Psych 156A/ Ling 150: Acquisition of Language II

Lecture 11
Phrases

Announcements

HW2 due today at the end of class

Review questions posted for phrases

HW3 available (due 5/29/12)

About Language Structure

Sentences are not just strings of words.

The girl danced with the elven 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 elven king.

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 elven king.

Noun Phrases (NP)

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 elven king.

Noun Phrases (NP)

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

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

She danced with him.

Noun Phrases (NP)

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

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 elven king.

Preposition Phrases (PP)

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 elven 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, based on their grammatical category.

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, based on their grammatical category.

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

Verb Phrases (VP)

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 elven 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, based on their grammatical category.

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, based on their grammatical category.

Det N V P Det Adj N

The girl danced with the elven king.

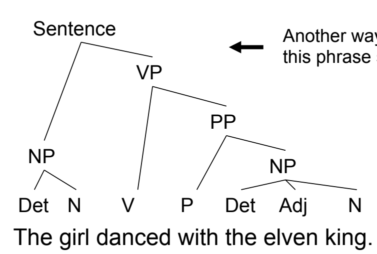
Verb Phrases (VP)

Preposition Phrases (PP)

Noun Phrases (NP)

About Language Structure

Sentences are not just strings of words.

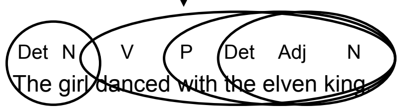


Another way to represent this phrase structure

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 elven 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

Prob("stlebe") < Prob("castle")
Prob("stlebe") < Prob("beyond")

to the castle beyond the goblin city

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 used to divide words into phrases?

the girl and the dwarf…

Posit a phrase boundary where the transitional probability is low between words (=~ group words together when their 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 using transitional probabilities.

Why?
TrProb(AB) = TrProb(BC) = TrProb(CD) = TrProb(DE) = TrProb(EF) = 1

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.

AB|CDEF

With the optional phrase left out, TrProb(BC) is less than 1 since sometimes B is followed by D instead of always being followed by C. A transitional probability learner later encountering ABCDEF might posit a phrase boundary between B and C because Tr(AB) and TrProb(CD) are still 1.

ABDEF

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.

AB|CDEF

Conclusion: AB is a unit, CDEF is a unit.
the goblin (= NP)
easily steals the child (= VP)

ABDEF

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

For ABDEF, Tr(AB) and Tr(DE) = 1, but TrProb(BD) < 1. So, a transitional probability learner will posit a boundary between B and D.
Conclusion: AB is a unit, DEF is a unit.
the goblin (= NP)
steals the child (= VP)

The goblin steals the child.

Artificial Language Experiments

Thompson & Newport 2007:
Adults (not children) listened to data from an artificial language for 20 minutes on multiple days

Assumption: Adults who are learning an artificial language will behave like children who are learning their first language since the adults have no prior experience with the artificial just as children have no prior experience with their first language

Is this a good assumption to make?

Adults in Artificial Language Experiments = Children in First Language?

Maybe yes, if children's brains behave like adults' brains. Then, the fact that adults can learn phrases from transitional probabilities means children should also be able to learn phrases from transitional probabilities.

Maybe no, if there are other factors that could interfere, such as adults having more cognitive resources to process information or using their native language experience to help them learn something about the artificial language. Then, just because adults succeed doesn't mean children will also succeed.

Some evidence that adults and children differ

Hudson Kam & Newport (2005): Adults and 5- to 7-year-old children differ in their willingness to make generalizations.

Adults and children were presented with an artificial language that used determiners (words like "the" and "a" in English) inconsistently in noun phrases. Sometimes, the determiner would appear (maybe 40%, 60% or 75% of the time) and sometimes it wouldn't.

Example of inconsistent use in English (rather than an artificial language):

"I want the pirate to win."

"I want pirate to win."

Some evidence that adults and children differ

Hudson Kam & Newport (2005): Adults and 5- to 7-year-old children differ in their willingness to make generalizations.

When presented with inconsistent input, adult learners matched the input and did not generalize determiner usage to all noun phrases. So, if they heard a determiner 60% of the time, they used a determiner 60% of the time when they produced sentences in this language.

Adult production:

"I want the pirate to win." (60%)

"I want pirate to win." (40%)

Some evidence that adults and children differ

Hudson Kam & Newport (2005): Adults and 5- to 7-year-old children differ in their willingness to make generalizations.

When presented with inconsistent input, child learners often generalized determiner usage to all noun phrases. So, if they heard a determiner 60% of the time, they used a determiner either 100% of the time when they produced sentences in this language - or 0% of the time (they didn't generalize the right way necessarily).

Child production:

"I want the pirate to win." (100%)

"I want pirate to win." (0%)

…but maybe not as much as we think

Hudson Kam & Newport (2009): Adults can be made to generalize too, when given inconsistent input.

When presented with inconsistent input but with one determiner being dominant (used 60% of the time as compared to others used 20% or less of the time), adult learners often generalized only the dominant determiner and used it nearly all the time (90%).

Adult production:

"I want the pirate to win." (90%)

"I want pirate to win." (10%)

…but maybe not as much as we think

Hudson Kam & Newport (2009): Children still differ from adults in *what* they generalize.

When presented with inconsistent input but with one determiner being dominant (used 60% of the time as compared to others used 20% or less of the time), child learners often generalized one determiner (even if it wasn't the dominant one) and used it nearly all the time (ex: 90%).

Child production:

"I want pirate to win." (10%)

"I want this pirate to win." (90%)

Artificial Language Similar To Real Language?

Properties of the artificial language that are similar to real language properties

optional phrases (the goblin chased a chicken *in the castle*)
PP is optional in the sentence

repeated phrases (NP Verb NP PP)
More than one NP is used in the sentence

moved phrases (In the castle the goblin chased a chicken)
PP is moved to the front of the sentence

Artificial Language Experiments

Baseline pattern: ABCDEF

real language parallel

A B C D E F
The goblin easily steals a 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

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.

	$A \to B$	$B \to C$	$C \to D$	$D \to E$	$E \to F$
Optional phrases	1.00	0.80	1.00	0.80	1.00
Optional control	0.90	0.90	0.90	0.90	0.90

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

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

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):
ABEFCD, CDABEF, CDEFAB, EFABCD, EFCDAB

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

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

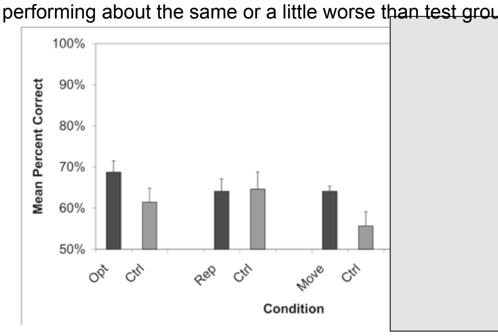
Learning a language with moved phrases

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

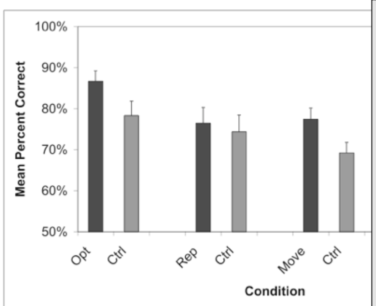
Artificial Language Learning: Categorization, Day 1

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



Artificial Language Learning: Categorization, Day 5

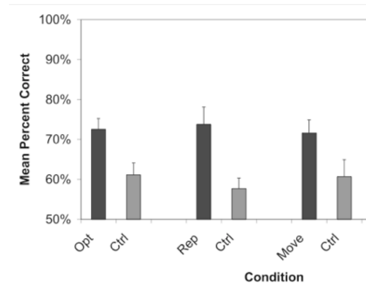
General improvement, though test groups still a little better than control groups. Still, subjects generally capable of categorization.



Mean % correct for all subjects is significantly above chance (which would be 50%)

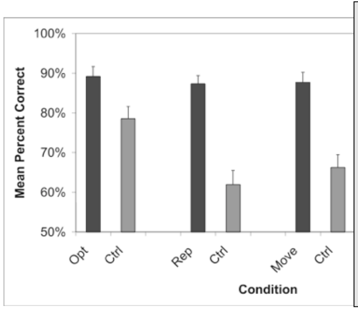
Artificial Language Learning: Phrases, Day 1

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.



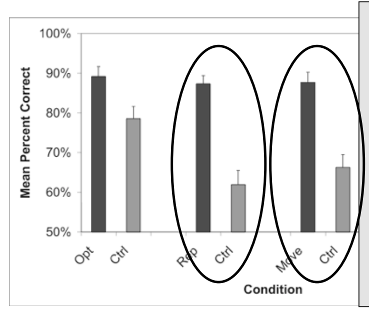
Artificial Language Learning: Phrases, Day 5

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



Artificial Language Learning: Phrases, Day 5

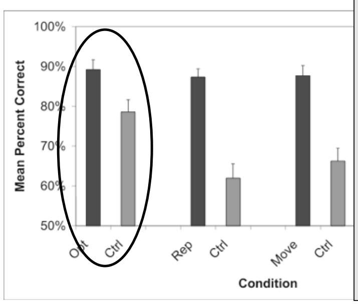
After 5 days of exposure (100 minutes), the difference between control subjects and test subjects becomes apparent.



Some properties seem easier to pick up on than others (repeated and movement language subjects are much better than control subjects).

Artificial Language Learning: Phrases, Day 5

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



Interestingly, control subjects in the optional phrase condition actually did really well - this is unexpected since the transitional probabilities were uninformative.

Learning a language with optional phrases, repeated phrases, and moved phrases

Baseline pattern: ABCDEF

Other patterns heard (phrases AB CD EF moved, repeated, or left out):
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.

Predictions for all-combined condition?

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 because there are so many patterns that are acceptable and even more potential patterns that are unacceptable.

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

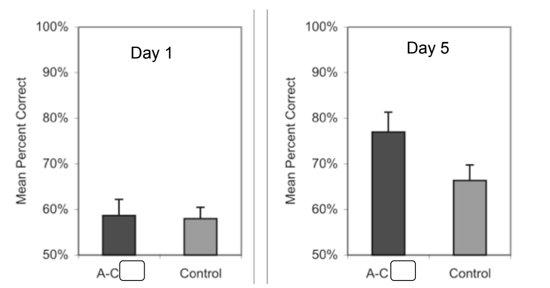
Second idea: The same, or easier.

Why? There are many more patterns that subjects’ minds can get information from. If even one of the variations (optional, repeated, moved phrases) is helpful, three of these will be even more helpful. This is reflected in the transitional probabilities, which are much lower across phrases than within phrases.

Prediction: Test subjects do much better than control subjects.

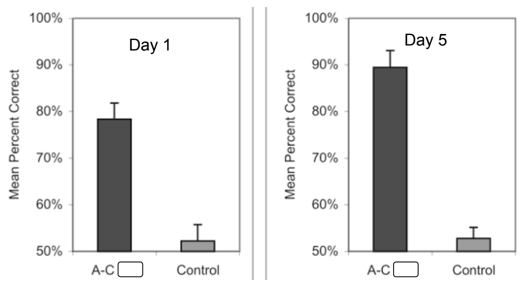
Artificial Language: Categorization

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 artificial words and realize some belong to the same category.



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.



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

Open Questions:

(1) 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 child-directed speech samples.) Is it actually useful?

(2) Will children be able to use transitional probabilities to find phrases? Is it useable?

Questions?

You should be able to do all the review questions for phrases and question 1 on HW3.