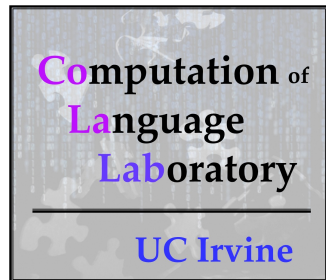# Syntactic categories derived from frequent frames benefit early language processing in English and ASL

Galia Barsever

Lisa Pearl

University of California, Irvine

Computation of Language Laboratory

UC Irvine

BUCLD
November 13, 2015

Introduction

Early acquisition strategies likely don't yield adult knowledge directly, but instead provide a stepping stone to later knowledge (Frank et al. 2009, Phillips & Pearl 2015).

???????

Linking to language processing

How can we link this developing knowledge to processing?

- How are developing knowledge representations used in language processing?

- How do different language modalities (which may be processed differently) affect acquisition strategies?

One way developing knowledge is used

One beneficial effect of preliminary knowledge could be that language input becomes easier for children to process, given their limited cognitive resources.

- more useful knowledge = easier to process

Different modalities

– Representationally, we think signed languages and spoken languages are represented in the same way (Lillo-Martin & Gajewski 2014)

Different modalities

– Representationally, we think signed languages and spoken languages are represented in the same way (Lillo-Martin & Gajewski 2014)

– Idea: If representations are learned the same way as spoken language, then a strategy that works for spoken language should work for sign language

Roadmap

- **Case study: Grammatical categorization**
  - focusing on the frequent frames (FFs) learning strategy (Mintz 2003)

- **Proposed metric:** assesses processing ease for a learner using category knowledge

- **Model implementation**

- **Take-home points**:
  - **Strategy works the same way** for a spoken language (English) and a signed language (ASL)
  - **FF-based categories make processing easier** than adult categories at early stage of development

Roadmap

- **Case study: Grammatical categorization**
  - focusing on the frequent frames (FFs) learning strategy (Mintz 2003)

- **Proposed metric:** assesses processing ease for a learner using category knowledge

- **Model implementation**

- **Take-home points**:
  - **Strategy works the same way** for a spoken language (English) and a signed language (ASL)
  - **FF-based categories make processing easier** than adult categories at early stage of development

Grammatical Categorization

– Foundational for syntax

– Basic intuition: If we recognize that different words belong to abstract categories, language becomes easier to predict because of the underlying structure.
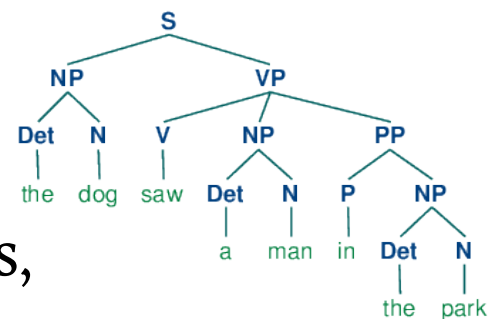
– This make language easier to process!

  • Formalized in surprisal theory of sentence processing (Levy 2008)

Grammatical Categorization

– Foundational for syntax

– Basic intuition: If we recognize that different words belong to abstract categories, language becomes easier to predict because of the underlying structure.
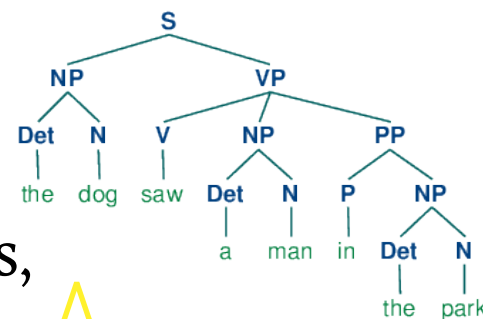
high probability = easier to process

 – This make language easier to process!
  • Formalized in surprisal theory of sentence processing (Levy 2008)

What are you learning in the early stages of categorization?

- Might not know all nouns (adult knowledge)

- Example: you know dog and kitty are the same type of thing but you don't know dog and idea are, too.
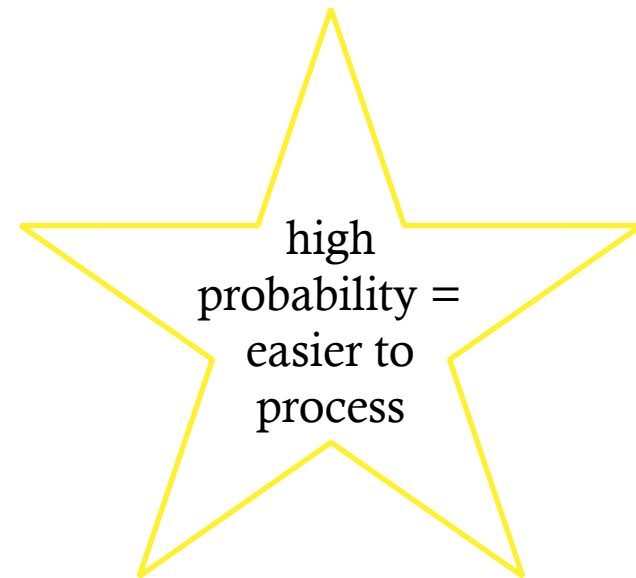
© Barry Bland / Barcroft USA

What's important:
How useful is this early knowledge?

One example:
Can the information about nouns that you learned help you process language faster or better?

Does this knowledge make language more predictable?

high probability = easier to process

Frequent Frames as a strategy: When and Why

– Frequent Frames (FFs) is a computationally inexpensive strategy intended for the beginning stages of grammatical categorization

  – Would be used before 12-14 months (Booth & Waxman 2003)

Frequent Frames as a strategy: When and Why

– Frequent Frames (FFs) is a computationally inexpensive strategy intended for the beginning stages of grammatical categorization

  – Would be used before 12-14 months (Booth & Waxman 2003)

– Basic intuition:

  • Frequently occurring frames tend to identify words that behave similarly in sentences

  • Behave similarly = they appear in the same context

    • Context = surrounding words

Frequent Frames: Frames

Example with words as frames:

For a sequence of words XYZ, X__Z is the frame, and Y is the thing identified.

I am hugging nice penguins. I am petting nice kitties.

am__nice captures: hugging and petting

Frequent Frames: Frequent

What's "Frequent"?

Intuition: Too many things are hard to pay attention to
—> Fewer is easier

Frequent things are likely to be salient, so a few very frequent things are likely to be in the child's intake (as opposed to all the things in the input).

Frequent Frames: Experimental Basis

– Children at early learning stages shown to be able to identify and use FFs (Mintz 2006), as well as non-adjacent dependencies (Lany & Saffran 2011)

Frequent Frames: Computational Justification

– Shown to work well on various spoken languages (Mintz 2003, Chemla et al. 2009, Weisleder & Waxman 2010,Wang et al. 2011)

Big Questions

Two questions:

1) Are FFs useful for early processing?

2) How well do FFs identify grammatical categories in ASL, a language with a different modality?

Roadmap

- **Case study: Grammatical categorization**
  - focusing on the frequent frames (FFs) learning strategy (Mintz 2003)

- **Proposed metric:** assesses processing ease for a learner using category knowledge

- **Model implementation**

- **Take-home points**:
  - **Strategy works the same way** for a spoken language (English) and a signed language (ASL)
  - **FF-based categories make processing easier** than adult categories at early stage of development

Frequent Frames: Measuring Success Before

- What does it mean for FFs to be successful?
  - Previously: Compared against adult knowledge
- Reason: (practical) It's what we had available

Frequent Frames:
Measuring Success Before

- What does it mean for FFs to be successful?

  - Previously: Compared against adult knowledge

- Reason: (practical) It's what we had available
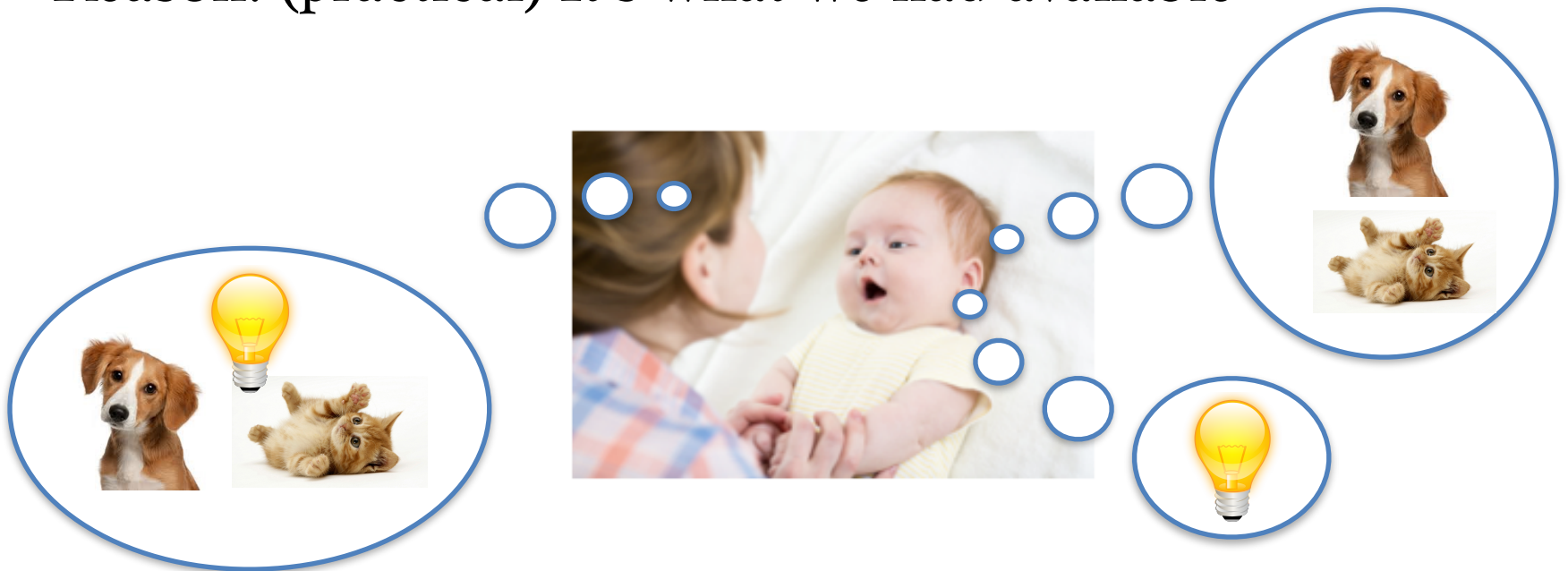
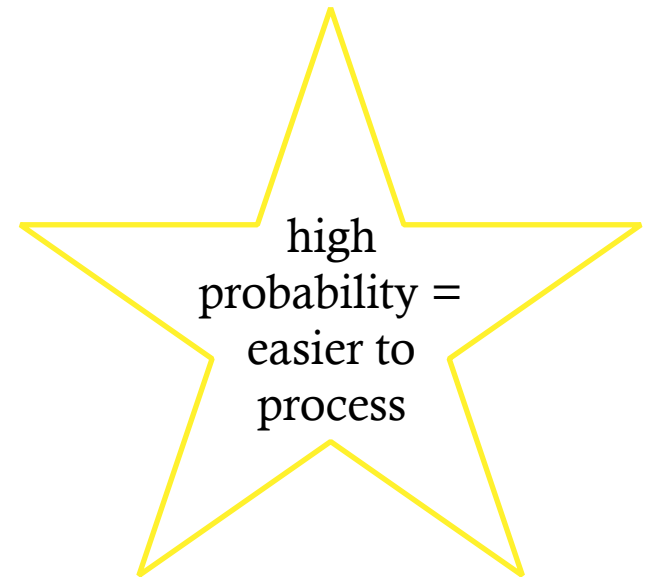- **But it means we expect *dog*, *kitty*, and *idea* to go together**

Introducing Our Utility Metric

Remember:

Making language easier to process =

Making language more predictable

A formal definition of more predictable:
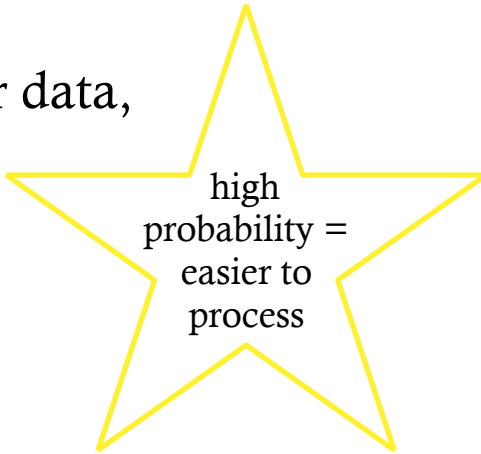Perplexity (Brown et al. 1992)

high probability = easier to process

Introducing the computational measure of perplexity

- We're getting this from computational linguists who use this to measure how effective their models are

- Perplexity is inversely related to probability, so low perplexity = higher probability of seeing the data

$$Perplexity(Words = w_1 w_2 ... w_N) = \sqrt[N]{\frac{1}{P(start - w_1 w_2 ... w_N - end)}}$$

If the probability is low because we didn't expect our data, then whatever we learned is not helpful.

high probability = easier to process

Introducing the computational measure of perplexity

$$Perplexity(Words = w_1 w_2 ... w_N) = \sqrt[N]{\frac{1}{P(start - w_1 w_2 ... w_N - end)}}$$

- ## Example of low perplexity:

$$Perplexity("I - like - nice - penguins") = \sqrt[4]{\frac{1}{probability(start - I - like - nice - penguins - end)}}$$

Number near 1!

Perplexity —> near 1!

Highly likely = low perplexity

Introducing the computational measure of perplexity

$$Perplexity(Words = w_1 w_2 ... w_N) = \sqrt[N]{\frac{1}{P(start - w_1 w_2 ... w_N - end)}}$$
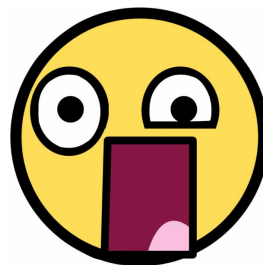
- Example of high perplexity:

$$Perplexity("Nice - I - penguins - like) = \sqrt[N]{\frac{1}{probability(start - Nice - I - penguins - like - end)}}$$
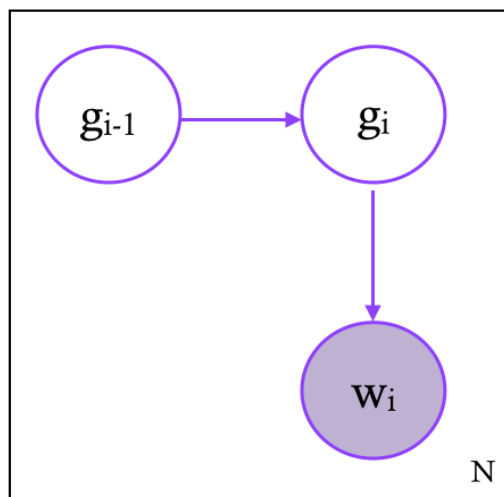
Small number!

Perplexity —> big number!

Not likely = high perplexity

How we calculate probability of a sentence

How do we get P(start-w_1…w_n-end)?

Bigram model:
current category depends
on previous category

Latent:
Grammatical categories
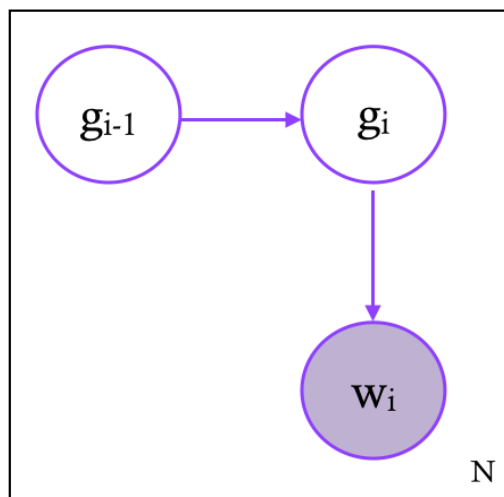
Observed:
Words

Naïve assumption of early language representation
which we think may be a reasonable approximation
of a child's developing structural knowledge.

How we calculate probability of a sentence

How do we get $P(\text{start-}w_1 \ldots w_n\text{-end})$?

Latent:
Grammatical categories

Observed:
Words

$P(\text{start-}w_1 \ldots w_n\text{-end}) = p(g_1|\text{start}) * p(w1|g_1)$
$\qquad * \; p(g_2|g_1) * p(w_2|g_2)$
$\qquad * \; p(g_3|g_2) * p(w_3|g_3)$
$\qquad * \; \ldots$
$\qquad * \; p(g_n|g_{n-1}) * p(w_n|g_n)$
$\qquad * \; p(\text{end}|g_n)$
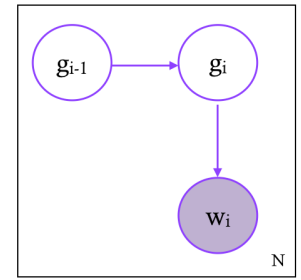
How do we get P(start-w_1…w_n-end)?

Sample calculation of the probability of "I like nice penguins"

Start → Pronoun

↓

I

# of times a pronoun follows start

P(start-I-like-nice-penguins-end) = p(pronoun|start) * p(I|pronoun)

# of times a "I" appears out of all pronouns

How do we get $P(\text{start-}w_1\ldots w_n\text{-end})$?

Sample calculation of the probability of "I like nice penguins"

Start → Pronoun → Verb
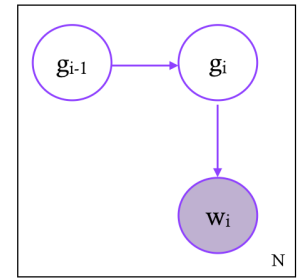
↓ ↓
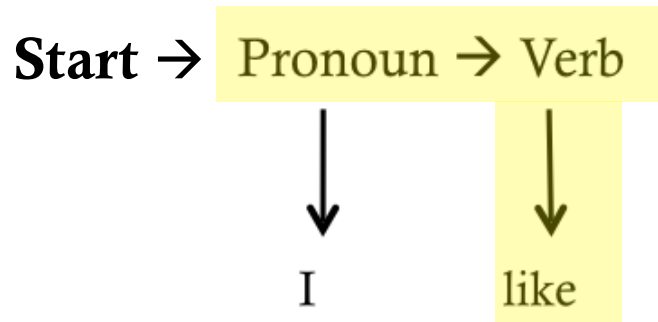
I like

$P(\text{start-I-like-nice-penguins-end}) = p(\text{pronoun}|\text{start}) * p(\text{I}|\text{pronoun})$
  $*p(\text{verb}|\text{pronoun}) * p(\text{like}|\text{verb})$

How do we get P(start-w_1…w_n-end)?

Sample calculation of the probability of "I like nice penguins"

Start → Pronoun → Verb → Adjective

↓   ↓   ↓

I   like   nice

P(start-I-like-nice-penguins-end) = p(pronoun|start) * p(I|pronoun)
      *p(verb|pronoun) * p(like|verb)
      *p(adjective|verb) * p(nice|adjective)

How do we get P(start-w_1...w_n-end)?

Sample calculation of the probability of "I like nice penguins"

Start → Pronoun → Verb → Adjective → Noun

↓ I ↓ like ↓ nice ↓ penguins

P(start-I-like-nice-penguins-end) = p(pronoun|start) * p(I|pronoun)
     *p(verb|pronoun) * p(like|verb)
     *p(adjective|verb) * p(nice|adjective)
     *p(noun|adjective) * p(penguins|noun)

How do we get $P(\text{start-}w_1\ldots w_n\text{-end})$?

Sample calculation of the probability of "I like nice penguins"
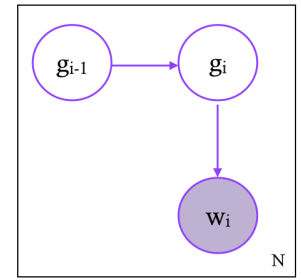
Start → Pronoun → Verb → Adjective → Noun → **End**
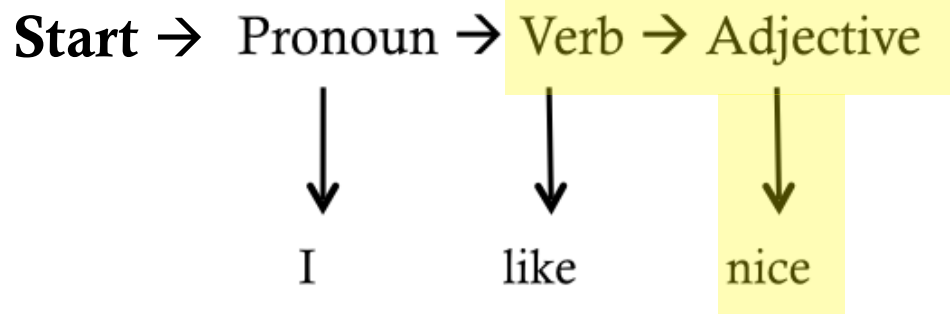
| I | like | nice | penguins |

$P(\text{start-I-like-nice-penguins-end}) = p(\text{pronoun}|\text{start}) * p(\text{I}|\text{pronoun})$
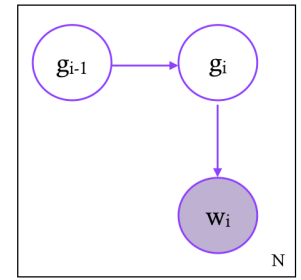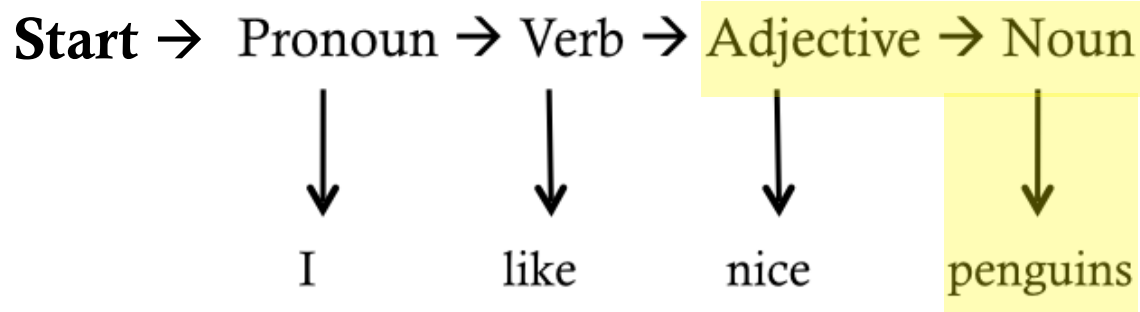$\quad *p(\text{verb}|\text{pronoun}) * p(\text{like}|\text{verb})$
$\quad *p(\text{adjective}|\text{verb}) * p(\text{nice}|\text{adjective})$
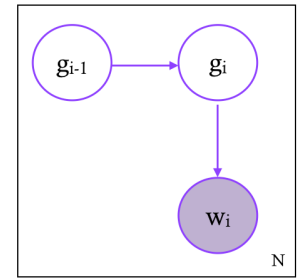$\quad *p(\text{noun}|\text{adjective}) * p(\text{penguins}|\text{noun})$
$\quad *p(\text{end}|\text{noun})$

Adult knowledge vs Processing ease

- Metrics of success:

- Standard way: Measuring against adult knowledge
- New way: Measuring processing ease as predictability

Adult knowledge vs Processing ease

Training:

Use word-level FFs to identify categories in English and ASL, using realistic data in both languages

Test:

Standard way:

Compare against adult grammatical categories

New way:

Compare FF-based categories and adult grammatical categories on their ability to predict language data (=comprehension)

Roadmap

- **Case study: Grammatical categorization**
  - focusing on the frequent frames (FFs) learning strategy (Mintz 2003)

- **Proposed metric:** assesses processing ease for a learner using category knowledge

- **Model implementation**

- **Take-home points**:
  - **Strategy works the same way** for a spoken language (English) and a signed language (ASL)
  - **FF-based categories make processing easier** than adult categories at early stage of development

Corpus Stats

English: Peter corpus (Bloom, Hood, & Lightbown, 1974; Bloom, Lightbown, & Hood, 1975) from CHILDES (MacWhinney, 2000)

ASL: BU ASLLRP (Neidle, C. & Vogler, C. 2012)

	tokens	*types*	*# utt*	*average utterance length*	*categories*
English (child-directed)	13039	930	3484	5.27 words	72 (as derived from %mor line tagging
ASL (adult-directed*)	10820	2321	1641	6.6 signs	34 (as derived from corpus) annotation)

*Ask me about how this impacts results!

Example of English Corpus

- One utterance in the English corpus:

*LOI: are you a fish ?
%mor: **cop**|be&PRES **pro**|you **det**|a **n**|fish ?

Example of ASL corpus

- One utterance in the ASL corpus:

Full gloss for roadtrip2.xml-20

A lot of things go into an ASL utterance that we don't consider when using an English or other spoken language corpus.

We used main gloss and Part of Speech tagging to run our model

Practical Decisions

- The choice points:
  - What counts as frequent?

  - What to do with words uncategorized by FFs?

  - Framing units

Choice point: Frequency

- We chose to implement Chemla et al.'s (2009) cutoff for frequency: they used the ones that grouped together at least 0.5% of types and 0.1% of tokens in the corpus (6 frames total)

    - Note: Differs from Mintz (2003), who used the top 0.13%

- Justification: Similar size of corpus between the French and ASL

Choice point: Words not in FFs

This was not a concern for previous implementations which were just comparing against adult categories. Why not?

Because they only cared about the accuracy of FF categories, so words not in those categories were ignored for purposes of evaluation.

Choice point: Words not in FFs

Our measure requires all words to be in a category, so that the probability of a sentence can be calculated.

$$Perplexity(Words = w_1 w_2 ... w_N) = \sqrt[N]{\frac{1}{P(start - w_1 w_2 ... w_N - end)}}$$

Choice point: Words not in FFs

- Choice: Individual category or one large one?

  - We chose each one = individual category.

  - Why? Intuition: Children will not think words are the same type of thing unless they have a reason to.

Choice point: Framing units

- Can use either words or morphemes
  - Unrealistic to use morphemes at this stage for ASL, so using "words" for both modalities
  - ASL: word = sign

Choice point: Framing units

- Can decide to include utterance boundaries in frames or not
    - Utterance boundaries have been shown to be highly salient (e.g., Seidl & Johnson 2006) so we chose to include them

Previous work: Category Precision

- Precision is a typical measure for category accuracy
  - Highly precise categories are comprised of a single target category (ex: a category made up mostly of nouns)
  - Note: Highly precise categories are thought to be more useful for early acquisition as compared to complete categories (ex: a category that includes all the nouns)

Complete!

Precise!

Previous work: Precision

- FF precision is generally good.
- Note: Frame unit can matter.

	English (Mintz 2003)	**French** (Chemla et al 2009)	**Spanish** (Waxman & Weisleder 2010)	**German** (Wang et al 2011)	**Turkish** (Wang et al 2011)
Avg precision	Word: 98%	Word: 100%	Word: 75%	Word: 86% Morpheme: 88%	Word: 47% Morpheme: 91%

Roadmap

- **Case study: Grammatical categorization**
  - focusing on the frequent frames (FFs) learning strategy (Mintz 2003)

- **Proposed metric:** assesses processing ease for a learner using category knowledge

- **Model implementation**

- **Take-home points**:
  - **Strategy works the same way** for a spoken language (English) and a signed language (ASL)
  - **FF-based categories make processing easier** than adult categories at early stage of development

Our Results: Precision

- How do FF categories compare with adult level knowledge?

Note: Precision scores range from 0.0 (worst) to 1.0 (perfect)

	FF-based: Precision
ASL	**0.415**
English	**0.248**

Our Results: Precision

- How do FF categories compare with adult level knowledge?

Note: Precision scores range from 0.0 (worst)
to 1.0 (perfect)

	FF-based: Precision
ASL	**0.415**
English	**0.248**

What happened?

Aren't FFs supposed to work well at identifying adult knowledge?

Things to note:

(1) Precision is the same for both English and ASL

- No modality difference.

Aren't FFs supposed to work well at identifying adult knowledge?

Things to note:

(1) Precision is the same for both English and ASL

- No modality difference.

(2) Bad at getting adult category knowledge.

- But what if toddlers don't need that? Instead, they want knowledge that helps them process language better.

Results: Perplexity

- How useful are the categories frequent frames identify at processing sentences by predicting what's coming?

- Remember: Perplexity =~ inverse probability of an utterance

1   infinity

low perplexity   high perplexity

	Perplexity:	**Perplexity:**
	FF-based categories	*Adult categories*
ASL	**9.8**	**45.5**
English	**122.6**	**607.9**

Results Summary

Categories inferred using the FF strategy, while not similar to adult level knowledge, ease processing for an early learner

Future Work

Evaluate FFs on child-directed ASL corpus (might reasonably impact results).

Ask me!

Future Work

If we do want early categories to resemble adult categories, maybe we don't need to get *all* of them.

- Maybe only need subset (noun, verb, adjectives) to be correct and not categories like determiners and auxiliaries

Ask me!

Future Work

Output validity: Does the learned knowledge scaffold future acquisition (Phillips & Pearl 2015)?

Grammatical categories scaffold syntactic knowledge (example: hierarchical structure (Perfors et al. 2011))

Ask me!

Bigger Take-Home Point

Linking representations to processing may give idea about why intermediate representations look the way they do.

- Serving function of language use and comprehension, even though ultimate goal is adult representation (which in theory will serve it even better).

thank you

- Thanks to Lisa Pearl, Chen Qin, Tiffany Ng, Chloe Haviland, Sebastian Reyes, Yvonne Kim, and Shawn Abrahamson