



How we can tell what children are actually learning about language and whether it's crazy

KATIE KHUU

FACULTY MENTOR: DR. LISA S. PEARL
DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF CALIFORNIA, IRVINE

How do infants learn language as well and as fast as they do?

WITHOUT SUPERVISION

- Learning by themselves, without someone to tell them what to say or how to say things

SYNTACTIC CATEGORY ACQUISITION

- A way of representing the **functions** of words (how we use them) in language

SYNTACTIC CATEGORY ACQUISITION



How does one learn syntactic categories?

“WUG”

“I have a wug.”

NOUN

“I wug cats.”

VERB

ADULT KNOWLEDGE

“I wug cats.”

Pronoun

Verb

Noun

“They chase rabbits.”

“He wants cookies.”

“She craves chicken.”

“WUG”

“I have a wug.”

???

“I wug cats.”

???



Infants, **aged 12-14 months**, have basic knowledge of syntactic categories.

(Booth and Waxman, 2003)

Infants, aged 12-14 months, have basic knowledge of syntactic categories.

- **How might they learn it?**
- **What cues do they use?**

(Booth and Waxman, 2003)

DISTRIBUTIONAL HYPOTHESIS

- Words that appear in similar contexts tend to have the same **functions**
- use the surrounding words – **the distribution of words** – to help determine what a particular word’s function is

DISTRIBUTIONAL HYPOTHESIS

“I have a wug.”

NOUN

“I wug cats.”

VERB

“I have a cat.”

“I hug cats.”

“I have a penguin.”

“I like cats.”

Distribution of words is a cue to help children learn syntactic categories.

Distribution of words is a cue to help children learn syntactic categories.



Word order is a kind of distributional cue.

Word order = distributional cue

Involves the words before and the words after – looks at the distribution of words

“**Wugs** are nice.”
 “I **wug** cats.”
 “I like **wugs**.”

Word order differs by **utterance type**.

How?



What is an utterance?

- Utterance: smallest, continuous unit of speech

What are utterance types?

- Utterances differ by communicative purposes (e.g. statement, question, command).

What are utterance types?

“Did you eat my candy?” **QUESTIONS**

“She ate my candy!” **DECLARATIVE**

“Don’t eat my candy!” **IMPERATIVE**

What are utterance types?

“Did you eat my candy?”

“She ate my candy!”

“Don’t eat my candy!”

Word order differs by utterance types.

How?



Utterance Types

“I like penguins.”	“Do you like penguins?”
“He wants cookies.”	“Does he want cookies?”
“She took a piece.”	“What did she take?”
“They need a hand.”	“What do they need?”

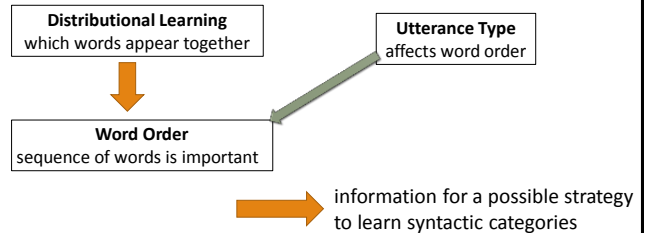
PRONOUN VERB (DETERMINER) NOUN Begins with: AUXILIARY VERB, WH-WORD...
 DECLARATIVE NOUN Ends with: question mark QUESTION -WORD...

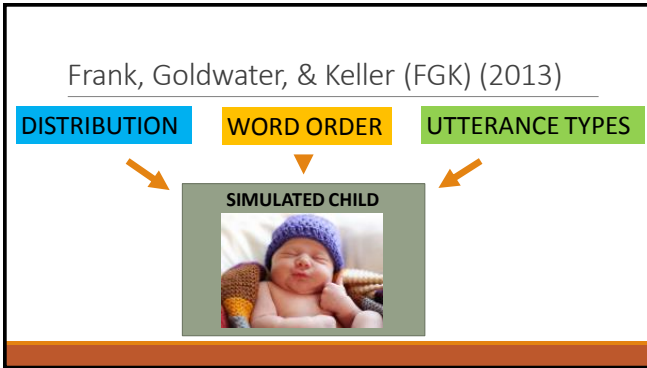
Utterance Types

“I like penguins.”	“Do you like penguins?”
“He wants cookies.”	“Does he want cookies?”
“She took a piece.”	“What did she take?”
“They need a hand.”	“What do they need?”

PRONOUN VERB (DETERMINER) NOUN Begins with: AUXILIARY VERB, WH-WORD...
 Ends with: question mark

Summary of syntactic category cues





How does FGK's strategy work?

- Input = child-directed speech

"Are you sleepy?"



How does FGK's strategy work?

- Input = child-directed speech

"Are you sleepy?"

(Linking)
Verb

(pro)
noun

adjective

ADULT CATEGORIES

How does FGK's strategy work?

- Input = child-directed speech

???

???

???

"Are you sleepy?"

(Linking)
Verb

(pro)
noun

adjective

ADULT CATEGORIES

INFERRED CATEGORIES

How does FGK's strategy work?

- Input = child-directed speech

2

8

7

"Are you sleepy?"

(Linking)
Verb

(pro)
noun

adjective

ADULT CATEGORIES

INFERRED CATEGORIES

How does FGK's strategy work?

- Input = child-directed speech

2

8

7

"Are you sleepy?"

(Linking)
Verb

(pro)
noun

adjective

ADULT CATEGORIES

INFERRED CATEGORIES

CATEGORY 8

Dog
You
Catherine
Yawn

CATEGORY 14

See
Need
Know
Run

How are we further testing this strategy?



Methods

- Age-appropriate dataset
- Evaluation methods
 - o - inferred category mappings
 - o - perplexity

Methods

- **Age-appropriate dataset**
- Evaluation methods
 - o - inferred category mappings
 - o - perplexity

Age-appropriate dataset

Age range of children receiving the input

- FGK: 18 months – 3 years old
- Me: 6 months – 12 months

Infants, **aged 12-14 months**, have basic knowledge of syntactic categories.

- How might they learn it?

(Booth and Waxman, 2003)

Age-appropriate dataset

Age range of children receiving the input

- FGK: 18 months – 3 years old
- Me: 6 months – 12 months

Why does age matter?

- Complexity of utterances changes as children get older.

Why does age matter?

6 months – 12 months

“Do you see the light?”

18 months – 3 years old

“Shall we build an arch for the tractor to drive through?”



Important: FGK’s strategy hasn’t yet been tested on realistic data of this kind.

Methods

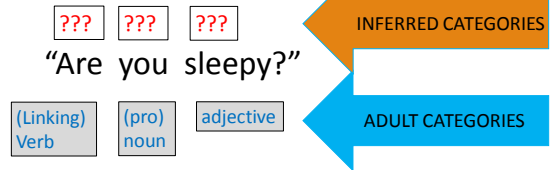
- Age-appropriate dataset
- **Evaluation methods**
 - Qualitative analysis: inferred category mappings
 - Quantitative analysis: perplexity

VM score

- how well did the simulated child match the adult categories?

How does FGK’s strategy work?

- Input = child-directed speech



How does FGK’s strategy work?

- Input = child-directed speech

2 8 7

“Are you sleepy?”

CATEGORY 8

Dog
Rabbits
Catherine
Yawn

CATEGORY 14

See
Need
Know
Run

VM score – don’t expect perfect

- We don’t expect perfect scores, because children at this age likely have preliminary category knowledge, rather than full adult knowledge.



NOUNS	VERBS
Dog	See
Rabbits	Need
Catherine	Know
Yawn	Run

NOUNS	VERBS
Dog	See
Rabbits	Need
Catherine	Know
Yawn	Run

CATEGORY 3

Yawn

NOUNS	VERBS
Dog	See
Rabbits	Need
Catherine	Know
Yawn	Run

CATEGORY 3

Yawn

CATEGORY 8

Rabbits
Catherine

NOUNS	VERBS
Dog	See
Rabbits	Need
Catherine	Know
Yawn	Run

CATEGORY 3

Yawn

CATEGORY 8

Rabbits
Catherine

CATEGORY 14

Run
Know
Dog

NOUNS	VERBS
Dog	See
Rabbits	Need
Catherine	Know
Yawn	Run

CATEGORY 3

Yawn

CATEGORY 8

Rabbits
Catherine

CATEGORY 14

Run
Know
Dog

CATEGORY 10

See
Need

VM score – how well did the simulated child match the adult categories?

Range: 0.0 – 1.0

Frank, Goldwater, & Keller (2013): **0.6-0.7**

My score: **0.48**

When using more realistic data, the strategy doesn't seem to do so well.

WHY? WHAT IS THE SIMULATED CHILD ACTUALLY DOING?



Methods

- Age-appropriate dataset
- Evaluation methods
 - o - Qualitative analysis: inferred category mappings
 - o - Quantitative analysis: perplexity

Qualitative analysis: inferred category mappings FINDINGS

It does match some adult categories well.

- matches some adult categories to exactly one inferred category

It does map some adult categories well.

- matches some adult categories to exactly one inferred category

ADJECTIVES
Happy
Cheerful
Little
Big

It does map some adult categories well.

- matches some adult categories to exactly one inferred category

ADJECTIVES		CATEGORY 5
Happy	→	Happy
Cheerful		Cheerful
Little		Little
Big		Big

It believes some adult categories belong together.

- Some verbs and interjections put together

VERB	INTERJECTION
See!	Yeah!
Look!	Huh!

It believes some adult categories belong together (and it makes sense).

- Verbs and interjections put together – words are utterance-initial exclamation markers

VERB	INTERJECTION											
See!	Yeah!	<table border="1"> <thead> <tr> <th colspan="2">CATEGORY 8</th> </tr> </thead> <tbody> <tr> <td>See!</td> <td>Yeah!</td> </tr> <tr> <td>Look!</td> <td>Huh!</td> </tr> <tr> <td>Yeah!</td> <td></td> </tr> <tr> <td>Huh!</td> <td></td> </tr> </tbody> </table>	CATEGORY 8		See!	Yeah!	Look!	Huh!	Yeah!		Huh!	
CATEGORY 8												
See!	Yeah!											
Look!	Huh!											
Yeah!												
Huh!												
Look!	Huh!											
Yeah!												
Huh!												

It splits an adult category into two.

NOUNS
you
I
we
friend
boy

It splits an adult category into two (and it's reasonable).

NOUNS													
you	<table border="1"> <thead> <tr> <th>CATEGORY 4</th> <th>CATEGORY 1</th> </tr> </thead> <tbody> <tr> <td>you</td> <td>friend</td> </tr> <tr> <td>I</td> <td>boy</td> </tr> <tr> <td>we</td> <td></td> </tr> <tr> <td>friend</td> <td></td> </tr> <tr> <td>boy</td> <td></td> </tr> </tbody> </table>	CATEGORY 4	CATEGORY 1	you	friend	I	boy	we		friend		boy	
CATEGORY 4		CATEGORY 1											
you		friend											
I		boy											
we													
friend													
boy													
I													
we													
friend													
boy													

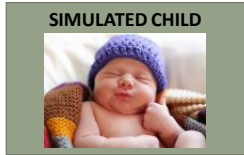
It has no clue what should belong to these other categories.

NOUN	INTERJECTION
kisses	heheheh
hat	gahb

It has no clue what should belong to these other categories.

NOUN	INTERJECTION											
kisses	heheheh	<table border="1"> <thead> <tr> <th colspan="2">CATEGORY 6</th> </tr> </thead> <tbody> <tr> <td>kisses</td> <td>heheheh</td> </tr> <tr> <td>hat</td> <td>gahb</td> </tr> <tr> <td>heheheh</td> <td></td> </tr> <tr> <td>gahb</td> <td></td> </tr> </tbody> </table>	CATEGORY 6		kisses	heheheh	hat	gahb	heheheh		gahb	
CATEGORY 6												
kisses	heheheh											
hat	gahb											
heheheh												
gahb												
hat	gahb											
heheheh												
gahb												

IMPORTANT: The simulated child is doing some reasonable (not crazy) things in trying to learn these syntactic categories.



Methods

- Age-appropriate dataset
- **Evaluation methods**
 - o - Qualitative analysis: inferred category mappings
 - o - **Quantitative analysis: perplexity**

Results: Perplexity Measure

- How predictable is the data **given the inferred categories**

Results: Perplexity Measure

- How predictable is the data **given the inferred categories**

*****Predictable data is easier to process (Levy 2008).**

Results: Perplexity Measure

- Range: $1 - +\infty$

Lower perplexity score = more predictable data

Higher perplexity score = more confusing data



Results: Perplexity Measure

Data = child-directed speech data

Using adult categories: **772**

Using **inferred** categories: **728**

Inferred categories make language processing easier!



Is what children are learning about language **crazy**?

Conclusion – **NOT CRAZY!!!!**

The inferred categories **ARE** useful (even if they're not the ones adults have).



Future Directions

- More sophisticated perplexity measure to evaluate how useful the categories are
- **Current limitation:** utterance types are currently derived from adult knowledge
 - **Current plan:** Use utterance types that 12-month-olds would use



Acknowledgements

- Professor Lisa S. Pearl, Ph.D.
- Stella Frank, Ph.D.
- Galia Barsever
- Computation of Language Acquisition (CoLa) Lab
- Undergraduate Research Opportunities Program (UROP)

QUESTIONS?
Comments?
Concerns?

