Psych229: Language Acquisition

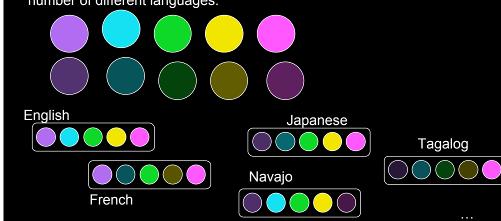
Lecture 20
Syntax Learning

Baker (2001): Complex Systems

Similarities & Differences: Parameters

Chomsky: Different combinations of different basic elements (parameters) would yield the observable languages.

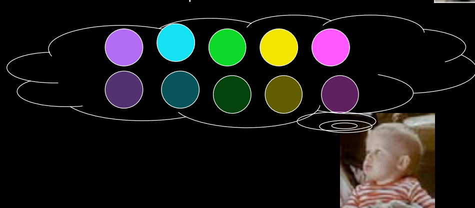
Idea: A relatively small number of parameters yields a large number of different languages.



Baker (2001): Complex Systems

Similarities & Differences: Parameters

Chomsky: Children are born knowing the parameters of variation. This is part of Universal Grammar. Input from the environment determines what values these parameters should have.



Baker (2001): Complex Systems

Similarities: Greenberg's Generalizations

Word Order Generalizations

Navajo	Japanese
Basic word order: Subject Object Verb	Basic word order: Subject Object Verb
Ashkii at'ééd yiyiiltsá Boy girl saw	Jareth-ga Hoggle-o butta Jareth Hoggle hit
The boy saw the girl	Jareth hit Hoggle

Baker (2001): Complex Systems

Similarities: Greenberg's Generalizations

Word Order Generalizations

Navajo	Japanese
Basic word order: Subject Object Verb	Basic word order: Subject Object Verb
Postpositions: Noun Phrase Postposition	Postpositions: Noun Phrase Postposition
'éé' biih náásdzá clothing into I-got-back I got back into (my) clothes.	Jareth-ga Sarah to kuruma da Jareth Sarah with car by London ni itta London to went Jareth went to London with Sarah by car.

Baker (2001): Complex Systems

Similarities: Greenberg's Generalizations

Word Order Generalizations

Navajo	Japanese
Basic word order: Subject Object Verb	Basic word order: Subject Object Verb
Postpositions: Noun Phrase Postposition	Postpositions: Noun Phrase Postposition
Possessor before Possessed Possessor Possession	Possessor before Possessed Possessor Possession
Chidí bi-jáád Car its-leg	Toby-no imooto-ga Toby's sister
the wheel of a car	Toby's sister

Baker (2001): Complex Systems

Similarities: Greenberg's Generalizations

Word Order Generalizations

Navajo	Japanese
Basic word order: Subject Object Verb	Basic word order: Subject Object Verb
Postpositions: Noun Phrase Postposition	Postpositions: Noun Phrase Postposition
Possessor before Possessed Possessor Possession	Possessor before Possessed Possessor Possession

Baker (2001): Complex Systems

Similarities: Greenberg's Generalizations

Word Order Generalizations

English	Edo (Nigeria)
Basic word order: Subject Verb Object	Basic word order: Subject Verb Object
Sarah found Toby.	Òzó mién Adésuwá Ozo found Adesuwa.

Baker (2001): Complex Systems

Similarities: Greenberg's Generalizations

Word Order Generalizations

English	Edo (Nigeria)
Basic word order: Subject Verb Object	Basic word order: Subject Verb Object
Prepositions: Preposition Noun Phrase	Prepositions: Preposition Noun Phrase
Jareth gave the crystal to Sarah.	Òzó rhié néné ebé né Adésuwá Ozo gave the book to Adesuwa.

Baker (2001): Complex Systems

Similarities: Greenberg's Generalizations

Word Order Generalizations

English	Edo (Nigeria)
Basic word order: Subject Verb Object	Basic word order: Subject Verb Object
Prepositions: Preposition Noun Phrase	Prepositions: Preposition Noun Phrase
Possessed before Possessor Possession Possessor	Possessed before Possessor Possession Possessor
quest of Sarah	Omo Ozó child Ozo
(alternative: Sarah's quest)	Ozo's child

Baker (2001): Complex Systems

Similarities: Greenberg's Generalizations

Word Order Generalizations

English	Edo (Nigeria)
Basic word order: Subject Verb Object	Basic word order: Subject Verb Object
Prepositions: Preposition Noun Phrase	Prepositions: Preposition Noun Phrase
Possessed before Possessor Possession Possessor	Possessed before Possessor Possession Possessor

Baker (2001): Complex Systems

Similarities: Greenberg's Generalizations

Word Order Generalizations

Point: Forty-five "universals" of languages found - patterns overwhelmingly followed by languages with unshared history (Navajo & Japanese, English & Edo)

Not all combinations are possible - some patterns rarely appear
Ex: Subject Verb Object language (English/Edo-like) + postpositions (Navajo/Japanese-like)

Baker (2001): Complex Systems

More Similarities & Differences

French vs. Italian

French	Italian
Subject Verb	Subject Verb
Jareth arrivera.	Jareth verrá.
Jareth will-come.	Jareth will-come.

Baker (2001): Complex Systems

More Similarities & Differences

French vs. Italian

French	Italian
Subject Verb	Subject Verb
*Verb Subject	Verb Subject
*Arrivera Jareth.	Verrá Jareth.
*Will-arrive Jareth.	Will-arrive Jareth.

Baker (2001): Complex Systems

More Similarities & Differences

French vs. Italian

French	Italian
Subject Verb	Subject Verb
*Verb Subject	Verb Subject
*Verb	Verb
*Arrivera	Verrá
He-will-come	He-will-come

Baker (2001): Complex Systems

More Similarities & Differences

French vs. Italian

French	Italian
Subject Verb	Subject Verb
*Verb Subject	Verb Subject
*Verb	Verb

Embedded Subject-question formation (easy to miss)

Tu veux que Marie épouse Jean.	Credi che Jareth verrá.
You want that Marie marries Jean.	You think Jareth will-come.
*Qui veux-tu que ___ épouse Jean?	Che credi che __ verrá?
Who want-you that marries Jean?	Who think-you that will-come?
Que veux-tu qui ___ épouse Jean?	

Baker (2001): Complex Systems

More Similarities & Differences

French vs. Italian

French	Italian
Subject Verb	Subject Verb
*Verb Subject	Verb Subject
*Verb	Verb

Embedded Subject-question formation (easy to miss)

*Qui veux-tu que ___ épouse Jean?	Che credi che __ verrá?
Who want-you that marries Jean?	Who think-you that will-come?
Que veux-tu qui ___ épouse Jean?	

Expletives

*Pleut	Piove.
It-rains.	It-rains.
Il pleut.	

Baker (2001): Complex Systems

More Similarities & Differences

French vs. Italian

French	Italian
Subject Verb	Subject Verb
*Verb Subject	Verb Subject
*Verb	Verb

Embedded Subject-question formation (easy to miss)

*Qui veux-tu que ___ épouse Jean?	Che credi che __ verrá?
Who want-you that marries Jean?	Who think-you that will-come?
Que veux-tu qui ___ épouse Jean?	

Expletives

All these involve the subject in some way - coincidence?

Idea: No! Parameter involving the subject.

Baker (2001): Complex Systems

The Value of Parameters for Learning: Learn the Hard Stuff from the Easy Stuff

French vs. Italian: Subject Parameter

French	Italian
Subject Verb	Subject Verb
*Verb Subject	Verb Subject
*Verb	Verb

Easy to notice

Hard to notice

Embedded Subject-question formation (easy to miss)

*Qui veux-tu que ___ épouse Jean?
Who want-you that marries Jean?
Que veux-tu qui ___ épouse Jean?

Che credi che ___ verrá?
Who think-you that will-come?

Expletives

*Pleut
It-rains.
Il pleut.

Piove.
It-rains.

Yang (2004): Learning Complex Systems

Language is a complex system

Only humans seem able to learn human languages
Something in our biology must allow us to do this. Chomsky: Universal Grammar = innate biases for learning language.

But obviously language is *learned*, not just prespecified beforehand.
Constrained variation across languages: phonology, lexicon, structure.

English

Navajo

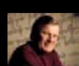
The point: need innate biases & probabilistic learning abilities
Need to explicitly integrate them with each other.

Yang (2004): Learning Complex Systems

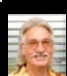
The linguist-psychologist breakdown

Linguists
Characterize "scope and limits of innate principles of Universal Grammar that govern the world's languages".

Noam Chomsky



David Lightfoot



Stephen Crain

Psychologists
Emphasize the "role of experience and the child's domain-general learning ability".

Michael Tomasello

Elizabeth Bates

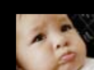
Brian MacWhinney

Yang (2004): Learning Complex Systems

Statistics for word segmentation (remember Gambell & Yang (2006))

"Modeling shows that the statistical learning (Saffran et al. 1996) does not reliably segment words such as those in child-directed English. Specifically, precision is 41.6%, recall is 23.3%. In other words, about 60% of words postulated by the statistical learner are not English words, and almost 80% of actual English words are not extracted. This is so even under favorable learning conditions".

Unconstrained (simple) statistics: not so good.



If statistical measure is constrained by language-specific knowledge (words have only one main stress), performance increases dramatically: 73.5% precision, 71.2% recall.

Constrained statistics - much better!

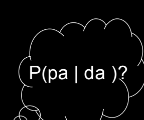
Yang (2004): Learning Complex Systems

Combining statistics with Universal Grammar

A big deal:
"Although infants seem to keep track of statistical information, any conclusion drawn from such findings must presuppose that children know *what kind of statistical information to keep track of*."

Ex: Transitional Probability

…of rhyming syllables?
…of syllables with nasal consonants?
…of syllables of the form CV (ba, ti)?



Yang (2004): Learning Complex Systems

Universal Grammar: Principles & Parameters

Principles: Apply to all human languages.
Ex: Language has hierarchical structure.
Smaller units are chunked into larger units.

sounds: g a b l ɪ n

syllables: gab lɪn

words: goblin

phrases: Noun Phrase (NP) The sneaky goblin; Verb Phrase (VP) stole the baby

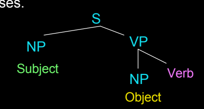
sentences: [S [NP The sneaky goblin] [VP stole [NP the baby]]]

Yang (2004): Learning Complex Systems

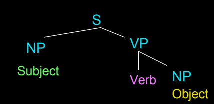
Universal Grammar: Principles & Parameters

Parameters: Constrained variation across languages. Child must learn which option native language uses.

Japanese/Navajo
- Basic word order: Subject Object Verb
- Postpositions: Noun Phrase Postposition
- Possessor before Possessed: Possessor Possession



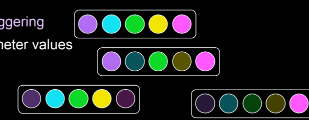
Edo/English
- Basic word order: Subject Verb Object
- Prepositions: Preposition Noun Phrase
- Possessed before Possessor: Possession Possessor



Yang (2004): Learning Complex Systems

Learning Parametric Systems: Triggering

Grammar = combination of parameter values



Trigger Learning:

At any given time, the child has in mind a single grammar.

If this current grammar can successfully analyze the current data, it stays. Otherwise, the child will shift to a completely new grammar by altering one or more parameter values. This new grammar will (hopefully) be able to analyze the current data.

Learning trajectory expectation: Sudden shifts in performance, not gradual. This is problematic.

Yang (2004): Learning Complex Systems

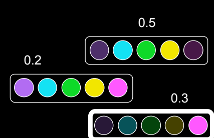
Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

The Learning Algorithm

For each data point *d* encountered in the input

Choose a grammar probabilistically from available grammars



Yang (2004): Learning Complex Systems

Learning Parametric Systems: Variational Learning

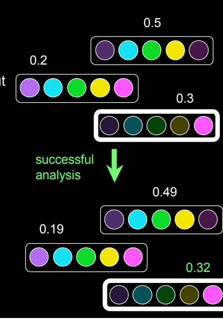
Grammars compete against each other to see which can best analyze the available data.

The Learning Algorithm

For each data point *d* encountered in the input

Choose a grammar probabilistically from available grammars

If this grammar can analyze the data point, increase its probability slightly (reward)



Yang (2004): Learning Complex Systems

Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

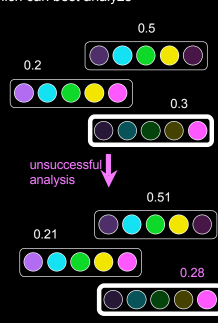
The Learning Algorithm

For each data point *d* encountered in the input

Choose a grammar probabilistically from available grammars

If this grammar can analyze the data point, increase its probability slightly (reward)

Else
decrease its probability slightly (punish)



Yang (2004): Learning Complex Systems

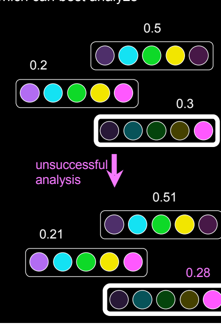
Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

The Basic Idea

If there is a single target grammar (the usual case), the non-target grammars will be chosen to analyze data at some point and be unsuccessful.

Each time this happens, they will lose some probability.



Yang (2004): Learning Complex Systems

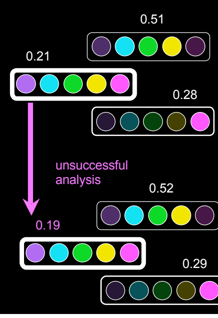
Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

The Basic Idea

If there is a single target grammar (the usual case), the non-target grammars will be chosen to analyze data at some point and be unsuccessful.

Each time this happens, they will lose some probability.



Yang (2004): Learning Complex Systems

Learning Parametric Systems: Variational Learning

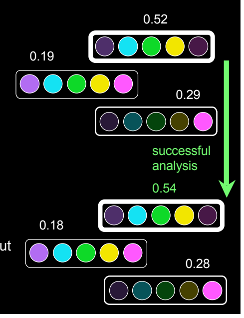
Grammars compete against each other to see which can best analyze the available data.

The Basic Idea

If there is a single target grammar (the usual case), the non-target grammars will be chosen to analyze data at some point and be unsuccessful.

Each time this happens, they will lose some probability.

The target grammar, in contrast, is always able to analyze the data and so will always increase in probability. It will eventually win out over the non-target grammars. (Probability =~ 1.0)



Yang (2004): Learning Complex Systems

Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

The Main Force

The crucial data is that which is unambiguous for the target grammar: this data is incompatible with non-target grammars.

The more unambiguous data there is, the faster the target grammar will win.

Added perk: Learning is then gradual (probabilistic).

Problem: Does unambiguous data exist for entire grammars? This requires data that is incompatible with every other possible parameter of every other possible grammar….

Yang (2004): Learning Complex Systems

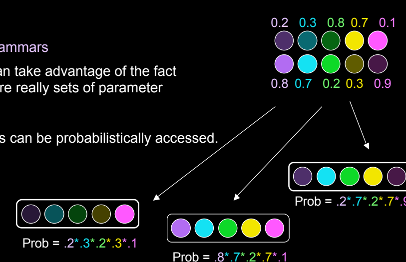
Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

Parameterized Grammars

This algorithm can take advantage of the fact that grammars are really sets of parameter values.

Parameter values can be probabilistically accessed.



Yang (2004): Learning Complex Systems

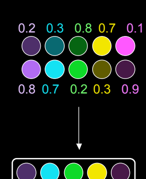
Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

The Learning Algorithm

For each data point *d* encountered in the input

Choose a grammar probabilistically from available grammars by probabilistically accessing the parameter values.



Yang (2004): Learning Complex Systems

Learning Parametric Systems: Variational Learning

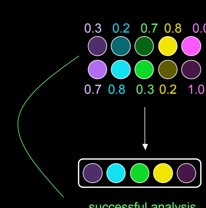
Grammars compete against each other to see which can best analyze the available data.

The Learning Algorithm

For each data point *d* encountered in the input

Choose a grammar probabilistically from available grammars by probabilistically accessing the parameter values.

If this grammar can analyze the data point, increase the probability of all participating parameters values slightly (reward)



Yang (2004): Learning Complex Systems

Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

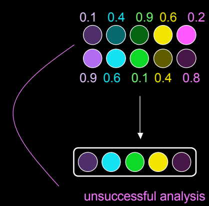
The Learning Algorithm

For each data point *d* encountered in the input

Choose a grammar probabilistically from available grammars by probabilistically accessing the parameter values.

If this grammar can analyze the data point, increase the probability of all participating parameters values slightly (reward)

Else
decrease the probability of all participating parameters values slightly (punish)



Yang (2004): Learning Complex Systems

Learning Parametric Systems: Variational Learning

Grammars compete against each other to see which can best analyze the available data.

The Main Force

The crucial data is that which is unambiguous for the target parameter values: this data is incompatible with non-target parameter values.

The more unambiguous data there is, the faster the target grammar will win.

Added perk remains: Learning is still gradual (probabilistic).

Problem ameliorated: unambiguous data much more likely to exist for individual parameter values instead of entire grammars.

Yang (2004): Learning Complex Systems

Variational Learning: Sample Case

Null subjects:

Parameter 1: Pro-drop, rely on unambiguous subject-verb agreement

Ex: Spanish, Italian (+pro-drop)

√ Yo puedo cantar.
I can-1st-sg sing-inf
'I can sing'

√ Puedo cantar.
can-1st-sg sing-inf
'I can sing'

√ Hay lluvia.
Is-3rd-sg rain
"There is rain"

Ex: English (-pro-drop)

√ I can sing

X * Can sing

X * Is rain

√ There is rain.

Yang (2004): Learning Complex Systems

Variational Learning: Sample Case

Null subjects:

Parameter 1: Topic-drop, drop subject/object if discourse topic

Ex: Chinese (+topic-drop)

(Topic = Jareth)

√ Mingtian guiji hui xiayu.
Tomorrow estimate will rain
'It is tomorrow that Jareth believes it will rain'

Ex: English (-topic-drop)

X *It is tomorrow that believes will rain.

Yang (2004): Learning Complex Systems

Variational Learning: Sample Case

Null subjects: 2 binary parameters, 4 grammars

+pro-drop, +topic-drop
Warlpiri, American Sign Language

+pro-drop, -topic-drop
Italian, Spanish

-pro-drop, +topic-drop
Chinese

-pro-drop, +topic-drop
English

What happens for an English-learning child?

Yang (2004): Learning Complex Systems

Variational Learning: Sample Case

Null subjects: 2 binary parameters, 4 grammars

+pro-dro… …drop
Warlpiri, A… Sign Language

+pro-dro… …opic-drop
Italia… …ish

-pro-drop, +topic-drop
Chinese

-pro-drop, +topic-drop
English

What happens for an English-learning child?

Pro-drop languages depend on rich subject-verb agreement morphology. English doesn't have that, which is something a child will easily notice. Knock out +pro-drop grammars.

Yang (2004): Learning Complex Systems

Variational Learning: Sample Case

Null subjects: 2 binary parameters, 4 grammars

+pro-drop, +to... drop
Warlpiri, A... Sign Language

+pro-dro... opic-drop
Italia... ish

-pro-drop, +topic-drop
Chinese

-pro-drop, +topic-drop
English

What happens for an English-learning child?

But this still leaves the +topic-drop option. What data will rule that out?

Answer: Expletive subjects. (Can't topic-drop them.)
"There's a goblin in the castle."
"It's raining outside."

But this only occurs in 1.2% of the data. (fairly rare)

Yang (2004): Learning Complex Systems

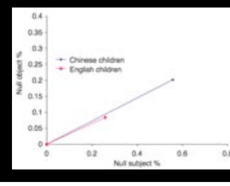
Variational Learning: Sample Case

Null subjects: Prediction if kids take awhile to notice English is -topic-drop

English kids use +topic-drop (Chinese-style) grammar until they encounter enough expletives to notice that English does not optionally drop topics.

Property of Chinese-style grammar: Can drop both subjects and objects

Prediction: When English children use +topic-drop grammar, they will drop subjects and objects at the same relative rate that +topic-drop (Chinese) children do



Same rate:
English children using Chinese grammar

Yang (2004): Learning Complex Systems

Variational Learning: General Predictions

The time course of when a parameter is set depends on how frequent the necessary evidence is in child-directed speech.

Parameters set early: more unambiguous data
Parameters set late: less unambiguous data
Parameters set at the same time: equal quantity of unambiguous data

Parameter	Target language	Requisite evidence	Input (%)	Time of acquisition
Wh fronting[a]	English	*Wh*-questions	25	very early [53]
verb raising[b]	French	verb adverb	7	1;8 [54]
obligatory subject	English	expletive subjects	1.2	3;0 [40,41]
verb second[c]	German/Dutch	OVS sentences [7,35]	1.2	3;0–3;2 [55]
scope marking[d]	English	long-distance *wh*-questions	0.2	4;0 + [56]

[a]English moves *Wh*-words in questions; in languages like Chinese, *Wh*-words stay put.
[b]In language like French, the finite verb moves past negation and adverbs ('Jean voit souvent/pas Marie'; 'Jean sees often/not Marie'), in contrast to English.
[c]In most Germanic languages, the finite verb takes the second position in the matrix clause, following one and exactly one phrase (of any type).
[d]In German, Hindi and other languages, long-distance *Wh*-questions leave intermediate copies of *Wh*-markers: '*Wer glaubst du wer Recht hat?*'; 'Who think you who right has?' (Who do you think has right?). For children to know that English doesn't use this the option, long-distance *Wh*-questions must be heard in the input. For many children, the German-option persists for quite some time, producing sentences like '*Who do you think who is the box?*' [56].

Thompson & Newport (2007): Statistically Learning Structure Rules

Transitional probability: segmenting words into phrases? Snapshot summary.

Artificial language paradigm, adult subjects, 20 minutes of exposure per session.
TPs: high within phrases, low across phrases

Properties of the artificial language: similar to real language properties
optional phrases (the goblin *in the castle* chased a chicken)
repeated phrases (Noun-Phrase Verb Noun-Phrase)
moved phrases (A chicken was chased by the goblin in the castle)
different-sized form classes (many nouns, few determiners)

Learning best when all of these properties are present ("structured complexity")

Thompson & Newport (2007): Statistically Learning Structure Rules

A look at real language properties

Optional phrases
Possible categories in a language:
Determiners ("the", "a"), Nouns ("goblin", "child"), Adverbs ("easily"), Verbs ("steals")

The goblin easily steals the child.

ABCDEF ← If the child only ever sees this, no way to know how the phrases break up.

ABCDEF

ABCDEF

ABCDEF

Thompson & Newport (2007): Statistically Learning Structure Rules

A look at real language properties in action with transitional probabilities

Example: Optional phrases
Possible categories in a language:
Determiners ("the", "a"), Nouns ("goblin", "child"), Adverbs ("easily"), Verbs ("steals")

The goblin easily steals the child.

AB|CDEF ← If the child only ever sees this, no way to know how the phrases break up.

ABCDEF

ABDEF ← With the optional phrase left out, the probability of (BC) is less than 1. Posit a phrase boundary there. AB is a unit, DEF is a unit.

ABDEF

The goblin steals the child.

Thompson & Newport (2007): Statistically Learning Structure Rules

Artificial language

Baseline pattern: ABCDEF

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)

Thompson & Newport (2007): Statistically Learning Structure Rules

Artificial language

Baseline pattern: ABCDEF

Optional language (remove one phrase at a time)

Phrases to be extracted: AB, CD, EF

Grammatical strings: ABCDEF, CDEF, ABEF, ABCD

Example strings heard:

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

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

	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 control language (remove one adjacent pair at a time)

Control strings: ABCDEF, BCDE, CDEF, ABEF, ABCF, ABCD

Thompson & Newport (2007): Statistically Learning Structure Rules

Assessment of learning

Sentence test: linear order correct? (extract ABCDEF pattern) [30 items]

Example: test between ABCDEF and random replacement ABCDCF

Sample items: kof hox jes sot fal ker vs. kof hox jes sot rel ker

Phrase test: phrase boundaries correct? (extract AB CD EF phrases) [18 items]

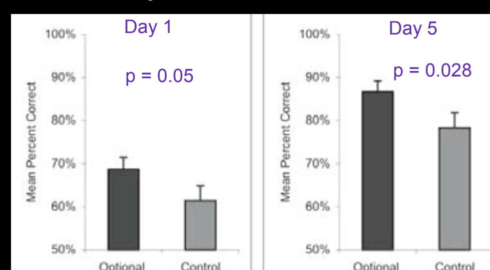
Example: test between AB and non-phrase BC

Sample items: kof hox vs. hox jes

Thompson & Newport (2007): Statistically Learning Structure Rules

Assessment of learning

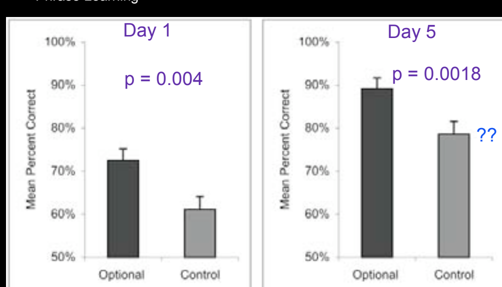
Sentence Learning



Thompson & Newport (2007): Statistically Learning Structure Rules

Assessment of learning

Phrase Learning



Thompson & Newport (2007): Statistically Learning Structure Rules

Artificial language

Baseline pattern: ABCDEF

Repeated phrases language (optionally repeat one phrase at the end, no word repeats)

Phrases to be extracted: AB, CD, EF

Grammatical strings: ABCDEF, ABCDEFAB, ABCDEFCD, ABCDEFEF

Example strings heard:

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
Half canonical: ABCDEF
Half distributed among other patterns

	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

Repeated control language (repeat one adjacent pair at a time)

Control strings: ABCDEF, ABCDEFAB, ABCDEFBC, ABCDEFCD, ABCDEFDE, ABCDEFFA

Thompson & Newport (2007): Statistically Learning Structure Rules

Artificial language

Baseline pattern: ABCDEF

Moved phrases language (1 of 6 legal permutations)
Phrases to be extracted: AB, CD, EF
Grammatical strings: ABCDEF, ABEFCD, CDABEF, CDEFAB, EFABCD, EFCDAB

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

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

	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

Moved control language (move one adjacent pair at a time)
Control strings: ABCDEF, ABEFCD, CDABEF, CDEFAB, EFABCD, EFCDAB, BCAFDE, AFDEBC, DEAFBC, DEBCAF

Thompson & Newport (2007): Statistically Learning Structure Rules

Artificial language

Baseline pattern: ABCDEF

Class size variation language (2 or 4 words per class)
Phrases to be extracted: AB, CD, EF
Grammatical strings: ABCDEF

Example strings heard:
kof neb jes zor fal nav   mer lev tid lum rud nav
daz neb rel zor taf sib   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)	

Thompson & Newport (2007): Statistically Learning Structure Rules

Artificial language

Baseline pattern: ABCDEF

Class size variation language (2 or 4 words per class)
Phrases to be extracted: AB, CD, EF
Grammatical strings: ABCDEF

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

word-level matters

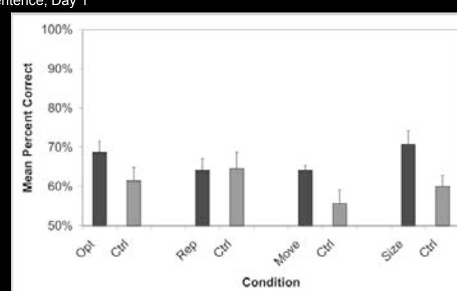
	A→B	B→C	C→D	D→E	E→F
Class size variation	1.00	1.00	1.00	1.00	1.00
Class size control	1.00	1.00	1.00	1.00	1.00

	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

Thompson & Newport (2007): Statistically Learning Structure Rules

Assessment of learning

Sentence, Day 1

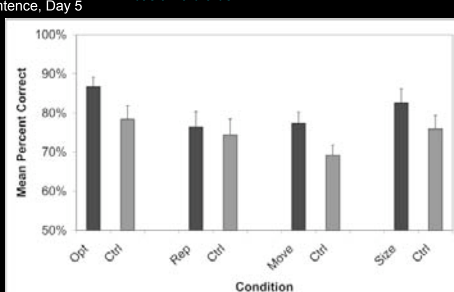


Thompson & Newport (2007): Statistically Learning Structure Rules

Assessment of learning

Sentence, Day 5

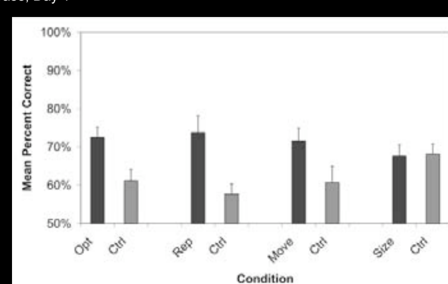
T&N say: Experimental groups better than control for basic word order



Thompson & Newport (2007): Statistically Learning Structure Rules

Assessment of learning

Phrase, Day 1

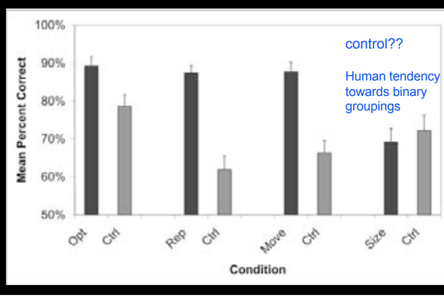


Thompson & Newport (2007): Statistically Learning Structure Rules

Assessment of learning

Phrase, Day 5

T&N say: Experimental groups better than control for basic word order, except for class size



Thompson & Newport (2007): Statistically Learning Structure Rules

Artificial language

Baseline pattern: ABCDEF

All-combined language (optional, repeated, moved, class size variation)

Phrases to be extracted: AB, CD, EF

Grammatical strings: ABCDEF, CDEF, ABCDEFAB, ABCDEFCD, CDABEF, CDEFAB, …

	$A \to B$	$B \to C$	$C \to D$	$D \to E$	$E \to F$
All-combined	1.00	0.56	1.00	0.52	1.00
All-combined control	0.80	0.83	0.74	0.77	0.74

More information….but many more members of language = harder to learn

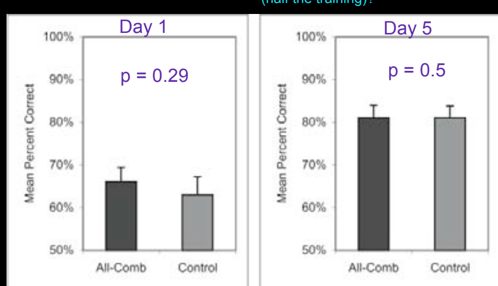
Language	*Sentence Types*	*Sentences*
Optional phrases	4	972
Repeated phrases	4	20,412
Moved phrases	6	4,374
Class size variation	1	512
All-combined	86	233,536

Thompson & Newport (2007): Statistically Learning Structure Rules

Assessment of learning

Sentence

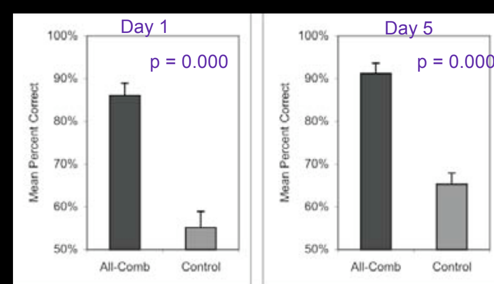
Control group performance: Due to memorization of canonical form (half the training)?



Thompson & Newport (2007): Statistically Learning Structure Rules

Assessment of learning

Phrase



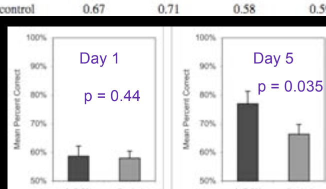
Thompson & Newport (2007): Statistically Learning Structure Rules

Idea for control subjects' sentence performance

Sentences: 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

Sentences



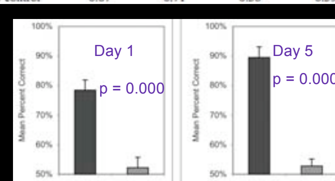
Thompson & Newport (2007): Statistically Learning Structure Rules

Idea for control subjects' sentence performance

Sentences: 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

Phrases



Thompson & Newport (2007): Statistically Learning Structure Rules

Discussion: Do we believe that this is strong evidence for the discovery of grammatical structure (and rules) via transitional probability?