Learning repetition and optionality

Meaghan Fowlie

UCLA Linguistics
Supervisor: Edward Stabler
mfowlie@ucla.edu
http://mfowlie.bol.ucla.edu

ISAIM 2014
Fort Lauderdale FL
January 7, 2014

Human language acquisition

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

Human language acquisition

Learning

Blah blah...

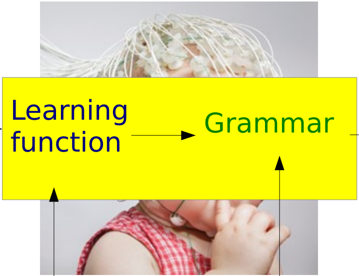
Blah blah...

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

Human language acquisition

Learning

Blah blah...



Blah blah...

Very little known; no good models of HL syntax learning

Project of much of linguistics

Overview

The phenomenon In human language, adjuncts are **optional** and often **repeatable**

The question How do formal learning models handle them?

Adjuncts

Generally adjectives, adverbs, prepositional phrases

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

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

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

Learners we'll look at today

1. 0-reversible learner (Angluin, 1982)
2. Substitutable CFGs (Clark, 2010)
3. PAC learner for PDFAs (Clark and Thollard, 2004)
4. (n-gram learner (García and Vidal, 1990))

Findings

1. 0-reversible learner (Angluin, 1982)
   - Optionality $\leftrightarrow$ Repetition
   - one repetition $\rightarrow$ indefinite repetition
   - need $ux^nv, ux^{n+1}v$
2. Substitutable CFGs (Clark, 2010)
   - repetition $\leftrightarrow$ optionality
   - one repetition $\rightarrow$indefinite repetition
   - need $ux^nv, ux^{n+1}v$
3. PAC learner for PDFAs (Clark and Thollard, 2004)
   - Representative sample $\rightarrow$learns repetition
   - Optionality $\not\leftrightarrow$ Repetition
4. (n-gram learner (García and Vidal, 1990))
   - Given up to n x's in a row in context (u, v), generalises to $ux^*v \subseteq L$
   - No generalisation from optionality to repetition or vice-versa

Learnability

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

Learnability

Definition (Language)

Given a finite set Σ, Σ^* is the set of all finite sequences of elements of Σ.
L is a language iff $L \subseteq \Sigma^*$

Definition (Learner)

A function from texts (samples from L) to grammars

Definition (Learn)

A learner learns a class of languages if it distinguishes them from each other

Learnability – a note

- A learner makes assumptions about the input strings and encodes them into its hypothesis grammar.
- So even if the input to a class X learner is not generated by a language of class X, the grammar that the learner hypothesises will be of class X.
- Thus the learner only correctly learns the language if it was from that class in the first place.
- This is what it means for a learner to learn a class of languages

Learnability

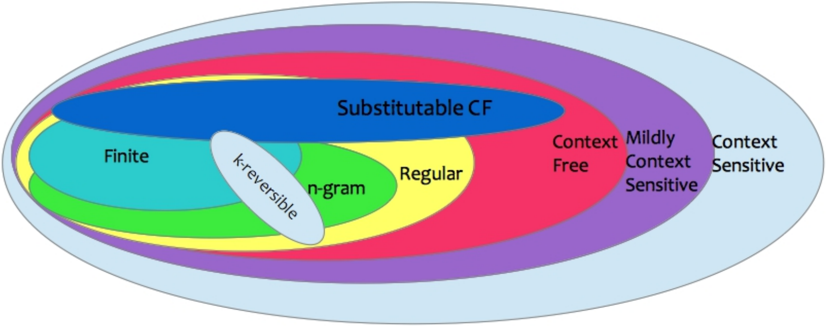
Definition (Gold learning)

(a.k.a. Identifiction in the limit from positive data) A language class is Gold-learnable if there is a function that will ultimately correctly converge on a grammar for every language in the class (Gold, 1967)

Definition (PAC learning)

$\forall 0 < \epsilon < 0.5, 0 < \delta < 0.5$ a Probably Approximately Correct learner outputs hypothesis grammars that are, with probability $1 - \delta$, ϵ-close to correct. (Valiant, 1984)

Chomsky hierarchy



Optionality and repetition

Definition (Optional)

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

Definition (Repeatable)

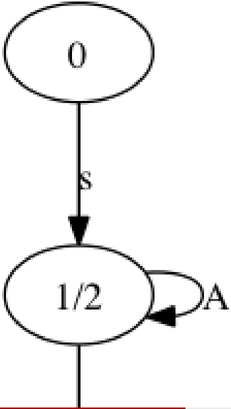
$x \in \Sigma^*$ is *repeatable in context u,v* iff $ux^+v \subseteq L$

0-reversible learner

Definition

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

If L is 0-reversible then for all strings u, v, if u and v share one suffix, they share all suffixes.



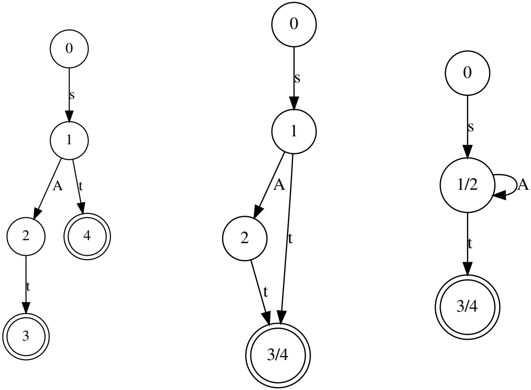
$L = sA^*t$
$s, sA, sAA, sAAA...$
all have suffixes
$t, At, AAt, AAAt...$

0-reversible learner

Optionality →Repetition

Sample: *st*, *sAt*

Prefix	Suffix
ϵ	st, sAt
s	**t**, At
st	ϵ
sA	**t**
sAt	ϵ



0-reversible learner

Theorem (Optionality → Repetition)

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

Proof.

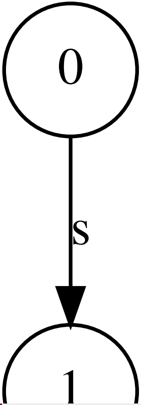
ux and u share suffix v.
u also has suffix xv
$\rightarrow ux$ also has suffix xv,
$\rightarrow uxxv \in L$.
uxx has suffix v as well
$\rightarrow uxx$ also must have suffix xv,
$\rightarrow uxxxv \in L$
etc. □

uv $\quad uxv$

$\quad uxv \quad \rightarrow uxxv$

0-reversible learner – repetition

Sample: *sAt*, *sAtt*



0-reversible learner: repetition $\rightarrow$optionality

Theorem (repetition $\rightarrow$optionality)

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

Proof.

$ux^kv, ux^{k+1} \in L$
u^{k-1}, ux^k share suffix xv
ux^k also has suffix v
$\rightarrow$ so does ux^{k-1}
$\rightarrow ux^{k-1}v \in L$
etc. □

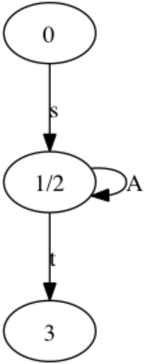
0-reversible learner - summary

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

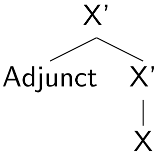
Interlude: States and repetition

- 0-reversible demonstrates something we want to capture: the notion that we should expect optionality and repetition of x in context C just in case *it doesn't matter whether* x *occurs in* C *or not*
- i.e. we're in the same state regardless of x's presence
- $\rightarrow$**We expect repetition and optionality to pattern together**

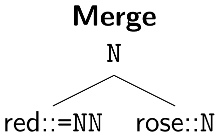
States and repetition



FSA



X-bar tree



Minimalist Grammar
Stabler (1997)
Kobele et al. (2007)

Substitutable Context Free learner

- CF equivalent of 0-reversible
- Learnable!

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 $ux^*v \subseteq L(G_i)$.

Proof.

By induction on the number of xs.
u, ux *share context* (ϵ, v).
u *also has context* (ϵ, xv) *so* ux *also must have this context.*
Therefore $uxxv \in L(G_i)$.
Suppose $ux^kv \in L$. *Then* ux^{k-1}, ux^k *share context* (ϵ, v).
ux^{k-1} *also has context* (ϵ, xv) *so* u^kx *also must have this context.*
Therefore $ux^{k+1}v \in L(G_i)$.

□

$uv \quad uxv$

$uxv \quad \rightarrow uxxv$

Substitutable Context Free learner

Theorem (Repetition → Optionality)

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

Proof.

By induction on the number of xs. ux^{n-1}, ux^n share context (ϵ, xv). ux^n also has context (ϵ, v) so ux^{n-1} also must have this context. Therefore $ux^{n-1}v \in L(G_i)$. Etc. □

uxv $\quad uxxv$

$uxv \quad \rightarrow uv$

Substitutable CF – summary

- repetition $\leftrightarrow$ optionality
- one repetition $\rightarrow$indefinite repetition
- need $ux^n v, ux^{n+1} v$

Human language

- HL $\subseteq$ Mildly Context Sensitive Languages (Joshi, 1985)
- Not all optional elements in HL are repeatable, e.g. most Adverbs in English
- But HL is not describable by strict substitution classes (à la Zellig Harris) anyway
- Humans need additional info, e.g. meaning

(4) a. The man in the yellow hat has a monkey
    b. He has a monkey

(5) a. The man laughed
    b. The cheerful man laughed

(6) a. The horse raced past the barn quickly.
    b. The horse raced past the barn fell.
    c. Sue fell.
    d. *Sue quickly

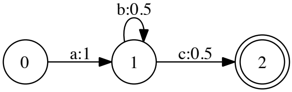
PDFA PAC learner

Clark and Thollard (2004)

- Learns Probabilistic Deterministic Finite State languages
- Requires huge sample sizes
- Similar to Angluin (1982)'s 0-reversible learner except states must share most of their suffixes (not just one) to be merged

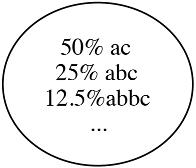
PDFA PAC learner

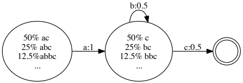
Source grammar



Sample will be approximately:
50% *ac*
25% *abc*
12.5% *abbc*
6.25% *abbbc*
etc.

PDFA PAC learner





PDFA PAC learner: summary

- Representative sample $\rightarrow$learns repetition
- Optionality $\not\leftrightarrow$ Repetition
- i.e. Learnable, but no shortcuts!

Artificial language learning

- **Experimental paradigm:** Expose people to a sample from a target language. See what they do with novel items from the language.
- **Experiment underway:**

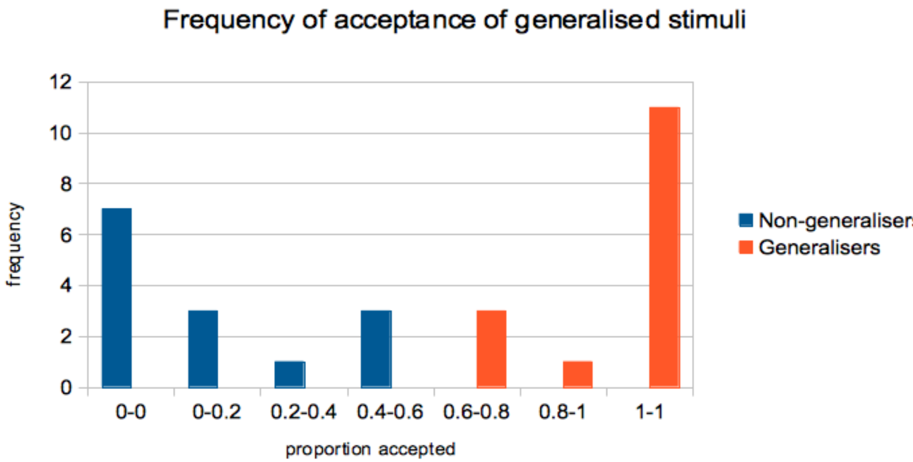
Pilot method On-line survey (Amazon's Mechanical Turk)

Exposure stimuli *ac abc abbc* (a,b,c word classes)

Testing stimuli *ac abc abbc abbbc abbbbc*

Pilot results participants divide into two groups: generalisers (accept *abbbc, abbbbc*) and non-generalisers (reject *abbbc, abbbbc*)

Artificial language learning



Future research

- k, l-substitutable CFLs Yoshinaka (2008)
- CFLs with finite context and finite kernal properties Clark et al. (2008)
- k, l-substitutable Muliple Context Free Languages Yoshinaka (2009)
- What else? Accepting suggestions!
- Experiment

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. 2008. A polynomial algorithm for the inference of context free languages. In *Grammatical inference: Algorithms and applications*, 29–42. Springer.

Clark, Alexander, and Franck Thollard. 2004. Pac-learnability of probabilistic deterministic finite state automata. *The Journal of Machine Learning Research* 5:473–497.

García, P., and E. Vidal. 1990. Inference of k-testable languages in the strict sense and application to syntactic pattern recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12:920–925.

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

Joshi, Aravind. 1985. How much context-sensitivity is necessary for characterizing structural descriptions. In *Natural language processing: Theoretical, computational and psychological perspectives*, ed. D. Dowty, L. Karttunen, and A. Zwicky, 206–250. New York: Cambridge University Press.

Kobele, Gregory M., Christian Retoré, and Sylvain Salvati. 2007. An automata-theoretic approach to minimalism. In *Model Theoretic Syntax at ESSLLI '07*, ed. J. Rogers and S. Kepser. ESSLLI.

Stabler, Edward. 1997. Derivational minimalism. *Logical Aspects of Computational Linguistics* 68–95.

Valiant, L. G. 1984. A theory of the learnable. *Commun. ACM* 27:1134–1142. URL http://doi.acm.org/10.1145/1968.1972.

Yoshinaka, Ryo. 2008. Identification in the limit of k, l-substitutable context-free languages. In *Grammatical inference: Algorithms and applications*, 266–279. Springer.

Yoshinaka, Ryo. 2009. Learning mildly context-sensitive languages with multidimensional substitutability from positive data. In *Algorithmic Learning Theory*, 278–292. Springer.

n-gram learner

Just makes of list of the n-grams it has seen. String is grammatical if it contains no bigrams not in the list and has boundary markers.

Example (Bigram learner)

$\Sigma = \{a, b\}$, boundary markers $\rtimes$, $\ltimes$

Sample: *a, ab, abb*

Bigrams: $\rtimes a$, *ab*, *bb*, $b\ltimes$, $a\ltimes$

Hypothesis language: ab^*

n-gram learner

Theorem

Let $u, v \in \Sigma^*$, $x \in \Sigma$. Let sample of L contain $uv, uxv, uxxv, ..., ux^{n-1}v, ux^nv$. Then an n-gram learner will conclude that $ux^*v \subseteq L$.

Example (Trigram learner)

aba<u>bbb</u>abab

[⋊ab] *abbbabab*	*abab* [bba] *bab*
[aba] *bbbabab*	*ababb* [bab] *ab*
a [bab] *bbabab*	*ababbb* [aba] *b*
ab [abb] *babab*	*ababbba* [bab]
aba [bbb] *abab*	*ababbbab* [ab⋉]

Trigrams: ⋊*ab*, *aba*, *bab*, *abb*, *bbb*, *bba*, *ab*⋉

n-gram learner

Hypothesis: $abab^* abab \subseteq L$

aba<u>*bbbb*</u>*abab*

trigrams: $\rtimes ab$, *aba*, *bab*, *abb*, *bbb*, *bba*, $ab\ltimes$

$\rtimes$ab *abbbbabab*
aba *bbbbabab*
a bab *bbbabab*
ab abb *bbabab*
aba bbb *babab*
abab bbb *abab*

ababb bba *bab*
ababbb bab *ab*
ababbbb aba *b*
ababbbba bab
ababbbbab ab$\ltimes$

n-gram learner

If the repeated element is repeated fewer than *n* times, repetition isn't generalised.

Trigram learner **Sample:** *aba<u>bb</u>abab*

⋊ab *abbabab*	*aba* bba *bab*
aba *bbabab*	*abab* bab *ab*
a bab *babab*	*ababb* aba *b*
ab abb *abab*	*ababba* bab
aba bbb *abab*	*ababbab* ab⋉

Trigrams: ⋊*ab*, *aba*, *bab*, *abb*, ~~*bbb*~~, *bba*, *ab*⋉

aba* **bbb **bb**...*abab*

n-gram learner summary

- Given up to n x's in a row in context (u, v), generalises to $ux^*v \subseteq L$
- No generlisation from optionality to repetition or vice-versa