Distributional Learning of Syntax

Alexander Clark

Department of Philosophy
King's College, London
alexsclark@gmail.com

5 November 2013
Manchester

Outline

- Methodology
- Substitutability
  - Semantics
  - Strong learning
- General theory
  - Queries and Features
  - Beyond Context-Free Grammars
- Conclusion
  - Probabilistic learning
  - Predictions

Outline

- Methodology
- Substitutability
  - Semantics
  - Strong learning
- General theory
  - Queries and Features
  - Beyond Context-Free Grammars
- Conclusion
  - Probabilistic learning
  - Predictions

Linguistics

Chomsky (1973)

The fundamental empirical problem of linguistics is to explain how a person can acquire knowledge of language

Linguistics

Chomsky (1973)

The fundamental empirical problem of linguistics is to explain how a person can acquire knowledge of language

nhornste (2013)

For a very very long time now anyone interested in Chomsky's work has understood that in his view the primary object of study in linguistics is not languages (e.g. English, French, Swahili, Piraha, etc) but the mental powers that make it possible for humans to acquire a language.

Linguistics

Chomsky (1973)

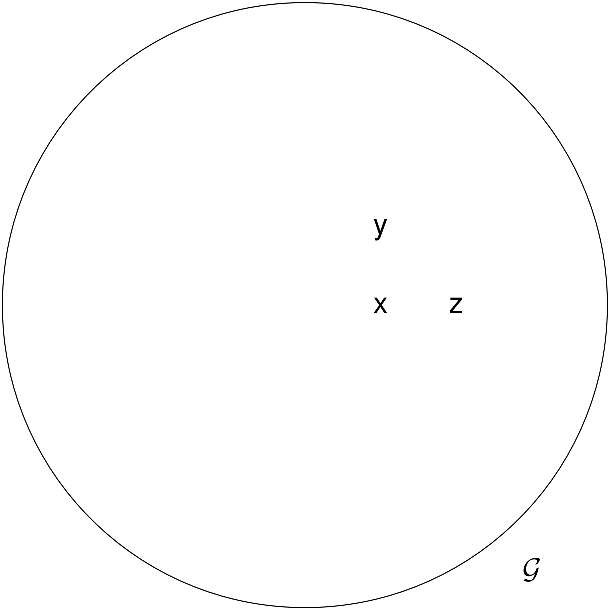
The fundamental empirical problem of linguistics is to explain how a person can acquire knowledge of language

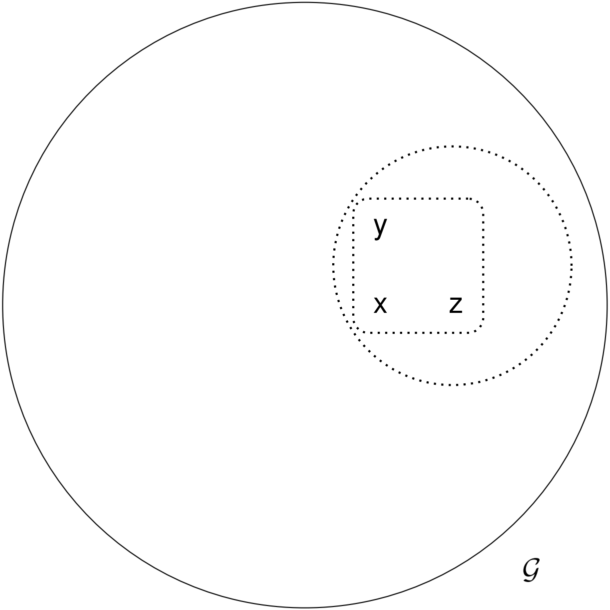
nhornste (2013)

For a very very long time now anyone interested in Chomsky's work has understood that in his view the primary object of study in linguistics is not languages (e.g. English, French, Swahili, Piraha, etc) but the mental powers that make it possible for humans to acquire a language. **Of course, a very good window into the latter is via investigation of the properties of particular languages and so linguists earn their bread and butter by detailed study of the properties of particular languages.**

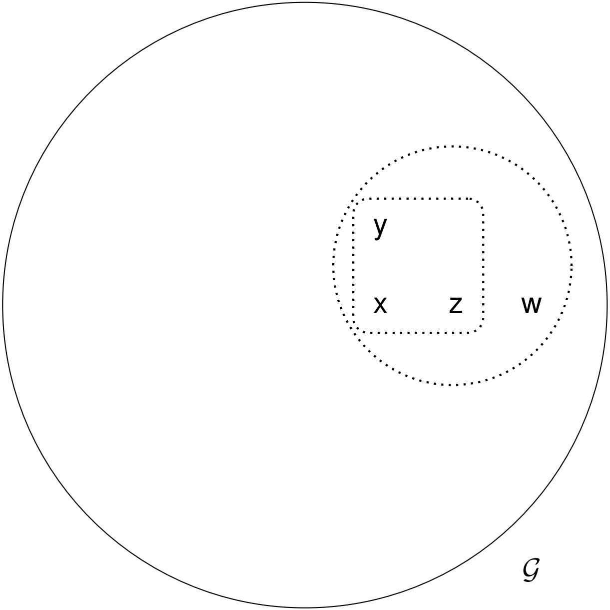
Standard view

Input $\longrightarrow$ $\boxed{\text{LAD}}$ $\longrightarrow$ I-language


y
x
z
$\mathcal{G}$



KING'S College LONDON Founded 1829

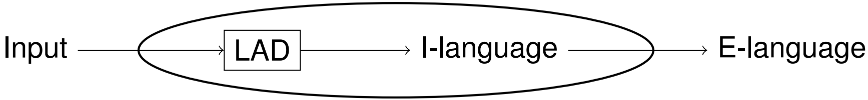

y
x
z
w
$\mathcal{G}$

Undetermination

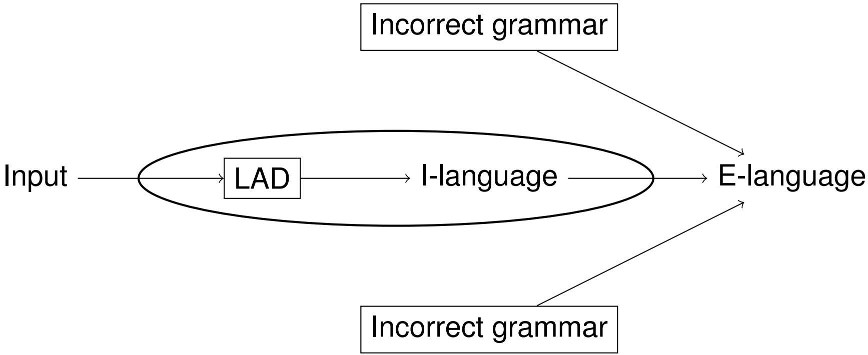
Chomsky (1966)

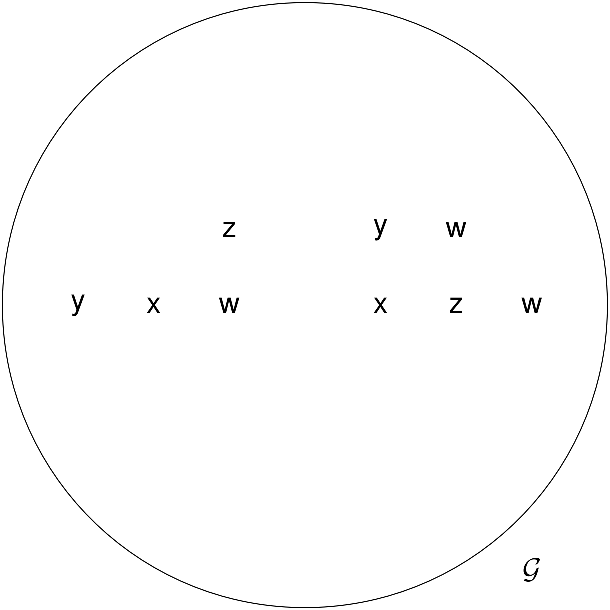
> *Choice of a descriptively adequate grammar for the language L is always much underdetermined (for the linguist, that is) by data from L.*

Standard view

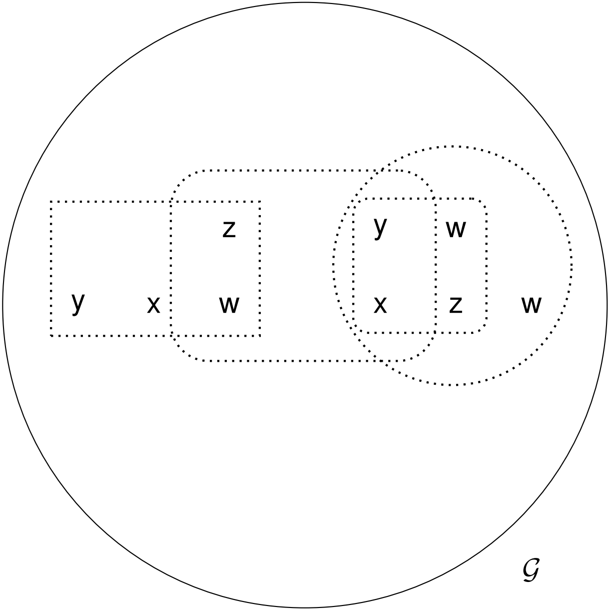


Correct and incorrect theories



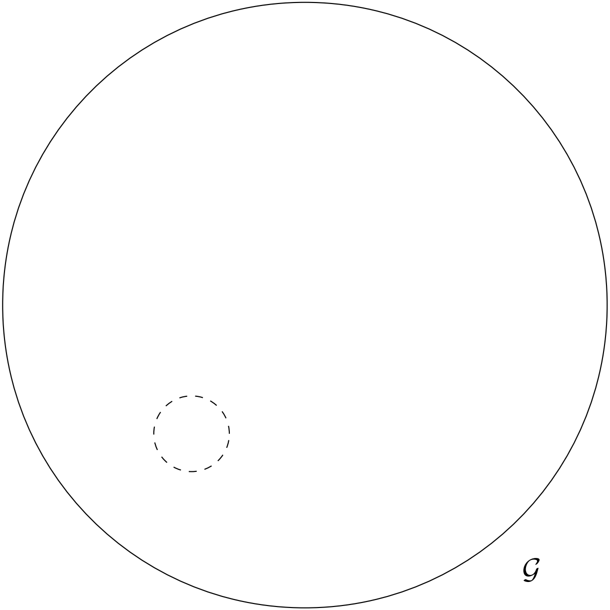


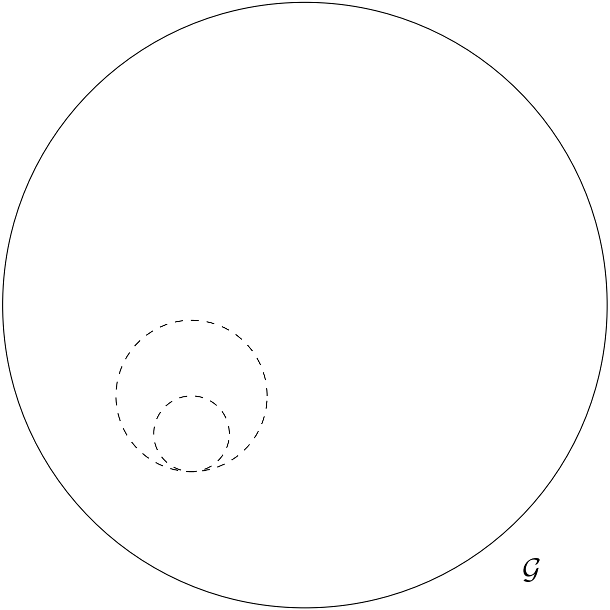
KING'S College LONDON Founded 1829


z
y
x
w
y
w
x
z
w
$\mathcal{G}$

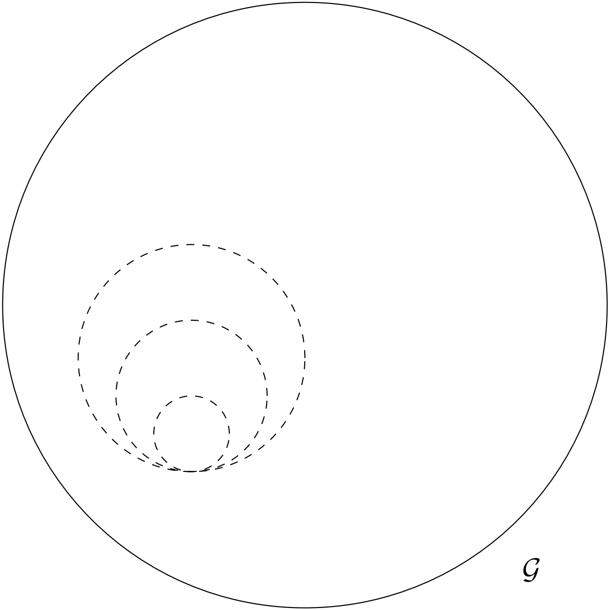
Chomsky, Hauser and Fitch (2005)

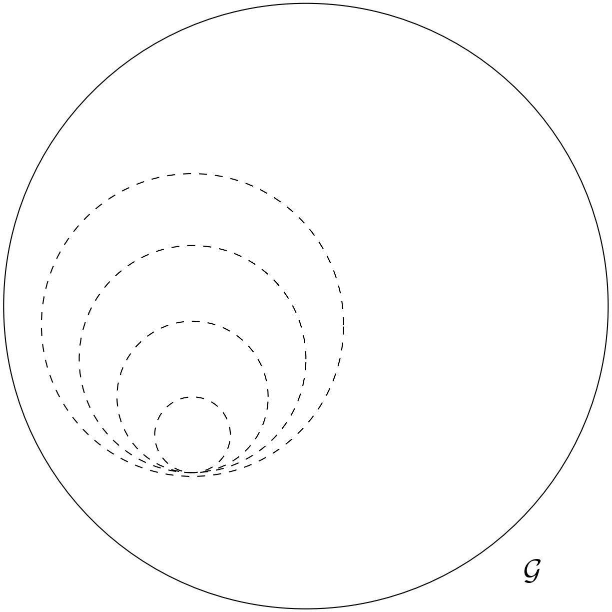
Two basic conditions that UG must satisfy are that it (1) accommodate the attainable I-languages, and (2) account for their acquisition.

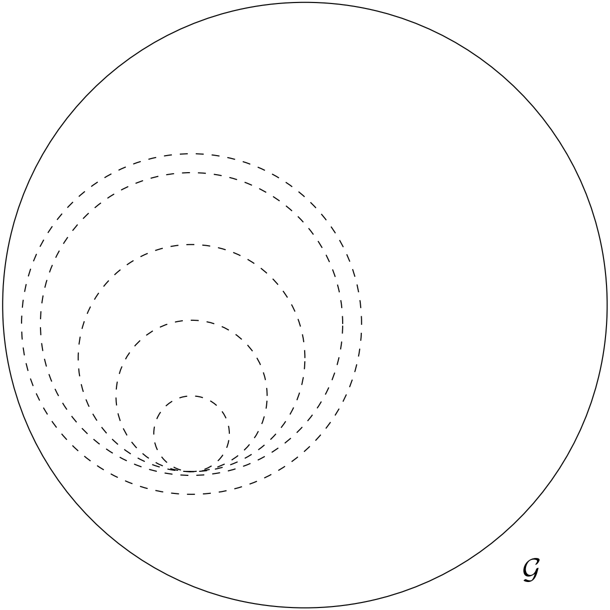

$\mathcal{G}$

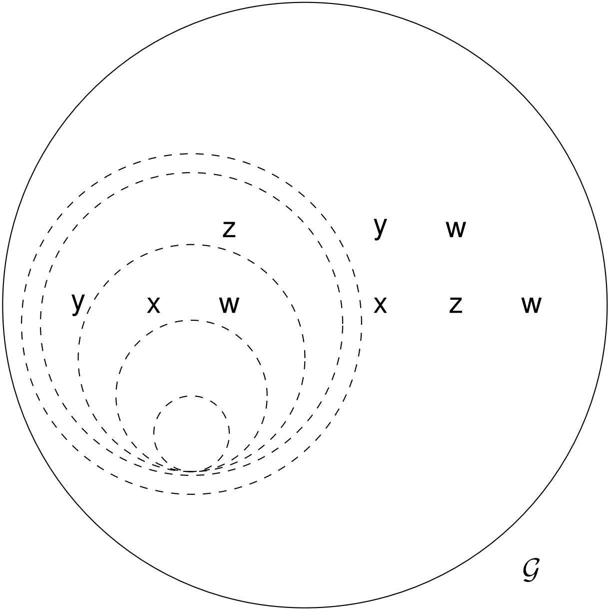


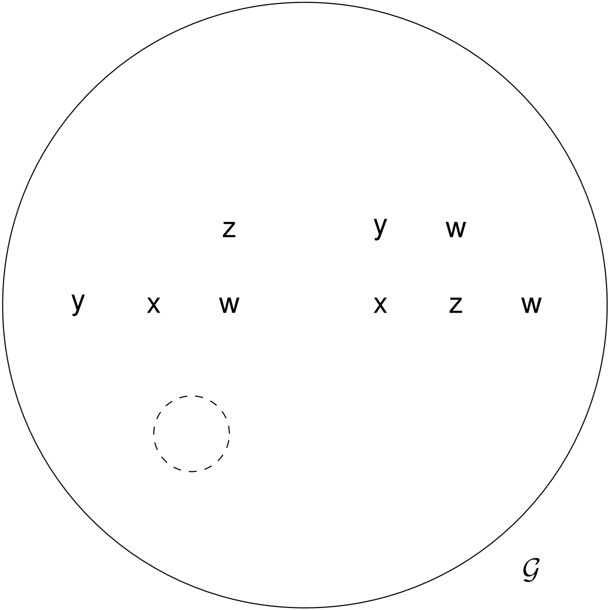
$\mathcal{G}$


$\mathcal{G}$

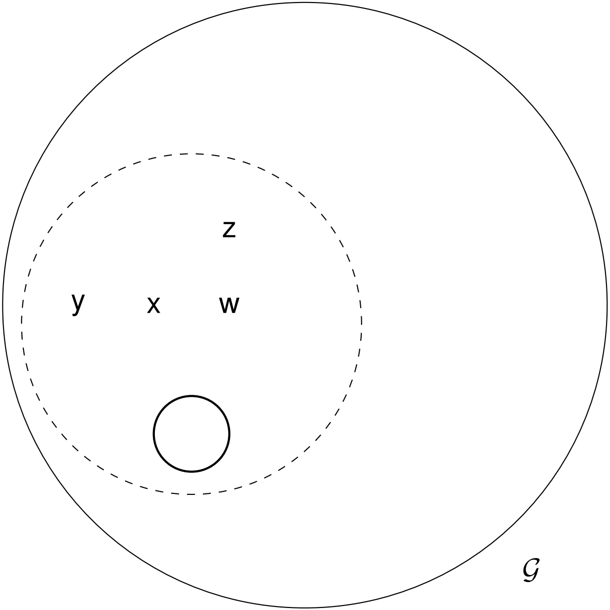

$\mathcal{G}$


$\mathcal{G}$


z
y
w
y
x
w
x
z
w
$\mathcal{G}$


z
y
w
y
x
w
x
z
w
$\mathcal{G}$

Outline

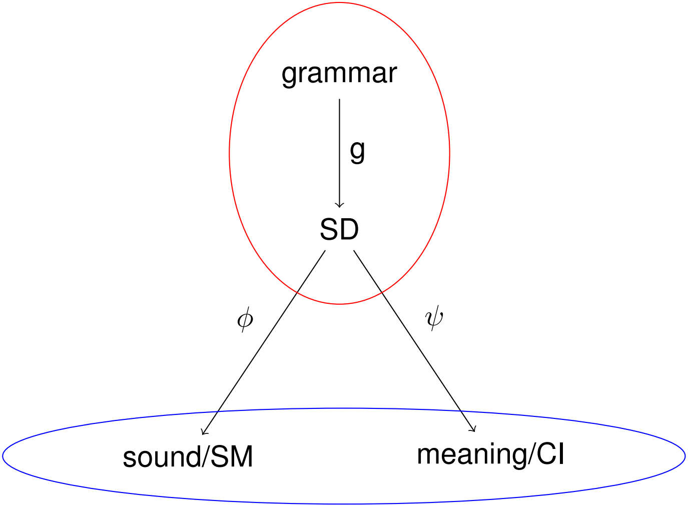


Learning approaches

1. A simple and complete answer but for a very small class of languages. [Clark(2013b)]
   - Unnecessarily strict learning model
   - Class of languages too small
2. A complex and incomplete model for a very large class of languages.[Clark and Yoshinaka(2013)]
   - Learning model too easy
   - Class of languages is perhaps unnecessarily large

Outline

Architecture



Running example

Propositional logic

Alphabet

propositional symbols	$A_1, A_2, \ldots$
binary connectives	$\wedge, \vee, \rightarrow, \leftrightarrow$
negation symbol	$\neg$
brackets	$($, $)$

- A_1
- $(A_1 \rightarrow A_3)$
- $(A_1 \rightarrow (\neg A_4))$

Running example

Propositional logic

Alphabet

propositional symbols	rain, snow, hot, cold, danger
binary connectives	and, or, implies, iff
negation symbol	not
brackets	open, close

- rain
- open snow implies cold close
- open snow implies open not hot close close

Classic model

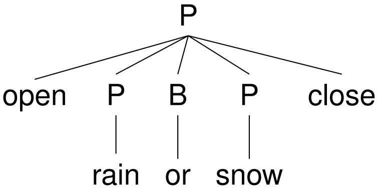
open rain or snow close

Classic model

open rain or snow close

Q	R	S	
T	T	T	T
T	T	F	T
T	F	T	T
...	...	...	
F	F	F	F

Classic model



open rain or snow close

Q	R	S	
T	T	T	T
T	T	F	T
T	F	T	T
...	...	...	
F	F	F	F

Simple example

“You must study the simplest system you think has the properties you are interested in”[Platt(1964)]

- semantically interpreted language
- Infinite non-regular language
- Hierarchically structured expressions
- Missing: ambiguity, vagueness, reference, illocutionary force…

Models of learning

Three classes of objects

strings, meanings, and trees.

Accessibility for the child:

- Strings – complete
- Meanings – partial
- Trees – no information at all

Weak learning

Learning model

	Inputs	Outputs
Weak	strings	strings

Goal

Learner must acquire a grammar that defines an infinite set of strings on the basis of a finite set of strings.

Distributional Learning

[Harris(1964)]

- Look at the dog
- Look at the cat

Distributional Learning

[Harris(1964)]

- Look at the dog
- Look at the cat
- That cat is crazy

Distributional Learning

[Harris(1964)]

- Look at the dog
- Look at the cat
- That cat is crazy
- That dog is crazy

English counterexample

- I can swim
- I may swim
- I want a can of beer

English counterexample

- I can swim
- I may swim
- I want a can of beer
- *I want a may of beer

English counterexample

- She is Italian
- She is a philosopher
- She is an Italian philosopher

English counterexample

- She is Italian
- She is a philosopher
- She is an Italian philosopher
- *She is an a philosopher philosopher

Logic example

Propositional logic is *substitutable*:

- open rain and cold close
- open rain implies cold close

Logic example

Propositional logic is *substitutable*:

- open rain and cold close
- open rain implies cold close

and is similar to implies

Logic example

Propositional logic is *substitutable*:

- open rain and cold close
- open rain implies cold close

and is similar to implies

- open snow implies open not hot close

Logic example

Propositional logic is *substitutable*:

- open rain and cold close
- open rain implies cold close

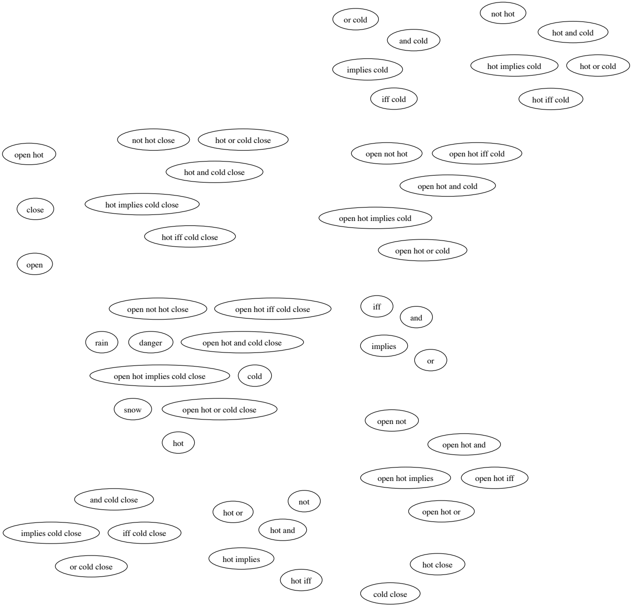
and is similar to implies

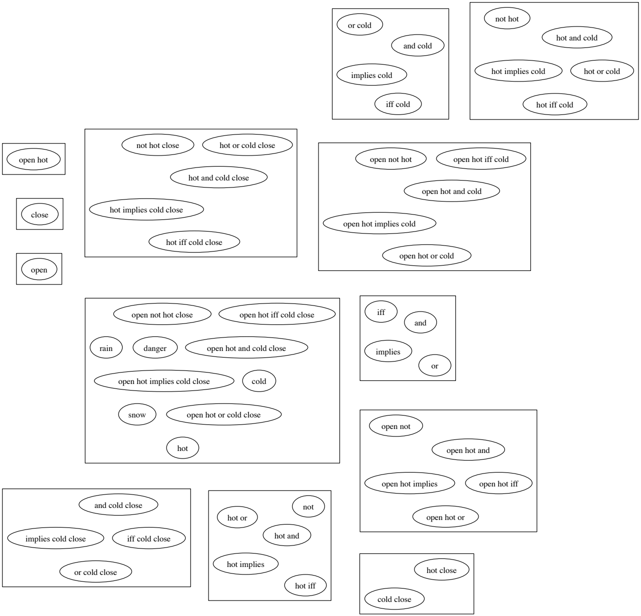
- open snow implies open not hot close
- open snow and open not hot close

Example

- hot
- cold
- open hot or cold close
- open not hot close
- open hot and cold close
- open hot implies cold close
- open hot iff cold close
- danger
- rain
- snow

Nonterminal for each substring



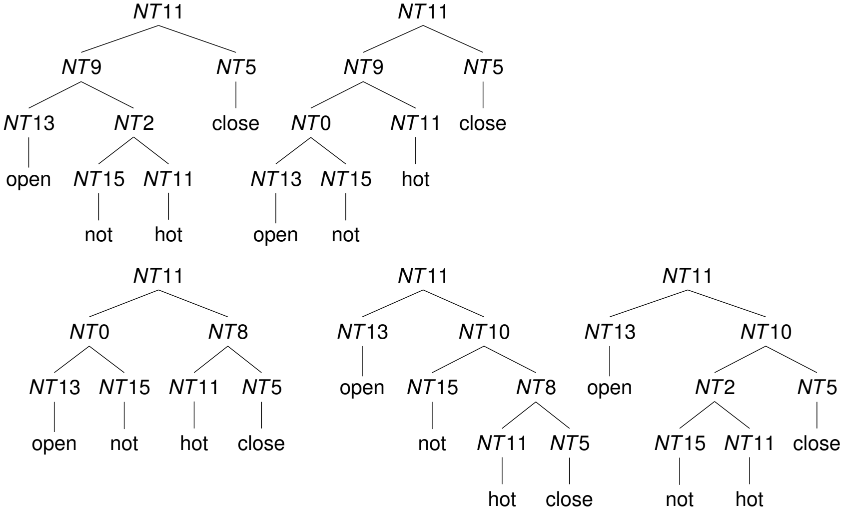

or cold
and cold
implies cold
iff cold
not hot
hot and cold
hot implies cold
hot or cold
hot iff cold
open hot
not hot close
hot or cold close
hot and cold close
hot implies cold close
hot iff cold close
close
open
open not hot
open hot iff cold
open hot and cold
open hot implies cold
open hot or cold
open not hot close
open hot iff cold close
rain
danger
open hot and cold close
open hot implies cold close
cold
snow
open hot or cold close
hot
iff
and
implies
or
open not
open hot and
open hot implies
open hot iff
open hot or
and cold close
implies cold close
iff cold close
or cold close
hot or
not
hot and
hot implies
hot iff
hot close
cold close

Weak learner

[Clark and Eyraud(2007)]

- If the language is a substitutable context-free language, then the hypothesis grammar will converge to a correct grammar.
- But the grammar may be different for each input data set.
- Parse trees will include every binary tree for strings in the training data.

open not hot close



Berwick, Chomsky critique

[Berwick et al.(2011)Berwick, Pietroski, Yankama, and Chomsky]

> *Put another way, language acquisition is not merely a matter of acquiring a capacity to associate word strings with interpretations. Much less is it a mere process of acquiring a (weak generative) capacity to produce just the valid word strings of a language. Idealizing, one can say that each child acquires a procedure that generates boundlessly many meaningful expressions, and that a single string of words can correspond to more than one expression.*

A second model

Two mathematically reasonable models

	Inputs	Outputs
Weak	strings	strings
Weak semantic	strings + meanings	strings + meanings

[Yoshinaka and Kanazawa(2011)]

Standard view: Semantic Bootstrapping

[Pinker(1995)]

> *Many models of language acquisition assume that the input to the child consists of a sentence and a representation of the meaning of that sentence, inferred from context and from the child's knowledge of the meanings of the words (e.g. Anderson, 1977; Berwick, 1986; Pinker, 1982, 1984; Wexler & Culicover, 1980). Of course, this can't literally be true – children don't hear every word of every sentence, and surely don't, to begin with, perceive the entire meaning of a sentence from context.*

Arguments against this view

1. Impossible for children actually to do this.
2. Assumes precisely the ability which needs to be explained.
3. Language would be unnecessary.
4. Language dependence of the semantic representation.
5. Language acquisition starts very early.
6. Blind children acquire language with only minor delay.

Why is this view held?

[Steedman(1996)]

> *As soon as it is recognized that the very earliest stages in acquiring syntax require some language-independent source of information about grammatical categories and grammatical relations, the only plausible source that has ever been identified is the semantic interpretation that underlies the utterance. . . . sooner or later, the child needs access to semantic interpretations in order to acquire syntactic competence.*

Alternative model

Two mathematically reasonable models

	Inputs	Outputs
Weak	strings	strings
Weak semantic	strings + meanings	strings + meanings

Alternative model

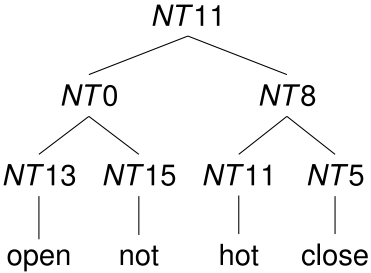
Two mathematically reasonable models

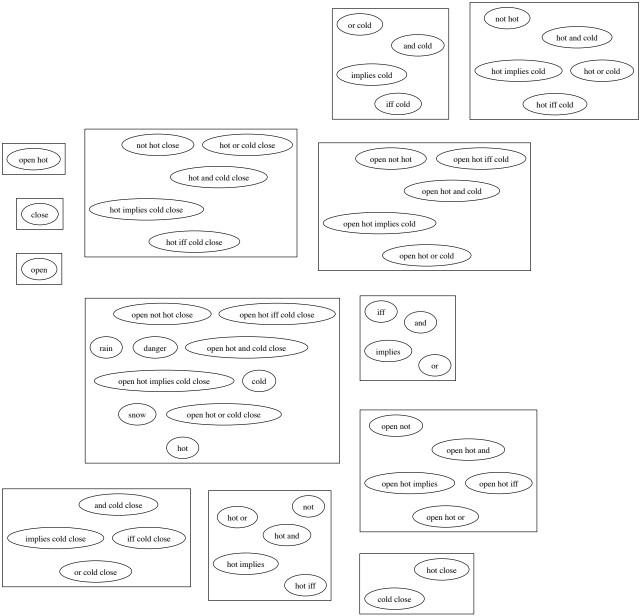
	Inputs	Outputs
Weak	strings	strings
Weak semantic	strings + meanings	strings + meanings

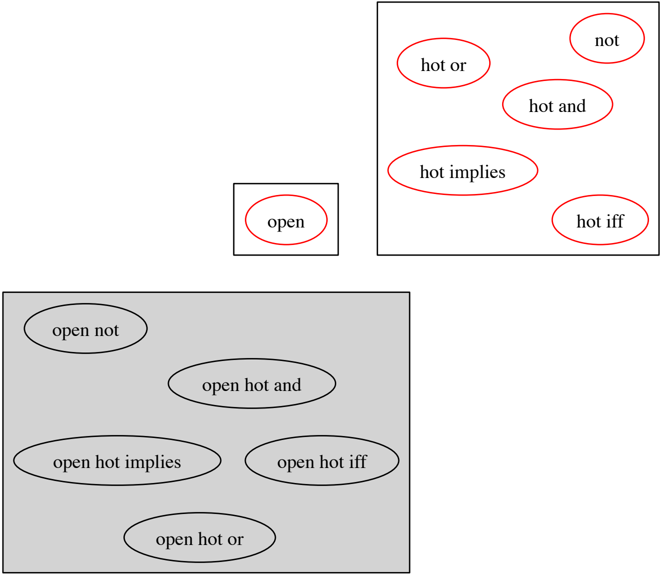
A mathematically unreasonable model

[Wexler and Culicover(1980)]

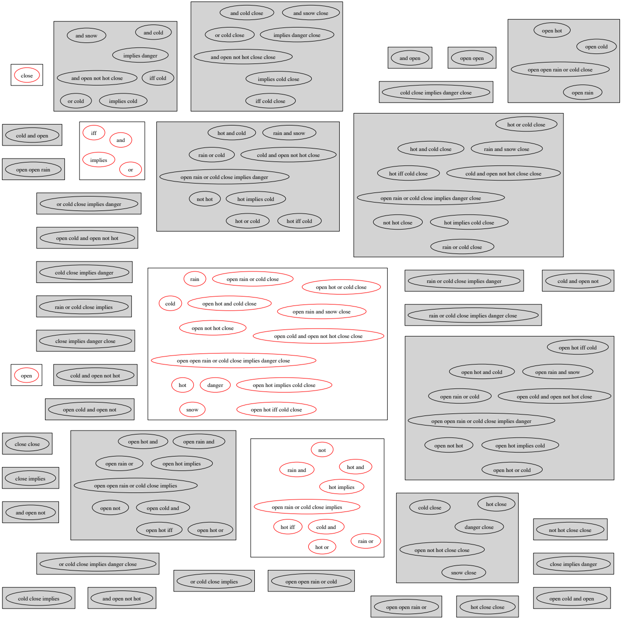
	Inputs	Outputs
Strong learning	strings	strings + trees


NT11
NT0
NT8
NT13
NT15
NT11
NT5
open
not
hot
close


or cold
and cold
implies cold
iff cold
not hot
hot and cold
hot implies cold
hot or cold
hot iff cold
open hot
close
open
not hot close
hot or cold close
hot and cold close
hot implies cold close
hot iff cold close
open not hot
open hot iff cold
open hot and cold
open hot implies cold
open hot or cold
open not hot close
open hot iff cold close
rain
danger
open hot and cold close
open hot implies cold close
cold
snow
open hot or cold close
hot
iff
and
implies
or
open not
open hot and
open hot implies
open hot iff
open hot or
and cold close
implies cold close
iff cold close
or cold close
hot or
not
hot and
hot implies
hot iff
hot close
cold close


hot or
not
hot and
hot implies
hot iff
open
open not
open hot and
open hot implies
open hot iff
open hot or


or cold
and cold
implies cold
iff cold
not hot
hot and cold
hot implies cold
hot or cold
hot iff cold
open hot
close
open
not hot close
hot or cold close
hot and cold close
hot implies cold close
hot iff cold close
open not hot
open hot iff cold
open hot and cold
open hot implies cold
open hot or cold
open not hot close
open hot iff cold close
rain
danger
open hot and cold close
open hot implies cold close
cold
snow
open hot or cold close
hot
iff
and
implies
or
open not
open hot and
open hot implies
open hot iff
open hot or
and cold close
implies cold close
iff cold close
or cold close
not
hot or
hot and
hot implies
hot iff
hot close
cold close



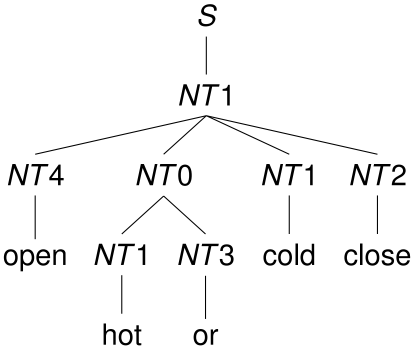
close
and snow
and cold
implies danger
and open not hot close
iff cold
or cold
implies cold
and cold close
and snow close
or cold close
implies danger close
and open not hot close close
implies cold close
iff cold close
and open
open open
cold close implies danger close
open hot
open cold
open open rain or cold close
open rain
cold and open
iff
and
implies
or
open open rain
hot and cold
rain and snow
rain or cold
cold and open not hot close
open rain or cold close implies danger
not hot
hot implies cold
hot or cold
hot iff cold
hot or cold close
hot and cold close
rain and snow close
hot iff cold close
cold and open not hot close close
open rain or cold close implies danger close
not hot close
hot implies cold close
rain or cold close
or cold close implies danger
open cold and open not hot
cold close implies danger
rain or cold close implies
close implies danger close
open
cold and open not hot
open cold and open not
rain
open rain or cold close
open hot or cold close
cold
open hot and cold close
open rain and snow close
open not hot close
open cold and open not hot close close
open open rain or cold close implies danger close
hot
danger
open hot implies cold close
snow
open hot iff cold close
rain or cold close implies danger
cold and open not
rain or cold close implies danger close
open hot iff cold
open hot and cold
open rain and snow
open rain or cold
open cold and open not hot close
open open rain or cold close implies danger
open not hot
open hot implies cold
open hot or cold
close close
close implies
and open not
open hot and
open rain and
open rain or
open hot implies
open open rain or cold close implies
open not
open cold and
open hot iff
open hot or
not
rain and
hot and
hot implies
open rain or cold close implies
hot iff
cold and
rain or
hot or
cold close
hot close
danger close
open not hot close close
snow close
not hot close close
close implies danger
or cold close implies danger close
or cold close implies
open open rain or cold
cold close implies
and open not hot
open open rain or
hot close close
open cold and open

Running example

(verbatim output from implementation)

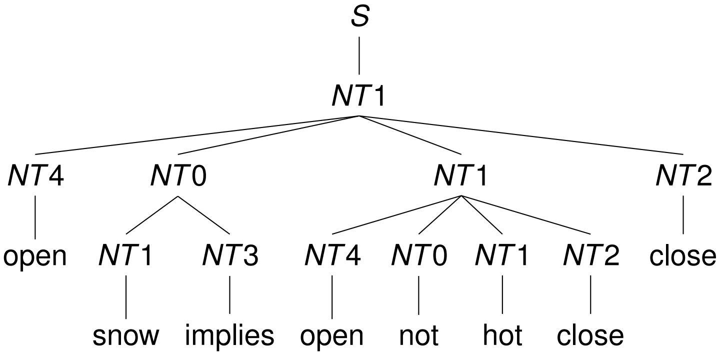
Running example

(verbatim output from implementation)



Running example

(verbatim output from implementation)



Result

Theorem: Clark, JMLR, 2013

This algorithm can learn all substituable languages with a finite number of primes:

- Rapidly, efficiently
- Strongly
- From positive data alone

Outline

Methodology

Substitutability

- Semantics
- Strong learning

General theory

- Queries and Features
- Beyond Context-Free Grammars

Conclusion

- Probabilistic learning
- Predictions

Problem decomposition

- A computationally efficient learner.

input	Allow Membership queries.
output	Only a weakly correct grammar.

Problem decomposition

- A computationally efficient learner.

  input Allow Membership queries.
  output Only a weakly correct grammar.
- Two additional components:

  input Converts the learner to use probabilities
  output Converts a weak learner to a strong learner.

Example

Is ’cat’ substitutable for ’dog’?

- The cat is over there.
- I want a dog for Christmas.
- I want a Siamese cat for Christmas.
- Put a cat-flap in the door to the kitchen.
- An Alsatian is a breed of dog.
- He continues to dog my footsteps.
- I would rather have a dog than a cat as a pet.

Example

Is 'cat' substitutable for 'dog'?

- The dog is over there.
- I want a cat for Christmas.
- I want a Siamese dog for Christmas.
- Put a dog-flap in the door to the kitchen.
- An Alsatian is a breed of cat.
- *He continues to cat my footsteps.
- I would rather have a dog than a dog as a pet.
- I would rather have a cat than a dog as a pet.

Shallow distributional features: privative/monovalued

- The _ is over there.
- I want a _ for Christmas.
- I want a Siamese _ for Christmas.
- Put a _-flap in the door to the kitchen.
- An Alsatian is a breed of _.
- He continues to _ my footsteps.
- I would rather have a _ than a dog as a pet.

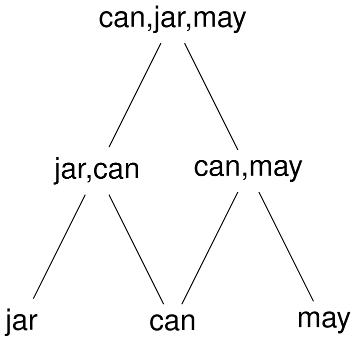
Ambiguity

Examples

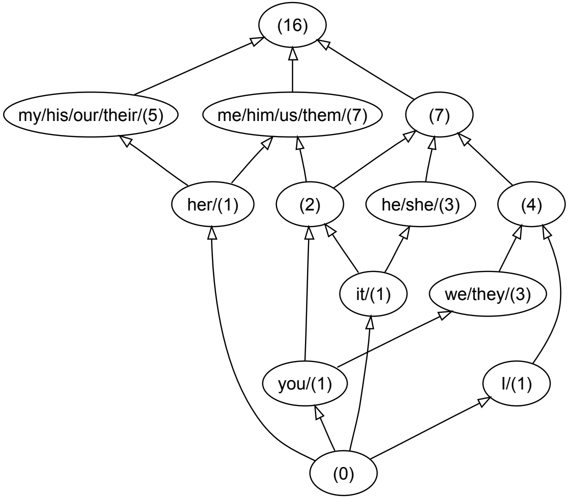
- Can I have a can of beans?
- May I have a jar of beans?

- "can" and "may" are different distributionally
- "can" and "jar" are different distributionally
- They are thus in different equivalence classes.

Partial substitutability



Abstractness of Representational Primitives



KING'S College LONDON Founded 1829

Chomsky hierarchy

Top down

$$
\begin{array}{c}
\text{UNRESTRICTED} \\
\uparrow \\
\text{CONTEXT SENSITIVE} \\
\uparrow \\
\text{CONTEXT FREE} \\
\uparrow \\
\text{REGULAR} \\
\uparrow \\
\text{FINITE}
\end{array}
$$

Chomsky hierarchy

Top down

UNRESTRICTED

↑

CONTEXT SENSITIVE

↑ ⟵ PTIME?

CONTEXT FREE

↑

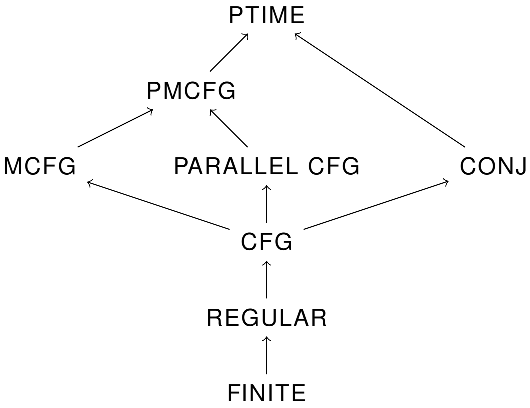
REGULAR

↑

FINITE

Chomsky hierarchy

Bottom up



Some claims

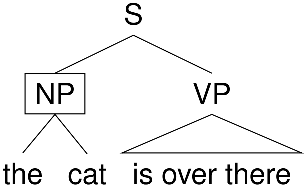
Weak claim

We have a family of distributional learning algorithms for each part of this hierarchy.

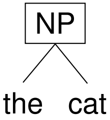
Stronger claim

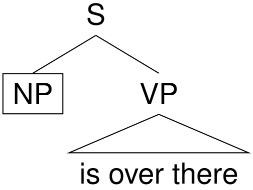
These are the *right* algorithms for these classes.

Syntactic environments



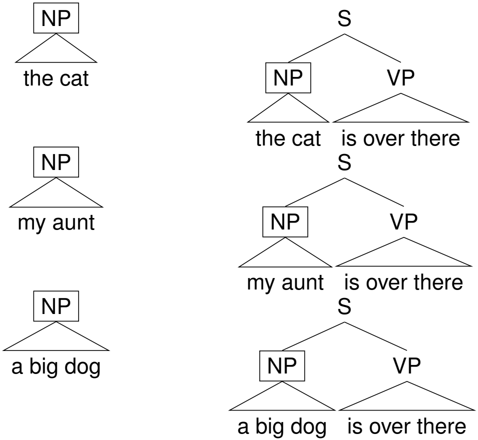
Syntactic environments





Functional view

Function from subtrees to whole trees.



Functional view

Mapped to the interface

the cat	$\rightarrow$	the cat is over there
my aunt	$\rightarrow$	my aunt is over there
a big dog	$\rightarrow$	a big dog is over there

Function

$f(x) = x$ is over there.

. . . is just a context

$\square$ is over there.

Features

The syntactic label just specifies a set of syntactic environments that it can occur in.

Assumption

- Any subtree with the same features can occur in the same set of environments.
- ‘Context-free’ derivation: we only look at a finite amount of local information

The features are merely a way of specifying this relation.

Table

Subtrees and tree contexts

	C_1	C_2	C_3	C_4	C_5
T_1	1	1	1		
T_2	1	1	1		
T_3	1	1	1		
T_4				1	1
T_5				1	1
T_6				1	1

Table

Subtrees and tree contexts

	C_1	C_2	C_3	C_4	C_5
T_1	N	N	N		
T_2	N	N	N		
T_3	N	N	N		
T_4				M	M
T_5				M	M
T_6				M	M

Table

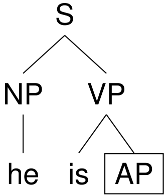
Subtrees and tree contexts

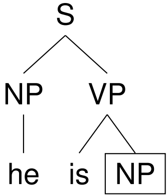
	C_1	C_2	C_3	C_4	C_5
T_1	1	1	1		
T_2	1	1	1		
T_3	1	1	1		
T_4				1	1
T_5				1	1
T_6				1	1

‘Ambiguity’ of subtrees and tree contexts

AUX	N
|	|
can	can

‘Ambiguity’ of subtrees and tree contexts





Table

Map to strings

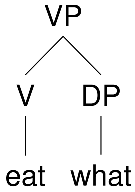
	$l_1 \Box r_1$	$l_2 \Box r_2$	$l_3 \Box r_3$	$l_4 \Box r_4$
w_1	1	1	1	
w_2	1	1	1	
w_3	1	1	1	1
w_4			1	1
w_5			1	1

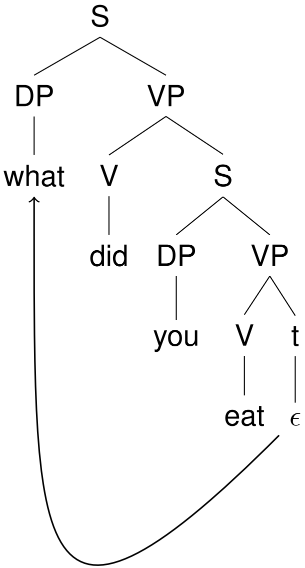
Minimal grammars

Theorem: [Clark(2013a)]

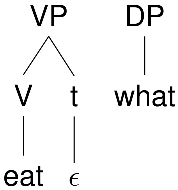
The smallest grammar that generates a given language will have categories that are *maximal rectangles* in the context substring table.

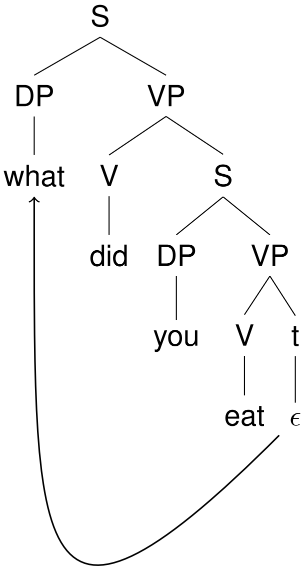
Movement





Movement


VP
DP
V
t
what
eat
ϵ


S
DP
VP
what
V
S
did
DP
VP
you
V
t
eat
ϵ

Pairs of strings

Subderivation

Yields a pair of strings $\langle$eat, what$\rangle$

Context

$\langle$eat, what$\rangle$ $\rightarrow$ 'what did you eat'
$f(x, y) = y$ did you x

Copying

Clark and Yoshinaka (2013)

- Suffixaufnahme
- Yoruba relative clauses
- Reduplication in morphology

$$\begin{array}{c} \text{Nplural(orangorang)} \\ | \\ \text{nsing(orang)} \end{array}$$

Context: $f(x) = xx$

Yoruba – Kobele (2006)

Example

Rira **aja nla ti o ge obinrin ti mo feran jẹ** ti Ade ra **aja nla ti o ge obinrin ti mo feran jẹ** ko da

$f(x) =$ 'Rira x ti Ade ra x ko da'

PMCFGs

Parallel Multiple Context-Free Grammars

Clark and Yoshinaka (2013) MLJ

A hierarchy of PMCFGs can be identified in the limit using positive data and membership queries.
$\mathbb{G}(p, q, r, s)$.

- This seems to contain all natural languages.
- More precisely: There are no arguments that the class of natural languages does not lie in one of these classes.
- Many simple CFGs do not lie in any of them.

$$\{a^n b^m \mid n \neq m\}$$

Hierarchy

For every level of the hierarchy we have a 'context':

- Regular: $f(x) = lx$
- CFG: $f(x) = lxr$
- MCFG: $f(x, y) = lxmyr$
- Parallel CFG: $f(x) = lxmxr \ldots$
- Parallel MCFG: $f(x, y) = lxmxyr \ldots$

Hierarchy

For every level of the hierarchy we have a 'context':

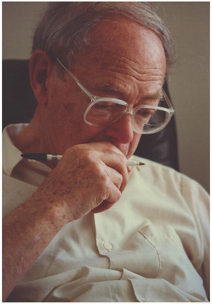
- Regular: $f(x) = lx$
- CFG: $f(x) = lxr$
- MCFG: $f(x, y) = lxmyr$
- Parallel CFG: $f(x) = lxmxr \ldots$
- Parallel MCFG: $f(x, y) = lxmxyr \ldots$

- Distribution $C_L(\vec{u}) = \{f \in \mathcal{F} | f(\vec{u}) \in L\}$

Distributional Learning

Zellig Harris (1949, 1951)

> *Here as throughout these procedures X and Y are substitutable if for every utterance which includes X we can find (or gain native acceptance for) an utterance which is identical except for having Y in the place of X*



- Contexts derived from CFGs.
- Complete equality of distribution.

Outline

Methodology

Substitutability
- Semantics
- Strong learning

General theory
- Queries and Features
- Beyond Context-Free Grammars

Conclusion
- Probabilistic learning
- Predictions

Overly idealised learning models?

- Ideal Bayesian learners (e.g. Chater and Vitanyi, 2007)
- Practical experience (e.g. TENJINNO, OMPHALOS, Lang, 1992)
- Equivalence models from probabilistic data (Carrasco and Oncina, 1994)
- Indirect negative evidence (Clark and Lappin, 2011)

Membership queries to probabilistic learning

Membership queries

- Possible string w
- Possible context $l \Box r$
- The learner can ask whether $lwr \in L$

Membership queries to probabilistic learning

Simplified version of probabilistic learning

- A string *w* that occurs with probability $> \epsilon_1$
- A context $l\Box r$ that occurs with probability $> \epsilon_2$
- Then *lwr* should occur more frequently than $f(\epsilon_1, \epsilon_2)$
- If we observe $\mathcal{O}(\frac{1}{f(\epsilon_1 \epsilon_2)})$ examples without seeing *lwr* then *lwr* is probably ungrammatical.

Computationally efficient probabilistic learning results

A problem in statistical learning of no linguistic interest

All Regular languages

Clark and Thollard (2004)

Some Context-free languages

Clark (2006)
Luque and Infante-Lopez (2010)
Shibata and Yoshinaka (2013)

Context-sensitive results?

No efficient ones yet

Computationally efficient probabilistic learning results

A problem in statistical learning of no linguistic interest

All Regular languages

Clark and Thollard (2004)

Some Context-free languages

Clark (2006)
Luque and Infante-Lopez (2010)
Shibata and Yoshinaka (2013)

Context-sensitive results?

No efficient ones yet
Inefficient: Angluin (1988), Chater and Vitányi (2007)

Some weak predictions

Weak learning

Class of languages is probably weakly adequate:
What subset of techniques are actually used?

Impossible languages

$$\{a^n b^m | n \neq m\}$$

Some strong predictions

For a word u let

- $Lex(u)$ be the set of lexical entries for u.
- $C_L(u)$ be the distribution of u

Strong identity implies weak identity

If $Lex(u) = Lex(v)$ then $C_L(u) = C_L(v)$

Some strong predictions

For a word u let

- $Lex(u)$ be the set of lexical entries for u.
- $C_L(u)$ be the distribution of u

Strong identity implies weak identity

If $Lex(u) = Lex(v)$ then $C_L(u) = C_L(v)$

Prediction: weak identity must imply strong identity

If $C_L(u) = C_L(v)$ then $Lex(u) = Lex(v)$

Some strong predictions

For a word u let

- $Lex(u)$ be the set of lexical entries for u.
- $C_L(u)$ be the distribution of u

Strong identity implies weak identity

If $Lex(u) = Lex(v)$ then $C_L(u) = C_L(v)$

Prediction: weak identity must imply strong identity

If $C_L(u) = C_L(v)$ then $Lex(u) = Lex(v)$
If $C_L(u) \supseteq C_L(v)$ then $Lex(u) \supseteq Lex(v)$

[Chomsky(1965)]

> *The only proposals that are explicit enough to support serious study are those that have been developed within taxonomic linguistics. It seems to have been demonstrated beyond reasonable doubt that quite apart from any questions of feasibility, methods of the sort that have been studied in taxonomic linguistics are intrinsically incapable of yielding the systems of grammatical knowledge that must be attributed to the speaker of a language.*

This is now clearly false.

Conclusions

- It is possible to learn sufficiently rich and abstract grammars by a process of inductive generalization from regularities in input data.
- Distributional learning is therefore a plausible theory of language acquisition.
- It is the only worked out alternative on the table now that parameter-based models in the classic sense have been abandoned.

Thanks to my colleagues

Ryo Yoshinaka, Rémi Eyraud, Shalom Lappin ...

Bibliography I

R.C. Berwick, P. Pietroski, B. Yankama, and N. Chomsky.
Poverty of the stimulus revisited.
Cognitive Science, 35:1207–1242, 2011.

Noam Chomsky.
Aspects of the Theory of Syntax.
MIT, 1965.

A. Clark.
The syntactic concept lattice: Another algebraic theory of the context-free languages?
Journal of Logic and Computation, 2013a.
doi: 10.1093/logcom/ext037.

KING'S College LONDON Founded 1829

Bibliography II

Alexander Clark.
Learning trees from strings: A strong learning algorithm for some context free grammars.
Journal of Machine Learning Research, 2013b.
to appear.

Alexander Clark and Rémi Eyraud.
Polynomial identification in the limit of substitutable context-free languages.
Journal of Machine Learning Research, 8:1725–1745, August 2007.

Alexander Clark and Ryo Yoshinaka.
Distributional learning of parallel multiple context-free grammars.
Machine Learning, pages 1–27, 2013.
ISSN 0885-6125.
doi: 10.1007/s10994-013-5403-2.

KING'S College LONDON Founded 1829

Bibliography III

URL
http://dx.doi.org/10.1007/s10994-013-5403-2.

Zellig Harris.
Distributional structure.
In J. A. Fodor and J. J. Katz, editors, *The structure of language: Readings in the philosophy of language*, pages 33–49. Prentice-Hall, 1964.

Steven Pinker.
Language acquisition.
In Daniel Osherson, Lila R. Gleitman, and Mark Liberman, editors, *An invitation to cognitive science: Volume 1 Language*, chapter 6, pages 135–182. MIT Press, 2nd edition, 1995.

John R Platt.
Strong inference.
Science, 146(3642):347–353, 1964.

Bibliography IV

M. Steedman.
Phrasal intonation and the acquisition of syntax.
In J.L. Morgan and K. Demuth, editors, *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, pages 331–342. Lawrence Erlbaum Associates, 1996.

K. Wexler and P. W. Culicover.
Formal Principles of Language Acquisition.
MIT Press, Cambridge, MA, 1980.

R. Yoshinaka and M. Kanazawa.
Distributional learning of abstract categorial grammars.
In Sylvain Pogodalla and Jean-Philippe Prost, editors, *Logical Aspects of Computational Linguistics*, pages 251–266. Springer Berlin Heidelberg, 2011.