Ling 151/Psych 156A: Acquisition of Language II

Lecture 14

Syntactic categorization II

Announcements

HW5 available (due 2/16/18)

Be working on review questions for syntactic categorization

Acquisition task

syntactic categorization

idea

glitter

unicorn

penguin

Noun

owl

kitty

Syntactic Categorization

idea

glitter unicorn

penguin **Noun** owl

kitty

Nouns behave similarly:

They can combine with certain types of words to make larger units (like Noun Phrases).

Syntactic Categorization

Determiner + Noun ("the kitty")

[NP → Det + N]

idea

glitter

unicorn

Noun

penguin

owl

Nouns behave similarly:

They can combine with certain types of words to make larger units (like Noun Phrases).

Syntactic Categorization

Determiner + Noun ("the ")

[NP → Det + N]

dax

idea

glitter

unicorn

Noun

penguin

owl

kitty

Rule with category Noun = new phrases with words of category Noun

Categories give us expressive power

This is very handy for generating new expressions we haven't heard before.

Syntactic Categorization

Determiner + Noun ("the dax ")

[NP → Det + N]

idea

glitter

unicorn

Noun

penguin

owl

kitty

Rule with category Noun = new phrases with words of category Noun

Categories give us expressive power

This is very handy for generating new expressions we haven't heard before.

Syntactic Categorization

Determiner + Noun ("the dax ")

[NP → Det + N]

idea

glitter

unicorn

Noun

penguin

owl

kitty

Rule with category Noun = new phrases with words of category Noun

Categories give us expressive power

Because we can do this as adults, we use this expressive power as evidence that we as adults have categories.

Syntactic Categorization

Determiner + Noun (“the dax ”)

[NP → Det + N]

idea

glitter

unicorn

Noun

penguin

owl

kitty

Rule with category Noun = new phrases with words of category Noun

Categories give us expressive power

This expressive power is sometimes called **productivity**.

Syntactic Categorization

idea
glitter unicorn
Noun owl
penguin
kitty

We have many categories in human language.

Some are open-class — it’s easy to add new words to them.

Syntactic Categorization

idea
glitter
unicorn
Noun
owl
penguin
kitty

We have many categories in human language.

Some are open-class — it's easy to add new words to them.

[VP → Negation + V]

It's not daxing
- it's dancing!

surprise
stand
Verb
find
dance
adore

Syntactic Categorization

idea
glitter unicorn
Noun owl
penguin
kitty

We have many categories in human language.

Some are open-class — it’s easy to add new words to them.

dax
surprise stand
find **Verb** dance
adore

Syntactic Categorization

idea
glitter unicorn
Noun owl
penguin
kitty

surprise stand
find **Verb** dance
adore

We have many categories in human language.

Some are closed-class — the words in them are fixed.

[VP → Negation + V]

It’s not daxing
- it’s dancing!

didn’t not
Negation
wouldn’t can’t
won’t

Syntactic Categorization

didn't

not

Negation

wouldn't

can't

won't

idea

glitter

unicorn

Noun

owl

penguin

kitty

surprise

stand

find **Verb** dance

adore

We have many categories in human language.

Some are closed-class — the words in them are fixed.

[VP → Auxiliary + V]

It would sing
if it could sing

might

would

could

Auxiliary

will

can

should

Syntactic Categorization

might would could

Auxiliary

will can should

didn’t not

Negation

wouldn’t can’t won’t

surprise stand

Verb

find dance adore

idea glitter unicorn

Noun

owl penguin kitty

There’s significant debate on when these categories develop.

Development of syntactic categories

didn't not

Negation

wouldn't can't won't

might would could

will **Auxiliary**

should can

idea glitter unicorn

Noun owl

penguin kitty

surprise stand

find **Verb** dance

adore

Some studies suggest that syntactic category knowledge may already be in place around the age of two

- Determiners (like "the"), Nouns: Valian 1986, Valian, Solt, & Stewart 2008
- Auxiliary verbs: Stromswold 1989, Rispoli, Hadley, & Holt 2009, Rissman, Legendre, & Landau 2013
- Verbs: Kowalski & Yang 2012

Development of syntactic categories

might would could will **Auxiliary** can should

didn't not **Negation** wouldn't can't won't

surprise stand find **Verb** dance adore

idea glitter unicorn **Noun** owl penguin kitty

Other studies suggest that it may appear significantly later:

- Determiners (like "the"), Nouns: Pine & Lieven 1997, Meylan et al. 2017
- Auxiliary verbs: Wilson 2003, Theakston & Lieven 2005, Theakston, Lieven, Pine, & Rowland 2005, Theakston & Lieven 2008, Theakston & Rowland 2009
- Verbs: Tomasello 1992, Tomasello 2006

How can we tell?

might would could will **Auxiliary** can should

didn't not **Negation** wouldn't can't won't

surprise stand find **Verb** dance adore

idea glitter unicorn **Noun** owl penguin kitty

Easy to observe: When children know individual words.

dance

How can we tell?

might would could
Auxiliary
will can should

didn't not
Negation
wouldn't can't won't

surprise stand
find **Verb** dance
adore

idea glitter unicorn
Noun owl
penguin kitty

Harder to observe: When children have recognized these words belong to categories.

surprise adore stand
find **Verb** ???
dance

How can we tell?

didn't not

Negation

wouldn't can't

won't

idea

glitter unicorn

Noun owl

penguin

kitty

might would

could

will **Auxiliary**

should can

surprise stand

find **Verb** dance

adore

One indicator:

Knowledge about how one word combines with other words is transferred within the category.

…could dance…

How can we tell?

might would could
Auxiliary
will can should

didn't not
Negation
wouldn't can't won't

idea glitter unicorn
Noun
penguin owl kitty

One indicator:

Knowledge about how one word combines with other words is transferred within the category.

…could dance…

surprise stand
find **Verb** dance
adore

How can we tell?

might would
could
will **Auxiliary**
can
should

didn't
not
Negation
wouldn't can't
won't

idea
glitter unicorn
Noun owl
penguin
kitty

One indicator:

Knowledge about how one word combines with other words is transferred within the category.

…could dance…

surprise stand
find **Verb** **dance**
adore

[VP → Aux + V]

How can we tell?

might would could
will **Auxiliary**
should can

didn't not
Negation
wouldn't can't
won't

idea
glitter unicorn
Noun owl
penguin
kitty

One indicator:

Knowledge about how one word combines with other words is transferred within the category.

...could surprise...

surprise stand
find **Verb** **dance**
adore

[VP → Aux + V]

How can we tell?

might would could

will **Auxiliary**

should can

didn't not

Negation

wouldn't can't

won't

idea

glitter unicorn

Noun owl

penguin

kitty

One indicator:

Knowledge about how one word combines with other words is transferred within the category.

…might surprise…

surprise stand

find **Verb** **dance**

adore

[VP → Aux + V]

How can we tell?

didn't not
Negation
wouldn't can't
won't

might would
could
will **Auxiliary**
should can

idea
glitter unicorn
Noun owl
penguin
kitty

This causes the child to combine words of the same category with similar words, so that there's overlap in usage within a category.

…will adore…
…might surprise…
…would stand…
…would find…
…will surprise…
…will dance…
…might dance…
…might stand…
…would dance…
…could adore…

surprise stand
find **Verb** dance
adore

[VP → Aux + V]

How can we tell?

might would could

Auxiliary

will can should

didn't not

Negation

wouldn't can't won't

idea glitter unicorn

Noun

owl penguin kitty

This overlap is something we can quantitatively assess to gauge productivity with respect to categories (Tomasello 1992, Pine & Lieven 1997, Naigles, Hoff, & Vear 2009, Yang 2010, 2011, 2013, Goldin-Meadow & Yang 2016).

…will adore…

…might surprise…

…would stand…

…would find…

…will surprise…

…will dance…

…might dance…

…might stand…

…would dance…

…could adore…

surprise stand

find **Verb** dance

adore

[VP → Aux + V]

How can we tell?

might would could
Auxiliary
will can should

didn't not
Negation
wouldn't can't won't

idea glitter unicorn
Noun
penguin owl kitty

Premise: If children's usage shows enough productivity, as measured by overlap, this suggests they have rules that are based on the more abstract symbols like *Noun, Verb*, *Auxiliary*, and *Negation*.

…will adore…
…might surprise…
…would stand…
…would find…
…will surprise…
…will dance…
…might dance…
…might stand…
…would dance…
…could adore…

surprise stand
find **Verb** **dance**
adore

[VP → Aux + V]

How can we tell?

Premise: If children's usage shows enough productivity, as measured by overlap, this suggests they have rules that are based on the more abstract symbols like *Noun, Verb*, *Auxiliary*, and *Negation*.

Productive rules based on categories

VP → Aux + V

…might surprise…

surprise stand

find **Verb** dance

adore

might would could

will **Auxiliary** can

should

How can we tell?

This contrasts with other alternatives for how to generate these combinations.

Semi-productive rules where some words come from categories and some words don't.

VP → Aux + surprise

…might surprise…

surprise
stand
adore
find
dance

might
would
could
will
Auxiliary
can
should

How can we tell?

This contrasts with other alternatives for how to generate these combinations.

Semi-productive rules where some words come from categories and some words don't.

VP → might + V

…might surprise…

surprise stand

find **Verb** dance

adore

might would could

will can

should

How can we tell?

This contrasts with other alternatives for how to generate these combinations.

Non-productive rules where word combinations are just amalgams memorized directly from the input because the child doesn't have categories.

VP → might + surprise

...might surprise...

surprise
stand
find
dance
adore
might
would
could
will
can
should

How can we tell?

Representation options

Productive rules

VP → Aux + V

Semi-productive rules

VP → Aux + surprise

VP → might + V

Non-productive rules

VP → might + surprise

How can we tell?

Representation options

VP → Aux + surprise

VP → might + surprise

VP → might + V

How much overlap do we expect to see if a child has category-based productive rules?

…will adore…

…might surprise…

…would stand…

…would find…

…will surprise…

…will dance…

…might dance…

…might stand…

…would dance…

…could adore…

VP → Aux + V

surprise stand

find **Verb** dance

adore

might would could

Auxiliary can

should

How can we tell?

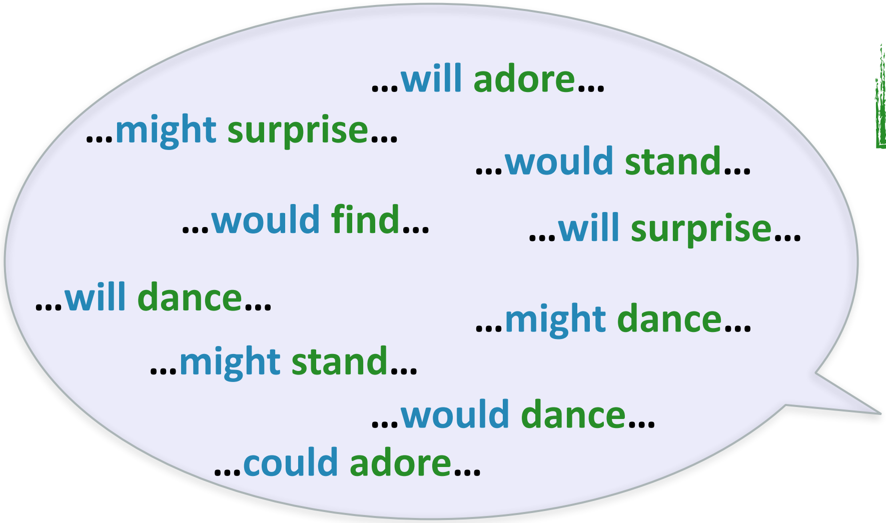
Representation options

VP → Aux + surprise

VP → might + surprise

VP → might + V

For example, should we expect every verb to combine with every auxiliary?



...will adore...

...might surprise...

...would stand...

...would find...

...will surprise...

...will dance...

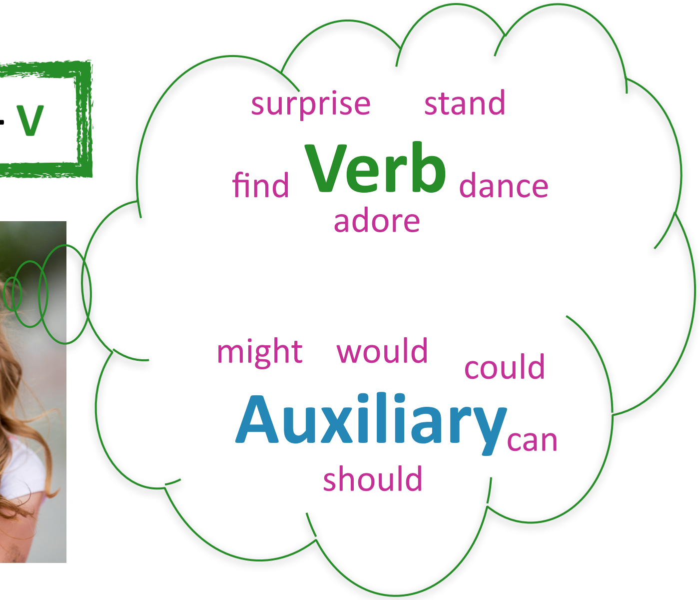
...might dance...

...might stand...

...would dance...

...could adore...

VP → Aux + V



How can we tell?

Representation options

VP → Aux + surprise

VP → might + surprise

VP → might + V

Probably not. We don't say everything we know when we speak – we say things to communicate our intended meaning at the time.

✗

For example, should we expect every verb to combine with every auxiliary?

…will adore…

…might surprise…

…would stand…

…would find…

…will surprise…

…will dance…

…might dance…

…might stand…

…would dance…

…could adore…

VP → Aux + V

surprise stand

find **Verb** dance

adore

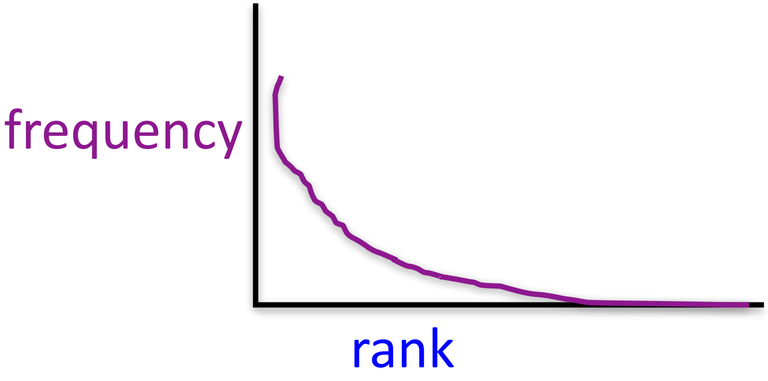
might would could

Auxiliary can

should

How can we tell?

In fact, it turns out naturalistic linguistic output shows power law behavior (a Zipfian distribution)...



verb	**freq**	**rank**
get	101	1
go	100	2
…		
feel	8	58
…		
dream	1	251
…		

…will adore…

…might surprise…

…would stand…

…would find…

…will surprise…

…will dance…

…might dance…

…might stand…

…would dance…

…could adore…

VP → Aux + V

surprise stand

find **Verb** dance

adore

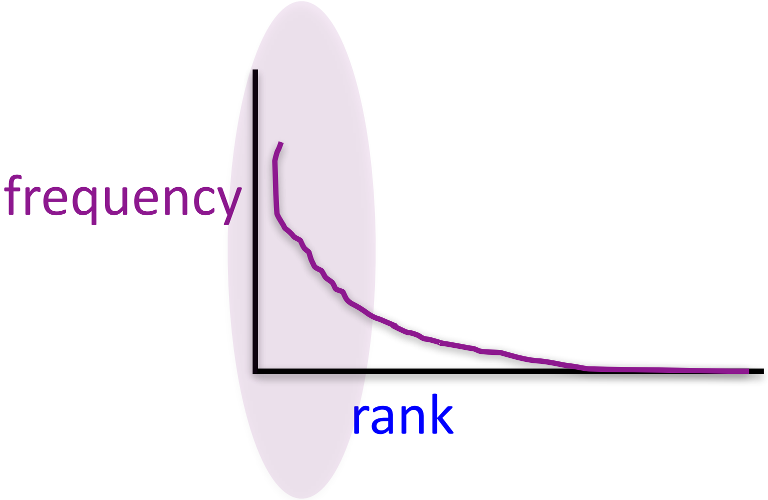
might would could

Auxiliary can

should

How can we tell?

…where a few things are said very frequently...



verb	**freq**	**rank**
get	101	1
go	100	2
…		
feel	8	58
…		
dream	1	251
…		

…will adore…

…might surprise…

…would stand…

…would find…

…will surprise…

…will dance…

…might dance…

…might stand…

…would dance…

…could adore…

VP → Aux + V

surprise stand

find **Verb** dance

adore

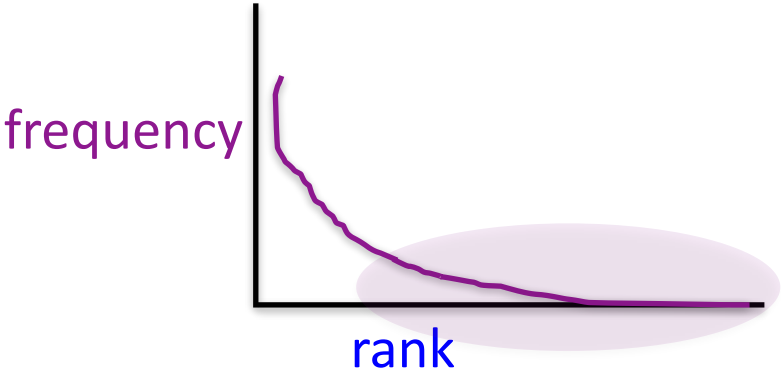
might would could

Auxiliary can

should

How can we tell?

... and most things are said very infrequently.



verb	freq	rank
get	101	1
go	100	2
…		
feel	8	58
…		
dream	1	251
…		

…will adore…

…might surprise…

…would stand…

…would find…

…will surprise…

…will dance…

…might dance…

…might stand…

…would dance…

…could adore…

VP → Aux + V

surprise stand

find **Verb** dance

adore

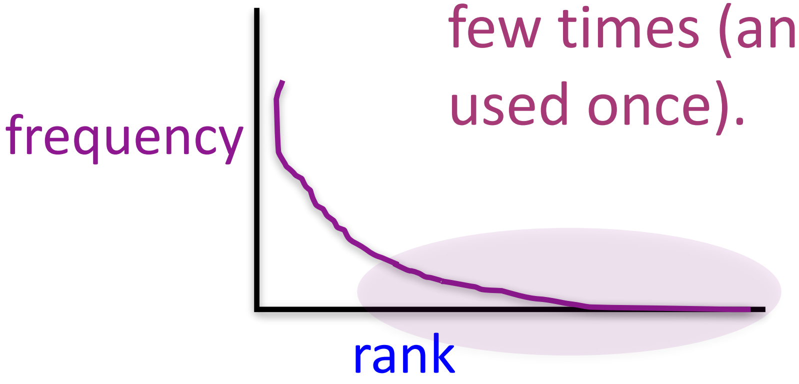
might would could

Auxiliary can

should

How can we tell?

One implication: We can’t expect much overlap in combinatorial usage for words that only are used a few times (and certainly not for those that are only used once).



...will adore...
...might surprise...
...would stand...
...would find...
...will surprise...
...will dance...
...might dance...
...might stand...
...would dance...
...could adore...

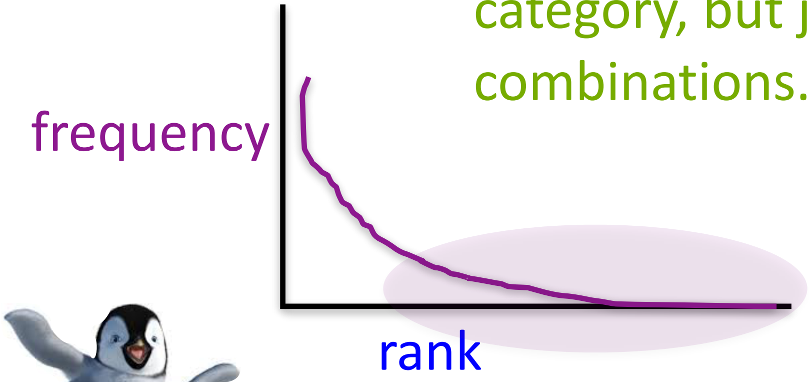
VP → Aux + V

surprise stand
find **Verb** dance
adore

might would could
Auxiliary can
should

How can we tell?

We need to somehow factor in that a child may know that combinatorial usage transfers to other words in the category, but just doesn't choose to say those other combinations.



…**will dance**…

…**will dance**…

…**will dance**…

…**will dance**…

VP → Aux + V

surprise stand

find **Verb** dance

adore

might would could

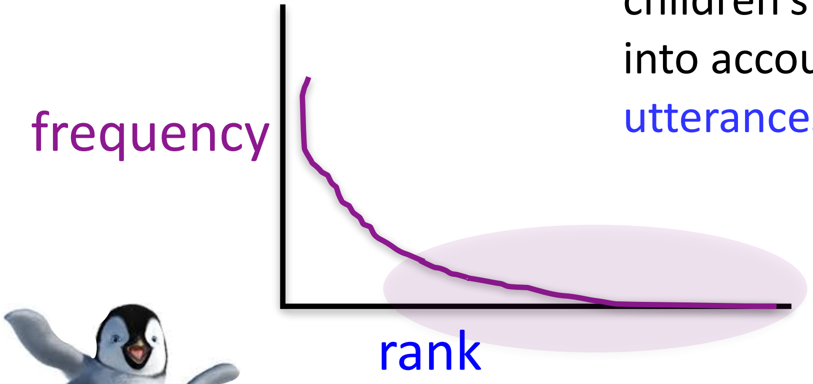
Auxiliary can

should

How can we tell?

Bates, Pearl & Braunwald, in prep.

What we can do: Computational-level analysis of children's productions, using formal metrics that take this into account and describe how children generate their utterances given their underlying representations



VP → Aux + dance

VP → will + V

VP → will + dance

VP → Aux + V

…will dance…

…will dance…

…will dance…

…will dance…

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → Aux + V

VP → will + dance

…will dance…

VP → will + V

VP → Aux + dance

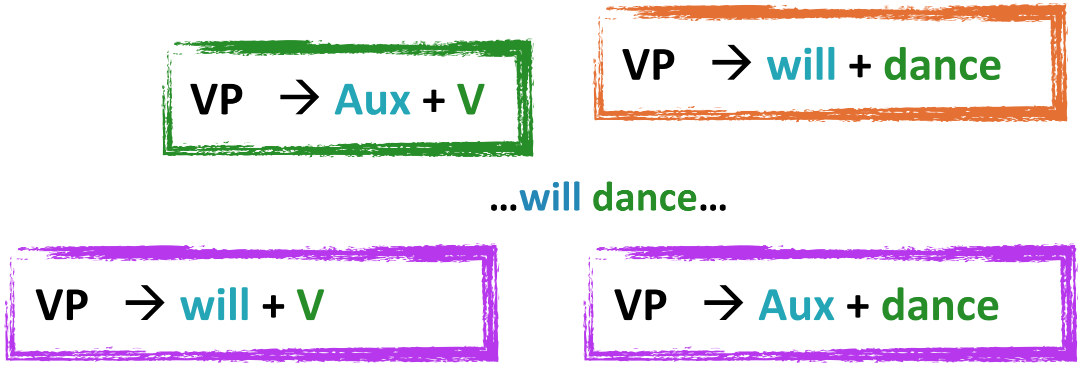
Basic idea:

Compare the observed overlap in children's produced combinations against the expected overlap if a specific underlying representation were what children used to generate that combination.

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations



VP → Aux + V

VP → will + dance

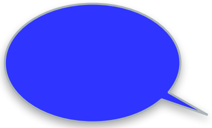
…will dance…

VP → will + V

VP → Aux + dance

Calculating observed overlap

This is based on the child's productions.



If a word combines with more than one lexical item (ex: a verb combining with more than one auxiliary verb), overlap for that word = 1.

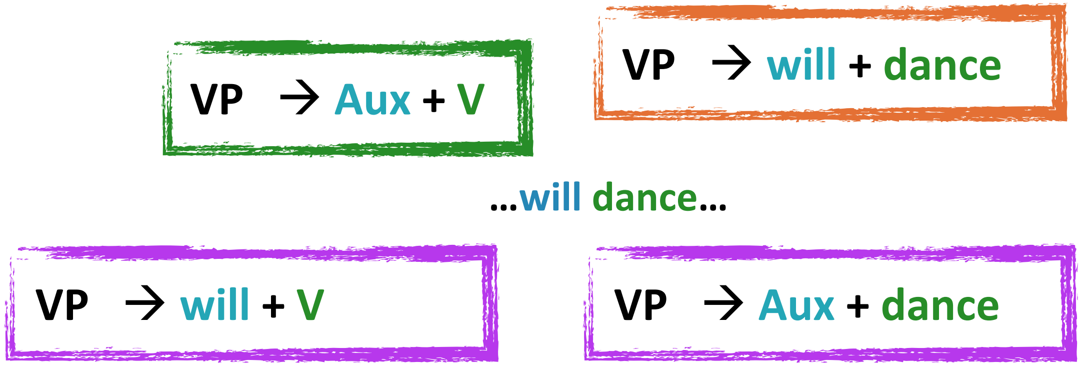
…will dance…
…**would** dance…
…will dance…
…will dance…

overlap for *dance* = 1

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations



VP → Aux + V

VP → will + dance

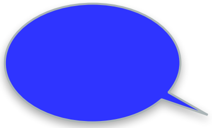
…will dance…

VP → will + V

VP → Aux + dance

Calculating observed overlap

This is based on the child's productions.



If a word combines with only one lexical item (ex: a verb combining with only one auxiliary verb), overlap for that word = 0.

…will dance…
…will dance…
…will dance…
…will dance…

overlap for *dance* = 0

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → Aux + V

VP → will + dance

...will dance...

VP → will + V

VP → Aux + dance

Calculating observed overlap

This is based on the child's productions.

$$\text{Observed overlap for words potentially from one category} = \frac{\text{total overlap from all words}}{\text{total number of words}}$$

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → Aux + V

VP → will + dance

…will dance…

VP → will + V

VP → Aux + dance

Calculating observed overlap

This is based on the child's productions.

Observed overlap for words potentially from one category

$$= \frac{\text{total overlap from all words}}{\text{total number of words}}$$

surprise stand

find **Verb** dance

adore

$$\frac{\text{overlap(surprise) + overlap(stand) + overlap(find) + overlap(adore) + overlap(dance)}}{5}$$

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → Aux + V

VP → will + dance

…will dance…

VP → will + V

VP → Aux + dance

Calculating observed overlap

This is based on the child's productions.

Observed overlap for words potentially from one category $= \frac{\text{total overlap from all words}}{\text{total number of words}}$

surprise stand

Verb

find dance

adore

$$\frac{1 + 1 + 0 + 1 + 0 = 3}{5}$$

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → Aux + V

VP → will + dance

…will dance…

VP → will + V

VP → Aux + dance

Calculating observed overlap

This is based on the child's productions.

$$\text{Observed overlap for words potentially from one category} = \frac{\text{total overlap from all words}}{\text{total number of words}}$$

surprise stand

Verb

find dance

adore

$$\frac{3}{5} = 0.6$$

How can we tell?

Bates, Pearl & Braunwald, in prep.

	Aux	Verb
observed overlap	0.5	0.6

Calculating expected overlap

This is based on what we think the child's underlying representations are.

Underlying representations

VP → Aux + V

VP → will + dance

…will dance…

VP → will + V

VP → Aux + dance

How can we tell?

Bates, Pearl & Braunwald, in prep.

observed overlap	**Aux**	**Verb**
	0.5	0.6

Calculating expected overlap

If both words come from categories, we can calculate expected overlap using a formula developed by Yang (2011) that takes into account how frequently the child produces words of these categories.

Underlying representations

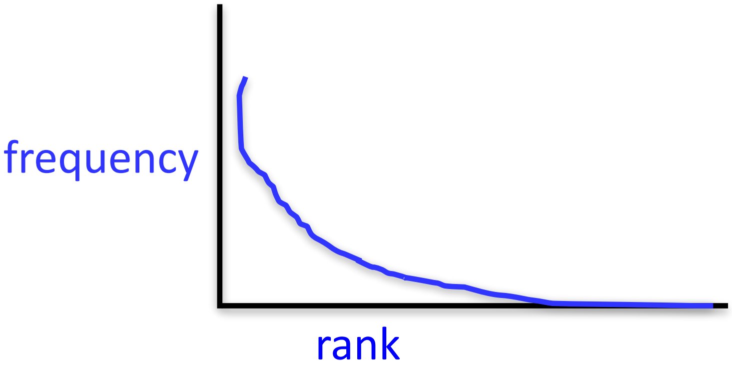
VP → will + dance

…will dance…

VP → will + V

VP → Aux + dance

VP → Aux + V



How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → will + dance

...will dance...

VP → will + V

VP → Aux + dance

VP → Aux + V

	Aux	Verb
observed overlap	0.5	0.6

Calculating expected overlap

If both words come from categories, we can calculate expected overlap using a formula developed by Yang (2011) that takes into account how frequently the child produces words of these categories.

$$Exp_{Prod} = \frac{\sum_{a_j \in Anch} expprod_{a_j}}{|Anch|}$$

$$expprod_{a_j} = 1 + (Unk| - 1)(1 - p_{a_{j_{out}}})^S - \sum_{w_i \in Unk} (p_{w_{i_{out}}} p_{a_{j_{out}}} + 1 - p_{a_{j_{out}}})^S$$

How can we tell?

Bates, Pearl & Braunwald, in prep.

	Aux	Verb
observed overlap	0.5	0.6

Calculating expected overlap

If both words come from categories, we can calculate expected overlap using a formula developed by Yang (2011) that takes into account how frequently the child produces words of these categories.

Underlying representations

VP → will + dance

…will dance…

VP → will + V

VP → Aux + dance

VP → Aux + V

	Aux	Verb
expected overlap	0.7	0.8

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → Aux + V

VP → will + dance

expected overlap

Aux 0.7

Verb 0.8

…will dance…

VP → will + V

VP → Aux + dance

observed overlap

Aux 0.5

Verb 0.6

Calculating expected overlap

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → Aux + V

VP → will + dance

expected overlap

Aux	Verb
0.7	0.8

…will dance…

observed overlap

Aux	Verb
0.5	0.6

VP → will + V

VP → Aux + dance

Calculating expected overlap

If one word comes from a category and one word doesn't, we can calculate expected overlap using a formula adapted from Yang's that takes into account how frequently the child produces words of the one category and how frequently she heard words of the other category combine with it.

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

	Aux	**Verb**
expected overlap	0.7	0.8

VP → Aux + V

VP → will + dance

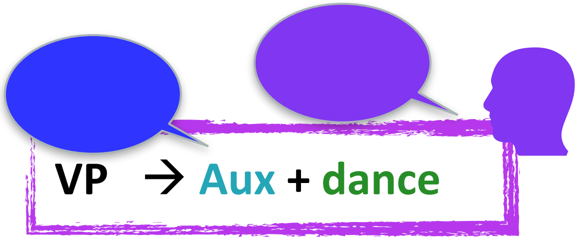
…will dance…

	Aux	**Verb**
observed overlap	0.5	0.6

VP → will + V

Calculating expected overlap

If one word comes from a category and one word doesn't, we can calculate expected overlap using a formula adapted from Yang's that takes into account how frequently the child produces words of the one category and how frequently she heard words of the other category combine with it.



VP → Aux + dance

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → Aux + V

VP → will + dance

expected overlap

Aux	**Verb**
0.7	0.8

…will dance…

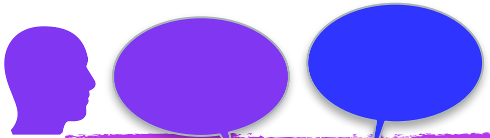
observed overlap

Aux	**Verb**
0.5	0.6

Calculating expected overlap

If one word comes from a category and one word doesn't, we can calculate expected overlap using a formula adapted from Yang's that takes into account how frequently the child produces words of the one category and how frequently she heard words of the other category combine with it.

VP → Aux + dance



VP → will + V

How can we tell?

Bates, Pearl & Braunwald, in prep.

expected overlap

Aux	Verb
0.7	0.8

observed overlap

Aux	Verb
0.5	0.6

Underlying representations

VP → Aux + V

VP → will + dance

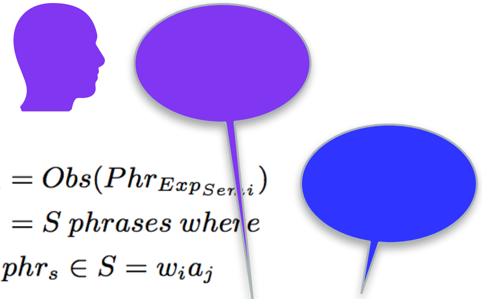
…will dance…

VP → will + V

VP → Aux + dance

Calculating expected overlap

If one word comes from a category and one word doesn't, we can calculate expected overlap using a formula adapted from Yang's that takes into account how frequently the child produces words of the one category and how frequently she heard words of the other category combine with it.



$$Exp_{Semi} = Obs(Phr_{Exp_{Semi}})$$

$$Phr_{Exp_{Semi}} = S\ phrases\ where$$

$$phr_s \in S = w_i a_j$$

$$with\ probability\ p_{w_{i_{in}}} * p_{a_{j_{out}}}$$

How can we tell?

Bates, Pearl & Braunwald, in prep.

expected overlap

observed overlap

	Aux	Verb
observed overlap	0.5	0.6

Calculating expected overlap

Underlying representations

VP → Aux + V

Aux	Verb
0.7	0.8

VP → will + dance

…will dance…

VP → will + V

Aux	Verb
0.9	0.7

VP → Aux + dance

Aux	Verb
0.5	0.6

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

expected overlap

VP → Aux + V

Aux 0.7 **Verb** 0.8

…will dance…

VP → will + V

Aux 0.9 **Verb** 0.7

VP → Aux + dance

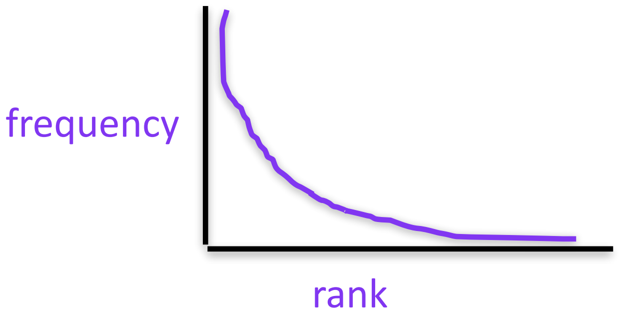
Aux 0.5 **Verb** 0.6

observed overlap **Aux** 0.5 **Verb** 0.6

Calculating expected overlap

VP → will + dance

If neither word comes from a category, we can calculate expected overlap using a formula from Yang (2010) that takes into account how frequently the child heard these words combine in the input.



How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → Aux + V

expected overlap

Aux 0.7 **Verb** 0.8

…will dance…

VP → will + V

VP → Aux + dance

observed overlap **Aux** 0.5 **Verb** 0.6

Aux 0.9 **Verb** 0.7

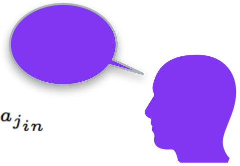
Aux 0.5 **Verb** 0.6

Calculating expected overlap

VP → will + dance

If neither word comes from a category, we can calculate expected overlap using a formula from Yang (2010) that takes into account how frequently the child heard these words combine in the input.

$$Exp_{Not} = Obs(Phr_{Exp_{Not}})$$
$$Phr_{Exp_{Not}} = S\ phrases\ where$$
$$phr_s \in S = w_i a_j$$
$$with\ probability\ p_{w_i a_{j_{in}}}$$



How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

expected overlap

VP → Aux + V

Aux	Verb
0.7	0.8

VP → will + dance

Aux	Verb
0.8	0.7

VP → will + V

Aux	Verb
0.9	0.7

VP → Aux + dance

Aux	Verb
0.5	0.6

observed overlap

Aux	Verb
0.5	0.6

…will dance…

The underlying representation whose **expected overlap best matches the observed overlap** is the most likely representation the child has.

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → Aux + V

Aux Verb
0.7 0.8

VP → will + dance

Aux Verb
0.8 0.7

VP → will + V

Aux Verb
0.9 0.7

	Aux	Verb
observed overlap	**0.5**	**0.6**
expected overlap	**0.5**	**0.6**

VP → Aux + dance

…will dance…

Here, it looks like a semi-productive representation where the **Aux word comes from a category** while the verb doesn't is the best match.

How can we tell?

Bates, Pearl & Braunwald, in prep.

Underlying representations

VP → Aux + V

Aux **Verb**
0.7 0.8

VP → will + dance

Aux **Verb**
0.8 0.7

VP → will + V

Aux **Verb**
0.9 0.7

	Aux	Verb
observed overlap	**0.5**	**0.6**
expected overlap	**0.5**	**0.6**

VP → Aux + dance

…will dance…

surprise stand
find dance
adore

might would could
Auxiliary can
will should

The development of syntactic categories

Bates, Pearl & Braunwald, in prep.

Focus: categories in **Verb Phrases**

might would

could

will **Auxiliary**

should can

didn’t not

Negation

wouldn’t can’t

won’t

surprise stand

find **Verb** dance

adore

idea

glitter unicorn

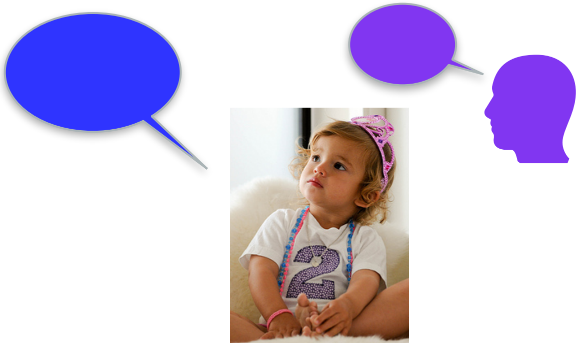
Noun owl

penguin

kitty

The development of syntactic categories

Bates, Pearl & Braunwald, in prep.



Focus: categories in **Verb Phrases**…that appeared sufficiently frequently in this child's productions

Auxiliary

Verb

Negation

Noun

Analyzing the utterances produced by a single American English child (L) between the ages of 20 and 24 months.

The development of syntactic categories

Bates, Pearl & Braunwald, in prep.

Focus: categories in **Verb Phrases**…that appeared sufficiently frequently in this child's productions

Auxiliary

Verb

Negation

Noun

Analyzing the utterances produced by a single American English child (L) between the ages of 20 and 24 months.

Child output: 2154 verb phrases

Child input: 2184 verb phrases from her mother

The development of syntactic categories

Bates, Pearl & Braunwald, in prep.

Focus: categories in **Verb Phrases**…that appeared sufficiently frequently in this child's productions

Auxiliary **Verb**

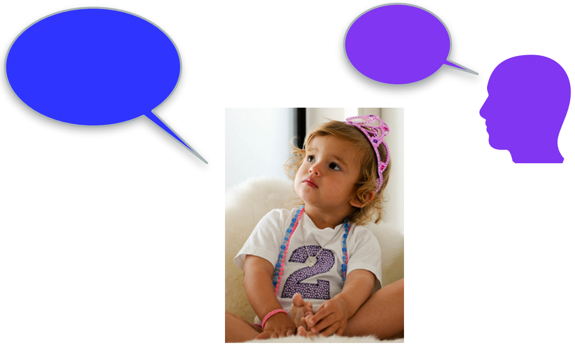
Negation **Noun**

Child output: 2154 verb phrases

L's verb usage appears to be typical, compared against a group of 93 children between 20 and 24 months from the American English CHILDES database

The development of syntactic categories

Bates, Pearl & Braunwald, in prep.



Focus: categories in **Verb Phrases**…that appeared sufficiently frequently in this child's productions

✓ Auxiliary

✓ Negation

✗ Verb

✗ Noun

Utterances most compatible with having adult-like closed-class categories, but not adult-like open-class categories.

VP → Aux + dance

VP → Neg + dance

VP → penguin + dance

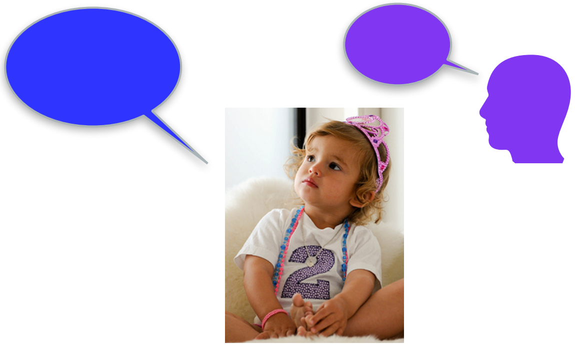
…will dance…

…won't dance…

…penguin dance…

The development of syntactic categories

Bates, Pearl & Braunwald, in prep.



Focus: categories in **Verb Phrases**…that appeared sufficiently frequently in this child’s productions

✓ Auxiliary

✓ Negation

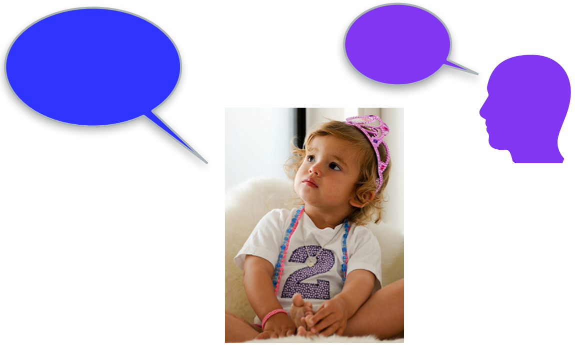
✗ Verb

✗ Noun

This is much earlier than when previous studies have thought children develop closed-class categories!

The development of syntactic categories

Yang 2010, 2011



Focus: categories in **Noun Phrases**

NP → Det Noun

…a penguin…

the

Determiner

a(n)

idea

glitter

unicorn

Noun

owl

penguin

kitty

The development of syntactic categories

Yang 2010, 2011

Focus: categories in **Noun Phrases**

NP → **Det Noun**

…a penguin…

the

Determiner

a(n)

idea

glitter

unicorn

Noun

owl

penguin

kitty

Data: Child-produced utterances from six American English corpora of the CHILDES database (age range 1;1 to 5;1).

First 100, 300, and 500 productions from all children to capture earliest stage of language production which should (presumably) be the least productive.

The development of syntactic categories

Yang 2010, 2011

Focus: categories in **Noun Phrases**

Yang evaluated a representation where both words come from categories.

NP → Det Noun

…a penguin…

the

Determiner

a(n)

idea

glitter

unicorn

Noun

owl

penguin

kitty

This matches pretty well!

	Observed Overlap	Expected Overlap
First 100 utterances	21.8	19.6
First 300 utterances	29.1	26.7
First 500 utterances	34.2	32.3

The development of syntactic categories

Yang 2010, 2011

Focus: categories in **Noun Phrases**

This contrasts with the representation where neither word comes from a category.

…a penguin…

the

Determiner

a(n)

idea

glitter

unicorn

Noun

owl

penguin

kitty

NP → a penguin

This doesn't match as well

	Observed Overlap	**Expected Overlap**	**Expected Overlap**
First 100 utterances	21.8	19.6	17.2
First 300 utterances	29.1	26.7	25.6
First 500 utterances	34.2	32.3	30.2

The development of syntactic categories

Yang 2010, 2011

Focus: categories in **Noun Phrases**

This suggests that for Noun Phrases, young children have created categories for both closed-class categories like Determiners and open-class categories like Nouns.

…**a penguin**…

NP → Det Noun

the

Determiner

a(n)

idea

glitter

unicorn

Noun

owl

penguin

kitty

The development of syntactic categories

Yang 2010, 2011

Focus: categories in **Noun Phrases**

Though it's probably worth evaluating representations where one word comes from a category and the other doesn't just to make sure…

…a penguin…

NP → a Noun

NP → Det penguin

Recap: Syntactic categorization

Productivity, as measured by the lexical overlap of words for a syntactic category, is one way to assess whether children seem to have knowledge of a particular syntactic category.

Natural language use seems to have a Zipfian distribution, where many combinations are rarely (or never) heard. This can make it hard to learn, and it can also make it hard to figure out what knowledge children have.

There are formal metrics for figuring out exactly how much overlap words should have if children have particular representations in their minds, given that language use has a Zipfian distribution.

Based on these metrics, it seems like children may attain knowledge of closed-class categories like Auxiliary, Negation, and Determiner earlier than previously thought.

Questions?

You should be able to do all the review questions for syntactic categorization and all of HW5.