Syntactic Categories Derived from Frequent Frames Benefit Early Language Processing in English and ASL

Galia Bar-Sever and Lisa Pearl

1. Introduction

Because language acquisition happens in stages, early language acquisition strategies probably don't yield adult knowledge directly. Instead, they're more likely to provide transitory representations that scaffold the acquisition of later knowledge (Frank, Goldwater and Keller, 2009, Connor, Gertner, Fisher and Roth, 2010, Connor, Fisher and Roth, 2013, Gutman, Dautriche, Crabbé and Christophe, 2014, Phillips and Pearl, 2015). For example, syntactic categories that twelve-month-olds have may not look like adult categories. Yet, these early syntactic categories may still be good enough for what twelve-month-olds need to do. However, if children's developing knowledge representations aren't adult-like, what *do* they look like and why might they look that way?

One idea is that children's developing language processing abilities help determine the form of their early representations (Lidz and Gagliardi, 2015, Omaki and Lidz, 2015). In particular, we can consider how developing knowledge representations might be used by children to help them process their native language. If the language input becomes easier to process, the representation enabling this might be considered "good enough" – and certainly useful to the child, even if it doesn't match the adult representation. We suggest that acquisition strategies yielding this sort of useful immature representation should be viewed as successful.

A related consideration involves the impact of different language modalities on language processing. Representationally, spoken and signed languages are similar (Lillo-Martin and Gajewski, 2014), though the method of transmission clearly differs. Mature processing of spoken and signed languages is also similar at an abstract level (Hickok, Love-Geffen and Klima, 2002). However, given the significant differences in transmission, it may well be that the immature representations easing immature language processing differ between spoken and signed languages. Therefore, early acquisition strategies that succeed at easing spoken language processing may not do so for signed language processing.

*Galia Bar-Sever, University of California, Irvine. gbarseve@uci.edu. Lisa Pearl, University of California, Irvine. lpearl@uci.edu. Thanks especially to Jill Lany, Daniel Swingley, Jesse Snedeker, Naomi Feldman, Diane Lillo-Martin, CoLaLab at UC Irvine, and the BUCLD 2015 audience for helpful comments and suggestions. Anything we got wrong isn't their fault.

 Proceedings of the 40th annual Boston University Conference on Language Development, ed. Jennifer Scott and Deb Waughtal, 32-46. Somerville, MA: Cascadilla Press.

We present a case study of this processing-sensitive approach for early syntactic categorization, focusing on the frequent frames (**FFs**) strategy (Mintz, 2003). We selected the FF strategy because of its success on a variety of spoken languages when evaluated using traditional metrics that expect the FF-based categories to match adult syntactic categories. We then propose a new evaluation metric, based on the information-theoretic measure *perplexity*, which assesses language processing ease for any syntactic category representation, whether immature or adult.

We subsequently discuss the FF strategy and perplexity implementations we use, including (i) the structural assumptions modeled learners have when we assess perplexity, (ii) the properties of the languages the FF strategy is assessed over (spoken: English, signed: American Sign Language (**ASL**)), and (iii) the practical details that matter for implementing FFs on language data. Our investigation yields two main findings. First, there is no difference by modality: the FF-based categories function similarly on English and ASL. This is intuitively satisfying as the FFs implementation doesn't necessarily differ by transmission method either, though it potentially can. Second, while FF-based categories don't match adult categories in either language, they make processing easier than the adult categories do in both languages if children also have immature assumptions about language structure. This suggests a synergy between early syntactic category representation and children's developing knowledge of language structure: language processing is less cognitively demanding when children have immature representations of both syntactic categories and the structures those categories are part of.

2. Case study: Early syntactic categorization and Frequent Frames

In essence, syntactic categories are clusters of individual lexical items that function similarly syntactically. For example, the adult COUNT-NOUN category includes lexical items like *kitty, penguin*, and *idea*, and each of these can be preceded by a DETERMINER like *the* or *a(n)* and used to create a NOUN PHRASE that can serve as the subject of a sentence. So, one purpose of syntactic categories is to more compactly represent the syntactic patterns of the language (i.e., a single rule NP $\rightarrow$ DETERMINER COUNT-NOUN will suffice, instead of multiple rules like NP $\rightarrow$ *the kitty*, NP $\rightarrow$ *a penguin*, etc.).

This representational economy relates to processing ease. If language users recognize that individual words are instances of a larger coherent category, it becomes easier to predict the underlying structure of the language input encountered, as implemented by the language's syntactic patterns. This is because the structural commonality across different utterances is more readily apparent (e.g., *the kitty is cute* and *a penguin is adorable* are both examples of DETERMINER COUNT-NOUN COPULA ADJECTIVE).

Given this, syntactic categories seem like a useful abstract representation to learn. But what do early syntactic categories look like? Experimental evidence suggests that the beginning stages of syntactic categorization occur before 12 to

14 months, when toddlers recognize linguistic markers for COUNT-NOUN and ADJECTIVE (Booth and Waxman, 2003). Given how early this is developmentally, it's likely the syntactic categories hypothesized at 12 months don't match adult categories. For example, young toddlers might not recognize all the nouns adults would identify as COUNT-NOUN – instead, toddlers might realize that *kitty* and *penguin* are the same kind of thing, without recognizing that *idea* is, too.

Frequent Frames (FFs) form the basis of an early categorization strategy that is both computationally inexpensive and linguistically-based. This strategy has yielded promising results for many spoken languages with different linguistic properties (e.g, English: Mintz 2003; Wang and Mintz 2008; French: Chemla, Mintz, Bernal and Christophe 2009; Spanish: Weisleder and Waxman 2010; German: Stumper, Bannard, Lieven and Tomasello 2011, Wang, Höhle, Ketrez, Küntay, Mintz, Danis, Mesh and Sung 2011; Dutch: Erkelens 2009; Turkish: Wang et al. 2011; and Mandarin Chinese: Xiao, Cai and Lee 2006).

The basic intuition is that young toddlers pay attention to frequently occurring frames, which identify linguistic units that behave similarly in utterances (i.e., appear in the same linguistic context, as implemented by the frame). For example, in the sentences *I am petting nice kitties* and *I am hugging nice penguins*, the word-level frame *am_ nice* identifies that *petting* and *hugging* have the same linguistic context and so are the same kind of word. Experimental evidence suggests that 12-month-olds are sensitive to word-level frames (Mintz, 2006). More generally, 12-month-olds can recognize the non-adjacent dependencies that frames rely on if the toddlers already know that adjacent dependencies exist between linguistic elements (Lany and Gómez, 2008).

The "frequent" part of the FFs strategy is meant to capture the intuition that young toddlers have limited attention. In particular, something that occurs frequently is likely to be salient to toddlers, and so the FFs strategy assumes that toddlers rely on a set of frames that are frequent enough to be noticed. More specifically, the intake for early categorization is a set of frequent frames and the output are clusters of linguistic elements captured by each frequent frame. In the cross-linguistic computational investigations mentioned above, these clusters have been compared against adult syntactic categories and generally found to be very accurate. For example, a frequent frame might cluster together many VERB items and exclude non-VERB items, and so be very accurate with respect to the adult VERB category.

3. A Processing-Based Evaluation Metric

One practical reason previous studies compared the categories created from FFs to adult categories is that this is a "gold standard" that's both available and fairly easy to agree on (at least, as implemented by syntactic category annotation in many corpora like CHILDES). However, as mentioned above, the problem is that 12-month-old syntactic categories may not match adult categories: toddlers might not (i) recognize all instances of a given category as belonging to that category (like COUNT-NOUN), and (ii) realize certain conceptually subtle categories even exist (like DETERMINER and AUXILIARY VERB).

We propose leveraging the intuition that useful acquired knowledge makes language easier to process, which in turn benefits subsequent acquisition processes like lexical acquisition (e.g., Weisleder and Fernald 2013, Fernald, Perfors and Marchman 2006, Fernald and Marchman 2012, Weisleder and Fernald 2013). Because children's ability to process speech improves rapidly between 15 months and two years (Fernald, Pinto, Swingley, Weinbergy and McRoberts, 1998), it's likely that very young children are gaining knowledge that enables them to improve their language processing efficiency. For syntactic categories, this means the exact categories (and members of those categories) don't matter. Instead, what matters is how knowledge of these categories helps toddlers deal with the language data they subsequently encounter.

A prominent approach for quantifying language processing efficiency is related to the predictability of upcoming data (e.g., the *surprisal theory* of language processing: Hale 2001, Levy 2008). The idea is that highly predictable things are easier to process and predictability can be quantified by probability. So, utterances that are more probable are therefore more predictable and thus easier to process.

We use a formal definition of utterance predictability based on probability, called *perplexity* (Brown, Pietra, Mercer, Pietra and Lai, 1992). In particular, perplexity is inversely related to probability (as shown in Equation 1), with the intuition that low probability utterances are highly perplexing. In contrast, high probability utterances are more predictable and so less perplexing.

$$Perplexity(U = w_1...w_n) = \sqrt[n]{\frac{1}{P(U = w_1...w_n)}} \quad (1)$$

In Equation 1, the perplexity of utterance U comprised of words $w_1...w_n$ is the geometric mean of the inverse probability of U. So, when the probability of U is low (e.g., a garbled utterance like *penguins I nice like*), the inverse probability is high and so U has a high perplexity. In contrast, when the probability is high (e.g., *I like nice penguins*), the inverse probability is low and so U has a low perplexity. Because probability ranges between 1 and 0, the inverse probability (and so perplexity) ranges between 1 and positive infinity.

Clearly, how we determine the probability of a sequence of words ($P(w_1...w_n)$) matters, since this is the heart of the perplexity calculation. We suggest two potentially plausible assumptions for how toddlers view language generation. First, words belong to underlying (i.e., latent) syntactic categories. This presumably motivates categorizing words in the first place. Second, toddler hypotheses about how language is structured are still developing. So, while they have yet to learn how their native language is truly structured, they likely recognize some local dependencies between syntactic categories (similar to how they recognize local dependencies more generally: e.g., Gómez and Lakusta 2004, Lany and Gómez 2008). One instantiation of this idea is that the current syntactic category depends on the previous category, i.e., a bigram generative model (Figure 1).

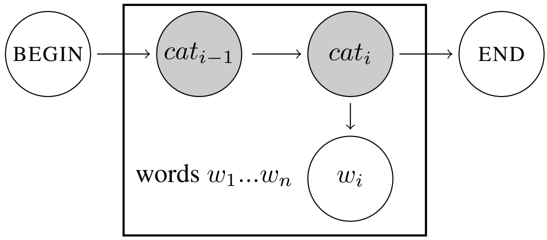

Figure 1: A bigram generative model for words $w_1...w_n$. Words are observed, as are the utterance boundaries indicated by BEGIN and END. Categories are latent.

In the bigram generative model in Figure 1, each word w_i is generated based on its latent category cat_i, which is conditioned on the previous word's latent category cat_{i-1}. To calculate the probability of any sequence $w_1...w_n$, we use Equation 2, which is the product of the probability of generating each word w_i in the utterance U. This involves the probability of generating w_i based on its latent category cat_i ($P(w_i|cat_i)$) multiplied by the probability of generating that latent category, given the previous category cat_{i-1} ($P(cat_i|cat_{i-1})$). The previous category for the first category is the utterance-initial boundary (BEGIN). Additionally, the probability of generating the utterance-final boundary (END) after the last category cat_n is included. We demonstrate this calculation for the utterance *I like nice penguins* in (3), assuming the utterance is represented by the syntactic category sequence PRONOUN VERB ADJ COUNT-N.

$$P(U = w_1...w_n) = \left(\prod_{w_i \in U} P(w_i|cat_i)P(cat_i|cat_{i-1})\right) P(\text{END}|cat_n) \qquad (2)$$

(3) Words: *I* *like* *nice* *penguins*

Categories: BEGIN PRONOUN VERB ADJ COUNT-N END

$$\begin{aligned} P(U = w_1...w_4) &= \left(\textstyle\prod_{w_i \in U}(P(w_i|cat_i)P(cat_i|cat_{i-1})\right) P(\text{END}|cat_n) \\ &= P(\textit{I}|\text{PRONOUN}) * P(\text{PRONOUN}|\text{BEGIN}) \\ &* P(\textit{like}|\text{VERB}) * P(\text{VERB}|\text{PRONOUN}) \\ &* P(\textit{nice}|\text{ADJ}) * P(\text{ADJ}|\text{VERB}) \\ &* P(\textit{penguins}|\text{COUNT-N}) * P(\text{COUNT-N}|\text{ADJ}) \\ &* P(\text{END}|\text{COUNT-N}) \end{aligned}$$

Using perplexity and an implementation of $P(U)$, we can compare different category representations because each will yield a perplexity score for an evaluation dataset. This allows us to quantify the processing ease for that dataset using different hypotheses about what the categories are and which words belong to each category. This means we can compare the processing efficiency impact on the data children encounter for both FF-based categories and adult categories. We posit that the category representation which eases data processing more is more useful to children at this stage of development.

We note that this is a comparative metric only, because a perplexity score is based on the predictability of a particular dataset. For example, a perplexity score of 608 isn't meaningful on its own; instead, it's only meaningful with respect to the dataset used to generate the perplexity score. So, if two category representations are used to generate a perplexity score for a specific dataset, these scores can be compared against each other, with a lower score indicating the data are less perplexing using that category representation. This would mean that that category representation increases language processing efficiency on those data compared to the other representation.

4. Implementation

There are two main implementation details when evaluating the FF categorization strategy: (i) which language data the strategy will be evaluated over, and (ii) practical considerations for the FFs instantiation. We discuss each in turn.

4.1. Language data

For English, we selected the Peter corpus (Bloom, Lightbown, Hood, Bowerman, Maratsos and Maratsos, 1975) from the CHILDES database (MacWhinney, 2000), one of the corpora used by Mintz (2003) which has syntactic categories annotated. This dataset contains 14977 utterances (71813 word tokens, 2227 word types, average utterance length=5.27 words) directed at a child between the ages of 1;9 and 2;4. The 72 syntactic categories in this dataset were derived from the %mor line annotations.

For ASL, we used the BU ASLLRP corpus (Neidle and Vogler, 2012), a newly developed corpus for ASL. This dataset contains 1641 utterances (10820 word tokens, 2321 word types, average utterance length=6.6 signs) directed at adults. The 34 categories in this dataset were derived from the POS annotation line.

While the Peter data may be reasonable as a sample of speech directed at children learning to syntactically categorize, the ASLLRP data likely aren't since they're adult-directed. This is because the differences between child-directed and adult-directed speech are non-trivial (see Ma, Golinkoff, Houston and Hirsh-Pasek 2011 for a review of differences at the prosodic, lexical, and structural levels). However, we are unaware of a corpus of child-directed signed language input currently available (in ASL or any other signed language), and so the adult-directed ASL corpus serves as a first step towards assessing categorization in ASL.

4.2. FF instantiation

The first consideration is what counts as "frequent" for a frequent frame. We chose to use the frequency cutoff used by Chemla et al. (2009) for their FF instantiation: a frame must include at least 0.5% of word types and 0.1% of word tokens to be counted as frequent. This seemed reasonable due to the similar corpus size for our ASL data (1641 utterances) and their French data (2006 utterances).

Another consideration concerns the words that are uncategorized by FFs. This actually wasn't a concern for previous studies that evaluated the accuracy of FF-based categories against adult categories because only words that were captured by FFs were evaluated (the rest were ignored). However, our proposed perplexity measure requires us to know the classification of every word, not just the words captured within the FFs. Two simple options are (i) collapsing all uncategorized words into a single category, or (ii) assuming each uncategorized word is its own individual category. We opted for the latter, based on the intuition that children won't treat things (e.g., words) as similar unless they have a reason to.

A third consideration is which units are the framing units. Previous studies have used either words or morphemes, with English studies typically using word-level frames (e.g., Mintz 2003, Chemla et al. 2009). We followed this for our English evaluation. For ASL, current corpus encoding makes it impractical to use morphemes (sub-sign units are not typically annotated). So, we use sign-level frames. Also, previous studies didn't allow utterance boundaries to be part of frames, while we do (e.g., the initial frame for *I like nice penguins* is START_*like* and the final frame is *nice*_END). This seems plausible, given young children's sensitivity to utterance boundaries (e.g., Longobardi, Rossi-Arnaud, Spataro, Putnick and Bornstein 2015).

A fourth consideration is the number of syntactic categories. Mintz (2003) collapsed the %mor annotation in CHILDES into categories corresponding roughly to "basic" linguistic categories like NOUN and VERB. In contrast, we chose to use the %mor annotations as is for the English corpus, since we were using the POS annotations as is for the ASL corpus. We do note that the number of true categories impacts evaluation both for the traditional approach and our proposed perplexity metric. In the traditional approach, more categories means there are likely to be fewer words in each category. So, FF-based categories may suffer in comparison if they don't make fine-grained enough category distinctions. For perplexity, the number of true categories impacts the probability of the utterance for the true category representation, where more categories means $P(cat_{i-1}|cat_i)$ is likely to be lower on average. However, due to there being fewer words on average per category, $P(w_i|cat_i)$ is likely to be higher. Importantly, perplexity can be assessed for the FF-based categories regardless of how many true categories there are supposed to be.

5. Results and Discussion

5.1. Standard evaluation metric: precision

Precision captures category accuracy and is one of the traditional metrics used to calculate how well inferred syntactic categories match adult categories (e.g., Mintz 2003). For example, an inferred category that has 5 members, all of which are adult category COUNT-NOUN, would have perfect precision. This contrasts with the recall measure, which focuses on category completeness. An inferred category that only has 5 count nouns is not very complete (there are many more words not included that are count nouns), and so would have low recall. We focus on precision, following the intuition of several previous FFs studies that highly precise initial categories are more useful to toddlers for bootstrapping subsequent acquisition processes (Mintz, 2003, Wang and Mintz, 2008, Chemla et al., 2009, Weisleder and Waxman, 2010, Stumper et al., 2011, Wang et al., 2011).

We use the pairwise precision instantiation of Mintz (2003), which calculates precision over the pairs of words in each FF-based category. In particular, for each FF-based category, an exhaustive list of all pairs of words in that category is made. For each word pair, if the two words are in the same adult category, they're counted as a *hit*; if the two words are not in the same adult category, they're counted as a *false alarm*. Pairwise precision for the category is equivalent to $\frac{hits}{hits+false\ alarms}$. So, the pairwise precision score ranges between 0 and 1, with 1.0 indicating perfect precision. This is done for each FF-based category, and the average is taken over all FF-based categories. Table 1 shows the precision results for our analysis of FFs on English and ASL, and also provides comparative results from previous studies on other languages.

Table 1: Precision results for Frequent Frames (FFs) compared against adult categories over different languages, based on previous studies and the current study. Framing units from which frames are constructed are indicated in parentheses (word, morpheme, or sign). Higher precision scores indicate more accurate FF-based categories.

Language	Study	Precision (framing unit)
French	Chemla et al. (2009)	1.00 (word)
Spanish	Weisleder and Waxman (2010)	0.75 (word)
German	Wang et al. (2011)	0.86 (word)
	Wang et al. (2011)	0.88 (morpheme)
Turkish	Wang et al. (2011)	0.47 (word)
	Wang et al. (2011)	0.91 (morpheme)
English	Mintz (2003)	0.98 (word)
	Bar-Sever and Pearl (2016)	0.68 (word)
ASL	Bar-Sever and Pearl (2016)	0.42 (sign)

One key result from Table 1 is that our FF results are much worse for English using word-level frames (precision=0.68) than previous findings. Moreover, when we turn to ASL, sign-level frames fare similarly poorly (precision=0.42). In fact, the sign-level ASL precision is similar to the word-level precision found for Turkish (Turkish=0.47). What can be made of this?

First, one interesting observation is that FF-based categories are similar irrespective of modality – they're just similarly *poor* with respect to matching the adult categories. Second, the implementation of FFs matters, beyond the unit used to construct frames. For English, our results may differ from those of Mintz (2003) because we allow utterance boundaries to be used as framing units. In addition, as discussed in section 4.2, to make the comparison as similar as possible between English and ASL, we relied on the categories annotated in the English corpus's %mor line, and this yielded more adult categories than Mintz (2003) used.

The simple point is that these FF-based categories don't match this version of adult categories in either English or ASL very well. However, perhaps these initial categories are useful in other ways – specifically, by helping toddlers process incoming language data more easily, which is measured by perplexity.

5.2. Perplexity evaluation

Recall from section 3 that the perplexity calculation relies on two kinds of probabilities: (i) the transition probabilities between categories ($P(cat_i|cat_{i-1})$), and (ii) the emission probabilities of words being generated by a specific category ($P(w_i|cat_i)$). To calculate perplexity on new language data, the learner must already have some idea of these probabilities given the previous data encountered. So, to estimate these probabilities, we split the language corpus into a training set (consisting of 90% of the data) and a test set (consisting of the remaining 10%). Transition and emission probabilities are estimated from the training set and used in the perplexity calculation of the test set. We calculate perplexity for each utterance in the test set and then take the average.[1]

We use 10-fold cross-validation, which rotates which section of the corpus is the test set. Here, this means there are ten different splits of the corpus into training and test set, with each tenth serving as the test set in one split. So, we get ten perplexity scores, one for each tenth of the corpus, and average those to get the perplexity score for the corpus.

Table 2 shows the perplexity scores for English and ASL using both the FF-based categories and the adult categories. Recall that perplexity is inversely related to probability, so lower perplexity scores indicate higher probability data that are less perplexing and so easier to process. We can see that for both languages the FF-based categories make new data less perplexing than the adult categories do (English: FFs=122.6 vs. Adult=607.9; ASL: FFs=9.8 vs. Adult=45.5).

[1] To prevent assigning a probability of 0, we use add-0.5 smoothing for transition or emission instances that were not observed in the training set but appear in the test set.

Table 2: Perplexity scores on English and ASL for the category representation derived from Frequent Frames (FFs) and the adult category representation (Adult). FF-based categories make new data less perplexing for both languages.

	Perplexity	
Category Representation	**English**	**ASL**
FFs	122.6	9.8
Adult	607.9	45.5

This result may seem counterintutive – why would adult categories not be best for processing language that's generated by adults? The answer has to do with the structural knowledge we give our modeled toddlers. In particular, in the perplexity calculation, the probability of an utterance ($P(U)$) is based on an immature assumption about how categories relate to each other, i.e., a bigram assumption. This structural knowledge will surely be refined as linguistic development continues, but it impacts what kind of categories are most useful to a toddler at this stage of development. In short, at this stage of development, the perplexity results indicate that FF-based categories are better than adult categories. In this sense, the FF categorization strategy is very successful indeed.

5.3. Future directions

These results suggest the possibility of an interesting synergy between children's developing representations of syntactic categories and their immature representations of syntactic structure. In particular, if toddlers only posit fairly local sequential relationships for syntactic structure, the syntactic categories that FFs yield are more useful for processing new language data than adult syntactic categories are. However, as these structural assumptions change, so too may the utility of adult-like syntactic categories.

Still, these results are preliminary for ASL at least, because of the current lack of a child-directed ASL corpus to evaluate the FFs strategy on. One useful comparative analysis is to evaluate FFs on an English adult-directed corpus with the same surface properties as the adult-directed ASL corpus (e.g., number of utterances). Then, we could observe the differences between the results on English adult-directed and child-directed speech, and extrapolate these to infer what results might be for child-directed ASL, based on our adult-directed ASL results.

Beyond this, we have only examined the performance of word/sign-level FFs. Because ASL shares linguistic properties with languages like Turkish (e.g., variable word order, rich morphology), it may be that ASL toddlers construct frames from sub-word units. In spoken languages, morphological units may form the basis for frames (e.g., in English, *She is eating chocolate* would include the frame *is_ing*), and morpheme-based FFs yield more adult-like categories in Turkish (Wang et al., 2011). In signed languages, an interesting implementation issue arises if we attempt to create frames from sub-sign units. In particular, unlike in spoken languages, the sub-sign units are simultaneously articulated. That is, a sign representing *eating* has (among others) a sub-sign unit representing EAT and a sub-sign unit representing IMPERFECTIVE-ASPECT (signaled in English by *-ing*), and these units occur at the same time, rather than sequentially. So, how would an ASL toddler construct sub-sign frames? That is, in the above example, which unit is part of the frame and which is the unit that's framed? More than one reasonable answer is possible.

In addition, we may wish to consider alternative evaluation metrics that are also sensitive to the goal of early categorization. In particular, because we know that toddlers seem to recognize certain adult categories earlier (e.g., COUNT-NOUN, ADJECTIVE: Booth and Waxman 2003), we may wish to use traditional comparison metrics (e.g., precision) for *only* those categories. That is, early categorization strategies may need to generate something like the adult COUNT-NOUN category but not the adult DETERMINER category.

The perplexity metric presented here can also be implemented with different ideas about the structural assumptions toddlers have at this stage of development. For example, perhaps toddlers assume words are contained in shallow syntactic skeletons derived from function words and prosodic boundaries (e.g., Christophe, Millotte, Bernal and Lidz 2008, Cauvet, Limissuri, Millotte, Skoruppa, Cabrol and Christophe 2014). Any concrete hypothesis about young children's structural representation of words and categories can easily be used in the perplexity metric in place of the bigram-based structural assumption we used here.

While perplexity assesses a particular category representation by how useful it is for processing language, we can also investigate how useful a representation is for directly scaffolding knowledge that depends on that representation (Phillips and Pearl, 2015). For example, syntactic categories are the basis for syntactic rules that capture language structure (e.g., NP $\rightarrow$ DETERMINER COUNT-NOUN). Perfors, Tenenbaum and Regier (2011) demonstrated that a modeled learner can infer than English syntactic rules are hierarchical rather than linear, given plausible child-directed speech data and adult-like syntactic category knowledge. Would this same inference be possible with the non-adult FF-based categories? If so, this suggests (a) the hierarchical structure inference can happen quite early (as soon as the toddler has some FF-based categories), and (b) FF-based categories are good enough to support this inference. In contrast, if the hierarchical inference fails when FF-based categories are used, this suggests that FF-based categories are not good enough to scaffold the inference; instead, this inference would only occur after children's syntactic category representations are more adult-like. Either way, we will have learned relevant information about both the FFs strategy and the time course of hierarchical structure inference.

Also, when exploring early categorization strategies more generally, we may wish to incorporate other cues toddlers are known to be sensitive to. For example, toddlers may consider semantic cues when forming categories, given their early recognition of nouns like body items and food items (Bergelson and Swingley, 2012). So, early categories that have more semantic coherency may be favored. Relatedly for the FFs strategy, framing elements may come not just from frequent words but also from words whose meaning is recognized early, with frames that are (i) frequent and (ii) contain these familiar words being even more salient for children. In a similar vein, given infant sensitivity to edge words, frames involving words at utterance edges may also be especially salient. We note that while our current FFs implementation allowed frames to consist of the utterance boundaries themselves, no special weight was given to frames involving edge words.

6. Conclusion

Linking children's developing knowledge representations to their developing language processing abilities may help explain why children's developing knowledge representations look the way they do. Here, using a new processing-based evaluation metric, we have found suggestive evidence that syntactic categories derived from Frequent Frames ease children's language processing by making new language data more predictable. Interestingly, this inferred category representation – while not matching adult categories very well – makes new language data more predictable than the adult representation would at this stage of development. This is true irrespective of language modality, suggesting it is a general property of the Frequent Frames early categorization strategy.

References

BERGELSON, ELIKA; and DANIEL SWINGLEY. 2012. At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences* 109(9).3253–3258.

BLOOM, LOIS; PATSY LIGHTBOWN; LOIS HOOD; MELISSA BOWERMAN; MICHAEL MARATSOS; and MICHAEL P MARATSOS. 1975. Structure and variation in child language. *Monographs of the society for Research in Child Development* 1–97.

BOOTH, AMY; and SANDRA WAXMAN. 2003. Mapping words to the world in infancy: On the evolution of expectations for nouns and adjectives. *Journal of Cognition and Development* 4(3).357–381.

BROWN, PETER F; VINCENT J DELLA PIETRA; ROBERT L MERCER; STEPHEN A DELLA PIETRA; and JENNIFER C LAI. 1992. An estimate of an upper bound for the entropy of English. *Computational Linguistics* 18(1).31–40.

CAUVET, ELODIE; RITA LIMISSURI; SÉVERINE MILLOTTE; KATRIN SKORUPPA; DOMINIQUE CABROL; and ANNE CHRISTOPHE. 2014. Function words constrain on-line recognition of verbs and nouns in french 18-month-olds. *Language Learning and Development* 10(1).1–18.

CHEMLA, EMMANUEL; TOBEN H MINTZ; SAVITA BERNAL; and ANNE CHRISTOPHE. 2009. Categorizing words using 'Frequent Frames': What cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science* 12(3).396–406.

CHRISTOPHE, ANNE; SÉVERINE MILLOTTE; SAVITA BERNAL; and JEFFREY LIDZ. 2008. Bootstrapping lexical and syntactic acquisition. *Language and Speech* 51(1-2).61–75.

CONNOR, MICHAEL; CYNTHIA FISHER; and DAN ROTH. 2013. Starting from scratch in semantic role labeling: Early indirect supervision. *Cognitive aspects of computational language acquisition*, 257–296. Springer.

CONNOR, MICHAEL; YAEL GERTNER; CYNTHIA FISHER; and DAN ROTH. 2010. Starting from scratch in semantic role labeling. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 989–998. Association for Computational Linguistics.

ERKELENS, MARIAN. 2009. *Learning to categorize verbs and nouns: studies on dutch*. Netherlands Graduate School of Linguistics.

FERNALD, ANNE; and VIRGINIA A MARCHMAN. 2012. Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development* 83(1).203–222.

FERNALD, ANNE; AMY PERFORS; and VIRGINIA A MARCHMAN. 2006. Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology* 42(1).98.

FERNALD, ANNE; JOHN P PINTO; DANIEL SWINGLEY; AMY WEINBERGY; and GERALD W MCROBERTS. 1998. Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science* 9(3).228–231.

FRANK, STELLA; SHARON GOLDWATER; and FRANK KELLER. 2009. Evaluating models of syntactic category acquisition without using a gold standard. *Proceedings of the 31st Annual Conference of the Cognitive Science Sociey*, 2576–2581.

GÓMEZ, REBECCA L; and LAURA LAKUSTA. 2004. A first step in form-based category abstraction by 12-month-old infants. *Developmental science* 7(5).567–580.

GUTMAN, ARIEL; ISABELLE DAUTRICHE; BENOÎT CRABBÉ; and ANNE CHRISTOPHE. 2014. Bootstrapping the syntactic bootstrapper: Probabilistic labeling of prosodic phrases. *Language Acquisition* 22.285–309.

HALE, JOHN. 2001. A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, 1–8. Association for Computational Linguistics.

HICKOK, GREGORY; TRACY LOVE-GEFFEN; and EDWARD S KLIMA. 2002. Role of the left hemisphere in sign language comprehension. *Brain and Language* 82(2).167–178.

LANY, JILL; and REBECCA L GÓMEZ. 2008. Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science* 19(12).1247–1252.

LEVY, ROGER. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3).1126–1177.

LIDZ, JEFFREY; and ANNIE GAGLIARDI. 2015. How Nature Meets Nurture: Universal Grammar and Statistical Learning. *Annual Review of Linguistics* 1(1).333–352.

LILLO-MARTIN, DIANE C; and JON GAJEWSKI. 2014. One grammar or two? Sign Languages and the Nature of Human Language. *Wiley Interdisciplinary Reviews: Cognitive Science* 5(4).387–401.

LONGOBARDI, EMIDDIA; CLELIA ROSSI-ARNAUD; PIETRO SPATARO; DIANE L PUTNICK; and MARC H BORNSTEIN. 2015. Children's acquisition of nouns and verbs in italian: contrasting the roles of frequency and positional salience in maternal language. *Journal of Child Language* 42(01).95–121.

MA, WEIYI; ROBERTA MICHNICK GOLINKOFF; DEREK M HOUSTON; and KATHY HIRSH-PASEK. 2011. Word learning in infant-and adult-directed speech. *Language Learning and Development* 7(3).185–201.

MACWHINNEY, BRIAN. 2000. *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

MINTZ, TOBEN. 2003. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90.91–117.

MINTZ, TOBEN. 2006. Finding the verbs: Distributional cues to categories available to young learners. *Action meets word: How children learn verbs*, ed. by Kathy Hirsh-Pasek and Roberta Golinkoff, 31–63. Oxford: Oxford University Press.

NEIDLE, CAROL; and CHRISTIAN VOGLER. 2012. A new web interface to facilitate access to corpora: Development of the ASLLRP data access interface (DAI). *Proceedings 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC*, Citeseer.

OMAKI, AKIRA; and JEFFREY LIDZ. 2015. Linking Parser Development to Acquisition of Syntactic Knowledge. *Language Acquisition* 22(2).158–192.

PERFORS, AMY; JOSHUA TENENBAUM; and TERRY REGIER. 2011. The learnability of abstract syntactic principles. *Cognition* 118.306–338.

PHILLIPS, LAWRENCE; and LISA PEARL. 2015. Utility-based evaluation metrics for models of language acquisition: A look at speech segmentation. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, Denver, Colorado: NAACL.

STUMPER, BARBARA; COLIN BANNARD; ELENA LIEVEN; and MICHAEL TOMASELLO. 2011. "frequent Frames" in German Child-Directed Speech: A Limited Cue to Grammatical Categories. *Cognitive science* 35(6).1190–1205.

WANG, HAO; BARBARA HÖHLE; NF KETREZ; AYLIN C KÜNTAY; TOBEN H MINTZ; N DANIS; K MESH; and H SUNG. 2011. Cross-linguistic distributional analyses with Frequent Frames: The cases of German and Turkish. *Proceedings of 35th Annual Boston University Conference on Language Development*, 628–640. Cascadilla Press Somerville, MA.

WANG, HAO; and TOBEN MINTZ. 2008. A dynamic learning model for categorizing words using frames. *Proceedings of the 32nd Annual Boston University Conference on Language Development [BUCLD 32]*, ed. by Harvey Chan, Heather Jacob and Enkeleida Kapia, 525–536. Somerville, MA: Cascadilla Press.

WEISLEDER, ADRIANA; and ANNE FERNALD. 2013. Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science* 24(11).2143–2152.

WEISLEDER, ADRIANA; and SANDRA R WAXMAN. 2010. What's in the input? Frequent frames in child-directed speech offer distributional cues to grammatical categories in spanish and english. *Journal of Child Language* 37(05).1089–1108.

XIAO, LING; XIN CAI; and THOMAS LEE. 2006. The development of the verb category and verb argument structures in Mandarin-speaking children before two years of age. *Proceedings of the Seventh Tokyo Conference on Psycholinguistics*, ed. by Yukio Otsu, 299–322. Tokyo: Hitizi Syobo.

Proceedings of the 40th annual Boston University Conference on Language Development

edited by Jennifer Scott
and Deb Waughtal

Cascadilla Press Somerville, MA 2016

Copyright information

ISSN 1080-692X
ISBN 978-1-57473-066-1 (2 volume set, paperback)
ISBN 978-1-57473-166-8 (2 volume set, library binding)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com