Adjuncts, Repetition, and Learnability

Meaghan Fowlie
m.fowlie@uu.nl
http://meaghanfowlie.com

Univertiteit Utrecht

SCiL 2020
Formal Languages Workshop Panel
January 5, 2020

Learning

Blah blah...

Blah blah...

Photograph by Andrew Hetherington, Scientific American
July 20 2011
http://www.scientificamerican.com/article.cfm?id=benasich-baby-brains-signal-later-language-problems

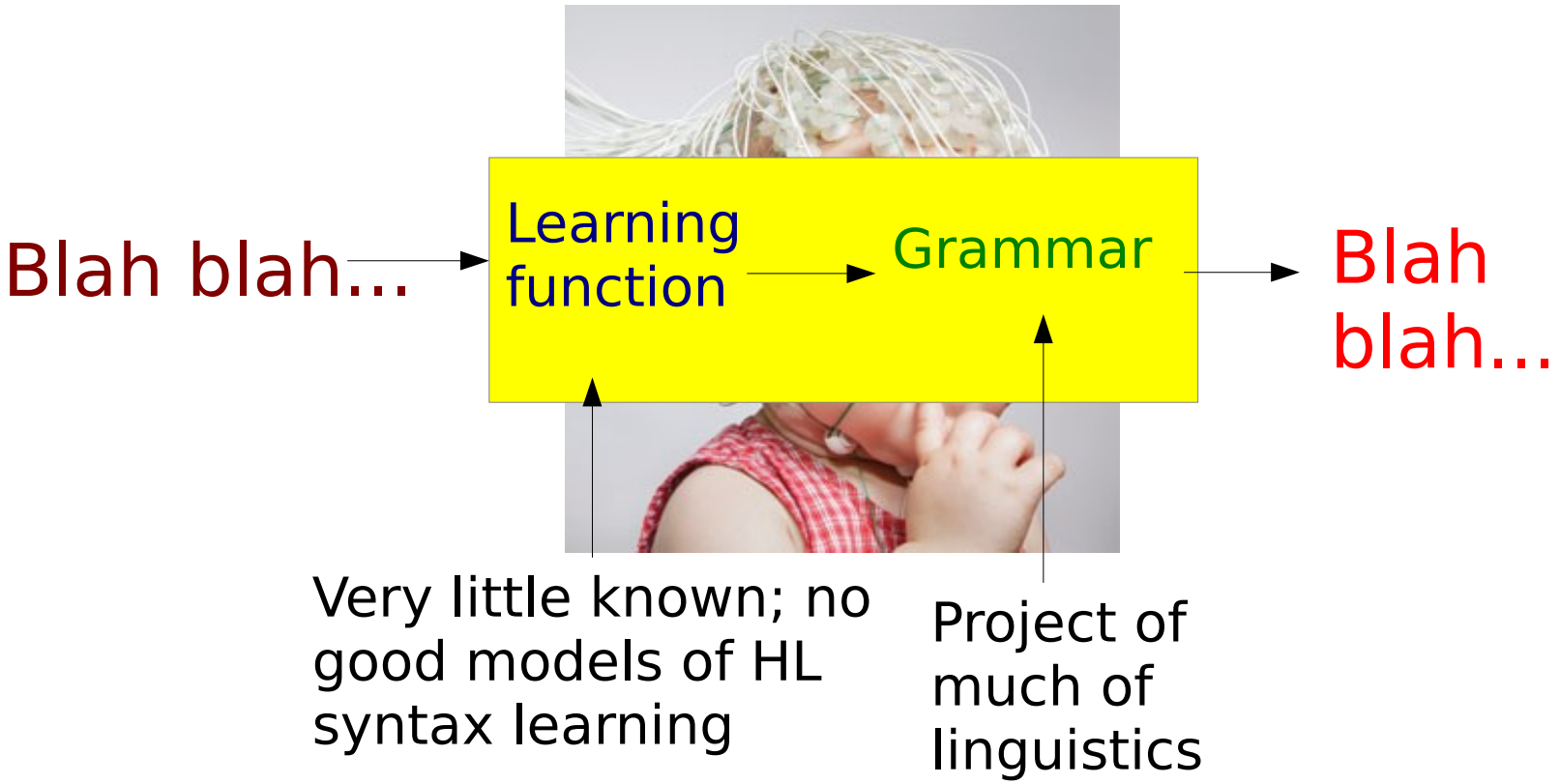
Learning

Blah blah...

Blah blah...

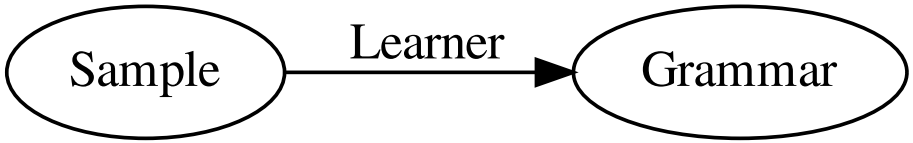
Photograph by Andrew Hetherington, Scientific American
July 20 2011
http://www.scientificamerican.com/article.cfm?id=benasich-baby-brains-signal-later-language-problems

Learning


Blah blah...
Learning function
Grammar
Blah blah...
Very little known; no good models of HL syntax learning
Project of much of linguistics

Overview

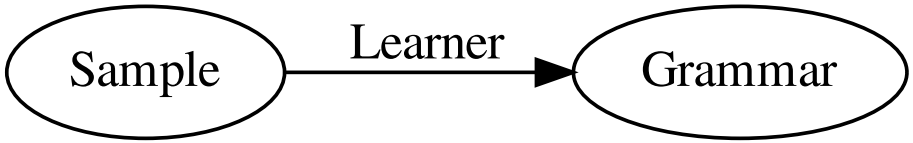
How do people learn adjuncts?



1. **Models**: How do formal models of language learning learn properties of adjuncts?

Overview

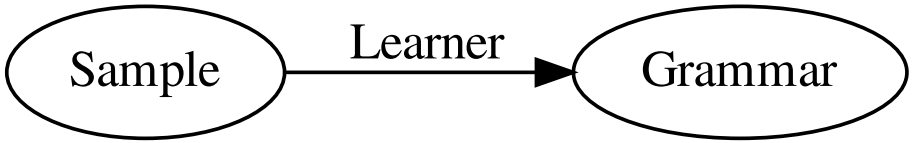
How do people learn adjuncts?



1. **Models**: How do formal models of language learning learn properties of adjuncts?
2. **People**: How do people learn properties of adjuncts?

Overview

How do people learn adjuncts?



1. **Models**: How do formal models of language learning learn properties of adjuncts?
2. **People**: How do people learn properties of adjuncts?
3. **Neural Networks**: How do neural networks learn properties of adjuncts?

Adjuncts

Generally adjectives, adverbs, prepositional phrases

(Almost) always **optional**, often **repeatable**

(1) a. My love is like a rose.
b. My love is like a red rose.
c. My love is like a red red rose.

(2) a. I'm tired!
b. I'm really tired!
c. I'm really really really really tired!

Adjuncts

Generally adjectives, adverbs, prepositional phrases

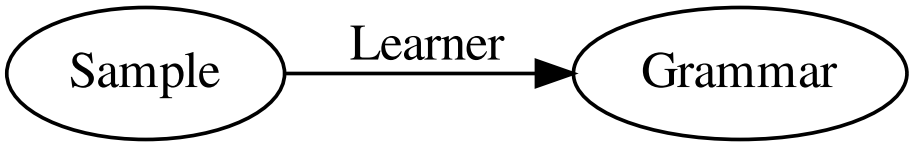
(Almost) always **optional**, often **repeatable**

(1) a. My love is like a rose.
b. My love is like a red rose.
c. My love is like a red red rose.

(2) a. I'm tired!
b. I'm really tired!
c. I'm really really really really tired!

(3) He suddenly (*suddenly suddenly) smiled.

Learnability



1. **Models**: How do formal models of language learning learn properties of adjuncts?
2. **People**: How do people learn properties of adjuncts?
3. **Neural Networks**: How do neural networks learn properties of adjuncts?

Learners we'll look at today

1. 0-reversible learner (Angluin, 1982)

Learners we'll look at today

1. 0-reversible learner (Angluin, 1982)
2. Substitutable CFGs (Clark, 2010)

Learners we'll look at today

1. 0-reversible learner (Angluin, 1982)
2. Substitutable CFGs (Clark, 2010)
3. CFGs (with finite kernel and finite context) (Clark et al., 2010)

Learnability

A very weak claim *For some definition of "learn" and some definition of "language", humans learn language*

Learnability

Definition ((String) Language ($L \subseteq \Sigma^*$))

A set of sequences of symbols, with the symbols taken from a finite set

eg: words are built out of phonemes $\rightarrow$ language = the words
eg: sentences are built of out morphemes $\rightarrow$ language = the sentences

Learnability

Definition ((String) Language ($L \subseteq \Sigma^*$))

A set of sequences of symbols, with the symbols taken from a finite set

eg: words are built out of phonemes $\rightarrow$ language = the words
eg: sentences are built of out morphemes $\rightarrow$ language = the sentences

Definition (Learner)

A function from samples of a language to grammars

Learnability

Definition ((String) Language ($L \subseteq \Sigma^*$))

A set of sequences of symbols, with the symbols taken from a finite set

eg: words are built out of phonemes $\rightarrow$ language = the words
eg: sentences are built of out morphemes $\rightarrow$ language = the sentences

Definition (Learner)

A function from samples of a language to grammars

Definition (Learn (in the limit from positive data) (Gold, 1967))

- A learner learns L if it eventually, after a finite number of samples, converges on a grammar that generates L
- A learner learns a class of languages if it distinguishes them from each other

Learnability

Definition ((String) Language ($L \subseteq \Sigma^*$))

A set of sequences of symbols, with the symbols taken from a finite set

eg: words are built out of phonemes $\rightarrow$ language = the words
eg: sentences are built of out morphemes $\rightarrow$ language = the sentences

Definition (Learner)

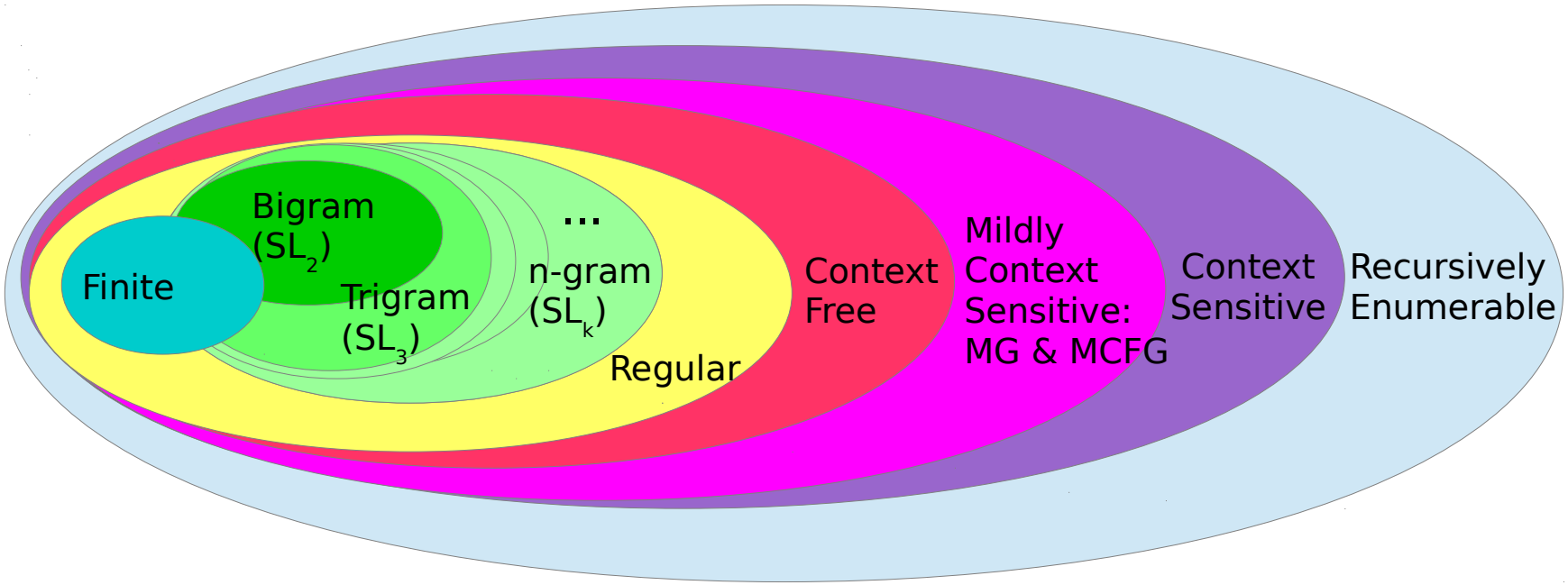
A function from samples of a language to grammars

Definition (Learn (in the limit from positive data) (Gold, 1967))

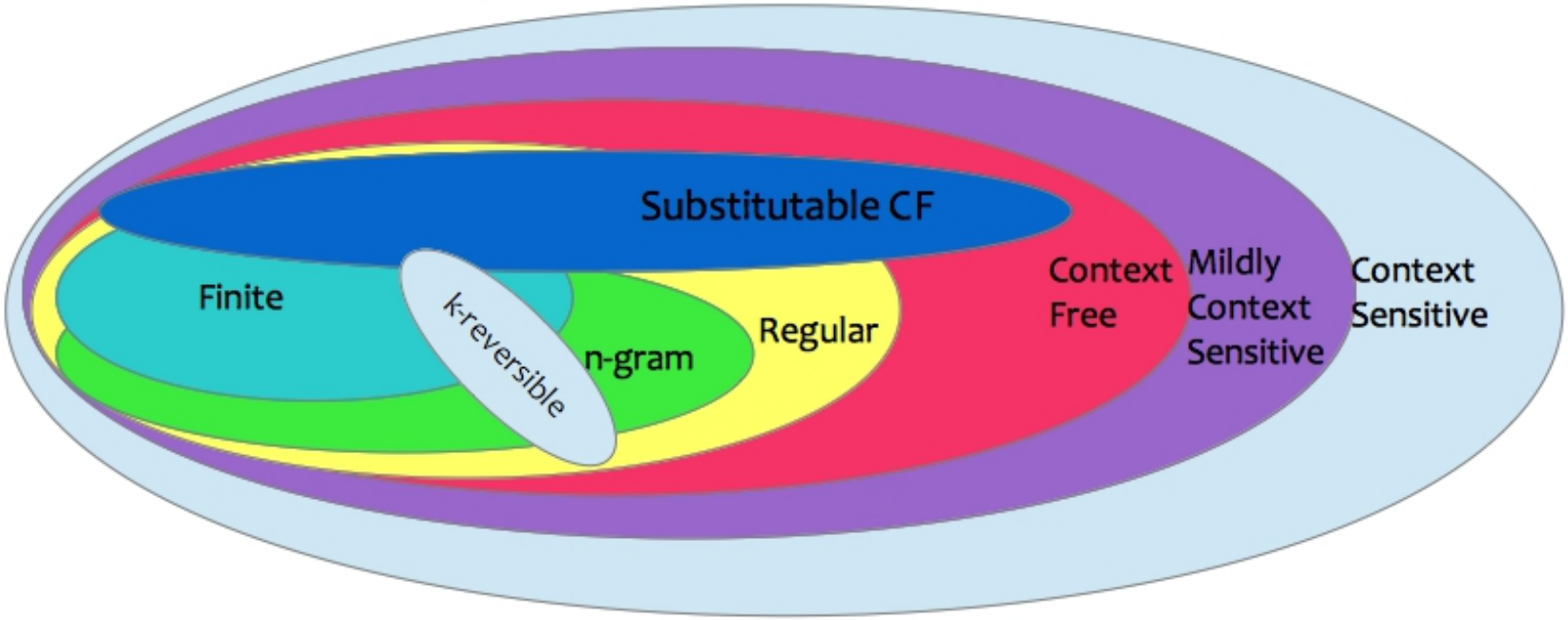
- A learner learns L if it eventually, after a finite number of samples, converges on a grammar that generates L
- A learner learns a class of languages if it distinguishes them from each other

Huge (really bad?) abstraction: just samples, no meaning or context

Refined Chomsky hierarchy



Refined Chomsky hierarchy vs Learnable classes



"Does this learner learn adjuncts?"

- But what does that even mean?

"Does this learner learn adjuncts?"

- But what does that even mean?
- What's an adjunct?

"Does this learner learn adjuncts?"

- But what does that even mean?
- What's an adjunct?
- How would a learner recognise something as an adjunct?

"Does this learner learn adjuncts?"

- But what does that even mean?
- What's an adjunct?
- How would a learner recognise something as an adjunct?
- How would I know whether it knew it was an adjunct?

"Does this learner learn adjuncts?"

- But what does that even mean?
- What's an adjunct?
- How would a learner recognise something as an adjunct?
- How would I know whether it knew it was an adjunct?

"Does this learner learn adjuncts?"

- But what does that even mean?
- What's an adjunct?
- How would a learner recognise something as an adjunct?
- How would I know whether it knew it was an adjunct?

$\rightarrow$ We need *surface properties* of adjuncts to look at

Optionality and repetition (bad definitions)

Definition (Optional (try 1))

$x \in \Sigma^*$ is *optional* iff sentences in L can have x but don't have to

$\rightarrow$ *cat* is optional in English because *The cat slept on the mat* $\in$ English and *The dog slept on the mat* $\in$ English

Optionality and repetition (bad definitions)

Definition (Optional (try 1))

$x \in \Sigma^*$ is *optional* iff sentences in L can have x but don't have to

$\rightarrow$ *cat* is optional in English because *The cat slept on the mat* $\in$ English and *The dog slept on the mat* $\in$ English

Definition (Repeatable (try 1))

$x \in \Sigma^*$ is *repeatable* iff you can say it more than once in a row

$\rightarrow$ *that* is repeatable in English because *I think that that is cool* $\in$ English

Optionality and repetition (good definitions)

Let $x, u, v \in \Sigma^*$

Definition (Optional)

x is *optional in context (u,v)* iff $uv \in L$ and $uxv \in L$

Optionality and repetition (good definitions)

Let $x, u, v \in \Sigma^*$

Definition (Optional)

x is *optional in context (u,v)* iff $uv \in L$ and $uxv \in L$

Definition (Repeatable)

x is *repeatable in context (u,v)* iff $ux^+v \subseteq L$

(4) Mary sniffed the (red red red) rose

We say *red* is repeatable and optional in context (Mary sniffed the, rose)

"How does this learner learn adjuncts?"

Lots of things this could mean, but I'm going to ask:

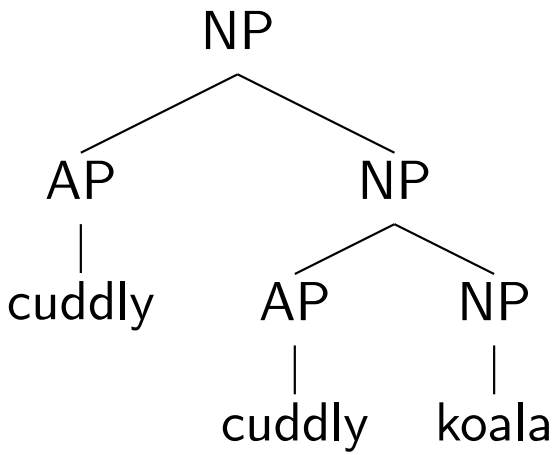
What generalisations will the learner make based on what samples?

Summary of learners

	0-rev	sub CFL	$CFL_{F,K}$
opt$\rightarrow$ rep	✓	✓	✗
rep$\rightarrow$ opt	✓	✓	✗
ac,abc,abbc$\rightarrow$ ab*c	✓	✓	✓
Learnable?	✓(Gold)	✓(Gold)	✓(MAT)
HL-like	no	somewhat	closer

Summary of learners

	0-rev	sub CFL	$CFL_{F,K}$
opt→ rep	✓	✓	✗
rep→ opt	✓	✓	✗
ac,abc,abbc→ ab*c	✓	✓	✓
Learnable?	✓(Gold)	✓(Gold)	✓(MAT)
HL-like	no	somewhat	closer

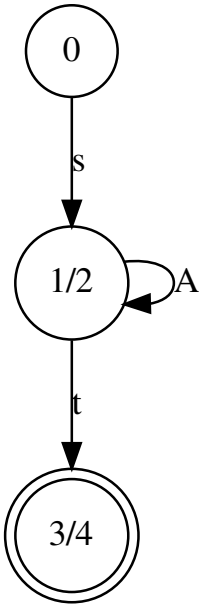


0-reversible learner (Angluin, 1982)

Definition

A FSA is 0-reversible iff it is deterministic both forward and backward

If L is 0-reversible then for all $u, v \in \Sigma^*$, if u and v share one suffix, they share all suffixes.



$L = sA^*t$

Note A is repeatable and optional in the context (s, t)

0-reversible learner

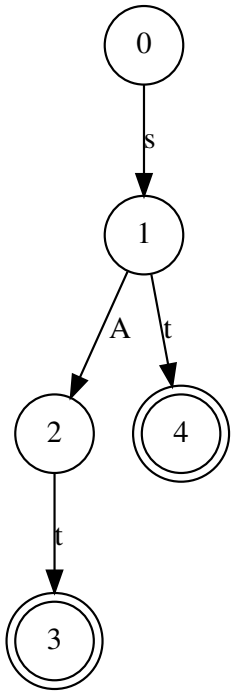
Optionality $\rightarrow$ Repetition

Sample: st, sAt

0-reversible learner

Optionality → Repetition

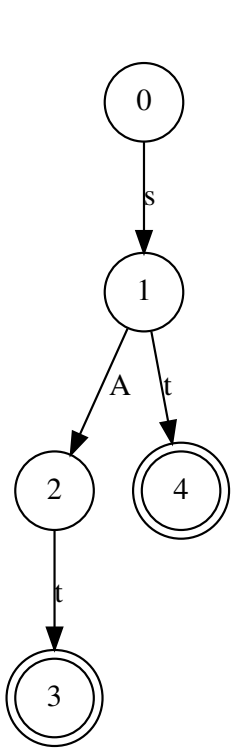
Sample: st, sAt

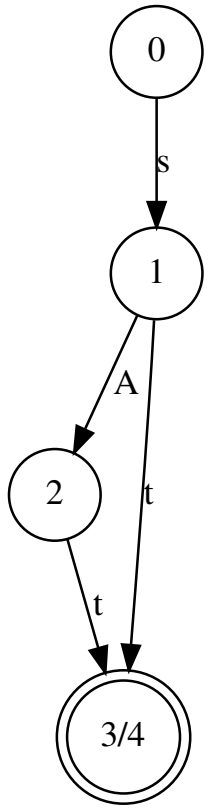


0-reversible learner

Optionality → Repetition

Sample: *st*, *sAt*

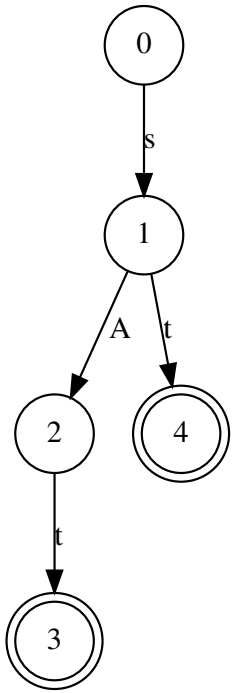


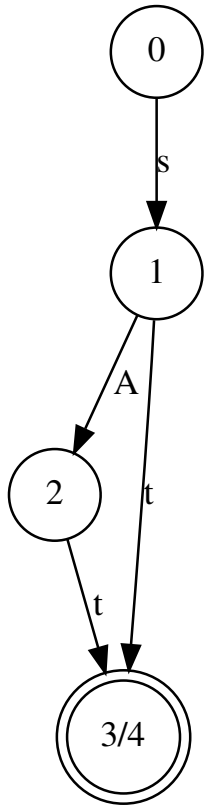


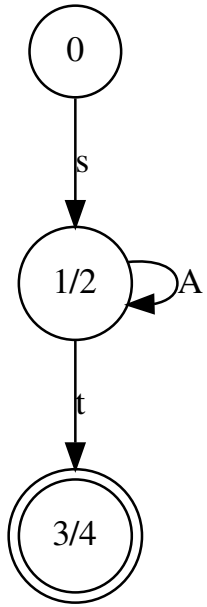
0-reversible learner

Optionality → Repetition

Sample: st, sAt







0-reversible learner

Theorem (Optionality → Repetition)

Let $u, v, x \in \Sigma^*$ and let $uv, uxv \in L$. Then $ux^*v \subseteq L$

0-reversible learner

Theorem (Optionality → Repetition)

Let $u, v, x \in \Sigma^*$ and let $uv, uxv \in L$. Then $ux^*v \subseteq L$

$uv \quad uxv$

$\quad\quad uxv \quad \rightarrow uxxv$

0-reversible learner: repetition $\rightarrow$ optionality

Theorem (repetition $\rightarrow$ optionality)

Let $ux^k v, ux^{k+1} \in L$ for some $k > 0$. Then $uv \in L$.

0-reversible learner: repetition $\rightarrow$ optionality

Theorem (repetition $\rightarrow$ optionality)

Let $ux^k v, ux^{k+1} \in L$ for some $k > 0$. Then $uv \in L$.

uxv $uxxv$

$uxv \rightarrow uv$

0-reversible learner - summary

- Optionality $\leftrightarrow$ Repetition

0-reversible learner - summary

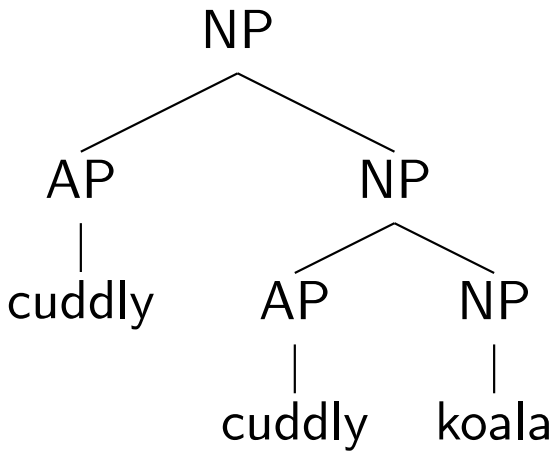
- Optionality $\leftrightarrow$ Repetition
- one repetition $\rightarrow$ indefinite repetition

0-reversible learner - summary

- Optionality $\leftrightarrow$ Repetition
- one repetition $\rightarrow$ indefinite repetition

0-reversible learner - summary

- Optionality $\leftrightarrow$ Repetition
- one repetition $\rightarrow$ indefinite repetition



HL is not 0-reversible

(5) a. The students slept
b. You slept
c. You were kicking yourself
d. *The students were kicking yourself

Summary of learners

	0-rev	sub CFL	$CFL_{F,K}$
opt→ rep	✓	✓	✗
rep→ opt	✓	✓	✗
ac,abc,abbc→ ab*c	✓	✓	✓
Learnable?	✓(Gold)	✓(Gold)	✓(MAT)
HL-like	no	somewhat	closer

Substitutable Context Free languages (Clark, 2010)

- CF equivalent of 0-reversible
- Learnable! (No time for the learner today, sorry!)

Definition (Substitutable context free language)

L is SCF iff for all $u, v, s, t, x_1, x_2 \in \Sigma^*$, if
$ux_1v \in L$ and
$ux_2v \in L$ and
$sx_1t \in L$ then
$sx_2t \in L$
i.e if two strings share one **context**, they share all contexts

Substitutable Context Free learner

Theorem (Optionality → Repetition)

Let $u, v, x \in \Sigma^*$ *and* $uv, uxv \in L$. *Then* $ux^*v \subseteq L(G_i)$.

Substitutable Context Free learner

Theorem (Optionality → Repetition)

Let $u, v, x \in \Sigma^*$ and $uv, uxv \in L$. Then $ux^*v \subseteq L(G_i)$.

uv $\quad$ uxv

$u\epsilon v$ $\quad$ $u\epsilon xv$

$\quad$ $\quad$ $u\epsilon xv \quad \rightarrow u\epsilon xxv$

$\quad$ $\quad$ $\quad \rightarrow uxxv$

Substitutable Context Free learner

Theorem (Repetition → Optionality)

Let $u, v, x \in \Sigma^*$ *and* $ux^n v, ux^{n+1} v \in L$ *Then* $uv \subseteq L(G_i)$.

Substitutable Context Free learner

Theorem (Repetition → Optionality)

Let $u, v, x \in \Sigma^*$ *and* $ux^n v, ux^{n+1} v \in L$ *Then* $uv \subseteq L(G_i)$.

uxv $uxxv$

Substitutable Context Free learner

Theorem (Repetition → Optionality)

Let $u, v, x \in \Sigma^*$ *and* $ux^n v, ux^{n+1} v \in L$ *Then* $uv \subseteq L(G_i)$.

uxv $uxxv$

$u\epsilon xv$ $u\epsilon xxv$

$u\epsilon xv \rightarrow u\epsilon v$

$\rightarrow uv$

Substitutable CF – summary

- repetition $\leftrightarrow$ optionality
- one repetition $\rightarrow$ indefinite repetition

Human Language is not substitutable CF

Intersubstitutability is a big part of syntactic categories:

(6)
a. The kids watched a movie
b. The kids found a movie
c. ϵ All the kids in the neighourhood watched a movie
d. ϵ We watched a movie

Human Language is not substitutable CF

Intersubstitutability is a big part of syntactic categories:

(6)
a. The kids watched a movie
b. The kids found a movie
c. ϵ All the kids in the neighourhood watched a movie
d. ϵ We watched a movie

But the "one common context $\rightarrow$ all common contexts" idea is too strong:

(7)
a. I hear you slept
b. I hear the kids slept
c. I hear you were kicking yourself
d. *I hear the kids were kicking yourself

Summary of learners

	0-rev	**sub CFL**	$CFL_{F,K}$
opt→ rep	✓	✓	✗
rep→ opt	✓	✓	✗
ac,abc,abbc→ ab*c	✓	✓	✓
Learnable?	✓(Gold)	✓(Gold)	✓(MAT)
HL-like	no	somewhat	closer

$CFG_{F,K}$ Clark et al. (2010)

Context-free languages with the finite kernel and finite context properties

- Loosely, CF languages such that you can make a context-free grammar using just sets of contexts a substring can appear in.

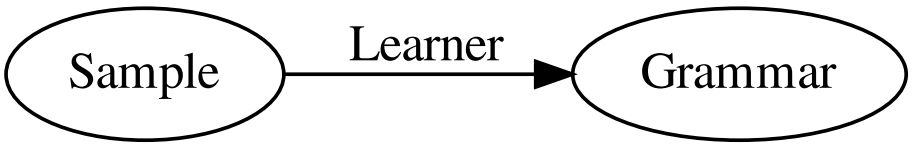
$CFG_{F,K}$ Clark et al. (2010)

Context-free languages with the finite kernel and finite context properties

- Loosely, CF languages such that you can make a context-free grammar using just sets of contexts a substring can appear in.
- Similar learning algorithm to Substitutable CFG

Summary of learners

	0-rev	sub CFL	$\mathbf{CFL_{F,K}}$
opt$\rightarrow$ rep	✓	✓	✗
rep$\rightarrow$ opt	✓	✓	✗
ac,abc,abbc$\rightarrow$ ab*c	✓	✓	✓
Learnable?	✓(Gold)	✓(Gold)	✓(MAT)
HL-like	no	somewhat	closer



1. **Models**: How do formal models of language learning learn properties of adjuncts?
2. **People**: How do people learn properties of adjuncts?
3. **Neural Networks:** How do NNs learn properties of adjuncts?

Artificial language learning

The paradigm:

Training phase Participants are exposed to grammatical items from the target language

Testing phase Participants are tested on new items to see what they learned. Data like reaction time and grammaticality judgments are gathered to infer what the participants learned

Artificial language learning experiment: Tagalog

(8) natulog ang babae
sleep D woman
'The woman is sleeping/slept'

Artificial language learning experiment: Tagalog

(8) natulog ang babae
sleep D woman
'The woman is sleeping/slept'

(9) natulog ang malaki babae
sleep D big woman
'The big woman is sleeping/slept'

Artificial language learning experiment: Tagalog

(8) natulog ang babae
sleep D woman
'The woman is sleeping/slept'

(9) natulog ang malaki babae
sleep D big woman
'The big woman is sleeping/slept'

(10) natulog ang babae malaki
sleep D woman big
'The big woman is sleeping/slept'

Artificial language learning experiment: Tagalog

(8) natulog ang babae
sleep D woman
'The woman is sleeping/slept'

(9) natulog ang malaki babae
sleep D big woman
'The big woman is sleeping/slept'

(10) natulog ang babae malaki
sleep D woman big
'The big woman is sleeping/slept'

(11) natulog ang malaki malaki babae
sleep D big big woman
'The big big woman is sleeping/slept'

Artificial language learning experiment: Tagalog

(8) natulog ang babae
sleep D woman
'The woman is sleeping/slept'

(9) natulog ang malaki babae
sleep D big woman
'The big woman is sleeping/slept'

(10) natulog ang babae malaki
sleep D woman big
'The big woman is sleeping/slept'

(11) natulog ang malaki malaki babae
sleep D big big woman
'The big big woman is sleeping/slept'

(12) natulog siguro ang babae
sleep maybe D woman
'Maybe the woman is sleeping/slept'

Artificial language learning experiment: Tagalog

(8) natulog ang babae
sleep D woman
'The woman is sleeping/slept'

(9) natulog ang malaki babae
sleep D big woman
'The big woman is sleeping/slept'

(10) natulog ang babae malaki
sleep D woman big
'The big woman is sleeping/slept'

(11) natulog ang malaki malaki babae
sleep D big big woman
'The big big woman is sleeping/slept'

(12) natulog siguro ang babae
sleep maybe D woman
'Maybe the woman is sleeping/slept'

G1:

V (Adv) D Adj* N

G2:

V (Adv) D N Adj*

Artificial language learning experiment: Tagalog

Research Question: In learning language, do people generalise from limited to indefinite repetition?

Training stimuli V (Adv) D (Adj) (Adj) (Adj) N

Testing stimuli also ungrammatical and V (Adv) D Adj^4 (Adj) N

To answer the research question: compare responses to **ungrammatical** and **generalised** stimuli. If they like generalised stimuli more than ungrammatical, they've generalised repetition

Artificial language learning experiment: Tagalog

Testing stimuli

- 100 grammatical sentences

Artificial language learning experiment: Tagalog

Testing stimuli

- 100 grammatical sentences
  - 62 new but familiar (no more than 3 adjectives)

Artificial language learning experiment: Tagalog

Testing stimuli

- 100 grammatical sentences
  - 62 new but familiar (no more than 3 adjectives)
  - 38 generalised (20 with 4 adjectives, 18 with 5)

Artificial language learning experiment: Tagalog

Testing stimuli

- 100 grammatical sentences
  - 62 new but familiar (no more than 3 adjectives)
  - 38 generalised (20 with 4 adjectives, 18 with 5)
- 78 ungrammatical sentences

Artificial language learning experiment: Tagalog

Testing stimuli

- 100 grammatical sentences
  - 62 new but familiar (no more than 3 adjectives)
  - 38 generalised (20 with 4 adjectives, 18 with 5)
- 78 ungrammatical sentences
  - 58 with the noun repeated instead of the adjective

Artificial language learning experiment: Tagalog

Testing stimuli

- 100 grammatical sentences
  - 62 new but familiar (no more than 3 adjectives)
  - 38 generalised (20 with 4 adjectives, 18 with 5)
- 78 ungrammatical sentences
  - 58 with the noun repeated instead of the adjective
  - 20 scrambled up grammatical sentences

Artificial language learning experiment 1: Tagalog

Training examples (Adj-medial grammar):

- natulog ang pusa'
- natulog siguro ang mapula kotse
- natulog siguro ang matanda matanda pusa'
- natulog ang masaya masaya masaya kotse

Artificial Language Learning experiment: Tagalog

Testing examples (Adj-medial grammar):
Generalised:

- umalis ang malaki malaki malaki malaki babae
- umalis ang malaki malaki malaki malaki malaki kotse
- umalis siguro ang mapula mapula mapula mapula babae

Repeated noun:

- *natulog ang malaki babae babae
- *umalis ang matanda kotse kotse kotse kotse

Scrambled:

- *siguro matanda matanda babe ang natulog
- *babae ang umalis

Artificial language learning experiment

Method:

- **Training:** Participants listen to randomised training stimuli over headphones

Artificial language learning experiment

Method:

- **Training:** Participants listen to randomised training stimuli over headphones
- While they listen they see a red dot on the screen

Artificial language learning experiment

Method:

- **Training:** Participants listen to randomised training stimuli over headphones
- While they listen they see a red dot on the screen
- They respond with a keypress according to which side of the screen it's on. (distractor task)

Artificial language learning experiment

Method:

- **Training:** Participants listen to randomised training stimuli over headphones
- While they listen they see a red dot on the screen
- They respond with a keypress according to which side of the screen it's on. (distractor task)
- **Testing:** Participants listen to new stimuli

Artificial language learning experiment

Method:

- **Training:** Participants listen to randomised training stimuli over headphones
- While they listen they see a red dot on the screen
- They respond with a keypress according to which side of the screen it's on. (distractor task)
- **Testing:** Participants listen to new stimuli
- They respond with a keypress whether they think it's a real or fake sentence of Tagalog (forced choice)

ALL experiment Results

- 51 UCLA undergrads

ALL experiment Results

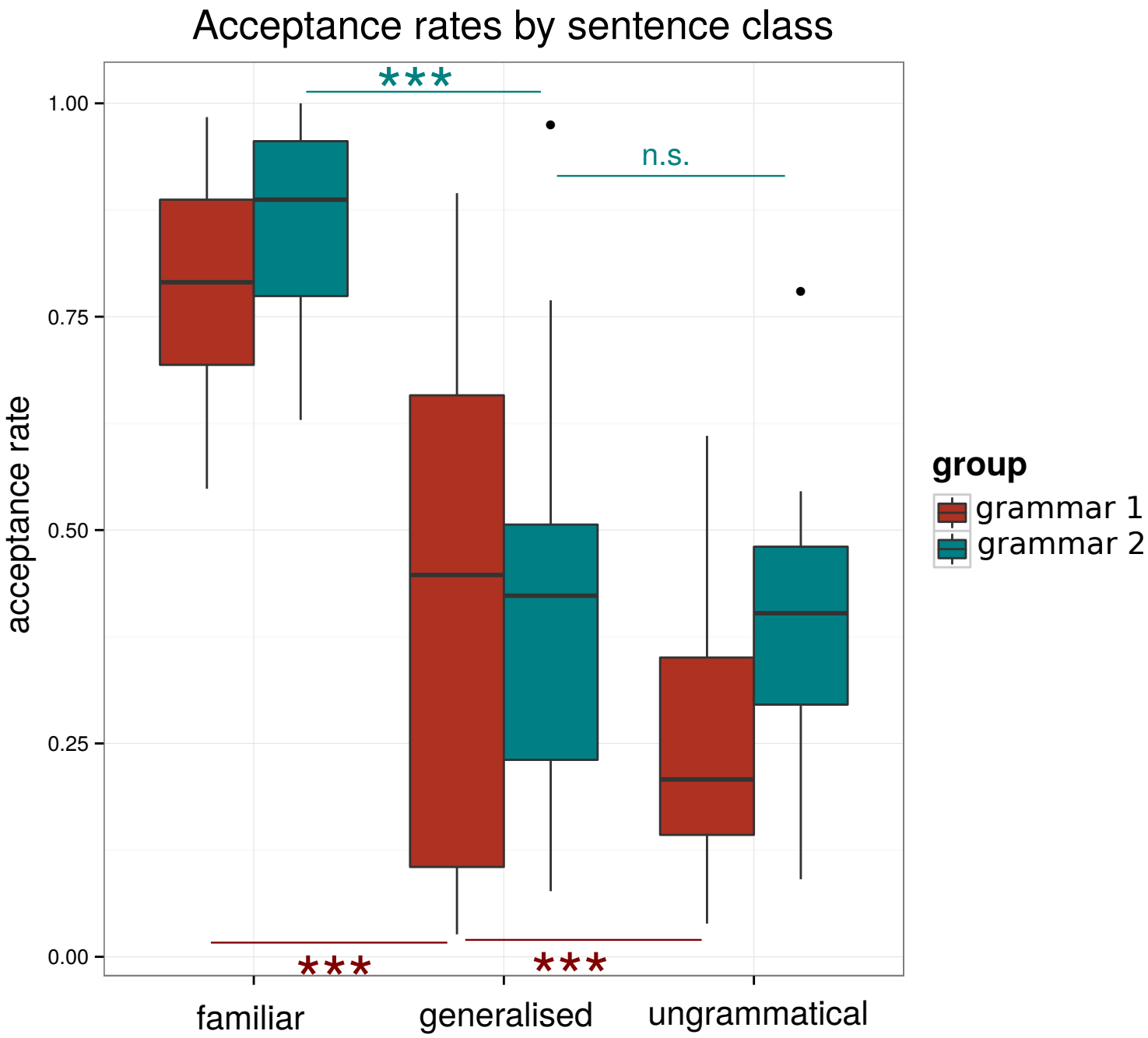
- 51 UCLA undergrads
- Dropped 6 peoples' results because they didn't learn the basics well enough (Accepted Familiar less than 0.15 more than Ungrammatical)

ALL experiment Results

- 51 UCLA undergrads
- Dropped 6 peoples' results because they didn't learn the basics well enough (Accepted Familiar less than 0.15 more than Ungrammatical)
- 29 Grammar 1: D A* N

ALL experiment Results

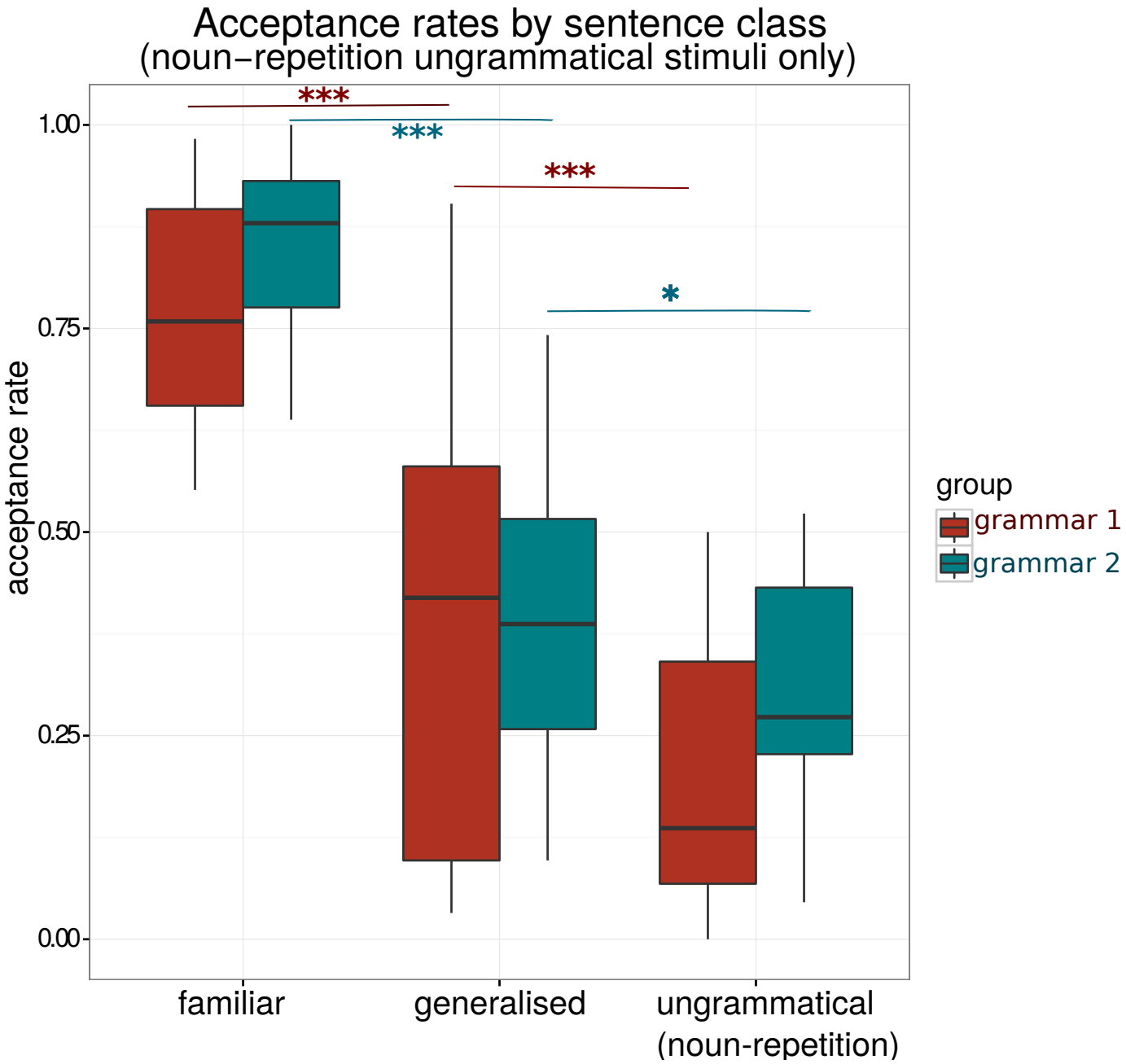
- 51 UCLA undergrads
- Dropped 6 peoples' results because they didn't learn the basics well enough (Accepted Familiar less than 0.15 more than Ungrammatical)
- 29 Grammar 1: D A* N
- 22 Grammar 2: D N A*

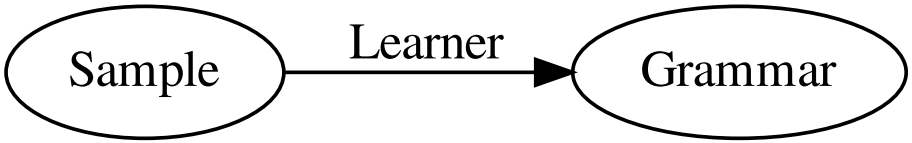

Acceptance rates by sentence class
***
n.s.
***
***
1.00
0.75
0.50
0.25
0.00
acceptance rate
familiar
generalised
ungrammatical
group
grammar 1
grammar 2

Ungrammatical types

- Repeated noun:
  - *natulog ang malaki babae babae
  - *umalis ang matanda kotse kotse kotse kotse
- Scrambled:
  - *siguro matanda matanda babe ang natulog
  - *babae ang umalis

- Adjective-final group's weird acceptance of ungrammatical items mostly driven by acceptance of Scrambled!
- Remove them, and everything's significant, but what on earth does that mean??


Acceptance rates by sentence class
(noun−repetition ungrammatical stimuli only)
***
***
***
*
1.00
0.75
0.50
0.25
0.00
acceptance rate
familiar
generalised
ungrammatical
(noun-repetition)
group
grammar 1
grammar 2



1. **Models**: How do formal models of language learning learn properties of adjuncts?
2. **People**: How do people learn properties of adjuncts?
3. **Neural Networks**: How do neural networks learn properties of adjuncts?

LSTMs generalise Tagalog repetition

1-layer LSTM encoder, trained to give grammaticality judgments

- **Training set:**
  - 200 grammatical (up to 3 adjectives)
  - 200 ungrammatical (128 Noun repetition, 72 scrambled)

LSTMs generalise Tagalog repetition

1-layer LSTM encoder, trained to give grammaticality judgments

- **Training set:**
  - 200 grammatical (up to 3 adjectives)
  - 200 ungrammatical (128 Noun repetition, 72 scrambled)
- **Development set:**
  - 56 grammatical (up to 3 adjectives)
  - 56 ungrammatical (28 noun rep, 28 scrambled)

LSTMs generalise Tagalog repetition

1-layer LSTM encoder, trained to give grammaticality judgments

- **Training set:**
  - 200 grammatical (up to 3 adjectives)
  - 200 ungrammatical (128 Noun repetition, 72 scrambled)
- **Development set:**
  - 56 grammatical (up to 3 adjectives)
  - 56 ungrammatical (28 noun rep, 28 scrambled)
- **Test set:**
  - all 1088 generalised to 4-20 adjectives
  - all 2176 ungrammatical (1088 scrambled generalised, 1088 4-20 noun rep)

LSTM Tagalog results

- $\approx 78 - 96\%$ accuracy on train, dev, and test

LSTM Tagalog results

- $\approx 78 - 96\%$ accuracy on train, dev, and test
- varying by initialisation, but always about the same for each set

LSTM Tagalog results

- $\approx 78 - 96\%$ accuracy on train, dev, and test
- varying by initialisation, but always about the same for each set
- Generalised to longer sentences and more repetition than it had seen before

LSTM Tagalog results

- $\approx 78 - 96\%$ accuracy on train, dev, and test
- varying by initialisation, but always about the same for each set
- Generalised to longer sentences and more repetition than it had seen before
- Errors don't seem to get worse with longer sentences

LSTM Tagalog results

- $\approx 78 - 96\%$ accuracy on train, dev, and test
- varying by initialisation, but always about the same for each set
- Generalised to longer sentences and more repetition than it had seen before
- Errors don't seem to get worse with longer sentences

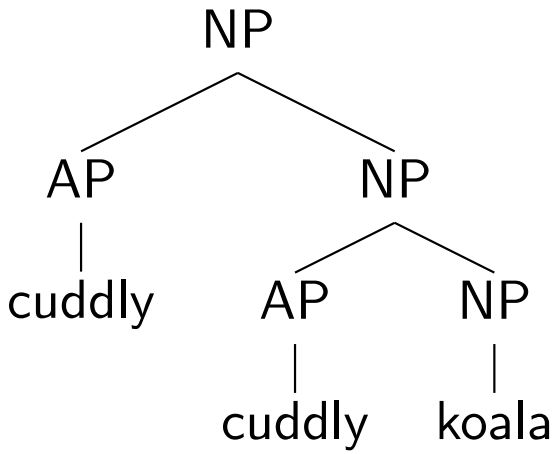
LSTM Tagalog results

- $\approx 78 - 96\%$ accuracy on train, dev, and test
- varying by initialisation, but always about the same for each set
- Generalised to longer sentences and more repetition than it had seen before
- Errors don't seem to get worse with longer sentences

What does this mean??

Summary

- Formal learners we looked at see optionality and repetition as the same thing



- people can (mostly) generalise from limited to indefinite repetition
- LSTMs can generalise really well from limited to indefinite repetition

Thank you!

References

Angluin, Dana. 1982. Inference of reversible languages. *Journal of the ACM (JACM)* 29:741–765.

Clark, Alexander. 2010. Learning context free grammars with the syntactic concept lattice. In *Grammatical inference: Theoretical results and applications*, ed. JoséM. Sempere and Pedro García, volume 6339 of *Lecture Notes in Computer Science*, 38–51. Springer Berlin Heidelberg. URL `http://dx.doi.org/10.1007/978-3-642-15488-1_5`.

Clark, Alexander, Rémi Eyraud, and Amaury Habrard. 2010. Using contextual representations to efficiently learn context-free languages. *The Journal of Machine Learning Research* 11:2707–2744.

Gold, E Mark. 1967. Language identification in the limit. *Information and Control* 10:447 – 474. URL `http://www.sciencedirect.com/science/article/pii/S0019995867911655`.

Opitz, Bertram, and Angela Friederici. 2007. Neural basis of processing sequential and hierarchical syntactic structures. *Human Brain Mapping* 585–592.

Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal learning and Verbal Behaviour* 6:855–863.

Artificial language learning

Question: Does ALL tell us anything at all about real language learning?
Hopeful answers:

1. Yes, because learning is largely unconscious/implicit (eg Reber (1967))

Artificial language learning

Question: Does ALL tell us anything at all about real language learning?
Hopeful answers:

1. Yes, because learning is largely unconscious/implicit (eg Reber (1967))
2. Yes, because of neuroimaging data (Opitz and Friederici, 2007)

Artificial language learning

Question: Does ALL tell us anything at all about real language learning?
Hopeful answers:

1. Yes, because learning is largely unconscious/implicit (eg Reber (1967))
2. Yes, because of neuroimaging data (Opitz and Friederici, 2007)
   - Participants were trained on human-language-like sentences

Artificial language learning

Question: Does ALL tell us anything at all about real language learning?
Hopeful answers:

1. Yes, because learning is largely unconscious/implicit (eg Reber (1967))
2. Yes, because of neuroimaging data (Opitz and Friederici, 2007)
   - Participants were trained on human-language-like sentences
   - In an MRI, participants listened to grammatical and ungrammatical sentences of the artificial language.

Artificial language learning

Question: Does ALL tell us anything at all about real language learning?
Hopeful answers:

1. Yes, because learning is largely unconscious/implicit (eg Reber (1967))
2. Yes, because of neuroimaging data (Opitz and Friederici, 2007)
   - Participants were trained on human-language-like sentences
   - In an MRI, participants listened to grammatical and ungrammatical sentences of the artificial language.
   - The higher their proficiency with the artificial language, the more the fMRI of their brains looked like they were processing their native language (Broca's area activation)

Experiment 1: Analysis

- I'm trying to get people to learn unconsciously, but maybe word repetition is just too salient

Experiment 1: Analysis

- I'm trying to get people to learn unconsciously, but maybe word repetition is just too salient
- In English, some adjuncts are indefinitely repeatable, acceptance drops the more there are.

Experiment 2: English survey

- Survey Monkey survey on English word repetition

Experiment 2: English survey

- Survey Monkey survey on English word repetition
- 30 sentences

Experiment 2: English survey

- Survey Monkey survey on English word repetition
- 30 sentences
- 5 choices:
  1. Doesn't sound like English
  2. Sounds pretty weird
  3. Sounds a bit weird
  4. I wouldn't say it, but it does sound like English
  5. Definitely English

Experiment 2: English survey

- 4: 2 Adjectives (*Can someone help me tear out this itchy itchy tag?*)

Experiment 2: English survey

- 4: 2 Adjectives (*Can someone help me tear out this itchy itchy tag?*)
- 2: 3 Adjectives (*What a stupid stupid stupid idea!*)

Experiment 2: English survey

- 4: 2 Adjectives (*Can someone help me tear out this itchy itchy tag?*)
- 2: 3 Adjectives (*What a stupid stupid stupid idea!*)
- 3: 4 Adjectives (*I haven't seen her in a long long long long time*)

Experiment 2: English survey

- 4: 2 Adjectives (*Can someone help me tear out this itchy itchy tag?*)
- 2: 3 Adjectives (*What a stupid stupid stupid idea!*)
- 3: 4 Adjectives (*I haven't seen her in a long long long long time*)
- 2: 5 Adjectives (*The big big big big big elephant stomped.*)

Experiment 2: English survey

- 4: 2 Adjectives (*Can someone help me tear out this itchy itchy tag?*)
- 2: 3 Adjectives (*What a stupid stupid stupid idea!*)
- 3: 4 Adjectives (*I haven't seen her in a long long long long time*)
- 2: 5 Adjectives (*The big big big big big elephant stomped.*)
- 1-6 *really*s (*I really* like her*)

Experiment 2: English survey

- 4: 2 Adjectives (*Can someone help me tear out this itchy itchy tag?*)
- 2: 3 Adjectives (*What a stupid stupid stupid idea!*)
- 3: 4 Adjectives (*I haven't seen her in a long long long long time*)
- 2: 5 Adjectives (*The big big big big big elephant stomped.*)
- 1-6 *reallys* (*I really* like her*)
- 1: 4 *sos* (*Marie ate so so so so much food!*)

Experiment 2: English survey

- 4: 2 Adjectives (*Can someone help me tear out this itchy itchy tag?*)
- 2: 3 Adjectives (*What a stupid stupid stupid idea!*)
- 3: 4 Adjectives (*I haven't seen her in a long long long long time*)
- 2: 5 Adjectives (*The big big big big big elephant stomped.*)
- 1-6 *reallys* (*I really* like her*)
- 1: 4 *sos* (*Marie ate so so so so much food!*)
- 1: 6 *evers* (*I'll never ever ever ever ever ever ever leave you*)

Experiment 2: English survey

- 4: 2 Adjectives (*Can someone help me tear out this itchy itchy tag?*)
- 2: 3 Adjectives (*What a stupid stupid stupid idea!*)
- 3: 4 Adjectives (*I haven't seen her in a long long long long time*)
- 2: 5 Adjectives (*The big big big big big elephant stomped.*)
- 1-6 *reallys* (*I really* like her*)
- 1: 4 *sos* (*Marie ate so so so so much food!*)
- 1: 6 *evers* (*I'll never ever ever ever ever ever ever leave you*)
- 11 fillers

Experiment 2: English survey results

People generally recognise lots of repetition as English, but the more repetition, the lower rating they give the sentence. (Lots of "4": *I wouldn't say it, but it does sound like English*)

Next

- Category repetition (*the big mean nasty bully*)

Next

- Category repetition (*the big mean nasty bully*)
- Embed repetition in a more complex grammar

$CFG_{F,K}$ learner

- Uses an *oracle*: the learner can ask *is this sentence ok?*

$CFG_{F,K}$ learner

- Uses an *oracle*: the learner can ask *is this sentence ok?*
- Clark et al conjecture that the oracle could be replaced by a probability distribution on the input

$CFG_{F,K}$ learner

- Uses an *oracle*: the learner can ask *is this sentence ok?*
- Clark et al conjecture that the oracle could be replaced by a probability distribution on the input
- Because of the oracle, the learner need only encounter the repeated string in the context in which it repeats to deduce repetition and optionality

$CFG_{F,K}$

- Get all the contexts in the sample (=**F**)

$CFG_{F,K}$

- Get all the contexts in the sample (=**F**)
- Get all the substrings in the sample (=**K**)

$CFG_{F,K}$

- Get all the contexts in the sample (=**F**)
- Get all the substrings in the sample (=**K**)
- For each substring s, stick s into each context in F

$CFG_{F,K}$

- Get all the contexts in the sample (=**F**)
- Get all the substrings in the sample (=**K**)
- For each substring s, stick s into each context in F
- Ask the oracle if you just built a grammatical sentence

$CFG_{F,K}$

- Get all the contexts in the sample (=**F**)
- Get all the substrings in the sample (=**K**)
- For each substring s, stick s into each context in F
- Ask the oracle if you just built a grammatical sentence
- if YES: make a rule:

$CFG_{F,K}$

- Get all the contexts in the sample (=**F**)
- Get all the substrings in the sample (=**K**)
- For each substring s, stick s into each context in F
- Ask the oracle if you just built a grammatical sentence
- if YES: make a rule:
  - Using the oracle, make a list of all the contexts from F in which s can appear. That's your LHS

$CFG_{F,K}$

- Get all the contexts in the sample (=**F**)
- Get all the substrings in the sample (=**K**)
- For each substring s, stick s into each context in F
- Ask the oracle if you just built a grammatical sentence
- if YES: make a rule:
  - Using the oracle, make a list of all the contexts from F in which s can appear. That's your LHS
  - if s is just one word, that's your RHS

$CFG_{F,K}$

- Get all the contexts in the sample (=**F**)
- Get all the substrings in the sample (=**K**)
- For each substring s, stick s into each context in F
- Ask the oracle if you just built a grammatical sentence
- if YES: make a rule:
  - Using the oracle, make a list of all the contexts from F in which s can appear. That's your LHS
  - if s is just one word, that's your RHS
  - if s is longer, for each way of splitting s into 2 substrings, s_1, s_2:

$CFG_{F,K}$

- Get all the contexts in the sample (=**F**)
- Get all the substrings in the sample (=**K**)
- For each substring s, stick s into each context in F
- Ask the oracle if you just built a grammatical sentence
- if YES: make a rule:
  - Using the oracle, make a list of all the contexts from F in which s can appear. That's your LHS
  - if s is just one word, that's your RHS
  - if s is longer, for each way of splitting s into 2 substrings, s_1, s_2:
  - Try sticking s_1 into all the contexts in F and asking the oracle if you made a grammatical sentence. The list of all the usable contexts is the left daughter

$CFG_{F,K}$

- Get all the contexts in the sample (=**F**)
- Get all the substrings in the sample (=**K**)
- For each substring s, stick s into each context in F
- Ask the oracle if you just built a grammatical sentence
- if YES: make a rule:
  - Using the oracle, make a list of all the contexts from F in which s can appear. That's your LHS
  - if s is just one word, that's your RHS
  - if s is longer, for each way of splitting s into 2 substrings, s_1, s_2:
  - Try sticking s_1 into all the contexts in F and asking the oracle if you made a grammatical sentence. The list of all the usable contexts is the left daughter
  - Try sticking s_2 into all the contexts in F and asking the oracle if you made a grammatical sentence. The list of all the usable contexts is the right daughter

$CFG_{F,K}$

- Input: abc

$CFG_{F,K}$

- Input: abc
- Contexts: $\{(\epsilon,\epsilon),(\epsilon,bc),(\epsilon,c),(a,c),(ab,\epsilon),(a,\epsilon)\}$

$CFG_{F,K}$

- Input: abc
- Contexts: $\{(\epsilon,\epsilon),(\epsilon,bc),(\epsilon,c),(a,c),(ab,\epsilon),(a,\epsilon)\}$
- Substrings: $\{abc,ab,bc,a,b,c\}$

$CFG_{F,K}$

- Input: abc
- Contexts: $\{(\epsilon,\epsilon),(\epsilon,bc),(\epsilon,c),(a,c),(ab,\epsilon),(a,\epsilon)\}$
- Substrings: $\{abc,ab,bc,a,b,c\}$
- Try abc in (ϵ,ϵ). Ask oracle $\rightarrow$ grammatical! (no others work)

$CFG_{F,K}$

- Input: abc
- Contexts: {(ϵ,ϵ),(ϵ,bc),(ϵ,c),(a,c),(ab,ϵ),(a,ϵ)}
- Substrings: {abc,ab,bc,a,b,c}
- Try abc in (ϵ,ϵ). Ask oracle $\rightarrow$ grammatical! (no others work)
  - Try a in (ϵ,ϵ) (no), (ϵ,bc) (YES), (ϵ,c) (YES), (a,c) (no), (ab,ϵ) (no), (a,ϵ) (no) (yes/no come from oracle)

$CFG_{F,K}$

- Input: abc
- Contexts: {(ϵ,ϵ),(ϵ,bc),(ϵ,c),(a,c),(ab,ϵ),(a,ϵ)}
- Substrings: {abc,ab,bc,a,b,c}
- Try abc in (ϵ,ϵ). Ask oracle $\rightarrow$ grammatical! (no others work)
  - Try a in (ϵ,ϵ) (no), (ϵ,bc) (YES), (ϵ,c) (YES), (a,c) (no), (ab,ϵ) (no), (a,ϵ) (no) (yes/no come from oracle)
  - Try bc in (ϵ,ϵ) (no), (ϵ,bc) (no),(ϵ,c) (no), (a,c) (no), (ab,ϵ) (YES), (a,ϵ) (YES)

$CFG_{F,K}$

- Input: abc
- Contexts: {(ϵ,ϵ),(ϵ,bc),(ϵ,c),(a,c),(ab,ϵ),(a,ϵ)}
- Substrings: {abc,ab,bc,a,b,c}
- Try abc in (ϵ,ϵ). Ask oracle $\rightarrow$ grammatical! (no others work)
  - Try a in (ϵ,ϵ) (no), (ϵ,bc) (YES), (ϵ,c) (YES), (a,c) (no), (ab,ϵ) (no), (a,ϵ) (no) (yes/no come from oracle)
  - Try bc in (ϵ,ϵ) (no), (ϵ,bc) (no),(ϵ,c) (no), (a,c) (no), (ab,ϵ) (YES), (a,ϵ) (YES)
  - New rule: {(ϵ,ϵ)}$\rightarrow$ {(ϵ,bc),(ϵ,c)} {(ab,ϵ),(a,ϵ)}

$CFG_{F,K}$

- Input: abc
- Contexts: {(ϵ,ϵ),(ϵ,bc),(ϵ,c),(a,c),(ab,ϵ),(a,ϵ)}
- Substrings: {abc,ab,bc,a,b,c}
- Try abc in (ϵ,ϵ). Ask oracle $\rightarrow$ grammatical! (no others work)
  - Try a in (ϵ,ϵ) (no), (ϵ,bc) (YES), (ϵ,c) (YES), (a,c) (no), (ab,ϵ) (no), (a,ϵ) (no) (yes/no come from oracle)
  - Try bc in (ϵ,ϵ) (no), (ϵ,bc) (no),(ϵ,c) (no), (a,c) (no), (ab,ϵ) (YES), (a,ϵ) (YES)
  - New rule: {(ϵ,ϵ)}$\rightarrow$ {(ϵ,bc),(ϵ,c)} {(ab,ϵ),(a,ϵ)}
  - ab and c get us the same rule

$CFG_{F,K}$

Hard-to-read Rules:

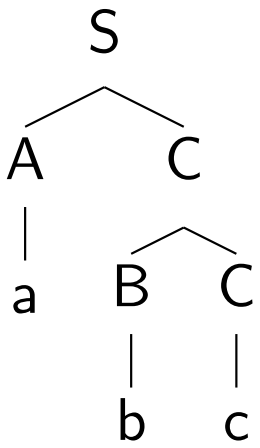
- $\{(\epsilon,c),(\epsilon,bc)\}\rightarrow a$
- $\{(a,c)\}\rightarrow b$
- $\{(a,\epsilon),(ab,\epsilon)\}\rightarrow c$
- $\{(\epsilon,c),(\epsilon,bc)\}\rightarrow \{(\epsilon,c),(\epsilon,bc)\}\ \{(a,c)\}$
- $\{(a,\epsilon),(ab,\epsilon)\}\rightarrow \{(a,c)\}\ \{(a,\epsilon),(ab,\epsilon)\}$
- $\{(\epsilon,\epsilon)\}\rightarrow \{(\epsilon,bc),(\epsilon,c)\}\ \{(ab,\epsilon),(a,\epsilon)\}$

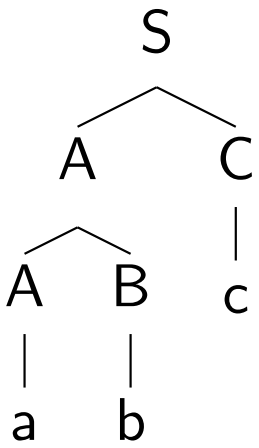
$CFG_{F,K}$

Equivalent Rules:

- A$\rightarrow$ a
- B$\rightarrow$ b
- C$\rightarrow$ c
- A$\rightarrow$ A B
- C$\rightarrow$ B C
- S$\rightarrow$ A C

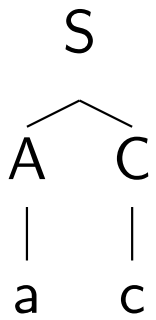
$CFG_{F,K}$





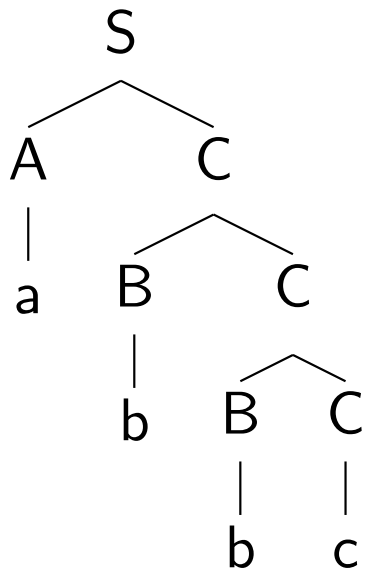
$CFG_{F,K}$

- A→ a
- B→ b
- C→ c
- A→ A B
- C→ B C
- S→ A C



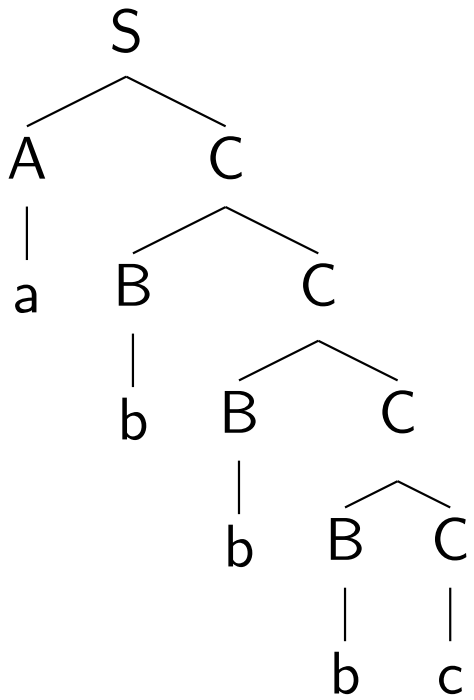
$CFG_{F,K}$

- A→ a
- B→ b
- C→ c
- A→ A B
- C→ B C
- S→ A C



$CFG_{F,K}$

- A→ a
- B→ b
- C→ c
- A→ A B
- C→ B C
- S→ A C



$CFG_{F,K}$

Already, just having heard abc, we guess our language is ab*c!

- bc is a substring of abc and (ab,ϵ) is a context of abc, so we got to ask the oracle "is abbc ok?"

$CFG_{F,K}$

Already, just having heard abc, we guess our language is ab*c!

- bc is a substring of abc and (ab,ϵ) is a context of abc, so we got to ask the oracle "is abbc ok?"
- In real life, this means that the baby must have heard abbc from time to time.

$CFG_{F,K}$

Already, just having heard abc, we guess our language is ab*c!

- bc is a substring of abc and (ab,ϵ) is a context of abc, so we got to ask the oracle "is abbc ok?"
- In real life, this means that the baby must have heard abbc from time to time.
- But, we never asked "is abbbbbbbc ok?" yet we still built a grammar that generates it.

$CFG_{F,K}$

Already, just having heard abc, we guess our language is ab*c!

- bc is a substring of abc and (ab,ϵ) is a context of abc, so we got to ask the oracle "is abbc ok?"
- In real life, this means that the baby must have heard abbc from time to time.
- But, we never asked "is abbbbbbbc ok?" yet we still built a grammar that generates it.
- Starting with just abc, the learner asks whether ac and abbc are okay.

$CFG_{F,K}$

Already, just having heard abc, we guess our language is ab*c!

- bc is a substring of abc and (ab,ϵ) is a context of abc, so we got to ask the oracle "is abbc ok?"
- In real life, this means that the baby must have heard abbc from time to time.
- But, we never asked "is abbbbbbbc ok?" yet we still built a grammar that generates it.
- Starting with just abc, the learner asks whether ac and abbc are okay.
- This means the baby only has to hear ac, abc, abbc to infer ab*c