# The linguistic relevance of MCFLs

Greg Kobele

University of Chicago

MCFG+ 2
Nara, Japan

# Introduction

Chomsky 1956

Suppose that for many languages there are certain clear cases of grammatical sentences and certain clear cases of ungrammatical sequences, e.g., (1) and (2), respectively, in English.

(1) John ate a sandwich

(2) Sandwich a ate John.

In this case, we can test the adequacy of a proposed linguistic theory by determining, for each language, whether or not the clear cases are handled properly by the grammars constructed in accordance with this theory. For example, if

# Introduction

- The ‘canonical’ datum of linguistics is of the form $w \in L$ or $w \notin L$.
- A theory of a language is a description of some $L$ which correctly classifies these data.
- A theory is good if concisely describes the data. (If the cost of encoding the actual data-cum-theory is low.)
- Sometimes using a grammar that generates a different language can provide a shorter description than could any other.

**$1024, 1048576, 59049 \in L$**

- $\langle 1024, 1048576, 59049 \rangle$?

- As the amount of data grows, the more benefit there is to treating it as a projection of an infinite set.

# Introduction

- The 'canonical' datum of linguistics is of the form $w \in L$ or $w \notin L$.
- A theory of a language is a description of some $L$ which correctly classifies these data.
- A theory is good if concisely describes the data. (If the cost of encoding the actual data-cum-theory is low.)
- Sometimes using a grammar that generates a different language can provide a shorter description than could any other.

**$1024, 1048576, 59049 \in L$**

- $\langle 1024, 1048576, 59049 \rangle$?
- $\langle f(x) = x^{10}, 2, 4, 3 \rangle$?

- As the amount of data grows, the more benefit there is to treating it as a projection of an infinite set.

# Introduction

- We are actually presented with data from different languages $(w \in L_1, u \notin L_2, v \in L_3, \ldots)$
- We can ask:

What kinds of properties do these $L$ share?

- We can then factor out these commonalities from the description of the individual $L$s, stating them just once.
- As the number of different languages we consider grows, the more benefit there is to treating them as a projection of an infinite set.

- $\langle\langle 1024, 1048576, 59049\rangle, \langle 9, 81\rangle, \langle 1, 2, 1\rangle\rangle$

# Introduction

- We are actually presented with data from different languages ($w \in L_1, u \notin L_2, v \in L_3, \ldots$)
- We can ask:

What kinds of properties do these $L$ share?

- We can then factor out these commonalities from the description of the individual $L$s, stating them just once.
- As the number of different languages we consider grows, the more benefit there is to treating them as a projection of an infinite set.

- $\langle\langle 1024, 1048576, 59049\rangle, \langle 9, 81\rangle, \langle 1, 2, 1\rangle\rangle$
- $\langle\langle x^{10}, 2, 4, 3\rangle, \langle x^2, 39\rangle, \langle x^1, 1, 2, 1\rangle\rangle$

# Introduction

- We are actually presented with data from different languages $(w \in L_1, u \notin L_2, v \in L_3, \ldots)$
- We can ask:

What kinds of properties do these $L$ share?

- We can then factor out these commonalities from the description of the individual $L$s, stating them just once.
- As the number of different languages we consider grows, the more benefit there is to treating them as a projection of an infinite set.

- $\langle\langle 1024, 1048576, 59049\rangle, \langle 9, 81\rangle, \langle 1, 2, 1\rangle\rangle$
- $\langle\langle x^{10}, 2, 4, 3\rangle, \langle x^2, 39\rangle, \langle x^1, 1, 2, 1\rangle\rangle$
- $\langle x^y, \langle 10, 2, 4, 3\rangle, \langle 2, 3, 9\rangle, \langle 1, 1, 2, 1\rangle\rangle$

# Introduction

- The more restricted the class of possible grammars is, the cheaper it, and the individual languages will be to describe.
- Clearly, we aren't (yet) computing the costs of various encoding schemes on data.
- Instead, we are looking at individual languages, and estimating how well we can encode them using various description methods.
- Consider the question

Is English regular?

# Introduction

**English contains sentences like**

People eat. Monkeys eat bananas. People monkeys eat die.
Bananas monkeys eat are yellow. People people eat eat. ...

- One option is to treat this as a finite set.
- Another is to treat this as a projection of an infinite language, ENG, which generates sentences of (among others) the form

$$N\ S^N\ V$$

where $S^N$ is an $S$ with an $N$ gap.

# Introduction

- Although the pattern of sentences of $\textsc{Eng}$ described previously uses non-regular notions, we can ask whether we can find a description of $\textsc{Eng}$ among the more restricted class of regular languages.
- We cannot:
  1. Assume for a contradiction: There is a regular description of $\textsc{Eng}$.
  2. The intersection of any two regular languages is again a regular language.
  3. $\textit{people}^* \textit{eat}^*$ is a regular language.
  4. $\textsc{Eng} \cap \textit{people}^* \textit{eat}^*$ is regular.
  5. $\textsc{Eng} \cap \textit{people}^* \textit{eat}^* = \textit{people}^n \textit{eat}^n$
  6. $\textit{people}^n \textit{eat}^n$ is not regular. $\perp$

# Introduction

- Although the pattern of sentences of ENG described previously uses non-regular notions, we can ask whether we can find a description of ENG among the more restricted class of regular languages.
- We cannot:
  1. Assume for a contradiction: There is a regular description of ENG.
  2. The intersection of any two regular languages is again a regular language.
  3. $people^* eat^*$ is a regular language.
  4. ENG $\cap$ $people^* eat^*$ is regular.
  5. ENG $\cap$ $people^* eat^* = people^n eat^n$
  6. $people^n eat^n$ is not regular. $\perp$

# Introduction

- The proof relies heavily on an analysis of the data.
- At best we can show that the analysis is or is not in the class in question.
- How convincing this will be depends on the perceived quality of the generalization.

# Introduction

- The proof relies heavily on an analysis of the data.
- At best we can show that the analysis is or is not in the class in question.
- How convincing this will be depends on the perceived quality of the generalization.
- Note that we cannot simply conclude based on the fact that $\mathit{people}^n\mathit{eat}^n \subseteq \textsc{Eng}$ that $\textsc{Eng}$ is not regular.
  - It is not in general true that a subset of a regular language will be regular.
  - $\Sigma^*$ is regular, but *every* language is a subset of it.

# Introduction

- We want to know whether our generalizations about language can be captured by means of a restrictive formal class.
- The more restrictive and natural the class from which we ultimately draw our descriptions of language, the cheaper it will be to encode.
- The general strategy will be to determine first what patterns are *not* part of the class under discussion, and second whether these patterns are a part of some natural language.
- ‘Part’ does *not* mean ‘subset of’, but something a little more complicated, depending on the closure properties of the class.

# Is NL Context-Free?

- The characteristic dependency of context-free languages is that of center embedding.
- A useful non-CF language is $w\overline{w}$, which intuitively requires arbitrarily many dependencies to cross.
- Like regular languages, CF languages are closed under homomorphisms and intersection with regular sets.

# (Swiss) German

## German — $ww^r$

- . . . wir Hans das Haus anstreichen lassen
  . . . we Hans the house paint let
  "we let Hans paint the house"

## Swiss German — $ww$

- . . . mer de Hans es huus lönd aastriiche
  . . . we Hans the house let paint
  "we want to let Hans paint the house"

## Swiss German

ACC: *laa* requires its object to be accusative:

- ... mer de/*em Hans es huus haend wela laa aastriiche
  ... we the Hans the house have wanted let paint

  "we wanted to let Hans paint the house"

DAT: *hälfe* requires its object to be dative:

- ... mer *de/em Hans es huus haend wela hälfe
  ... we the Hans the house have wanted help
  aastriiche
  paint
  "we wanted to help Hans paint the house"

# Swiss German

- Describing Swiss German as an infinite set, it seems natural to say that the nouns and verbs are in a 1-1 relation. (Each verb selects exactly one object, which must be present.)
- Moreover, the case on the object must match the case required by the verb.
- Most importantly, this crossing-style word order remains possible no matter how many verbs and objects there are...

- ...mer d'chind em Hans es huus haend wela laa
  ...we the children the Hans the house have wanted let
  hälfe aastriiche
  help paint
  "we wanted to let the children help Hans paint the house"

# Swiss German

- Assume for a contradiction: SWISS is context-free.
- The intersection of any context-free language and regular language is a context-free language.
- $L =$
  $\ldots$ *mer d′chind*$^*$ (*em Hans*)$^*$ *es huus haend wela laa*$^*$ *hälfe*$^*$ *aastriich*
  is a regular language.
- SWISS $\cap\ L$ is context-free.
- SWISS $\cap\ L =$
  $\ldots$ *mer d′chind*$^i$ (*em Hans*)$^j$ *es huus haend wela laa*$^i$ *hälfe*$^j$ *aastriiche*
- This is not context-free. $\perp$

# Swiss German

- Assume for a contradiction: SWISS is context-free.
- The intersection of any context-free language and regular language is a context-free language.
- $L =$
  $\ldots$ *mer d′chind*$^*$ (*em Hans*)$^*$ *es huus haend wela laa*$^*$ *hälfe*$^*$ *aastriich*
  is a regular language.
- SWISS $\cap\ L$ is context-free.
- SWISS $\cap\ L =$
  $\ldots$ *mer d′chind*$^i$ (*em Hans*)$^j$ *es huus haend wela laa*$^i$ *hälfe*$^j$ *aastriiche*
- This is not context-free. $\perp$

# Background

- As natural languages are not contained within the context free languages, the next step in the Chomsky hierarchy are the context sensitive languages (type 1).
- But the context sensitive languages already have all the complexities of the recursively enumerable languages. . . (Savitch)
  - Let $L$ be an arbitrary r.e. language, and $M$ a deterministic turing machine with $L(M) = L$.
  - For every string $w \in L$, let $M(w)$ denote the number of steps $M$ takes to recognize $w$.
  - Then the language $L' := \{0^{M(w)}1w : w \in L\}$ is context-sensitive.
- Are there *any* formal constraints on possible natural languages?

# Not everything is possible

- We still have at least the intuition that the kinds of patterns we see in languages are all 'simple' in some sense...
- Joshi tried to make this more precise:

## "Mild" context-sensitivity

- no 'complex' patterns → PTIME
- expressions are built by combing other expressions, and by adding to them a fixed amount of pronounced material → constant growth /semilinearity
- limited numbers of crossing dependency types
- (extends the context-free languages)

# Constant Growth / Semilinearity

- There is a constant $k$ such that for any string $w$, there is another string $u$ such that $|w| < |u| \leq kn$
- The language $a^{2^n}$ is not of constant growth (but $a^{2^n}b^*$ is).
- Semilinearity is a better approximation of the intuition about how expressions are 'constructed'.

**A language is semilinear iff**

it is *letter equivalent* to a regular language

- Two languages are letter equivalent ($L_1 \approx L_2$) iff each of their sentences are, modulo word order, in the other

**For example:**

| $a^n b^n c^n$ | $\approx (abc)^*$ |
| --- | --- |
| $abc$ | $abc$ |
| $aabbcc$ | $abcabc$ |
| $aaabbbccc$ | $abcabcabc$ |
| $\vdots$ | |

# Semilinearity

- the parikh image of a string $w$ is a finite sequence of integers (a parikh vector), which indicates how many tokens of each letter occur in $w$
- a set $L$ of parikh vectors is linear iff:

$$L = \{\vec{x} + n_1\vec{y}_1 + \cdots + n_m\vec{y}_m : n_1, \ldots, n_m \in \mathbb{N}\}$$

- a semilinear set is a finite union of linear sets

**A language is semilinear iff**

its parikh image is a semilinear set.

**Intuition**

A linear set‘represents’ a single

- path ($\vec{x}$) with loops ($\vec{y}_i$)
- derivation tree ($\vec{x}$) with pumps ($\vec{y}_i$)

# Casting a semilinear shadow (I)

**Question:**

What property of languages does semilinearity reflect?

**Answer:**

None. (!!!)

**Reason:**

Every set of strings over an alphabet with at least two letters can be (straightforwardly) encoded as a semilinear set.

$$sl(L) := (01 \cdot L) \cup (10 \cdot \Sigma^*)$$

In other words: If a language is semilinear, we don't know whether this is because it has a simple structure, or because its complex structure has been hidden by other operations.

# Casting a semilinear shadow (II)

**Question:**

What property of *classes* of languages does semilinearity reflect?

**Answer:**

A non-trivial one!

**Reason:**

If a grammar formalism only generates semilinear languages, we can suspect that its basic combinatorics are 'concatenative'!

# Limited Cross-serial Dependencies

- For fixed $k$, $ww^k$ is ok.
  - An MCFG of dimension $k$ can derive $ww^{k-1}$
- the language $ww^+$ is not – this is the case where the number of crossing dependency types (the number of copies of $w$) can grow without bound.
  - Note that semilinearity already rules out $ww^+$ (constant growth does not – strings of every even length are in this set).

# Mildly Context Sensitive Grammar Formalisms

- Although lots of possible classes with these properties, it is usually taken to mean one of the below (in order of proper inclusion):

# Mildly Context Sensitive Grammar Formalisms

- Although lots of possible classes with these properties, it is usually taken to mean one of the below (in order of proper inclusion):
  - 2-$MCFL_{wn}$ $\equiv$ Linear Indexed Languages $\equiv$ Tree Adjoining Languages

# Mildly Context Sensitive Grammar Formalisms

- Although lots of possible classes with these properties, it is usually taken to mean one of the below (in order of proper inclusion):
  - 2-$MCFL_{wn}$ $\equiv$ Linear Indexed Languages $\equiv$ Tree Adjoining Languages
  - $MCFL_{wn}$ $\equiv$ simple Macro languages $\equiv$ $yCFTL_s$ $\equiv$ ACG(2,3)

# Mildly Context Sensitive Grammar Formalisms

- Although lots of possible classes with these properties, it is usually taken to mean one of the below (in order of proper inclusion):
  - 2-$MCFL_{wn}$ $\equiv$ Linear Indexed Languages $\equiv$ Tree Adjoining Languages
  - $MCFL_{wn}$ $\equiv$ simple Macro languages $\equiv$ $yCFTL_{s}$ $\equiv$ ACG(2,3)
  - MCFL $\equiv$ $yDT_{fc}$(REG) $\equiv$ OUT(DTWT) $\equiv$ STR(CFHG) $\equiv$ Minimalist Languages $\equiv$ MCTALs $\equiv$ ACG(2,4)

# Are NLs MCS?

- Just as *ww* is a simple pattern which is a non-CFL, $a^{2^n}$ is a non-MCFL (and non-semilinear).
- $a^{2^n}$ can be derived by allowing oneself to copy recursively:
  - $S(a)$. ($a$ is an $S$)
  - $S(xx) :- S(x)$. (if $x$ is an $S$, so is $xx$)
- So we can try to find constructions in NL which seem to involve copying,
- and determine whether we can embedd them in one another.

# The Verbal Relative Clause Construction

Consider the following sentences (of Yoruba, a language of Nigeria).

1. Jimo ra adie
   Jimo buy chicken
   "Jimo bought a chicken."
2. Adie ti Jimo ra kere
   chicken that Jimo buy little
   "The chicken that Jimo bought is little."
3. Rira ti Jimo ra adie ko da
   buying that Jimo buy chicken not good
   "The way/fact that Jimo bought the chicken wasn't good."
4. Rira adie ti Jimo ra adie ko da
   buying chicken that Jimo buy chicken not good

# The Verbal Relative Clause Construction

Consider the following sentences (of Yoruba, a language of Nigeria).

1. Jimo ra adie
   Jimo buy chicken
   "Jimo bought a chicken."
2. Adie **ti** Jimo ra kere
   chicken that Jimo buy little
   "The chicken that Jimo bought is little."
3. Rira **ti** Jimo ra adie ko da
   buying that Jimo buy chicken not good
   "The way/fact that Jimo bought the chicken wasn't good."
4. Rira adie **ti** Jimo ra adie ko da
   buying chicken that Jimo buy chicken not good

# The Verbal Relative Clause Construction

Consider the following sentences (of Yoruba, a language of Nigeria).

1. Jimo ra adie
   Jimo buy chicken
   "Jimo bought a chicken."
2. Adie **ti** Jimo ra kere
   chicken that Jimo buy little
   "The chicken that Jimo bought is little."
3. Rira **ti** Jimo ra adie ko da
   buying that Jimo buy chicken not good
   "The way/fact that Jimo bought the chicken wasn't good."
4. Rira adie **ti** Jimo ra adie ko da
   buying chicken that Jimo buy chicken not good

# Copying in VRels

1. *Jije ti Jimo ra adie
   eating that Jimo buy chicken
2. *Rira nkan ti Jimo ra adie
   buying something that Jimo buy chicken
3. *Rira adie ti Jimo ra nkan
   buying chicken that Jimo buy something

# Verbal Relative Clauses and Typology

$$S\ [V_1\ O\ V_2]_{VP}$$

- Yoruba (Nigeria):
  copying of V, $V_1+V_2$, and VP
- Wolof (Senegal):
  copying of V, $V_1+V_2$
- Twi (Ghana):
  copying of V

# The copied material can be arbitrarily large (I)

## Serial verbs

- Jimo ra adie se
  Jimo buy chicken cook
  "Jimo bought the chicken to cook."
- Rira adie se **ti** Jimo ra adie se ko da
  buying chicken cook that Jimo buy chicken cook not good
- Jimo ra adie se je
  Jimo buy chicken cook eat
  "Jimo bought the chicken to cook and eat."
- Rira adie se je **ti** Jimo ra adie se je
  buying chicken cook eat that Jimo buy chicken cook eat
  ko da
  not good

⋮

# The copied material can be arbitrarily large (II)

## Relative clauses

- Olu ra adie ti o go
  Olu buy chicken that 3s dumb
  "Olu bought the stupid chicken"
- Rira adie ti o go **ti** Olu ra adie ti o
  buying chicken that 3s dumb that Olu buy chicken that 3s
  go ko da
  dumb not good
- *Rira adie ti o **go** **ti** Olu ra adie ti o
  buying chicken that 3s dumb that Olu buy chicken that 3s
  **kere** ko da
  small not good

# The basic generalization

**There is a general process in Yoruba**

- which produces NPs from Ss
- by copying a VP within the S

**The copied VP can be arbitrarily large, because**

- VPs can contain NPs (e.g. relative clauses)
- VPs can contain VPs (serial verbs)

# YORUBA is not multiple context-free

**Theorem (Seki et al)**

MCFLs are closed under

- intersection with regular sets
- homomorphism

$h(\text{YORUBA} \cap R) = \{b^{2^n} : n > 2\}$, where

- $R = a^*(xcxdca)(xcxd^*ca^*xcxdca)^*(xcx)d^*e$
  where:
  - $a = rira$
  - $b = adie$
  - $c = je\ ti\ Jimo\ ra$
  - $d = je$
  - $e = ko\ da$
  - $x = abcbd$
- $h(\sigma) = \begin{cases} b & \text{if } \sigma = adie \\ \epsilon & \text{otherwise} \end{cases}$

# But is Yoruba?

- The assumptions we have made about YORUBA (that copies can be embedded in copies) are very indirectly supported.
- No sentence with even one instance of such an embedding is judged acceptable!
- Compare the situation in English:
  - $x =$ War or no war, I'm joining the army.
  - claim that $x$ or no claim that $x$, he's not joining the army.

  To the extent that we can even figure out what is going on, what do we think???
- Note that
  - *War or no battle, . . .
  - Claim that John is dead or no claim that John is dead, . . .

# Conclusion

- While there are arguments for the non-MCFL nature of natural language, these are less convincing than those for the non-CFL nature thereof.
- If we *do* accept them, the next obvious class is the one of *parallel* MCFLs, which allow recursive copying, while maintaining many of the nice properties of MCFLs.
- If we do *not*, we must find some other generalization about the data.