Psych 215: Language Sciences (Language Acquisition)

Lecture 13
Learning Phrases

About Language Structure

Sentences are not just strings of words.

The girl danced with the goblin king.

About Language Structure

Sentences are not just strings of words.

Words cluster into larger units called phrases, based on their grammatical category.

Noun (N) = girl, goblin, dream, laughter, …

Determiner (Det) = a, the, an, these, …

Adjective (Adj) = lovely, stinky, purple, …

Verb (V) = laugh, dance, see, defeat, …

Adverb (Adv) = lazily, well, rather, …

Preposition (P) = with, on, around, towards, …

About Language Structure

Sentences are not just strings of words.

Words cluster into larger units called phrases, based on their grammatical category.

Det N V P Det Adj N
The girl danced with the Goblin King.

About Language Structure

Sentences are not just strings of words.
Words cluster into larger units called phrases.

Det N V P Det Adj N
The girl danced with the Goblin King.

Noun Phrases (NP)

About Language Structure

Sentences are not just strings of words.
Words cluster into larger units called phrases.

Det N V P Det Adj N
The girl danced with the Goblin King.

Noun Phrases (NP)

Can be replaced with pronouns like “he”, “she”, or “it”

About Language Structure

Sentences are not just strings of words.
Words cluster into larger units called phrases.

Det N V P Det Adj N
She danced with him.

Noun Phrases (NP)

Can be replaced with pronouns like “he”, “she”, and “it”

About Language Structure

Sentences are not just strings of words.
Words cluster into larger units called phrases.

Det N V P Det Adj N
The girl danced with the Goblin King.

Preposition Phrases (PP)

About Language Structure

Sentences are not just strings of words.
Words cluster into larger units called phrases.

Det N V P Det Adj N
The girl danced with the Goblin King.

Preposition Phrases (PP)

Can be replaced with words like “here” and “there”

About Language Structure

Sentences are not just strings of words.
Words cluster into larger units called phrases.

Det N V P Det Adj N
The girl danced there.

Preposition Phrases (PP)

Can be replaced with words like “here” and “there”

About Language Structure

Sentences are not just strings of words.
Words cluster into larger units called phrases.

Det N V P Det Adj N
The girl danced with the Goblin King.

Verb Phrases (VP)

About Language Structure

Sentences are not just strings of words.
Words cluster into larger units called phrases.

Det N V P Det Adj N
The girl danced with the Goblin King.

Verb Phrases (VP)

Can be replaced with words like “do so” and “did so”

About Language Structure

Sentences are not just strings of words.
Words cluster into larger units called phrases.

Det N V P Det Adj N
The girl did so.

Verb Phrases (VP)

Can be replaced with words like “do so” and “did so”

About Language Structure

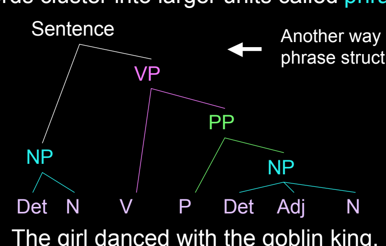
Sentences are not just strings of words.
Words cluster into larger units called phrases.

Det N V P Det Adj N
The girl danced with the goblin king.

Verb Phrases (VP)
Preposition Phrases (PP)
Noun Phrases (NP)

About Language Structure

Sentences are not just strings of words.
Words cluster into larger units called phrases.



Another way to represent phrase structure

About Language Structure

Things that phrases can do:

Have pro-forms replace them

pro-forms: words that have minimal specific meaning and which can stand in for phrases (“he”, “she”, “there”, “here”, “do so”)

The girl who ate the peach and forgot everything saved Hoggle in the goblin city.

About Language Structure

Things that phrases can do:

Have pro-forms replace them

pro-forms: words that have minimal specific meaning and which can stand in for phrases (“he”, “she”, “there”, “here”, “do so”)

She saved Hoggle in the goblin city.

The girl who ate the peach and forgot everything saved Hoggle there.

The girl who did so saved Hoggle in the goblin city.

About Language Structure

Things that phrases can do:

Have pro-forms replace them

pro-forms: words that have minimal specific meaning and which can stand in for phrases (“he”, “she”, “there”, “here”, “do so”)

* She Hoggle in the goblin city. (she saved ≠ phrase)

* The girl who ate the peach and forgot everything saved Hoggle in the it. (goblin city ≠ phrase)

The girl who did so Hoggle in the goblin city. (ate the peach and forgot everything saved ≠ phrase)

About Language Structure

Things that phrases can do:

Be conjoined to other phrases of the same kind: use “and”

The girl who ate the peach and forgot everything saved Hoggle.

About Language Structure

Things that phrases can do:

Be conjoined to other phrases of the same kind: use “and”

The girl who ate the peach and forgot everything saved Hoggle.

Ludo saved Hoggle.
He saved Hoggle.
Ludo = NP

About Language Structure

Things that phrases can do:

Be conjoined to other phrases of the same kind: use “and”

Ludo and the girl who ate the peach and forgot everything saved Hoggle.

Ludo = NP
The girl who ate the peach and forgot everything = NP

About Language Structure

Things that phrases can do:

Be conjoined to other phrases of the same kind: use “and”

The girl who and Ludo ate the peach and forgot everything saved Hoggle.

Ludo = NP
The girl who ≠ NP

About Language Structure

Things that phrases can do:

Move around in the sentence without making the sentence sound too odd

The girl who ate the peach and forgot everything saved Hoggle in the goblin city.

About Language Structure

Things that phrases can do:

Move around in the sentence without making the sentence sound too odd

In the goblin city, the girl who ate the peach and forgot everything saved Hoggle.

In the goblin city = PP

About Language Structure

Things that phrases can do:

Move around in the sentence without making the sentence sound too odd

* Who ate the, the girl peach and forgot everything saved Hoggle in the goblin city.

who ate the ≠ phrase

About Language Structure

Things that phrases can do (summary):

Be replaced by very generic single word forms (pro-forms)

Be conjoined to other phrases of the same kind

Move around in the sentence without making the sentence sound too odd

Computational Problem

How do children figure out which words belong together (as phrases) and which words don't?

Det N V P Det Adj N
The girl danced with the goblin king.

↓

Det N V P Det Adj N
The girl danced with the goblin king.

Learning Phrases

One way we've seen that children can learn things is by tracking the statistical information available.

Saffran, Aslin, & Newport (1996):
Transitional Probability is something 8-month olds can track

who's afraid of the big bad wolf

Posit a word boundary at the minimum of the transitional probabilities between syllables

Learning Phrases

One way we've seen that children can learn things is by tracking the statistical information available.

Thompson & Newport (2007):
Transitional Probability to divide words into phrases?

$\boxed{\text{the girl}}$ and $\boxed{\text{the dwarf}}$...

Posit a phrase where the transitional probability is high?

A look at real language properties in action with transitional probabilities

Example: Optional phrases

A B C D E F
The goblin easily steals the child.

A look at real language properties in action with transitional probabilities

Example: Optional phrases

A B C D E F
The goblin easily steals the child.

ABCDEF ← If the child only ever sees this order of categories, there's no way to know how the words break up into phrases.

A look at real language properties in action with transitional probabilities

Example: Optional phrases

A B C D E F
The goblin easily steals the child.

ABCDEF But suppose C is an optional word/phrase. (easily is an adverb that can be left out)

ABDEF Data without C sometimes will appear.

The goblin steals the child.

A look at real language properties in action with transitional probabilities

Example: Optional phrases

A B C D E F

The goblin easily steals the child.

ABCDEF

ABDEF

With the optional phrase left out, the transitional probability of (BC) is less than 1. A transitional probability learner posits a phrase boundary there. Conclusion: AB is a unit, CDEF is a unit.

the goblin (= NP)

easily steals the child (= VP)

The goblin steals the child.

A look at real language properties in action with transitional probabilities

Example: Optional phrases

A B C D E F

The goblin easily steals the child.

ABCDEF

ABDEF

With the optional phrase put in, the transitional probability of (BD) is less than 1. A transitional probability learner posits a phrase boundary there. Conclusion: AB is a unit, DEF is a unit.

the goblin (= NP)

steals the child (= VP)

The goblin steals the child.

Artificial Language Experiments

Adults listened to data from an artificial language for 20 minutes on multiple days

Properties of the artificial language: similar to real language properties

optional phrases (the goblin chased a chicken *in the castle*)
repeated phrases (NP Verb NP)
moved phrases (In the castle the goblin chased a chicken)

Artificial Language Experiments

Baseline pattern: ABCDEF

real language parallel

A B C D E F

The goblin easily steals the child.

Nonsense Words Assigned to Each Form Class

A Words	B Words	C Words	D Words	E Words	F Words
KOF (oaf)	HOX (box)	JES (dress)	SOT (coat)	FAL (pal)	KER (her)
DAZ (has)	NEB (web)	REL (fell)	ZOR (core)	TAF (waif)	NAV (have)
MER (her)	LEV (rev)	TID (bid)	LUM (bum)	RUD (bud)	SIB (bib)

Artificial Language Phrases

AB CD EF

How do we tell if learning happened?

Baseline assessment: Can subjects actually realize all these nonsense words belong to 6 distinct categories? Can they categorize?

kof hox jes sot fal ker   is the same as
daz neb tid zor rud sib

How do we tell if learning happened?

Baseline assessment: Can subjects actually realize all these nonsense words belong to 6 distinct categories? Can they categorize?

kof hox jes sot fal ker   is the same as
daz neb tid zor rud sib

See if they can tell the difference between the correct order they were exposed to (ABCDEF) and some other pattern they never heard (ABCDCF)

kof hox jes sot fal ker   is right
kof hox jes sot rel ker   is wrong

How do we tell if learning happened?

Phrase learning assessment: If they can categorize, do they learn what the phrases are (AB CD EF)?

Example: test between AB and non-phrase BC

Sample test item - which one do they think belongs together?

kof hox   vs.   hox jes

Learning a language with optional phrases

Baseline pattern: ABCDEF

Other patterns heard (phrases AB CD EF missing):
CDEF, ABEF, ABCD

kof hox jes sot fal ker
rel zor taf nav
mer neb rud sib
daz lev tid lum

Stimuli: 96 of possible 972
48 canonical: ABCDEF
48 distributed among other patterns

Control subjects:
Control language (remove one adjacent pair at a time)
Additional control patterns heard:
BCDE, ABCF, ADEF

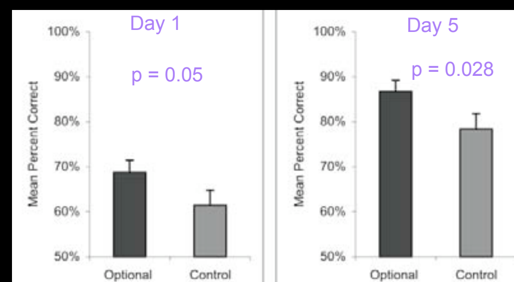
Learning a language with optional phrases

Transitional Probabilities in the Optional Phrase language and the Control language are different. The Optional Phrase language has lower probability across phrase boundaries than within phrases. The control language has the same probability no matter what.

	A→B	B→C	C→D	D→E	E→F
Optional phrases	1.00	0.80	1.00	0.80	1.00
Optional control	0.90	0.90	0.90	0.90	0.90

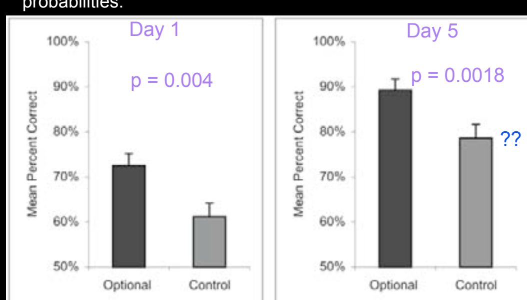
Optional Language Learning: Categorization

Above chance performance, improvement with more exposure to language, similar performance for test group as for control group



Optional Language Learning: Phrases

Test group with informative transitional probabilities generally doing better than the control group with uninformative probabilities.



Learning a language with repeated phrases

Baseline pattern: ABCDEF

Other patterns heard (phrases AB CD EF repeated):
ABCDEFAB, ABCDEFCD, ABCDEFEF

kof hox jes sot fal ker
kof hox rel zor taf nav daz neb
mer neb jes zor rud sib tid sot
daz lev tid lum fal nav taf ker

Stimuli: 68
34 canonical: ABCDEF
34 distributed among other patterns

Control subjects:
Control language (repeat one adjacent pair at a time)
Additional control patterns heard:
ABCDEFBC, ABCDEFDE, ABCDEFFA

Learning a language with repeated phrases

Transitional Probabilities in the Repeated Phrase language and the Control language are different. The Repeated Phrase language has lower probability across phrase boundaries than within phrases. The control language has almost the same probability no matter what.

	A→B	B→C	C→D	D→E	E→F
Repeated phrases	1.00	0.86	1.00	0.86	1.00
Repeated control	0.92	0.94	0.92	0.94	0.93

Learning a language with moved phrases

Baseline pattern: ABCDEF

Other patterns heard (phrases AB CD EF moved):
ABCDEF, ABEFCD, CDABEF, CDEFAB,
EFABCD, EFCDAB

Example strings heard:
kof hox jes sot fal ker
daz neb taf nav rel zor
…

Stimuli: 80
40 canonical: ABCDEF
40 distributed among other patterns

Control subjects:
Control language (move one adjacent pair at a time)
Additional control patterns heard:
BCAFDE, AFDEBC, DEAFBC, DEBCAF

Learning a language with moved phrases

Transitional Probabilities in the Moved Phrase language and the Control language are different. The Moved Phrase language has lower probability across phrase boundaries than within phrases. The control language has the same probability no matter what.

	A→B	B→C	C→D	D→E	E→F
Moved phrases	1.00	0.60	1.00	0.60	1.00
Moved control	0.78	0.78	0.78	0.78	0.78

Learning a language with class size variation

Baseline pattern: ABCDEF

Phrases AB CD EF: Difference is 2 words vs. 4 words per class

Example strings heard:
kof neb jes zor fal nav
daz neb rel zor taf sib
…

Stimuli: 80 ABCDEF
mer lev tid lum rud nav
hox lev sot lum ker sib

A words	B words	C words	D words	E words	F words
KOF (oaf)		JES (dress)		FAL (pal)	
DAZ (has)	NEB (web)	REL (fell)	ZOR (core)	TAF (waif)	NAV (have)
MER (her)	LEV (rev)	TID (bid)	LUM (bum)	RUD (bud)	SIB (bib)
HOX (box)		SOT (coat)		KER (her)	

Learning a language with variable class size

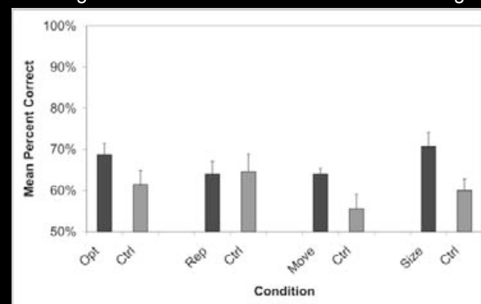
Transitional Probabilities in the Class Size Variation language and the Control language are different. The Class Size Variation language has different probability between individual words within the classes, based on class size. The control language has the same probability no matter what. Both the Class Size Variation language and the control language have the same probability between classes, however.

	DAZ→NEB	NEB→REL	REL→ZOR	ZOR→TAF	TAF→NAV
Class size variation	.50	.25	.50	.25	.50
Class size control	.33	.33	.33	.33	.33

	A→B	B→C	C→D	D→E	E→F
Class size variation	1.00	1.00	1.00	1.00	1.00
Class size control	1.00	1.00	1.00	1.00	1.00

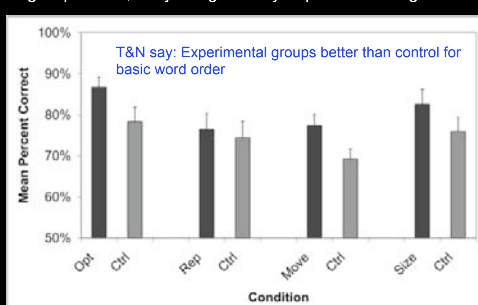
Artificial Language Learning: Categorization, Day 1

Generally above chance performance (50%), control group performing about the same or a little worse than test groups.



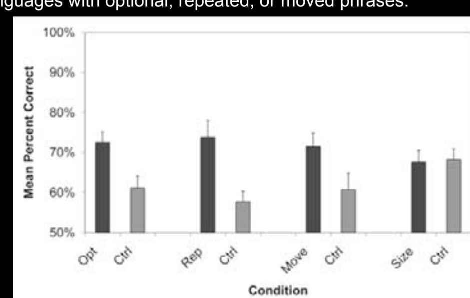
Artificial Language Learning: Categorization, Day 5

General improvement, though test groups still a little better than control groups. Still, subjects generally capable of categorization.



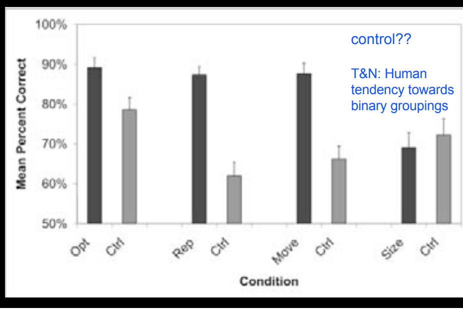
Artificial Language Learning: Phrases, Day 1

In each case, even after only 20 minutes of exposure (day 1), test subjects are better than control subjects for each of the languages with optional, repeated, or moved phrases.



Artificial Language Learning: Phrases, Day 5

After 5 days of exposure (100 minutes), the difference between control subjects and test subjects becomes apparent.



Learning a language with optional phrases, repeated phrases, moved phrases, & class size variation

Baseline pattern: ABCDEF

Other patterns heard (phrases AB CD EF moved):
CDEF, ABEF, ABCD, ABCDEFAB, ABCDEFCD, ABCDEFEF, ABCDEF, ABEFCD, CDABEF, CDEFAB, EFABCD, EFCDAB

	A→B	B→C	C→D	D→E	E→F
All-combined	1.00	0.33	1.00	0.22	1.00
All-combined control	0.67	0.71	0.58	0.59	0.47

Transitional Probabilities in the “All-combined” language and the Control language are different. The “All-combined” language has lower probability across phrase boundaries than within phrases. The control language probabilities are more uniform, though they do vary.

Learning a language with optional phrases, repeated phrases, moved phrases, & class size variation

Baseline pattern: ABCDEF

Other patterns heard (phrases AB CD EF moved):
CDEF, ABEF, ABCD, ABCDEFAB, ABCDEFCD, ABCDEFEF, ABCDEF, ABEFCD, CDABEF, CDEFAB, EFABCD, EFCDAB

However, keep in mind that the number of valid sentence types is much larger…not to mention the total number of sentences in the language.

Language	*Sentence Types*	*Sentences*
Optional phrases	4	972
Repeated phrases	4	20,412
Moved phrases	6	4,374
Class size variation	1	512
All-combined	86	233,536

Predictions for all-combined?

One idea: Harder

Why? There are many more patterns that are acceptable for the artificial language. Even if transitional probability is informative, it's a lot of information to track.

Prediction: Test subjects don't do much better than control subjects.

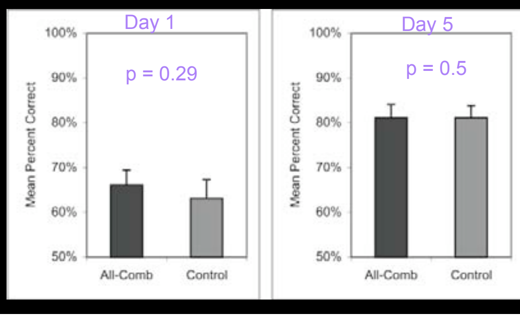
Second idea: The same, or easier.

Why? There are many more patterns that subjects' minds can catch. If even one of the variations (optional, repeated, moved phrases) is helpful, three of these will be even more helpful.

Prediction: Test subjects do much better than control subjects.

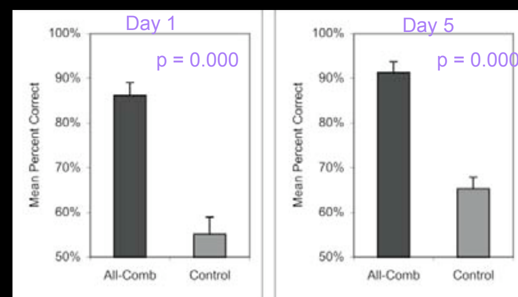
Artificial Language: Categorization

Control subjects do just as well as test subjects! Oops…
T&N: Perhaps due to memorization of the canonical form.



Artificial Language: Phrases

Test subjects outperforming control subjects on this measurement.



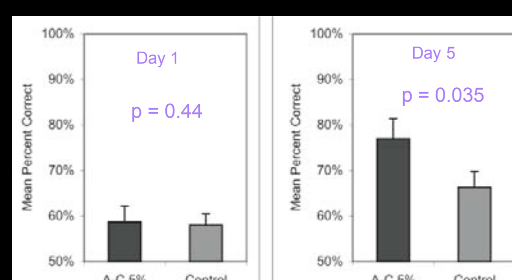
Idea for Control Subjects' Categorization Performance

What if only 5% of the data are of the canonical form? No memorization possible. But the transitional probability peaks and valleys are still constant, so experimental condition subjects should still do well.

	$A \to B$	$B \to C$	$C \to D$	$D \to E$	$E \to F$
All-combined 5%	1.00	0.33	1.00	0.22	1.00
All-combined 5% control	0.67	0.71	0.58	0.59	0.47

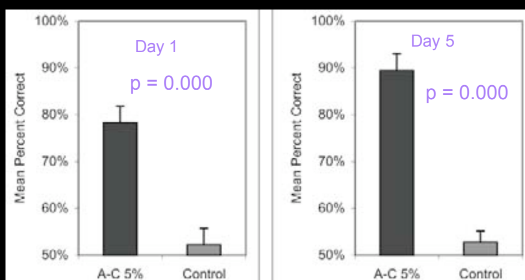
Artificial Language: Categorization

Test subjects do about as well as control subjects for being able to categorize. This is good, since it means subjects can abstract across the novel words.



Artificial Language: Phrases

Test subjects much better than control subjects. Second prediction is supported: finding phrases is easier when more variations are available, even though there are more patterns to learn.



Statistical Learning of Phrases

Thompson & Newport (2007): Adults can learn phrases in artificial languages if there are “sentences” that show the kinds of variation real sentences can have.

Interesting: When there are more variation types (optional, repeated, *and* moving phrases), adults are even better at unconsciously identifying phrases.

Open Question: How well will this work for real language data? (Remember Gambell & Yang (2006) found that transitional probabilities don't work so well for word segmentation when the data is realistic.)