Expressive power

Learnability and Language Acquisition

Alexander Clark

Department of Computer Science
Royal Holloway, University of London

Tuesday

Outline

The data

Chomsky hierarchy

- Swiss German

Bottom up

Problems

Conclusion

(Ed Stabler, Greg Kobele, Makoto Kanazawa)

Outline

The data

Chomsky hierarchy
- Swiss German

Bottom up

Problems

Conclusion

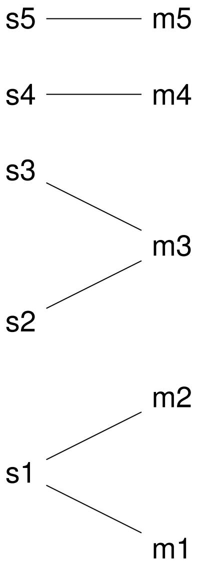
What are the linguistic data?

- We observe sound/meaning pairs.

What are the linguistic data?

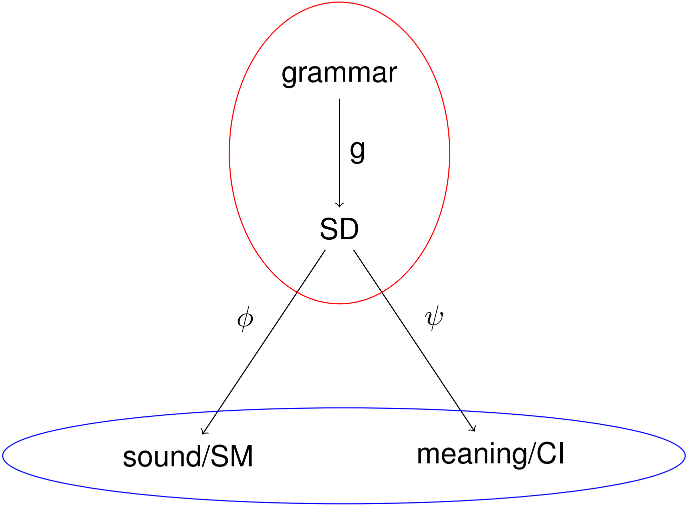
- We observe sound/meaning pairs.
- More precisely, we observe sounds/gestures.
- We don't observe the meanings, but we are aware of entailment relations between propositions. (or something similar)
- The sound/meaning relation is not one-to-one, but there is a lot of structure. (knowing what was said tells you a lot about the meaning!)

Sound/meaning pairs



Structural descriptions

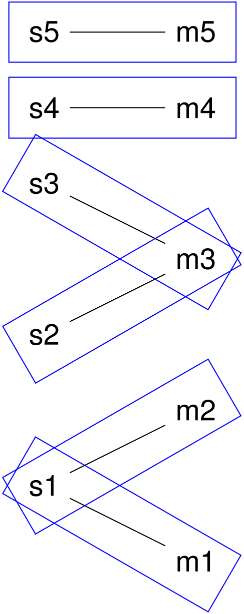
Traditional view



Example

- Grammar might be a context-free grammar, augmented with some compositional semantics.
- Structure might be a labeled derivation (parse) tree.
- ϕ is just the yield of the tree (a function).
- ψ just computes the semantics (another function).

Standard model



Issues

Description

If you are just describing the data, then you can stipulate this as a requirement.
If you are *learning* then you may not be able to learn grammars that satisfy this stipulation.
So you need to decide whether this stipulation is necessary.

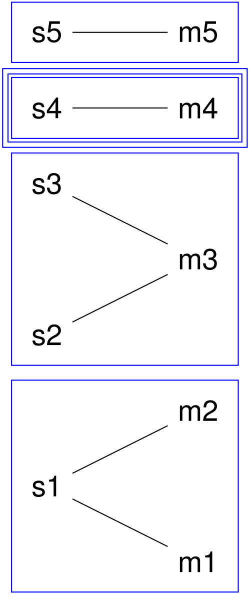
Issues

Description

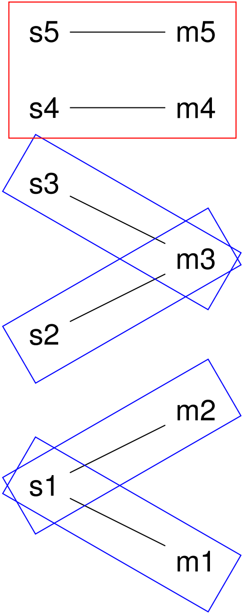
If you are just describing the data, then you can stipulate this as a requirement.
If you are *learning* then you may not be able to learn grammars that satisfy this stipulation.
So you need to decide whether this stipulation is necessary.

- Everybody here speaks two languages.
- Either way is fine.
- I gave him a cake and her a biscuit/I gave him a cake.
- John likes but Marys hates Sibelius/John likes Sibelius

Alternative model



Invalid structures



Weak and Strong

Weak generative capacity

Set of strings that a grammar can generate
Mathematically tractable

Strong generative capacity

Set of structures that a grammar can generate
Linguistically interesting but less tractable

From a learnability point of view, this is not so crucial: the data that the child sees consists of semantically well formed utterances.

Outline

Chomsky hierarchy of rewriting systems

Chomsky, 1959

Rewriting rules of the form $\alpha \to \beta$

Type 0	Unrestricted rewriting systems
Type 1	Context sensitive grammars
Type 2	Context free grammars
Type 3	Regular grammars/finite automata

- Classes of *representations*
- Purely "syntactic" hierarchy

Rules

Terminal symbols

Σ: a, b, c observed

nonterminal symbols

V: N, P, Q unobserved

Rules: $\alpha \rightarrow \beta$

$N \rightarrow abc$
$N \rightarrow Pbc$
$N \rightarrow aPb$
$N \rightarrow PaQ$
$aNb \rightarrow aPQb$
$NabP \rightarrow QcRS$

Finite languages

Grammar rules

Simple list of all sentences in the language

$N \rightarrow abc$

analogy: Monkey vocalisations

Language class

All finite languages

Languages with a finite number of sentences

Regular languages

Grammar rules

$N \rightarrow aP$

Only one nonterminal on the rhs, and all terminals on the left.

Language class

All regular languages

DFAs, NFAs, regular languages, generated by a regular language, finite Myhill-Nerode congruence, ...

(Predates the Chomsky Hierarchy)

Linear languages

Grammar rules

$N \rightarrow aPbc$

Only one nonterminal on the rhs

Language class

All linear languages

Between regular languages and context-free languages

Example $L = \{a^n b^n | n > 0\}$

Context-free grammars

Grammar rules

$N \rightarrow aPbQc$

Only one nonterminal on the lhs, any number on the rhs

Language class

Context-free languages
Polynomially parsable
regular grammars applied to trees

Examples of context-free languages

$(ab)^+$

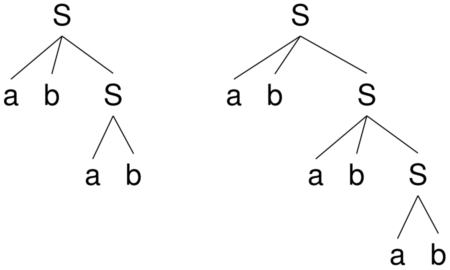
Examples of context-free languages

$(ab)^+$

Grammar 1

$S \rightarrow ab$

$S \rightarrow abS$



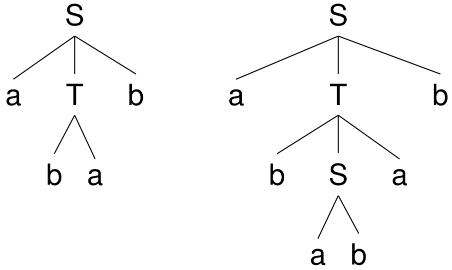
Examples of context-free languages

$(ab)^+$

Grammar 2

$S \rightarrow ab$ and $S \rightarrow aTb$

$T \rightarrow ba$ and $T \rightarrow bSa$

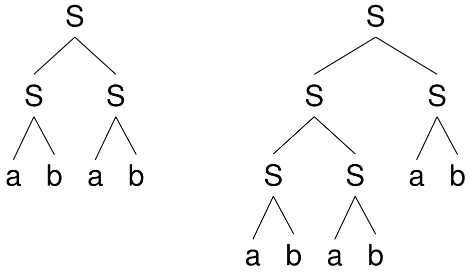


Examples of context-free languages

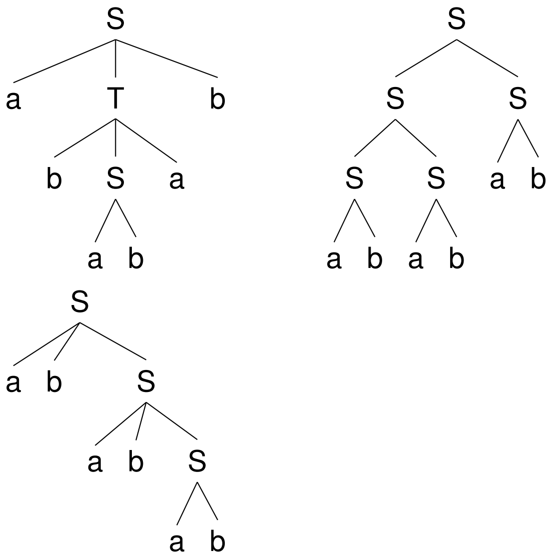
$(ab)^+$

Grammar 3

$S \rightarrow ab$ and $S \rightarrow SS$



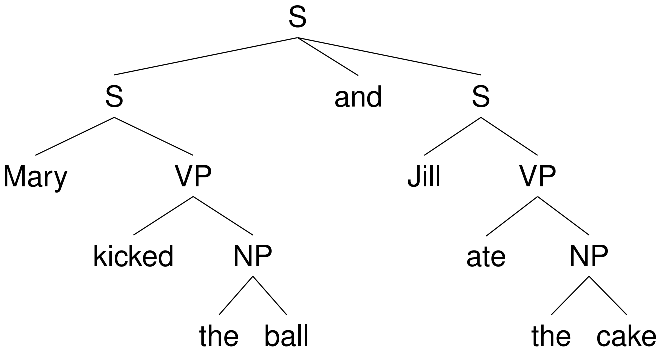
Which is the 'right' structure?



English example

Which structure gives you the right dependencies?

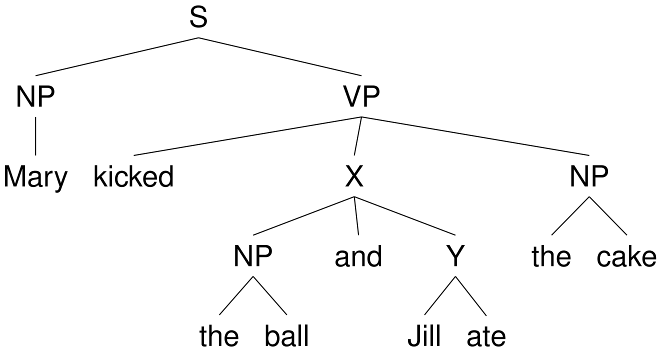
- Mary kicked the ball and Jill ate the cake



English example

Which structure gives you the right dependencies?

- Mary kicked the ball and Jill ate the cake



Semantically well-formed

- Mary kicked the ball and Jill ate the cake
- # Mary fed the ball and Jill ate the cake
- # Mary kicked the cake and Jill ate the cake
- # Mary fed the ball and Jill ate the cat
- Mary fed the cat and Jill ate the cake

Crucial to consider semantic wellformedness, not just syntactic wellformedness, especially when we come to think about learnability.

Examples of (properly) context-free languages

$a^n b^n | n > 0$

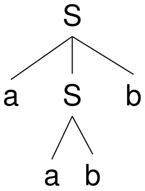
Examples of (properly) context-free languages

$a^n b^n | n > 0$

Grammar 1

$S \rightarrow ab$

$S \rightarrow aSb$

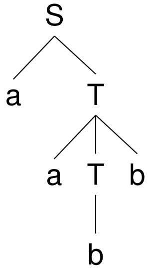


Examples of context-free languages

$a^n b^n | n > 0$

Grammar 2

$S \rightarrow ab$ and $S \rightarrow aT$
$T \rightarrow b$ and $T \rightarrow aTb$



Argument that a language is not regular

Facts:

- If L and R are regular languages, then $L \cap R$ is also regular.

General structure of correct argument

We want to show that L is not regular, but L is very complex
Take a simple regular language R
$L \cap R$ is a simple but infinite subset of L
Show that $L \cap R$ is not regular
Therefore L is not regular

English is not regular

Example – Shuly Wintner

A white male hired another white male

A white male – whom a white male hired – hired another white male

Define R

$R =$
$\{\text{A white male(whom a white male)}^*\text{(hired)}^*\text{hired another white male}\}$

Subset of English

$L \cap R =$
$\{\text{A white male(whom a white male)}^n\text{(hired)}^n\text{hired another white male} | n > 0\}$

Argument that a language is not context free

Facts:

- If L is CF and R is regular, then $L \cap R$ is also CF.

General structure of correct argument

We want to show that L is not CF, but L is very complex
Take a simple regular language R
$L \cap R$ is a simple but infinite subset of L
Show that $L \cap R$ is not CF
Therefore L is not CF

Swiss German

Shieber, 1985

...	das	mer	d'chind	em Hans	es huus	lönd	hälfe	aastriiche
...	that	we	the children-ACC	Hans-DAT	house-ACC	let	help	paint

'... that we let the children help Hans paint the house'

Proof

This language is not CF

abcd, aabccd, aaabcccd, abbcdd, . . .
number of *a*s equals number of *c*s and
number of *b*s equals number of *d*s
$a^n b^m c^n d^m$

Proof

This language is not CF

abcd, aabccd, aaabcccd, abbcdd, . . .
number of *a*s equals number of *c*s and
number of *b*s equals number of *d*s
$a^n b^m c^n d^m$

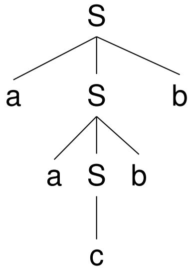
Pick regular language

a is accusative noun, b is dative noun
c is verb that needs accusative, d is verb that needs dative
$R = a^* b^* c^* d^*$

Outline

Derivation tree

Top-down versus Bottom-up



Top down versus bottom up

Kanazawa (2011)

Top down

Chomsky (1956, 1959)

- $N \rightarrow PQ$
- Rewrite N as PQ

Bottom up derivations

Smullyan (1961), Chomsky (1995), Stabler (1997)

- $N \leftarrow PQ$
- Combine P and Q to make N
- $N(xy) := P(x), Q(y)$

Minimalist program and MERGE

External merge

Combine X and Y to make a new object.

Internal merge

Y is part of X

Minimalist Grammars (Stabler, 1997)

Tractable formalisation of this approach.

Mildly context sensitive languages

Joshi

Slightly vague and informal definition:

Definition

A class of languages is MCS if:

- Not too complex: Efficiently parsable (in P)
- Constant growth property/semilinearity
- Includes some classes of crossing dependencies
- (includes all context-free languages)

Convergence of MCS formalisms

Joshi et al. (1990)

A crucial result for linguistics: all of the then current proposals turned out to be weakly equivalent:

- Linear indexed grammars
- Tree adjoining grammars
- Head grammars
- Combinatory categorial grammars

The derivation structures can all be described using LCFRs (we will use Multiple Context Free Grammars instead)

Movement

MCFGs and MGs

See handout.

Outline

Argument that a language is not MCF

Facts (Seki et al.):

- If L is MCF and R is regular, then $L \cap R$ is also MCF.

General structure of correct argument

We want to show that L is not MCF, but L is very complex
Take a simple regular language R
$L \cap R$ is a simple but infinite subset of L
Show that $L \cap R$ is not MCF
Therefore L is not MCF

Old Georgian

Kracht and Michaelis

Suffixaufnahme/Stacking in Old Georgian

A word can have a number of case-suffixes that depends on how deeply embedded it is
WME- word marking English
Verb 'run', and noun 'son', case marker 'gen'

- son run
- son songen run
- son songen songengen run
- son songen songengen songengengen run.

See Joshi's rebuttal and Pesetsky's recent work

Chinese number names

Radzinski, 1991

Data

wu = 5, zhao $= 10^{12}$

wu zhao zhao wu zhao $= 5 \times 10^{24} + 5 \times 10^{12}$

*wu zhao wu zhao zhao

Language intersected with $\{\text{wu}, \text{zhao}\}^*$

$\{wuzhao^{k_1} wuzhao^{k_2} \ldots wuzhao^{k_n} | k_1 > k_2 \cdots > k_n\}$

This is not semilinear

Yoruba

Kobele, 2006 (and MCFG2 2011)

Relative clauses

- Olu ra adie ti o go
- Olu buy chicken that 3s dumb
- Rira adie ti o go to Olu ra adie ti o go ko da
- buying chicken that 3s dumb that Olu buy chicken that 3s dumb not good
- * Rira adie ti o go to Olu ra adie ti o kere ko da

Recursive copying: intersection with appropriate R gives 2^n copies of 'adie'

Outline

Conclusion

- Broad consensus that somewhere in the MCFG hierarchy is correct
- Some problematic cases but all a bit marginal
- Possibly we need some true copying operation as well
  $N_1(uu) \leftarrow P_1(u)$
  Also for full stem reduplication in morphology?
- Alternatively there might be some nontrivial operation at Spellout.
- We will make the default assumption here that MCFGs are adequate.