

# Formal Learning Theory 1

## 1 Context-Free Grammars

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

1.  $V$  is a set of variables/non-terminals/categories.
2.  $\Sigma$  is a terminal/vocabulary symbols.  $V$  and  $\Sigma$  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.  $S \rightarrow aSb$
  2.  $S \rightarrow \epsilon$

The famous professor language:

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

- $R =$
1.  $S \rightarrow DP VP$
  2.  $DP \rightarrow D AP$
  3.  $AP \rightarrow A NP$
  4.  $NP \rightarrow N$
  5.  $VP \rightarrow V DP$
  6.  $D \rightarrow a$
  7.  $A \rightarrow famous$
  8.  $N \rightarrow professor$
  9.  $V \rightarrow hired$

**Classes of languages in the Chomsky Hierarchy:**

1. Finite
2. Regular
3. Context-Free
4. Context-Sensitive
5. Recursively enumerable (r.e)
6. Non-recursively enumerable (non-r.e.)

## 2 The Gold Learning Paradigm

Traditional important questions in linguistics<sup>1</sup>:

1. Is language acquisition more like learning from evidence or more like growth or triggering?
2. Are specific properties of human syntax innately given at the outset in every human learner?
3. How do properties of the language learning process shape the language we speak?
4. How can linguistic ability be acquired as rapidly and reliably as it is, from readily available data?

These questions are super important for linguists, but they are quite vague.

- Can we have a more precise formulation of them (and defs. of terms like *learning*) which will allow us to sharpen them up and allow us to provide more exact answers?

The Gold Paradigm (Gold (1967) and Blum & Blum (1975)).

- First (most influential) paradigm for the formal study of language learning.
- The GP yielded a formal result (**Gold's theorem**) which was taken to have important significance for our traditional important questions.
  - Many cognitive scientists treat it as providing support for **rationalism**: substantial innate knowledge or constraints are needed to facilitate language acquisition. (vs **empiricism**).
  - Gold's result is also frequently misinterpreted in the literature (cf. Johnson 2004).
- Today we see the description of the paradigm and some positive results about learnability, and next week we see some negative results and discuss their implications for cognitive science.

Languages are sets of expressions together with a non-expression #, where an expression is an element of  $\Sigma^*$ .

A **(positive) text T** is an infinite sequence whose elements are expressions and #.

- An infinite sequence is a function with domain  $\mathbb{N}$ , assigns to each expression a natural number.
- $T(n)$  is the  $n$ th member of T, counting up from 0.
- $T[n]$  is the length  $n$  initial sequence of T.

$$(1) \quad T[n] = \langle T(0), T(1), \dots, T(n-1) \rangle$$

The **content of text T**,  $\text{content}(T)$  is the set of expressions that appear in T.

- T is a text for language L iff  $\text{content}(T) = L$ .
- # is never part of the content.
- For every text T and every  $x \in \text{content}(T)$ , there is some finite  $i$  such that  $x = T(i)$ .
- A **learner** is a (possibly partial) function  $\phi$  from finite sequences of expressions and # to grammars. (recall  $L(G) =$  the language generated by G)

---

<sup>1</sup>The presentation of the Gold paradigm and the Gold theorem are taken from Stabler (2014) (refs. on the website).

## 2.1 A def. of learning: Identification from positive text

If  $x \in \text{Dom}(\phi)$ , we say  $\phi$  is defined on  $x$  (notation  $\phi(x) \downarrow$ ).

- (2)  $\phi$  **converges** on  $T$ ,  $\phi(T) \downarrow$ , iff for some  $i$ ,  $\phi(T[j]) = \phi(T[i])$  for all  $j \geq i$ .  
  - In this case, we define  $\phi(T)$  to be  $\phi(T[i])$ .
- (3)  $\phi$  identifies  $T$  iff  $\phi(T) \downarrow$  and  $\text{content}(T) = L(\phi(T))$ .
- (4)  $\phi$  identifies  $L$  iff  $\phi$  identifies all texts for  $L$ .
- (5)  $\phi$  identifies a class of languages  $\mathcal{L}$  iff for all  $L \in \mathcal{L}$ ,  $\phi$  identifies  $L$ .
- (6) A class of languages  $\mathcal{L}$  is **identifiable/learnable** iff some learner identifies  $\mathcal{L}$ .

## 3 Identification by enumeration

### 3.1 Enumerations

An **enumeration** is a sequence, a list, a function whose domain is the positive integers or  $\mathbb{N}$ .

- An enumerable set is one whose members can be enumerated (and also  $\emptyset$ ).
  - For any vocabulary  $\Sigma$ ,  $\Sigma^*$  is enumerable. (sort into groups based on length of string, and alphabetical order).
- (7) **Fact:** The set of all grammars is enumerable.
    1. Assume that the elements of  $\Sigma$  are alphabetic, and order them alphabetically.
    2.  $V$  is finite and disjoint from  $\Sigma$ , and we can order it alphabetically (for example).
    3. We enumerate the elements of  $\Sigma \cup V$  by ordering  $\Sigma$  before  $V$ .
    4. We can then order the left sides of the rewrite rules in a similar way ( $(\Sigma \cup V^* \times V \times (\Sigma \cup V^*))$ ), as well as the right sides ( $\Sigma \cup V$ ).
    5. Assuming that the categories  $V$ , vocabulary  $\Sigma$  and start symbol  $S$  are given, we can then enumerate grammars with 0 symbols, grammars with 1 symbol, grammars with 2 symbols, and so on, to get all grammars.

### 3.2 $FIN$ is identifiable/learnable

**Théorème 3.1** *The class  $FIN$  is identifiable (Gold 1967).*

**Proof:** Suppose that we have an enumeration of grammars for all the r.e. languages.

Define a learner  $\phi_e$  as follows: for any text  $T$ ,  $\phi_e(T[i])$  is the first grammar  $G$  in the enumeration such that  $L(G) = \text{content}(T[i])$

- Consider any  $L \in FIN$  and any text  $T$  for  $L$ . Then there is some  $i$  such that  $\text{content}(T[i]) = L$ . In this case,  $\phi_e(T[j]) = \phi_e(T[i])$ , for all  $j \geq i$ , so  $\phi_e$  converges on  $T$ , and  $L(\phi_e(T)) = \text{content}(T)$ .

Notice also that nothing in the definition of learnability implies that learners are necessarily computable functions, and so we did not need to explain how the function  $\phi_e$  could be computed (what is the algorithm for obtaining the grammar?). If we are interested in a computable learner, we could let  $\phi_e(T[i])$  be the grammar  $\{S \rightarrow |x \in \text{content}(T[i])|\}$ . (for example)