Théorie des langages
(grammaires formelles)

origines linguistiques

Chomsky 1955 MIT
(linguistes) ⟶ compilation

syntactic structures

1ère formalisation de la grammaire

grammaires formelles
de réécriture.

codage (réécriture / étapes
machine de Turing
système de Post

($\Sigma$) $T$ terminaux (en linguistique des mots 200 000 300)

souvent noté $\Sigma$

en nombre fini

Langages $\mathcal{L} \subset T^*$

en MINUSCULES $T^*$ suites finies de terminaux

la suite vide est possible

($T^+$ suites finies non vide de terminaux)

$\mathcal{L}$ ensemble pas forcément fini de suites de terminaux (mots)

Un langage peut être décrit par un ensemble <u>fini</u> de règles.

Non Terminaux : NT
(en nombre fini) 200 NT

dont 1 NT particulier S
start / sentence

en MAJUSCULES

| | phrase | sentence |
|---|---|---|
| syntagme nominal | groupe nominal | noun <u>phrase</u> |
| syntagme verbal | groupe verbal | verb <u>phrase</u> |
| syntagme prep | groupe prépositionnel | prepositional <u>phrase</u> |

# Règle de réécriture

$$\underbrace{w\ X\ w'}_{\substack{\text{des T ds NT} \\ \text{au moins un NT}}} \longrightarrow \underbrace{w''}_{\text{des T et des NT}}$$

$$w,\ w',\ w'' \in (T \cup NT)^*$$

$$a\,X\,b\,a\,b\,Z\,b \in (T \cup NT)^*$$

$XXZ \in NT^* \subset (T \cup NT)^*$ $\qquad$ $abaa \in T^* \subset (T \cup NT)^*$

si toutes les règles sont de la forme :

$$W \, X \, W' \longrightarrow W \, M \, W'$$

$$W, W', M \in (T \cup NT)^*$$

$X$ se réécrit en $M$
dans le contexte $W \_ W'$

alors on dit que la grammaire
est contextuelles.

si toutes les règles sont de la forme

$$w \, X \, w' \longrightarrow w''$$

$w, w', w'' \in (T \cup NT)^*$

avec $|w \, X \, w'| < |w''|$

alors la grammaire est dite

length increasing (croissante)

Kuroda (1967?)

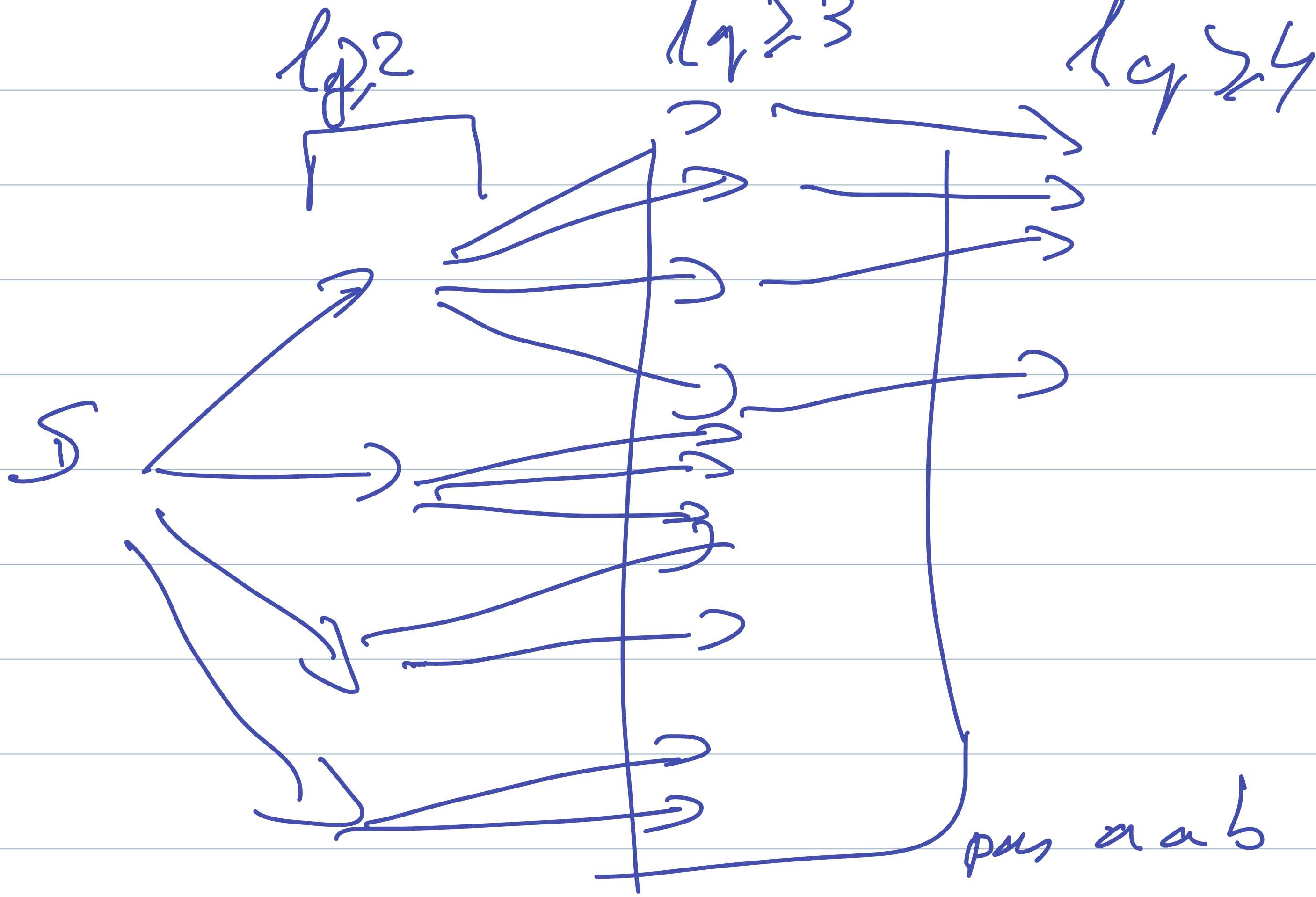
Les langages produits par des grammaires contextuelles sont les mêmes que les langages produits par des grammaires length increasing

L'appartenance d'une phrase à un langage produit (reconnu) par une grammaire contextuelle (ou croissante) est décidable.

$|a\,a\,b| = 5$

$a\,a\,b\,a\,a \in \mathcal{L}(G)$

croissante



Si toutes les règles sont de la forme

$$X \longrightarrow w''$$

$X \in NT$ $\qquad w'' \in (T \cup NT)^*$

alors la grammaire est dite hors-contexte non contextuelles (algébrique)

si toutes les règles sont de la forme

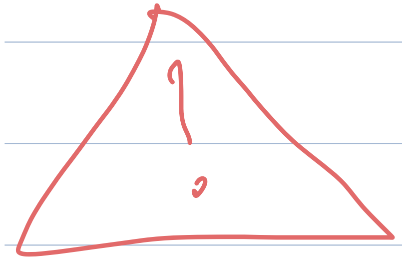
$$\text{ou} \begin{cases} X \longrightarrow aY \\ X \longrightarrow a \end{cases}$$

avec $X, Y \in NT$ $\quad a \in T$

alors la grammaire est dite

régulière (droite)

le langage est dit rationnel



$$S \to aX$$
$$S \to ab$$
$$\cancel{X} \to Sb$$

NT à droite

NT à gauche

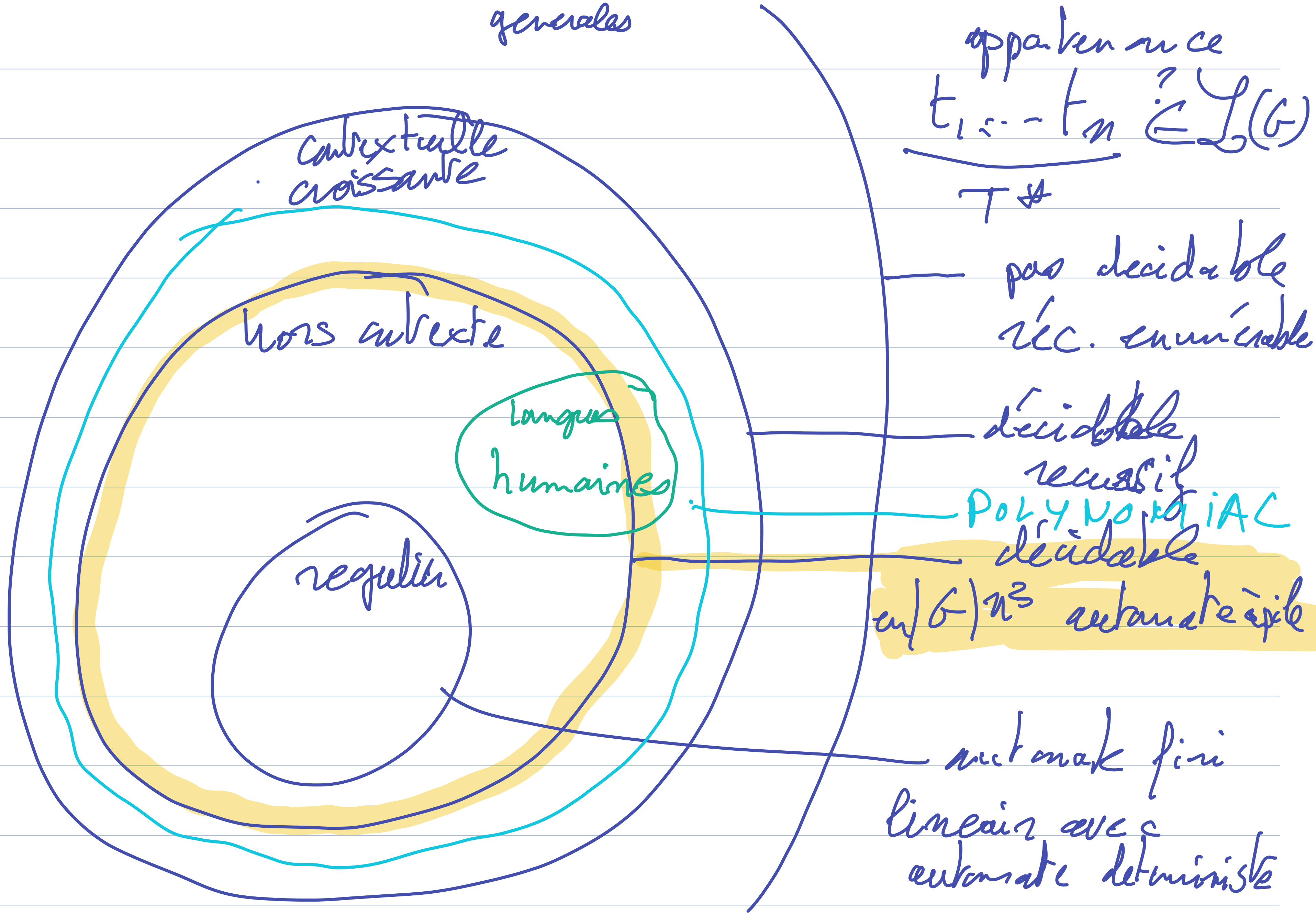
<u>pas</u> régulière

le langage

$\mathcal{L}(G) \subset T^*$ est l'ensemble des suites de terminaux produites à partir du non terminal "spécial" S

Un langage est dit de type 0, 1, 2, 3 lorsque il existe une grammaire régulière, hors contexte, contextuelle, générale

| langage | | grammaires |
| --- | --- | --- |
| régulier, rationnel | 0 | régulière |
| hors contexte, algébrique | 1 | hors contexte, non contextuelle |
| contextuels | 2 | contextuelle, croissante |
| r.e. énumérable | 3 | générale, réécriture |


generales
contextuelle croissante
hors contexte
langues humaines
regulier
appartenance
$\underbrace{t_1 \dots t_n}_{T^*} \overset{?}{\in} \mathcal{L}(G)$
pas decidable
réc. enumérable
décidable recursif
POLYNOMIAL
décidable
en $|G| n^3$ automate à pile
automate fini
linéaire avec automate déterministe

Exos Grammaires formelles
(Deug MIAS (L2) Nantes 2001)

Donnez le type de grammaire ?

si hors contexte, langage engendré ?

$T = \{a, b\}$ $NT = \{S, A, B\}$

$S \to AB$

$B \to SA$

$bB \to a$

$Ab \to SBb$

- $\underline{Aa} \to SBb$

type 3 / générale

G2) $T = \{ \_/‾ , ‾\backslash\_ \}$   $NT = \{S, A, U, V\}$

$S \to UAV \mid UV$   (2 règles)

ou

$A \to VSU \mid \underline{VU}$

$U \to \_/‾$

$V \to ‾\backslash\_$

$G_2$ : hors contexte (tjrs 1 NT à gauche $\to$)

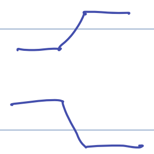
$S' \to UAV \mid UV$

ou

$A \to VSU \mid VU$

$U \to$

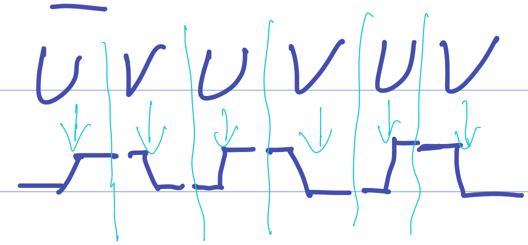
$V \to$

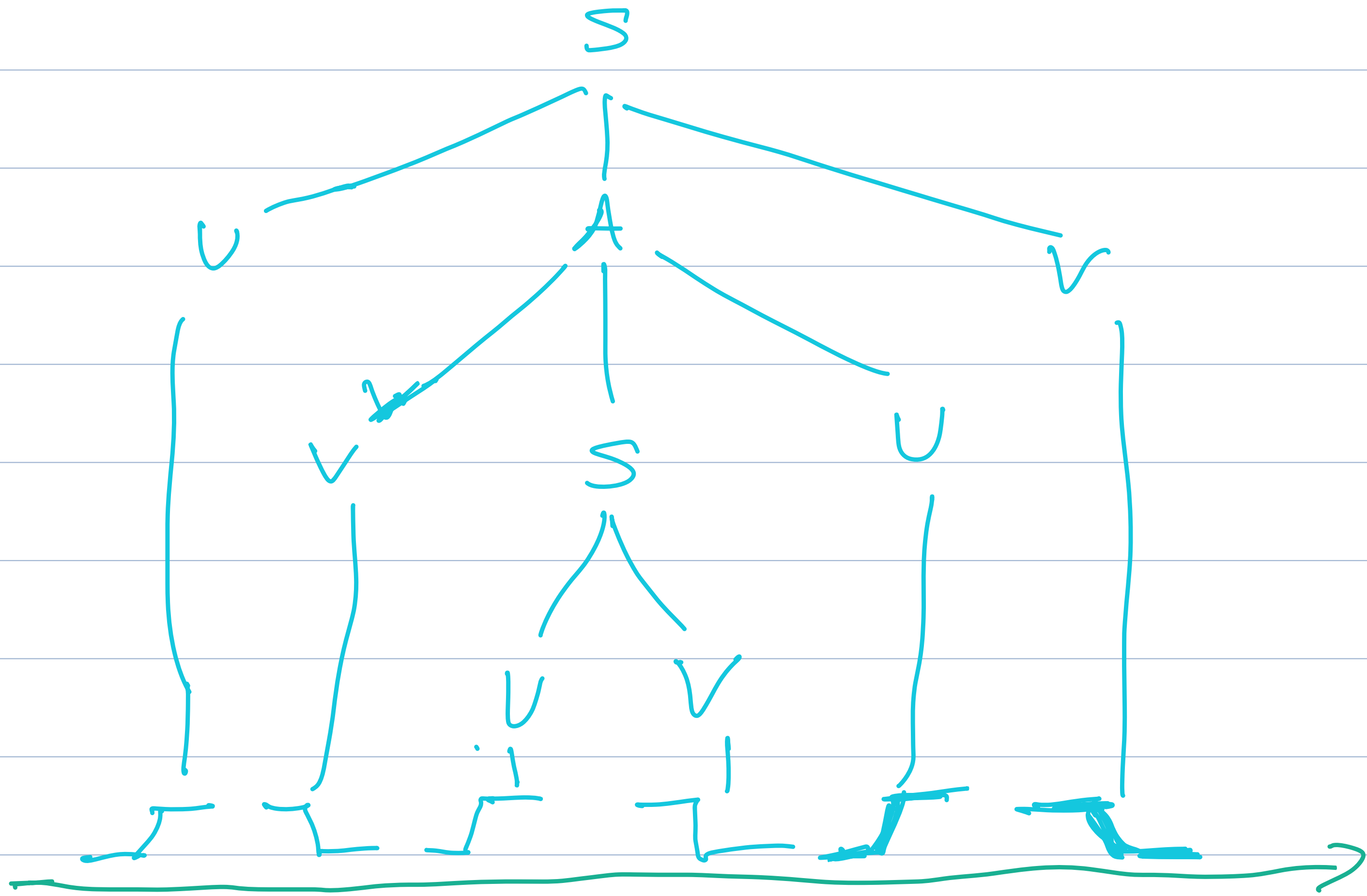


$S \to UAV \to UVSUV \to \dots$

$\dots \to UVUVUV$

$\dots \to$





Quel est le langage engendré

$$\mathcal{L}(G_2) = \{ (\ulcorner \ \urcorner \!\lrcorner)^n \ / \ n \in \mathbb{N}^* \}$$

Comment montrer que $\mathcal{L}(G_2) = \{(\ulcorner \ \urcorner \!\lrcorner)^n / n \in \mathbb{N}^*\}$

① $\mathcal{L}(G_2) \subset \{ (\ulcorner \ \urcorner \!\lrcorner)^n \ / \ n \in \mathbb{N}^* \}$

② $\{ (\ulcorner \ \urcorner \!\lrcorner)^n \ / \ n \in \mathbb{N} \} \subset \mathcal{L}(G^2)$

$S$

1 $U \; A \; V$ impair $U\,(VU)^p A\,(VU)^p V$

2 $U \; \underline{VSU} \; V$ pair $U \cdots V\, S\, U \cdots V$

3 $U \; V \; UAV \; U \; V$ impair

$E_p = 1.2 = 4$ $U \; V \; U\,VSU\,V \; U \; V$

$(UV)^2 \; S \; (UV)^2$

2pième étape $(UV)^{2p} \, S \, (UV)^{2p}$

2p+1 étape $(UV)^{2p}\, U\, A\, V\, (UV)^{2p}$

Si $\begin{cases} 2p\text{ième étape } (UV)^{2p}\, S\, (UV)^{2p} \text{ réalisable} \\ 2p+1 \text{ une étape } (UV)^{2p}\, UAV\, (UV)^{2p} \end{cases}$

$$\Big\downarrow A \to VSU$$

alors $\begin{cases} 2p+2 \text{ étape } (UV)^{2p+2}\, S\, (UV)^{2p+2} \\ 2p+3 \text{ étape } (UV)^{2p+2}\, \overset{\downarrow UAV}{UAV}\, (UV)^{2p+2} \end{cases}$

par récurrence

$G_{a^n b^n}$ $\quad S \to ab \mid aSb$

$a^n b^n \quad n > 0$

$S \to aSb \to aaSbb$

$\to aaabbb$

$\forall n > 0 \quad a^n b^n \in \mathcal{L}(G_{a^nb^n})$

pour $n = 1$ c'est vrai $S \to ab$

supposons que $S \to a^n b^n$

montrons que $S \to a^{n+1} b^{n+1}$

$$S \to aSb \xrightarrow{\text{hyp ind } S \to a^n b^n} a\,a^n b^n b$$

$$= a^{n+1} b^{n+1}$$

si $S \to w$ alors $w$ est de la forme $a^n S b^n$ ou $a^p b^p$

par induction sur ~~S→w~~ le nbre d'étape de $S \to w$

montrons qu'en $N$ étapes $S$ donne $a^p S b^p$ $a^n b^n$

en supposant que pour $p < N$ $S$ ne donne que $a^p S b^p$ ou des $a^n b^n$

si j'ai une derivation de W en N étapes

$$S \xrightarrow{N-1 \text{ étapes}} W' \xrightarrow{N^{\text{e}} \text{ étape}} W$$

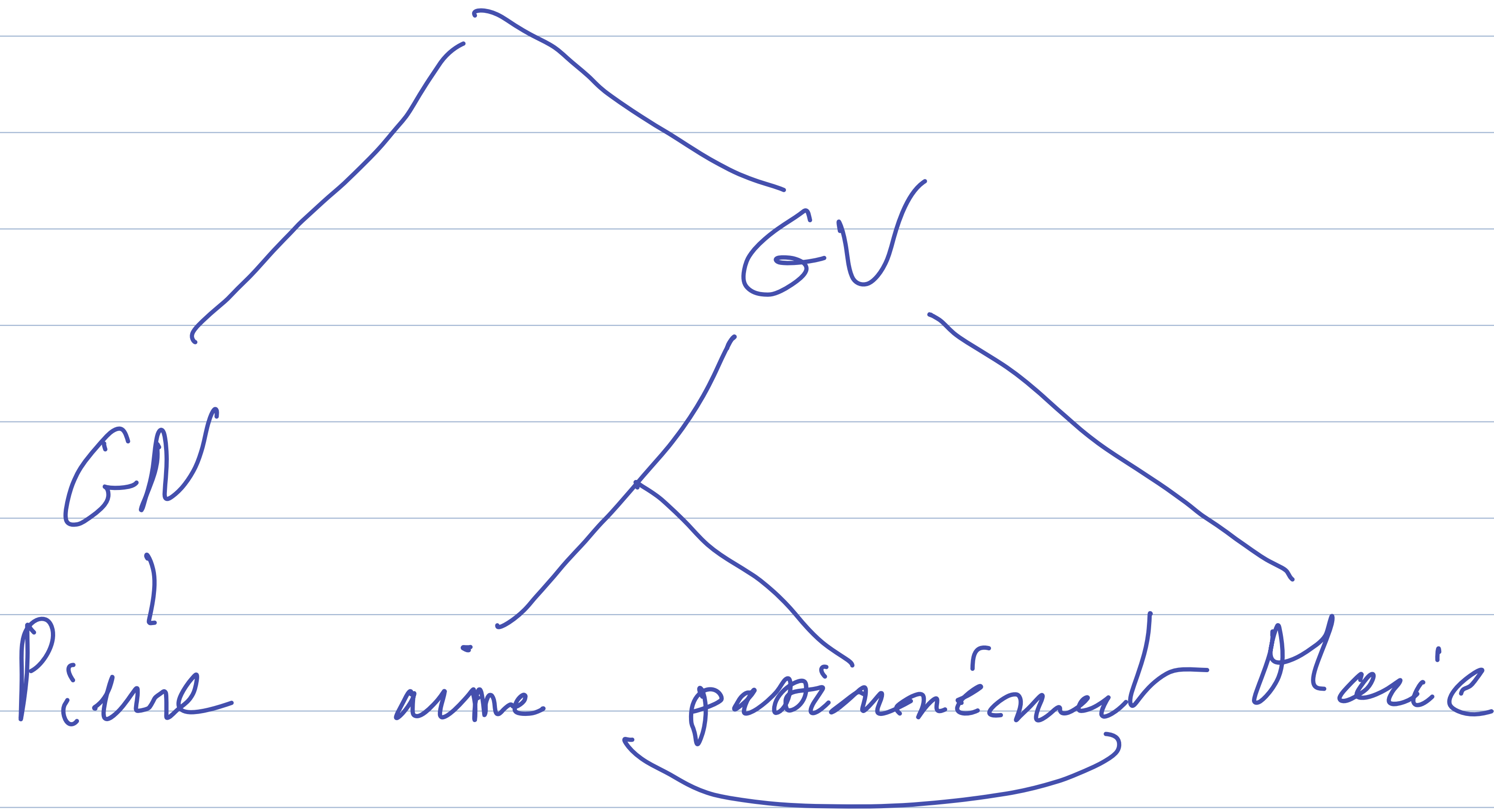
par hypothèse de recurrence

$W'$ est de la forme $a^p S b^p$ ~~ou $a^k b^k$~~

le 2ᵉ cas est exclu

donc $W'$ est $a^p S b^p$ et $W$ est $a^{p+1} S b^{p+1}$ et $a^{p+1} b^{p+1}$ tout est ok

$S \to GN \; GV$



Adv portant sur toute la phrase et non juste sur le GV sont les adverbes exprimant le sentiment du locuteur.

Malheureusement, l'aspirateur est jaune.

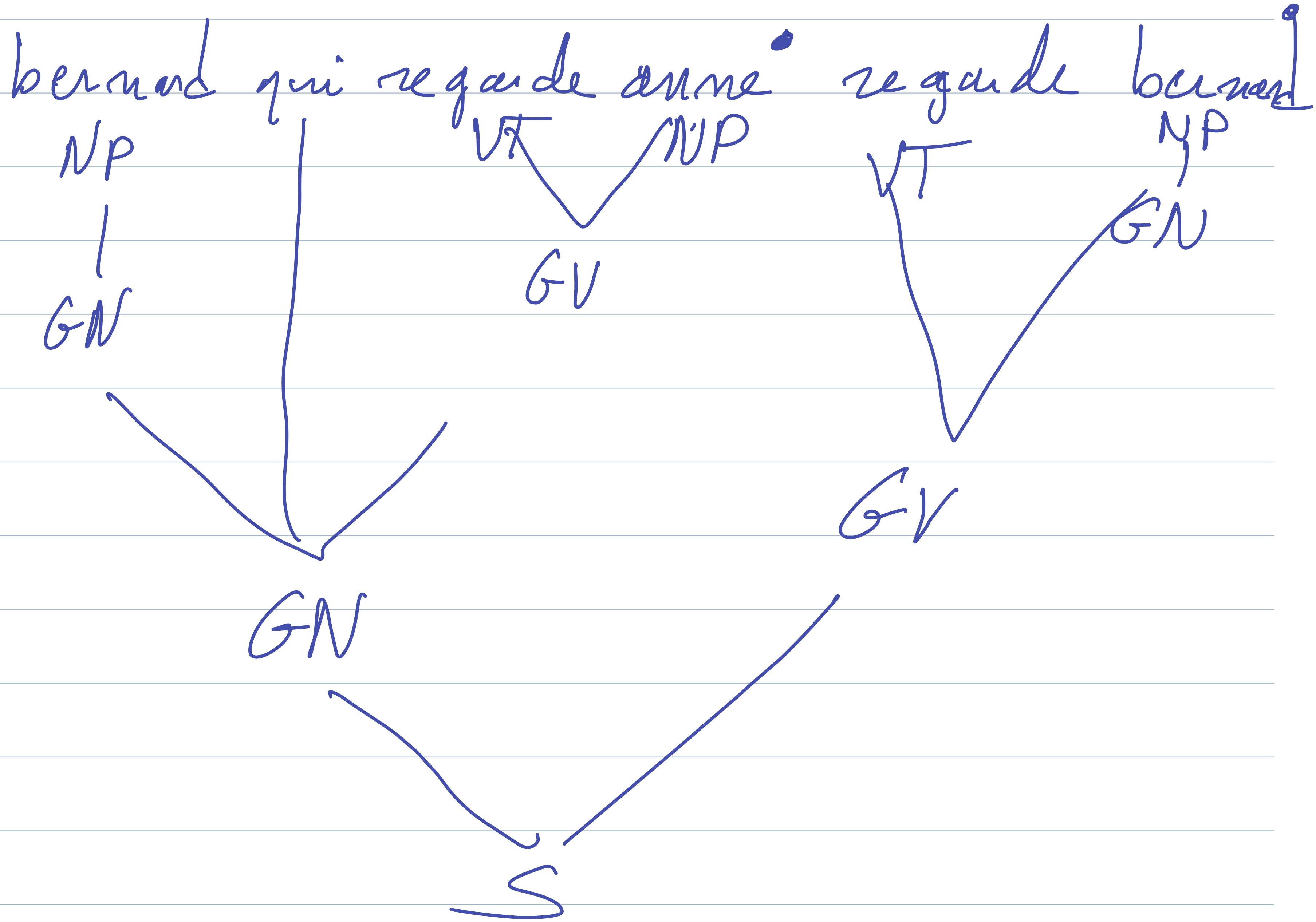
S $\to$ GN GV

GN $\to$ NP | NP qui GV

GV $\to$ VT NP

NP $\to$ anne, bernard

VT $\to$ regarde


bernard qui regarde anne regarde bernard
NP
VT
NP
VT
NP
GN
GV
GN
GN
GV
S

$NP(\text{qui } VT\ NP)^*$ $\quad VT\ (\text{qui } VT\ NP)^*$

→ répliquer

Automates analyse des mots

chanterons ⟶ chanter
1 p pl
futur

ils chanteront quelques succès des années 70

3e pers pl
pl chanter 1p pl futur

≠

grammaire hors contexte

algo de ~~pa~~ parsing

extension pour avoir l'accord

* il dorment → pb

DCG

50%

TP1 20% TP2 30% Ex

règle du max

DCG

HMM

Parsing