Chapitre 9 : Grammaires non contextuelles

January 26, 2025

1 Grammaire non contextuelle (ou hors-contexte)

1.1 Vocabulaire

Définition 9.1 - *grammaire au sens général, grammaire de type 0*

Une *grammaire* est défini par un quadruplet (Σ, V, P, S) où :

- Σ est un alphabet fini de *symboles terminaux*, dit aussi *alphabet terminal*
- V est un alphabet fini de *symboles non terminaux* (ou *variables*), dit aussi *alphabet non terminal*
- $P \subset (\Sigma \cup V)^* \times (\Sigma \cup V)^*$ est un ensemble de *règles de production*.
  Une règle de production $(w_1, w_2) \in P$, notée $w_1 \to w_2$ est un couple de mots écrits avec des symboles terminaux et non terminaux.
- $S \in V$ est un symbole non terminal avec un statut particulier de *symbole initial* (ou *axiome*, *variable initiale*)

Une grammaire sans propriété particulière est dite *type 0*.

Remarque 9.2 - *grammaires*

On note usuellement par des majuscules les symboles non terminaux, et en minuscule les terminaux.

Exemple 9.3 - *de grammaire de type 0*

Pour $\Sigma = \{\mathtt{a}\}$, $S = \mathtt{S}$, $V = \{S, \mathtt{D}, \mathtt{F}, \mathtt{X}, \mathtt{Y}, \mathtt{Z}\}$ et $P = \{\mathtt{S} \to \mathtt{DXaD}$, $\mathtt{Xa} \to \mathtt{aaX}$, $\mathtt{XF} \to \mathtt{YF}$, $\mathtt{aY} \to \mathtt{Ya}$, $\mathtt{DY} \to \mathtt{DX}$, $\mathtt{XZ} \to \mathtt{Z}$, $\mathtt{aZ} \to \mathtt{Za}$, $\mathtt{DZ} \to \epsilon\}$, $G = (\Sigma, V, P, S)$ est une grammaire de type 0.

Définition 9.4 - *dérivabilité immédiate*

Soit $G = (\Sigma, V, P, S)$ une grammaire. Soit α et β deux mots de $(\Sigma \cup V)^*$. On dit que *α se dérive immédiatement en β* lorsqu'il existe $(\alpha_1, \beta_1) \in P$ tel que :

$$\exists (u, v) \in \Big((\Sigma \cup V)^*\Big)^2, \begin{cases} \alpha = u\alpha_1 v \\ \beta = u\beta_1 v \end{cases}$$

Le cas échéant, on note $\alpha \Rightarrow \beta$. On parle de *dérivation immédiate*. Moralement, la règle de production (α_1, β_1) remplace le facteur α_1 par le facteur β_1.

Définition 9.5 - *clôture reflexive et transitive*

On note $\Rightarrow^*$ la *clôture reflexive et transitive* de la relation $\Rightarrow$ de dérivabilité immédiate.

$\Rightarrow^*$ est définie comme la plus petite relation au sens de l'inclusion tel que :

- $\forall \alpha \in (\Sigma \cup V)^*$, $\alpha \Rightarrow^* \alpha$
- $\forall (\alpha, \beta) \in \Big((\Sigma \cup V)^*\Big)^2, (\alpha \Rightarrow \beta) \implies (\alpha \Rightarrow^* \beta)$
- $\forall (\alpha, \beta, \gamma) \in \Big((\Sigma \cup V)^*\Big)^3$, $(\alpha \Rightarrow^* \beta$ et $\beta \Rightarrow^* \gamma) \implies (\alpha \Rightarrow^* \gamma)$

Autrement dit, $\alpha \Rightarrow^* \beta$ lorsqu'il existe $(\alpha = \alpha_0, \ldots, \alpha_k = \beta)$ une suite de mots dans $(\Sigma \cup V)^*$ telle que :

$$\forall i \in [\![0, k-1]\!], \alpha_i \Rightarrow \alpha_{i+1}$$

Définition 9.6 - *clôture reflexive et transitive de $\Rightarrow$*

Soit $G = (\Sigma, V, P, S)$ une grammaire. Soit α et β deux mots de $(\Sigma \cup V)^*$. On note *$\Rightarrow^*$ la clôture reflexive et transitive de la relation $\Rightarrow$*. Ainsi $\alpha \Rightarrow^* \beta$ lorsqu'il existe $(\alpha = \alpha_0, \ldots, \alpha_k = \beta)$ une suite de mots dans $(\Sigma \cup V)^*$ telle que :

$$\forall i \in [\![0, k-1]\!], \alpha_i \Rightarrow \alpha_{i+1}$$

Exemple 9.7 - *de dérivation*

Dans la grammaire précédemment introduite :
Pour $\Sigma = \{a\}$ et $V = \{S, D, F, X, Y, Z\}$ et $P = \{\texttt{S} \to \texttt{DXaD}, \texttt{Xa} \to \texttt{aaX}, \texttt{XF} \to \texttt{YF}, \texttt{aY} \to \texttt{Ya}, \texttt{DY} \to \texttt{DX}, \texttt{XZ} \to \texttt{Z}, \texttt{aZ} \to \texttt{Za}, \texttt{DZ} \to \epsilon\}$, $G = (\Sigma, V, P, S)$ est une grammaire de type 0.

$$
\begin{aligned}
S &\Rightarrow \texttt{DXaF} \\
&\Rightarrow \texttt{DaaXF} \\
&\Rightarrow \texttt{DaaYF} \\
&\Rightarrow \texttt{DaYaF} \\
&\Rightarrow \texttt{DYaaF} \\
&\Rightarrow \texttt{DXaaF} \\
&\Rightarrow \texttt{DaaXaF} \\
&\Rightarrow \texttt{DaaaaXF} \\
&\Rightarrow \texttt{DaaaaZ} \\
&\Rightarrow \texttt{DaaaZa} \\
&\Rightarrow \texttt{DaaZaa} \\
&\Rightarrow \texttt{DaZaaa} \\
&\Rightarrow \texttt{DZaaaa} \\
&\Rightarrow \texttt{aaaa}
\end{aligned}
$$

D'où $S \Rightarrow^* \texttt{aaaa}$

Définition 9.8 - *langage engendré, langage élargi engendré par une grammaire depuis un mot*

Soit $G = (\Sigma, V, P, S)$ une grammaire et $\alpha \in (\Sigma \cup V)^*$.

- le *langage engendré par G depuis α* est l'ensemble des mots de Σ^* que l'on peut obtenir par dérivation de α en utilisant les règles de production de G :

$$\mathcal{L}_G(\alpha) = \{\beta \in \Sigma^*,\, \alpha \Rightarrow^* \beta\}$$

- le *langage élargi engendré par G depuis α* est l'ensemble des mots de $(\Sigma \cup V)^*$ que l'on peut obtenir par dérivation de α en utilisant les règles de production de G :

$$\widehat{\mathcal{L}_G(\alpha)} = \{\beta \in (\Sigma \cup V)^*,\, \alpha \Rightarrow^* \beta\}$$

Définition 9.9 - *langage engendré par une grammaire*

Soit $G = (\Sigma, V, P, S)$ une grammaire et $\alpha \in (\Sigma \cup V)^*$. Le *langage engendré par G* désigne $\mathcal{L}_G(S)$ le langage engendré par G depuis son symbole initial S.

Exemple 9.10

Pour la grammaire de l'exemple, on pourrait montrer que $\mathcal{L}_G(S) = \{a^{2^n}, n \in \mathbb{N}^*\}$ On montre que ce langage n'est pas régulier (absurde + lemme de l'étoile tmtc)

Définition 9.11 - *langage de type 0*

On dit qu'un *langage est de type 0* s'il peut être engendré par une grammaire de type 0.

Théorème 9.12 - *de Chomsky (HP)*

Les langages de type 0 sont exactement les langages récursivement énumérables, c'est-à-dire les langages reconnaissables par une machine de Turing.

Définition 9.13 - *grammaire contextuelle (HP)*

Une grammaire $G = (\Sigma, V, P, S)$ de type 0 est appelée *grammaire contextuelle (ou de type 1 ou monotone)* lorsque :

$$\forall(\alpha, \beta) \in P \setminus \{(S, \epsilon)\}, \, |\alpha| \leq |\beta|$$

Moralement, tous les facteurs "produits" sont plus long que les facteurs remplacés.

Exemple 9.14 - *de grammaire qui n'est pas "contextuelle"*

la grammaire exemple définie par $\Sigma = \{a\}$ et $V = \{S, D, F, X, Y, Z\}$ et $P = \{\texttt{S} \to \texttt{DXaD}, \texttt{Xa} \to \texttt{aaX}, \texttt{XF} \to \texttt{YF}, \texttt{aY} \to \texttt{Ya}, \texttt{DY} \to \texttt{DX}, \texttt{XZ} \to \texttt{Z}, \texttt{aZ} \to \texttt{Za}, \texttt{DZ} \to \epsilon\}$ puis $G = (\Sigma, V, P, S)$ n'est pas contextuelle :

$$(\texttt{DZ} \to \epsilon) \in P$$

mais :

$$|DZ| = 2 > |\epsilon| = 0$$

Remarque 9.15 - *indépendance des caractères "contextuel" et "non contextuel"*

Nous le reverrons, mais une grammaire est "contextuelle" indépendamment de son caractère "non contextuel".

Remarque 9.16

Sans l'autorisation d'avoir $(S, \epsilon) \in P$, les langages engendrés par des grammaires contextuelles ne peuvent pas contenir ϵ.
En effet, on a initialement S de longueur 1 et les règles de productions ne peuvent qu'augmenter la longueur du mot. 0 serait donc une longueur inexistente.

Définition 9.17 - *grammaire non contextuelle*

Une grammaire $G = (\Sigma, V, P, S)$ est dite *non contextuelle* (ou *hors contexte* ou encore *algébrique* ou *de type 2*) lorsque :

$$P \subset V \times (\Sigma \cup V)^*$$

Moralement, les règles de production ne permettent de remplacer que des symboles non terminaux (et pas des mots) : des majuscules.

Exemple 9.18 - *de règle de production non valide pour une grammaire "non contextuelle"*

`DZ` $\rightarrow \epsilon$ n'est pas une règle possible pour une grammaire hors contexte.

Exemple 9.19 - *de grammaire "non contextuelle"*

$G_1 = (\Sigma_1, V_1, P_1, S_1)$ définie par :

- $\Sigma_1 = \{a, b\}$
- $V_1 = \{S_1\}$
- $P_1 = \{S_1 \rightarrow aS_1b,\ S_1 \rightarrow \epsilon\}$

est une grammaire hors contexte.

Remarque 9.20

On dit "hors contexte" car les symboles non terminaux

PASSE 1 ajouter à l'exemple :
$\widehat{\mathcal{L}_{G_1}(S_1)} = \{a^n S_1 b^n,\ n \in \mathbb{N}\} \cup \{a^n b^n,\ n \in \mathbb{N}\}$ pour comprendre le chapeau.

Voir démo 1

Exemple 9.21 - *Langage de Dyck*

Le langage de Dyck (bon parenthésage) est engendré par :

- $\Sigma = \{\texttt{(}, \texttt{)}\}$
- $S \rightarrow (S)S$
- $S \rightarrow \epsilon$

Remarque 9.22 - *grammaires "non contextuelles" uniquement définies par les règles de production*

En général, on donne une grammaire hors-contexte uniquement avec les règles de production.
On ne précise le symbole initial que si ce n'est pas S.
Les symboles non terminaux sont exactement ceux que l'on rencontre à gauche des règles et les terminaux sont les autres qu'on rencontre.

Remarque 9.23 - *"hors-contexte" $\not\Rightarrow$ "régulier"*

Les exemples montrent qu'il existe des langages hors contexte qui ne sont pas réguliers.

Remarque 9.24 - *abréviation d'une famille de règles de production*

Pour noter rapidement un ensemble de règles de production utilisant le même symbole non terminal comme départ :

$$\{\texttt{X,}\alpha_1 \rightarrow \texttt{,} \ \ldots, \texttt{X} \rightarrow \alpha_k\} \subset P$$

avec $X \in V$, on note :

$$X \rightarrow \alpha_1 | \cdots | \alpha_k$$

Ce n'est qu'une notation, on a bien k règles de production derrière ça.

Exemple 9.25 - *exploitant ce raccourci*

G_1 s'écrit :

$$S_1 \rightarrow aS_1b|\epsilon$$

1.2 Langages réguliers et langages hors-contexte

Un résultat à connaître et savoir redémontrer.

Proposition 9.26 - *"régulier" $\implies$ "hors-contexte"*

Soit Σ un alphabet, Tout langage sur Σ régulier est "hors-contexte".

Voir Démo 2 (elle est incomplète)

Remarque 9.27 - *concernant ce résultat*

L'inclusion réciproque est fausse si $|Sigma| \geq 2$:

$$\{a^n b^n,\, n \in \mathbb{N}\}$$

est un langage hors-contexte mais non régulier.

cette définition sera rappelée, mais il faut vite la comprendre ! FC où on donne la déf, et on doit interpréter !

Définition 9.28 - *grammaire linéaire (HP)*

$G = (\Sigma, V, P, S)$ est dite :

- *linéaire* si :
$$P \subset V \times (\Sigma^* \cup \Sigma^* V \Sigma^*)$$
Moralement, on a toujours au plus un symbole non terminal (voir Fig.1)
- *linéaire gauche* lorsque :
$$P \subset V \times \Big(\Sigma^* \cup (V\Sigma^*)\Big)$$
Le facteur "produit" commence par un unique symbole non terminal ou n'en a aucun.
- *linéaire droit* lorsque :
$$P \subset V \times \Big(\Sigma^* \cup (Sigma^* V)\Big)$$

Remarque 9.29 - *chaine d'implications*

linéaire droit implqiue linéaire implique hors contexte linéaire gauche implique linéaire implique hors contexte.

Définition 9.30 - *langages linéaire, linéaire droit, linéaire gauche*

Un langage est dit linéaire (resp. simple, droit, gauche) lorsqu'il peut être engendré par une grammaire linéaire (resp. simple, linéaire droite).

Remarque 9.31 - *existence langages linéaires non réguliers*

$$\texttt{S} \rightarrow \texttt{aSb}|\epsilon$$

est linéaire et engendre $\{a^n b^n,\ n \in \mathbb{N}\}$

Proposition 9.32 - *pour un langage, régulier $\iff$ linéaire droit $\iff$ linéaire gauche*

Soit L un langage. Les propriétés suivantes sont équivalentes.

1. L est régulier

2. L est linéaire droit

3. L est linéaire gauche

Remarque 9.33 - *sur la démo*

En montrant régulier implqiue linéaire droit, on montre à nouveau que régulier implqiue hors contexte, ceci constitue une autre démo que la propriété précédente.

Démo 3.