

Teoria dei Linguaggi

Indice

Introduzione	3
--------------------	---

Parte I — Gerarchia di Chomsky	4
---------------------------------------	----------

1. Breve ripasso	5
------------------------	---

Introduzione

In questo corso studieremo dei **sistemi formali** che descrivono dei linguaggi: ci chiederemo cosa sono in grado di fare, ovvero cosa sono in grado di descrivere in termini di **linguaggi**.

Ci occuperemo anche delle **risorse utilizzate**, come il **numero di mosse** eseguite da una macchina che deve riconoscere un linguaggio, oppure il **numero di stati** che sono necessari per descrivere una macchina a stati finiti, oppure ancora lo **spazio utilizzato** da una macchina di Turing.

Un **linguaggio** è «uno strumento di comunicazione usato da membri di una stessa comunità», ed è composto da due elementi:

- **sintassi**: insieme di simboli (o parole) che devono essere combinati con una serie di regole;
- **semantica**: associazione frase-significato.

Per i linguaggi naturali è difficile dare delle regole sintattiche: vista questa difficoltà, nel 1956 **Noam Chomsky** introduce il concetto di **grammatiche formali**, che si servono di regole matematiche per la definizione della sintassi di un linguaggio.

Il primo utilizzo dei linguaggi formali risale agli stessi anni con il **compilatore Fortran**. Anche se ci hanno messo l'equivalente di 18 anni/uomo, questa è la prima applicazione dei linguaggi formali. Con l'avvento, negli anni successivi, dei linguaggi Algol, ovvero linguaggi con strutture di controllo, la teoria dei linguaggi formali è diventata sempre più importante.

Oggi la teoria dei linguaggi formali è usata nei compilatori di compilatori, dei tool usati per generare dei compilatori per un dato linguaggio fornendo la descrizione di quest'ultimo.

Parte I — Gerarchia di Chomsky

1. Breve ripasso

Prima di addentrarci nello studio della gerarchia di Chomsky facciamo un breve ripasso delle basi che ci serviranno durante lo studio dei linguaggi formali.

Definizione 1.1 (Alfabeto): Un **alfabeto** è un **insieme non vuoto e finito di simboli**, di solito indicato con le lettere greche maiuscole

$$\Sigma \quad | \quad \Gamma.$$

Definizione 1.2 (Stringa): Una **stringa**, o **parola**, è una **sequenza finita** di simboli appartenenti all'alfabeto Σ . Viene indicata con la lettera x e la possiamo scrivere come

$$x = a_1 \dots a_n \quad | \quad a_i \in \Sigma.$$

Data una stringa x , indichiamo il **numero di caratteri** di x con la notazione

$$|x|$$

e il **numero di occorrenze di un carattere** di Σ in x con la notazione

$$|x|_a \quad | \quad a \in \Sigma.$$

Una stringa/parola molto importante è la **parola vuota**, che possiamo indicare in vari modi:

$$\varepsilon \quad | \quad \lambda \quad | \quad \Lambda.$$

Come dice il nome, questa parola non ha simboli, ovvero è l'unica parola tale che

$$|\varepsilon| = 0.$$

L'insieme di tutte le possibili parole che possiamo formare usando l'alfabeto Σ si indica con Σ^* . Questo insieme, ovviamente, è un **insieme infinito**.

Con le parole possiamo definire una serie di operazioni, ma la più importante tra tutte è la **concatenazione**, o **prodotto**. Date due stringhe

$$x, y \in \Sigma^* \quad | \quad x = x_1 \dots x_n \wedge y = y_1 \dots y_m$$

allora la concatenazione di x e y è la stringa

$$w = x \cdot y = x_1 \dots x_n y_1 \dots y_m.$$

L'operazione di concatenazione **non è commutativa** ma è **associativa**, quindi la struttura

$$(\Sigma^*, \cdot, \varepsilon)$$

è un **monoide libero** generato da Σ .

Vediamo, per (quasi) finire, alcune proprietà che possiamo dare alle stringhe/parole.

Definizione 1.3 (Prefisso): La stringa $x \in \Sigma^*$ si dice **prefisso** di w se

$$\exists y \in \Sigma^* \mid w = xy.$$

In poche parole, x è prefisso di w se riusciamo a scomporre la stringa w in due parti, dove x è la prima di queste due. Abbiamo due tipi di prefisso:

- **proprio** se $y \neq \varepsilon$;
- **non banale** se $x \neq \varepsilon$.

Il **numero** di prefissi di una stringa w è $|w| + 1$.

Definizione 1.4 (Suffisso): La stringa $y \in \Sigma^*$ si dice **suffisso** di w se

$$\exists x \in \Sigma^* \mid w = xy.$$

In poche parole, vale quanto scritto prima, ma in questo caso y è la seconda delle due parti. Anche qui abbiamo tipi di suffisso:

- **proprio** se $x \neq \varepsilon$;
- **non banale** se $y \neq \varepsilon$.

Anche il **numero** di suffissi di una stringa w è $|w| + 1$.

Definizione 1.5 (Fattore): La stringa $y \in \Sigma^*$ si dice **fattore** di w se

$$\exists x, z \in \Sigma^* \mid w = xyz.$$

In poche parole, vale quanto scritto prima, ma in questo caso dividiamo la stringa w in tre parti e y è la centrale di queste.

Il **numero** di fattori di una stringa w è

$$\leq \frac{|w||w+1|}{2} + 1$$

per via dei possibili doppioni che possiamo trovare.

Definizione 1.6 (Sottosequenza): La stringa $x \in \Sigma^*$ si dice **sottosequenza** di w se x è ottenuta eliminando 0 o più caratteri da w . L'eliminazione può avvenire in maniera non contigua: posso eliminare qualsiasi carattere, ma la stringa risultante che leggiamo deve contenere i caratteri nello stesso ordine di partenza.

Possiamo dire che un **fattore** è una **sottosequenza contigua**.

Per finire veramente, diamo forse la definizione più importante, quella di **linguaggio**.

Definizione 1.7 (Linguaggio): Un **linguaggio** L , definito su un alfabeto Σ , è un qualunque sottoinsieme di Σ^* , ovvero

$$L \subseteq \Sigma^*.$$

Ora che abbiamo fatto un ripasso siamo pronti per vedere la gerarchia di Chomsky, per poi addentrarci nello studio di alcune classi di questa gerarchia.