to children for choosing words to express new referents. Our finding suggests that meaning extension is a cognitive ability identifiable in species about 320 million years apart from humans.

Keywords: word meaning extension; overextension; animal semantics; probabilistic model; comparative cognition

Introduction

A key function of the lexicon is to support the expression of new or emerging ideas using a limited vocabulary. One commonly observed solution is the attribution of multiple meanings to a single word, a process known as word meaning extension (abbreviated as WME) (Bréal, 1897; Ramiro, Srinivasan, Malt, & Xu, 2018; Brochhagen, Boleda, Gualdoni, & Xu, 2023). For example, the word *face* originally denoting a body part was extended to express both a visage and the side of a cube.

How exactly words acquire new meanings is not random, because applying disparate meanings to a word may result in the word being unintelligible or uninformative for language users. In fact, existing work has argued that WME in the historical development of English and across languages tends to operate by relating similar items or based on *semantic relatedness* (Ramiro et al., 2018; Xu, Duong, Malt, Jiang, & Srinivasan, 2020; Brochhagen & Boleda, 2022). In other words, the meanings attributed to a given word tend to have some conceptual connectedness.

Humans are not alone in their ability to label entities with distinct vocalizations, as demonstrated in marmoset monkeys (Oren et al., 2024). Furthermore, a similar preference for conceptual connectedness has been found in nonhuman animals where researchers found that baboons learned rules faster when there was some degree of connectedness in the stimuli (Chemla, Dautriche, Buccola, & Fagot, 2019), which demonstrates the importance of conceptual relatedness as a pre-linguistic cognitive ability inherited from our evolutionary ancestors. These studies together allude to advanced cognitive and linguistic capabilities in nonhumans.

We build on these ideas by connecting occurrences of WME in human language to the preference for conceptual relatedness in nonhumans. Our question is whether humans and nonhumans might have a shared capacity for word meaning extension. To find an answer, our first task is to look at WME instances that can be compared across species. Adult humans have an established connection between their conceptual understanding and their use of language, making it difficult to compare their language use to that of nonhumans. In contrast, children are a population whose conceptual mappings are in development. We therefore focus on a comparison of WME found in human children: child overextension.

Child overextension is a well-documented phenomenon commonly observed in humans between the ages of one and two. Overextension refers to the extended use of a single word form to refer to novel concepts outside its conventional range of use (Clark, 1978). An example would be a child using the word 'cat' (the utterance) to refer to all four-legged furry mammals (the referents). Moreover, overextension is different from misspeaking or babbling because, as Clark asserted, children could comprehend the correct referent labels despite not using them in production. This production-comprehension asymmetry is an indication that the child is actively using their conceptual understanding of the world to attribute multiple meanings to one specific word.

Vygotsky, an early observer of child overextension, presented the idea of a chain complex, where one word is used to refer to different concepts that are chained together in some logical way in conceptual space (Vygotsky, 1986). Building on these previous studies, later research described three main modalities along which concepts are typically related in child overextension (Rescorla, 1980):

- **Categorical**: wherein the child uses a word for items within a category or in a systematic taxonomy, e.g., calling all fruits an 'orange'.

- **Perceptual or analogical**: where the child uses a word for objects that are perceptually similar, but not necessarily in

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

the same category, e.g., using the word 'ball' to refer to a balloon.

- **Predicate-based**: wherein the child uses a word to express some associative relation between the utterance and the referent. For example, the child may say 'dada' when indicating her father's shoes not because of a categorical or analogical link but because the shoes belong to her father. Predicate-based overextensions are generally in the place of a fuller sentence, e.g., 'these shoes belong to dada.'

We compare animal overextension with human child overextension by taking a formal approach following recent work on child overextension (Pinto Jr & Xu, 2021):

1. Build a model that uses Rescorla's modalities to predict child overextension, particularly the word choices that children make when referring to new referents.
2. Collect data from an animal whose utterances constitute potential overextensions.
3. Apply the model to the said data and assess the degree to which it accounts for the animal overextension data, better than alternative baselines.

As an initial step, we implement the child overextension prediction model from existing work (Pinto Jr & Xu, 2021). This model approximates Rescorla's modalities with categorical information from WordNet (Miller, 1994), visual analogical information from ImageNet (Deng et al., 2009), and predicate information from a word association dataset called the Small World of Words (SWoW) (De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2018). This model was trained on child overextension data and is therefore tuned to predict how *children* overextend words.

Our choice of comparative nonhuman agent is Apollo, a Grey Parrot featured in his owners' YouTube Channel 'Apollo and Frens' (Lacey & Mason, 2024). The channel contains years worth of videos of Apollo's training, self-babbling, and linguistic development. He is trained to identify various properties, and the prompt objects are frequently changed to prevent operant conditioning. We hypothesize that a model based on Rescorla's semantic modalities and trained solely on child overextension data might predict the overextensions of a Grey Parrot. We define success as the model's semantic modalities outperforming the random and frequency baselines.

To preview our analysis, we apply a child overextension model to the performance of Apollo the Grey parrot. We compare performance of the model and the dominant modalities between the agents and show that there is evidence for child-like overextension capabilities in parrots.

Computational methodology

We build our model based on existing work on child overextension (Pinto Jr & Xu, 2021) which has roots in the chain complex described in (Vygotsky, 1986), where a child learns the word 'quah' for a duck, and then extends 'quah' to the water in which a duck floats, and then to any liquid including milk in the bottle. The way the child attributes more meanings to the word 'quah' is achieved by the chaining of one related concept to the next. This process can be incorporated into a Bayesian probabilistic model. By integrating semantic distance into the likelihood calculation, we allow for a computed multi-dimension semantic space to express the conceptual space of children and, theoretically, parrots.

Model Formulation

With vocabulary V, the probability that the agent utters a word w when trying to express a concept c can be modeled with the Bayesian rule, as follows:

$$p_{prod}(w|c) = \frac{p(c|w)p(w)}{\sum_{w' \in V} p(c|w')p(w')}$$

Here

- $p(w)$ is the prior probability of the word.
- $p(c|w)$ is the probability that a concept c is referred to by a given word w.

In this study, we take $p(w)$ from a uniform distribution in Apollo's vocabulary. This is because of the lack of comprehensive data on parrot vocabulary frequency. The way we formulate $p(c|w)$ is by computing the similarity score of the concept c with the concept c_w that the uttered word traditionally refers to. For example, we would desire $p(\texttt{BALLOON}|ball) > p(\texttt{BALLOON}|cup)$ because the concept `BALLOON` is semantically closer to the concept `BALL` than to the concept `CUP`.

The similarity between two concepts in the multimodal semantic space is expressed as a decaying exponential function of the distance. For the multimodal model, the decay is modulated by a single learned parameter, h. Each unimodal model also has its own parameter h_m. See (Pinto Jr & Xu, 2021) for more details.

$$p(c|w) = f_{sim}(c, c_w) = \exp -\frac{d(c, c_w)^2}{h}$$

For the final part of our formulation, the distance in the semantic space is the L^2 norm between the concepts on each of the modalities: categorical, visual, and predicate-based/associative. We use all modalities in the multimodal calculation (formulation below) and also calculate the unimodal distance. We explain how we constructed the modality spaces in Section .

$$d(c, c_w) = \sqrt{\sum_m d_m(c, c_w)^2}$$

Here m belongs to a multimodal semantic space including categorical, visual, and associative representations yet to be specified.

Model Specifications The input to the model is a string indicating the referent concept. The output of the model $L = \{a_0, a_1, ..., a_m\}$ is a set of descending ranked lists, one for each modality. Each element of $a_m = (w_{m,0}, w_{m,1}, ...w_{m,n})$ is a word from Apollo's vocabulary. They are arranged in decreasing probability for the given referent and modality, i.e. $p(w_{m,k}|c_{ref}) \geq p(w_{m,k+1}|c_{ref})$. Note that while the number of words is n, the number of rankings is $n_r \leq n$, since candidates can be equidistant from the referent concept. For example, suppose the categorical distances between candidates `PIG` and `DOG` from referent `COW` are identical, as both are non-bovine mammals. This would result in a tie in the categorical ranking of the words when the referent word is "cow". It is this ranking of the word, r_w, which corresponds to its probability, not its index in the output list.

The parameters for h used for this model are the same as those used by (Pinto Jr & Xu, 2021). That original model from was trained on child overextension data, so the application of his model on parrot overextension data will give us comparative information about a parrot's capacity for meaning extension.

To reflect that higher scores should be associated with lower ranks, and to decouple the results from vocabulary size, we min-max scale the rank of the given word, r_w, across all ranks in the modality list, and then subtract from 1. This gives us a ranking metric $R(r) = 1 - r$ for $r \in [1, n_r]$ such that $R \in (0, 1)$.

Baselines For each of the words in the vocabulary, we calculate three baselines: uniform/random, frequency, and sound. We hypothesize that these baselines should be dominant in cases where the observed overextension was non-deliberate, i.e. the parrot did not truly extend the meaning of his uttered word to the referent.

In the random case, the expected ranking of a given word is halfway down the ranked list giving $R_{random} = 0.5$.

In the frequency case, we take relative word frequency estimates from the samples in the data and rank the words in V by their frequency. Each word's frequency baseline ranking metric, $R_{frequency}$, depends on this frequency-ranked list.

The last of our baselines is sound similarity which captures phonetic relations among concepts. For this, we employed the Carnegie Mellon University Pronouncing Dictionary (Lenzo, 2024) to look up a phonetic ARPABET representation of all the words in V. We then construct a matrix that records the Levenshtein Distance between any two words in V (Levenshtein, 1965).

For a given referent c, the sound baseline ranking metric of any candidate word w is given by $R^c_{sound}(w)$, where R^c_{sound} is the ranking of words by their pronunciation Levenshtein distance from c. As an example with the toy vocabulary $V = \{crack, plant, block, black\}$, $R^{plant}_{sound}(block) = 0$, $R^{crack}_{sound}(block) = 0.5$, and $R^{black}_{sound}(block) = 1$.

We should expect the sound baseline to be a strong predictor in cases where the agent made an error in utterance, such as mixing up production mechanisms for similar consonants. In the case of overextension where the target word is outside of its vocabulary, we expect semantic modalities to dominate all baselines if the agent is overextending in a manner similar to children. In the next section, we describe how the semantic modalities are calculated.

Semantic Space Implementation The computational model of the semantic modalities has inevitable simplifications which present a departure from (Rescorla, 1980). The details of the modality implementations are as follows:

Categorical. (Rescorla, 1980) gives no specific categorization framework, so we replicate existing work (Pinto Jr & Xu, 2021) by using WordNet's hierarchical synset structure. The categorical distance between two words is given by the WordNet synset-to-synset Wu-Palmer Similarity (`wup_similarity`).

Analogical $\rightarrow$ **Visual**. Due to technological limitations, distance in the allegorical dimension is simplified to only account for visual factors. The allegorical modality is impoverished in our model, as both children and parrots would have other cues upon which to draw (e.g. tactile, auditory). Using the images in each word's ImageNet2012 synset folder, we retrieve a visual embedding for each concept using the output of the first ReLU layer of the VGG19 convolutional model.

Predicate-based $\rightarrow$ **Associative**. We approximate the predicate-based distance between concepts by their word-association distance. This strongly simplifies the connection found by (Rescorla, 1980), who considered many different types of relationships to make predicate connections between two concepts, and considered idiosyncracies in his evaluation. For example, while the general public would not associate the word 'doll' with 'house', a child who is playing with a dollhouse may use the former to refer to the latter. For general associativity, we use the Small World of Words English dataset as found in (De Deyne et al., 2018). We build a graph of word associations and run a decaying random walk to approximate the associative similarity of two concept words.

Data Collection

Apollo is an African Grey parrot whose owners upload videos of him labeling objects onto their YouTube channel, 'Apollo and Frens' (Lacey & Mason, 2024). In these videos, one of his trainers produces an object (referent) and prompts Apollo with one of four questions:

- What is this [object]? EXISTENTIAL
- What colour [is this object]? PERCEPTUAL, COLOUR
- What is [this object] made of? PERCEPTUAL, MATERIAL
- What am I doing? PERCEPTUAL, ACTION

Apollo then produces a word from his vocabulary (utterance). We collected 244 referent-utterance pairs (RUPs) from these labeling videos. Of these, 45 are cases in which Apollo

Table 1: OoV Overextension Referent-Utterance Pairs. †: Apollo's understanding of 'cord' is that of phone chargers. ‡: This is an approximation for ImageNet compatibility. ††: A toy white egg with green polka-dots.

Novel Referent	True Word	Apollo's Choice
Red plastic bucket	Bucket	Cup
Plastic bead	Bead	Ball
Frayed rope	Rope	Cord †
Stack of tissues	Tissue ‡	Book
Yoshi egg††	Egg	Ball
Egg cartons	Carton	Box
Cooking pot	Pot	Bowl
Interior wooden door	Door	Block
Bathroom wall material	Drywall	Rock

made a mistake labeling the prompt object, i.e. the appropriate label was inside his vocabulary, but he used a different word. These cases are in-vocabulary (iV) overextensions, and analyzing why Apollo uses one word instead of another can also give us insight into his conceptual mapping for familiar objects.

From the collected RUPs, we extracted a relative frequency for each word.

One video, called 'Is He Wrong Though?' consisted of one of Apollo's handlers showing objects that Apollo had labeled by himself. There are ten such cases, shown in Table 1. These ten cases are out-of-vocabulary (OoV) overextensions because the objects' appropriate labels are not in Apollo's vocabulary, so he by necessity overextended when labeling them. The OoV overextensions are valuable insights into the parrot's conceptual mapping, as a well-performing child overextension model would suggest that Apollo's concept-chaining works in a similar fashion to that of human children.

Processing

Each word w in the lexicon V_L is hand-labeled with an appropriate WordNet synset. The reason for hand-labeling the data was to ensure that the final synset accurately reflected what the parrot was exposed to. All words with no appropriate WordNet synsets are dropped from V_L. The set of ImageNet synsets available in ImageNet2012 are cross-referenced with each of the WordNet synsets, and the closest results are taken as the corresponding ImageNet synsets. Any words that did not have an appropriate corresponding ImageNet synset were dropped from V_L.

Results

Examples of the model output are visualized in Figures 1 and 2. Figure 1 visualizes ranked model output when given the input RUP $bead, ball$, as seen in Table 1. The visual modality is dominant, ranking 'ball' in first place. All other semantic modalities performed worse than the frequency baseline, and the categorical modality performed worse than chance. The model indicates that *if* Apollo is overextending, he is extending the word 'ball' to the concept `BEAD` because of a *perceived visual similarity*.

Figure 2 visualizes the ranked model output when given the RUP $book, bowl$. This is an instance in which Apollo made a mistake during training, i.e. his trainer presented him with an object for which the correct label was in Apollo's vocabulary, but Apollo uttered the incorrect word. The notable difference between this case and that seen in Figure 1 is the poor performance of the semantic modalities when compared to the sound and frequency baselines. Since 'ball' and 'bowl' are close sound-wise, only changing one vowel, the excellent performance of the sound prediction is expected; however, the poor semantic performance indicates that he is not extending the word 'bowl' to the concept `BOOK`.

Performance by data type

We want to verify that the model, which is trained on child overextension data, can predict the language use of Apollo. If it can, this points to a capacity for creativity and the similarity between the way parrots use language creatively and the way young children use language creatively. To evaluate model performance, we calculate the Mean Reciprocal Rank (MRR) across referents. MRR is a common metric for evaluating the effectiveness of a ranking algorithm, typically used to optimize search engines. For a single given list L, which consists of relevant (correct) and irrelevant (incorrect) results:

$$MRR(L) = \frac{1}{rank_i}$$

Here, $rank_i$ is the rank of the *first correct result* in L. A higher MRR indicates a more precise prediction from the model. Figure 3 shows a chart of the MRR Scores for different subsets of RUPs. The second column represents the performance of the model in OoV RUPs, as shown in Table 1. The third column (iV) represents the model's performance across the iV overextensions that Apollo made in training. The first column shows the model's performance across both data subsets.

Notably, the multimodal model performs best when applied to the entire dataset. All semantic modalities except categorical perform better than frequency, the best-performing baseline (compared to random and sound error). This indicates that Apollo considered semantic information before producing an utterance. This phenomenon is magnified, seen clearly in the OoV column, which denotes the alleged overextensions. We see here that the semantic modalities strongly outperform the baselines as predictors of an utterance given a referent. The error cases show a countertrend, with frequency being the best-performing dimension, although the multimodal and predicate modalities are still decent predictors. This difference between the semantic modalities' OoV prediction power and their error prediction power suggests that the utterance choice in the OoV cases more closely follows the process behind children's overextension, while the error cases are more strongly influenced by frequency.

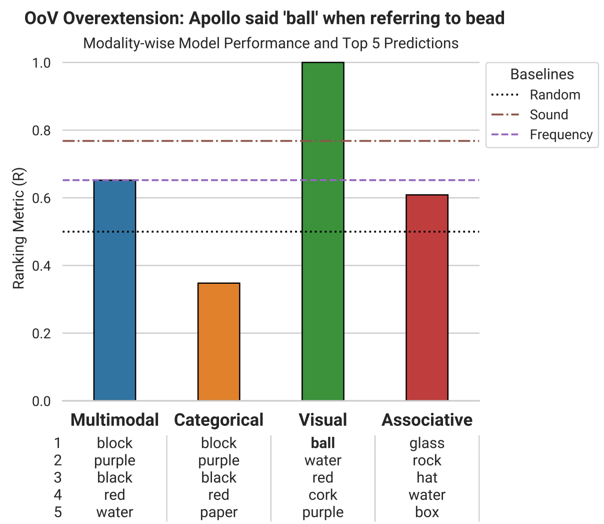

Figure 1: Modality and baseline results from an alleged overextension referent-utterance pair.

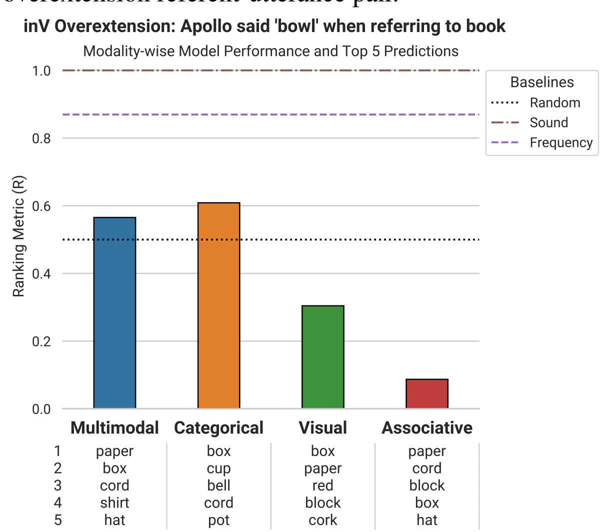

Figure 2: Modality and baseline results from an error referent-utterance pair.

	All	OoV	iV
Random	0.0833	0.0833	0.0833
Frequency	0.2496	0.1661	0.2739
Sound	0.1559	0.1	0.1721
Categorical	0.316	0.5021	0.2619
Visual	0.2226	0.3073	0.198
Associative	0.2787	0.5737	0.1931
Multimodal	0.309	0.4838	0.2583

Figure 3: A summary of the baseline and model performances, measured with mean reciprocal rank. The subsets of data are shown column-wise. A higher mean reciprocal rank indicates better predictive accuracy.

We now verify whether Apollo is overextending. If the RUP results in low semantic scores but high baseline scores, we can reasonably be certain that his utterance was unintentional. However, a high score in even one semantic modality could indicate that he is overextending a word to a different concept based on that semantic dimension. To see if there is a distributional difference between the alleged overextensions from Table 1 and his mistakes during training, we summarize the best-performing, or dominant, modalities across the different subsets of data. We can also compare Apollo's dominant-modality distributions to that of children's.

In all RUPs, the dominant modality in prediction (including sound-based error) is visualized in Figure 4 as a winner-takes-all pie-chart. There are three dominant-modality distributions: those of the children in Rescorla et al. (Rescorla, 1980) at the far left, those of Apollo's out-of-vocabulary referents in the middle, and those of Apollo's training errors (in-vocabulary overextensions) in the far right. Both parrot cases have a much higher percentage of their overextension attributed to frequency than the child case, indicating that, even in the alleged overextension cases, it is likely that Apollo has also had some babble-luck.

We observe that Apollo's utterances are best predicted by semantic modalities, similarly to children's cases, which are dominantly predicted by semantic modalities as opposed to sound-based error. While children skew towards categorical and visual representations, Apollo's dominant semantic modalities are based on associative and visual representations. It is also worth noting that, in out-of-vocabulary overextension, Apollo relied almost exclusively on semantic modalities. In contrast, sound-based errors appeared to be more frequent in the in-vocabulary case.

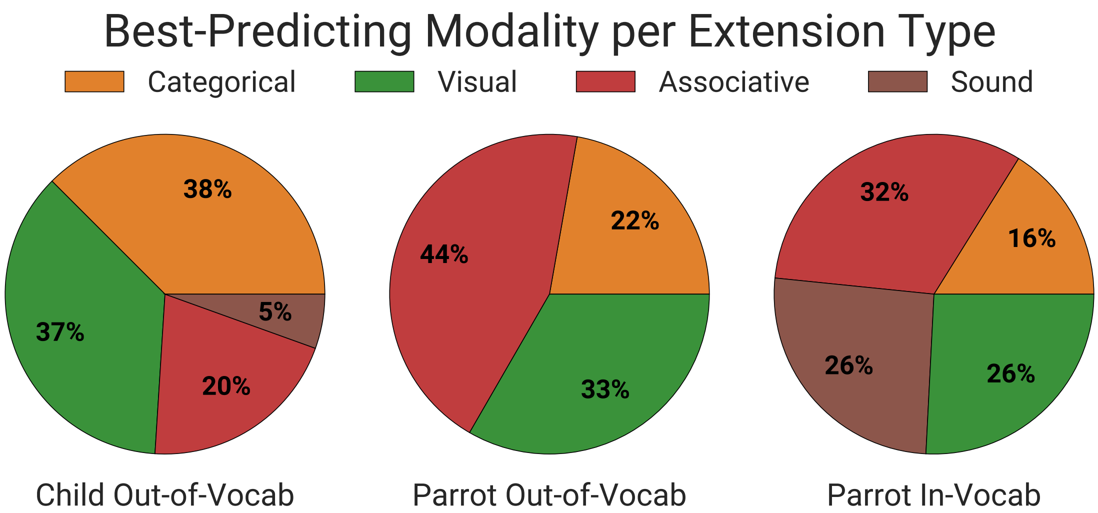

Figure 4: The winner-takes-all dominant modality distributions for three subsets of RUPs: child, parrot OoV, and parrot iV. The prominence of the sound modality in child overextension is closer to that of the OoV parrot cases than that of the iV parrot cases.

Discussion

We present a novel dataset that contains the observed and attested use of language by a Grey parrot named Apollo. We collected these data from the YouTube channel of Apollo's trainers, called *Apollo n Frens*. We then ran the collected referent-utterance pairs through a child overextension model, which ranked the words in Apollo's vocabulary from most to least likely according to the overextension probabilistic formulation we described.

Our results show that, while the frequency baseline is a decent predictor for parrot utterance, the semantic modalities (multimodal, categorical, visual, and associative) are consistently good predictors. In the attested overextension cases, in which the referent words are out-of-vocabulary (OoV), the semantic modalities are overwhelmingly dominant. Semantic modality dominance indicates that the cognitive process behind Apollo's utterances somewhat parallels that of human children, at least when he chooses a word to refer to an object for which he has no in-vocabulary label.

Sound is a common modality in the iV cases, but not in the OoV cases. This indicates that some of Apollo's mistakes were made because he tried to produce one word but mixed up the sounds that constituted his desired utterance; an example is shown in Figure 2. Although sound-based error might be thought of as a kind of overextension, it has little relevance to semantics per se and instead reflects phonetic confusion among the words that Apollo has acquired. Due to our definition of related concepts in overextension being limited to the *semantic* space, we do not categorize these as overextensions.

Limitations and Possible Extensions

There is a distinct difference between the distributions of the most dominant modalities (as seen in Figure 4) and their predictive power (as seen in Figure 3). Taking the OoV subset, in particular, the best-predicting modalities are visual and associative. However, the dominance of the visual and associative modalities is equal to that of the categorical modality; this indicates that the dominance measure may not be a good indicator of the performance of those modalities and, consequently, their dominance in the conceptual map of Apollo. To get better modality-dominance estimations, we would require more data.

A small sample space also means that our frequency estimates may be impoverished. The frequency estimates we used were taken over the entire sample space, whereas to more holistically understand Apollo's baseline accuracy, we would need to have frequency estimates that are conditioned on the type of prompt. Currently, we do not know if being asked 'What is this made of?' makes Apollo randomly select between 'metal' and 'glass', or if he selects appropriately based on perception and other cognitive processes.

Conclusion

We present evidence that the capacity for word meaning extension in the Grey parrot is similar to that of young children. A Bayesian probabilistic model trained on child overextension data performs well on out-of-vocabulary parrot overextension data, indicating that the multimodal process for overextension in children is similar to that of the test subject, Apollo. Although the categorical, perceptual, and associative systems of a parrot are undoubtedly different than those of humans, the ability of the model to predict utterances of a parrot without further training points to similarities between the cognitive-conceptual systems of species 320 million years apart.

Acknowledgements

We thank Guillaume Thomas and Thomas Brochhagen for helpful feedback on the paper. SC was supported by a York University Liberal Arts & Professional Studies Postdoctoral Fellowship. YX is supported by NSERC Discovery Grant RGPIN-2018-05872.

References

Bréal, M. (1897). *Essai de sémantique*. Paris: Hachette.

Brochhagen, T., & Boleda, G. (2022). When do languages use the same word for different meanings? The Goldilocks principle in colexification. *Cognition*, *226*, 105179.

Brochhagen, T., Boleda, G., Gualdoni, E., & Xu, Y. (2023). From language development to language evolution: A unified view of human lexical creativity. *Science*, *381*(6656), 431–436.

Chemla, E., Dautriche, I., Buccola, B., & Fagot, J. (2019, July). Constraints on the lexicons of human languages have cognitive roots present in baboons (papio papio). *Proceedings of the National Academy of Sciences*, *116*(30), 14926–14930. Retrieved from http://dx.doi.org/10.1073/pnas.1907023116 doi: 10.1073/pnas.1907023116

Clark, E. V. (1978, December). Strategies for communicating. *Child Development*, *49*(4), 953. Retrieved from http://dx.doi.org/10.2307/1128734 doi: 10.2307/1128734

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2018, October). The "small world of words" english word association norms for over 12, 000 cue words. *Behavior Research Methods*, *51*(3), 987–1006. Retrieved from http://dx.doi.org/10.3758/s13428-018-1115-7 doi: 10.3758/s13428-018-1115-7

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 ieee conference on computer vision and pattern recognition* (pp. 248–255).

Lacey, V., & Mason, D. (2024). *Apollo and Frens.* https://www.youtube.com/@ApolloandFrens. ([Accessed 11-07-2024])

Lenzo, K. (2024). *The CMU Pronouncing Dictionary — speech.cs.cmu.edu.* http://www.speech.cs.cmu.edu/cgi-bin/cmudict. ([Accessed 12-07-2024])

Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics. Doklady*, *10*, 707-710. Retrieved from https://api.semanticscholar.org/CorpusID:60827152

Miller, G. A. (1994). WordNet: A lexical database for English. In *Human Language Technology: Proceedings of a workshop held at Plainsboro, New Jersey, March 8-11, 1994.* Retrieved from https://aclanthology.org/H94-1111

Oren, G., Shapira, A., Lifshitz, R., Vinepinsky, E., Cohen, R., Fried, T., ... Omer, D. (2024). Vocal labeling of others by nonhuman primates. *Science*, *385*(6712), 996-1003.

Pinto Jr, R. F., & Xu, Y. (2021). A computational theory of child overextension. *Cognition*, *206*, 104472.

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*(10), 2323–2328.

Rescorla, L. A. (1980, June). Overextension in early language development. *Journal of Child Language*, *7*(2), 321–335. Retrieved from http://dx.doi.org/10.1017/S0305000900002658 doi: 10.1017/s0305000900002658

Vygotsky. (1986). *Vygotsky: Thought & language 2ed (cloth).* London, England: MIT Press.

Xu, Y., Duong, K., Malt, B. C., Jiang, S., & Srinivasan, M. (2020). Conceptual relations predict colexification across languages. *Cognition*, *201*, 104280.