Formal Language Theory 3

1 Regular Grammars

FSAs don't look very much like grammars linguists use. So we can adopt an equivalent formalism: **Regular** (Type 3) grammars. (also known as *left/right linear* grammars).

Définition 1.1 *Let G be a regular grammar* $\langle V, \Sigma, R, S\rangle$, *where*

1. *V is a set of variables/non-terminals/categories.*
2. *Σ is a terminal/vocabulary symbols. V and Σ are disjoint.*
3. *R is a finite set of rules.*
4. $S \in V$ *is the start variable.*

(1) Condition on rules:
  a. Right linear grammar: All rules in R are one of the following:
    1.B $\rightarrow a$, where $B \in V$ and $a \in \Sigma$.
    2.B $\rightarrow aC$, where $B, C \in V$ and $a \in \Sigma$.
  b. Left linear grammar:
    1.$B \rightarrow a$
    2.$B \rightarrow Ca$

Example: G = $\langle V, \Sigma, S, R\rangle$, where

1. V = {S, A, B}
2. $\Sigma = \{a, b\}$

Sample derivation? How to do this with FSA?

2 Context-Free Grammars

Définition 2.1 *Let G be a context-sensitive grammar* $\langle V, \Sigma, R, S\rangle$, *where*

1. *V is a set of variables/non-terminals/categories.*
2. *Σ is a terminal/vocabulary symbols. V and Σ are disjoint.*
3. *R is a finite set of rules.*
4. $S \in V$ *is the start variable.*

Each rule has the form: Variable $\rightarrow$ string of variables and/or non-terminals.

$G_{a^n b^n} = \langle \{S\}, \{a, b\}, R, S\rangle$

- R = {

1. S → aSb
2. S → ϵ }

The famous professor language:

G = $\langle\{S, AP, DP, VP, NP\}, \{a, famous, professor, hired\}, S, R\rangle$
R =

1. S → DP VP
2. DP → D AP
3. AP → A NP
4. NP → N
5. VP → V DP
6. D → a
7. A → famous
8. N → professor
9. V → hired

Fun facts about CF languages:

1. There is a pumping lemma for CF languages, which is more complicated than the pumping lemma for regular languages.

   (2) $a^n b^n c^n \notin CF$

2. CF langs are closed under union, but **not** intersection.
3. CF langs are closed under intersection with **regular** languages.

Consider the copying language: $\{xx | x \in \{a, b\}^*\}$

- This language features cross-serial dependencies, and unbounded copying.
- It is also not context free.

How to show a natural language is not context free:

1. We know that CF langs are closed under intersection with regular languages.
2. So we show that the intersection of a regular language and our natural language is not context-free.

A common way to do it: show that the intersection has the same structure as the copying language.

3 Argument that Bambara is not context-free (Culy 1985)

Bambara: a Mandé language spoken in Mali.

- Bambara has a free choice construction that involves copying.

(3) a. wulu 'dog' — wulu o wulu 'whichever dog'
b. malo 'rice' — malo o malo 'whichever rice'
c. *wulu o malo
d. *malo o wulu

There is also an agentive nominalization construction: N + V_{trans}+la.

(4) a. wulu + nyini + la = wulunyinina 'one who searches for dogs'
b. wulu + filè+ la = wulufilèla 'one who watches dogs'

The agentive construction can appear in the free choice construction:

(5) a. wulunyinina o wulunyinina 'whoever searches for dogs'
b. wulufilèla o wulufilèla 'whoever watches dogs'

(6) a. wulunyininanyinina o wulunyininanyinina 'whoever searches for dog searchers'
b. wulufilèlafilèla o wulufilèlafilèla 'whoever watches dog watchers'
c. wulufilèlanyinina o wulufilèlanyinina 'whoever searches for dog watchers'

Let R = {wulu(filèla)^h(nyinina)^i o wulu(filèla)^j(nyinina)^k|$h, i, j, k \geq 1$} **regular language**.

- BAM∩R = {wulu(filèla)^n(nyinina)^m o wulu(filèla)^n(nyinina)^m|$m, n \geq 1$}, which has a similar structure to the copying and counting languages, which can be shown to be non context free.

Conclusion: NL $\not\subseteq CF$.