

# Korection deux frases incorecte an frases corecte

Hurot Eliott - 28537

2025

Comment peut-on écrire un programme  
qui corrige des phrases en français ?

## 1 Objectif

## 2 Présentation de l'algorithme

- Stockage des données
- Analyse lexicale
- Analyse syntaxique
  - Grammaire
- Vérification && Correction

## 3 Gestion des résultats

## 4 Annexe

---

## **Algorithm 1** Correction de phrases

---

**Require:** Phrase à corriger - string

**Ensure:** Phrase corrigée - string

---

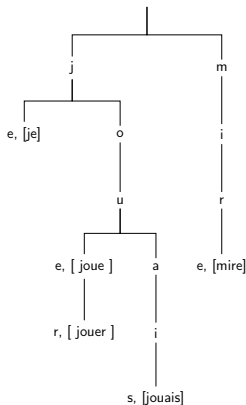
# Étapes :

- 1 Analyse lexicale (Lexing)
- 2 Analyse syntaxique (Parsing)
- 3 Vérification de la phrase
- 4 Correction de la phrase

# Stockage des données - Trie

Structure de donnée qui prend avantage de la similitude entre les clés

je - joue - jouer - mire - jouais



Taille : 983809    Hauteur : 39

---

## Algorithm 2 Analyse lexicale

---

**Require:** Phrase à corriger - string

**Ensure:** Liste de tokens - token list list

---

Token : Classe grammaticale, valeur, informations

Pourquoi token list list ?

Plusieurs sens possibles pour un même mot

# Exemple

le petit chat rouge joue

[[D : le, Ov : le], [A : petit, N : petit], [N : chat], [A : rouge, N : rouge], [V : joue, V : joue]]

Complexité ?

$O(n * s)$

n : nombre de mots dans la phrase

s : taille du mot le plus long de la phrase



---

## Algorithm 3 Analyse syntaxique

---

**Require:** Liste de tokens - token list list

**Ensure:** Arbre syntaxique - syntax\_tree list

---

LR parser, LL parser

# Définition

Quadruplet  $(V, T, \Sigma, S)$

- $V$  Ensemble de symboles non terminaux
- $T$  Ensemble de symboles terminaux
- $R$  Ensemble de règles de production de la forme :

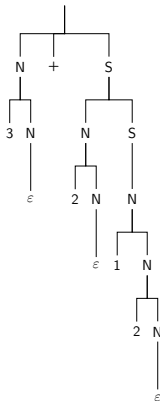
$$X \longrightarrow \alpha, \quad X \in V, \quad \alpha \in (V \cup T)$$

- $S$  Symbole de départ

# Exemple

3 + 2 + 12

$S \rightarrow N + S \mid N$   
 $N \rightarrow [0-9] N \mid \epsilon$



# Grammaire du français (restrictive)

$S$	$\longrightarrow$	$S_u V_g$	$V_g$	$\longrightarrow$	$P V C$
$S_u$	$\longrightarrow$	$N_g   P_{ps}$	$P$	$\longrightarrow$	Pronom
$N_g$	$\longrightarrow$	$D A N A$	$V$	$\longrightarrow$	verbe   $A_V$
$P_{ps}$	$\longrightarrow$	Pronom personnel sujet	$A_V$	$\longrightarrow$	Auxiliaire Verbe
$D$	$\longrightarrow$	Déterminant	$C$	$\longrightarrow$	$C_{OD}   C_{OI}$
$A$	$\longrightarrow$	Adjectif $A   \varepsilon$	$C_{OD}$	$\longrightarrow$	$N_g   V$
$N$	$\longrightarrow$	Nom	$C_{OI}$	$\longrightarrow$	$P_r N_g$
			$P_r$	$\longrightarrow$	Pronom

---

## Algorithm 4 Vérification & Correction

---

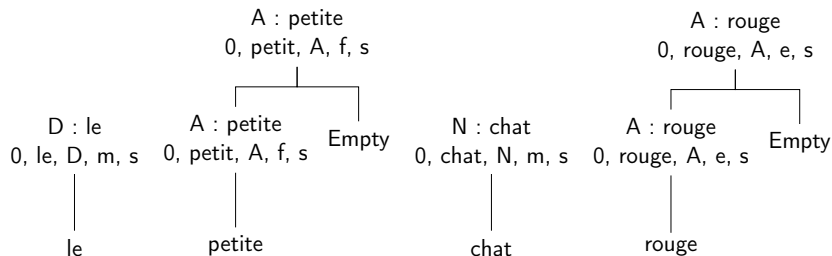
**Require:** Arbre syntaxique d'un élément - `syntax_tree list`

**Ensure:** Arbre syntaxique de l'élément corrigé - `syntax_tree list`

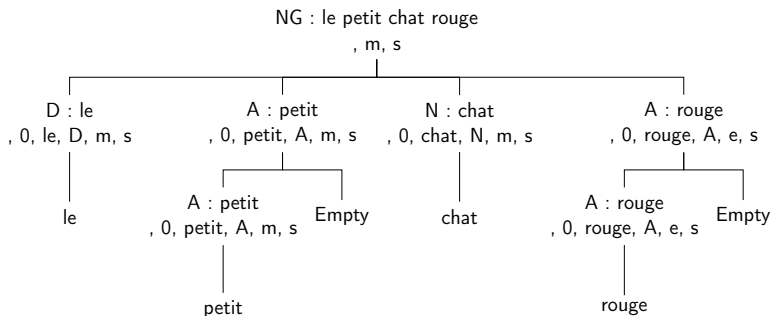
---

2 types d'erreurs : Erreur syntaxique et Fautes de frappe

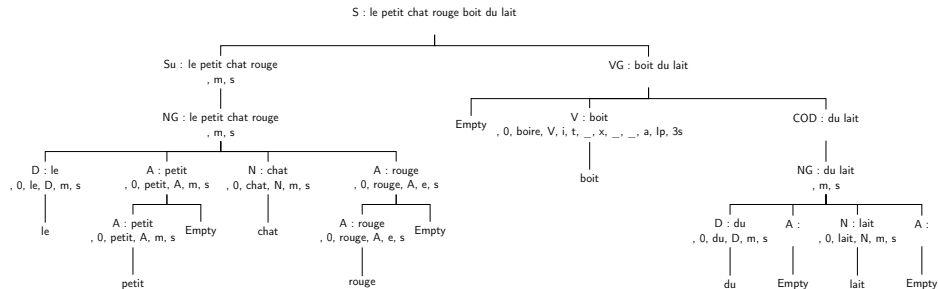
Exemple : Le petite chat rouge boit du lait



# Vérification & Correction



# Vérification & Correction



Problème : Trop de corrections possible

Exemple :

la petit chat roug bois du lait  $\longrightarrow$  28 corrections possibles



# Distance de Levenshtein

Distance entre deux chaînes de caractères

$$\text{lev}(a, b) = \begin{cases} \max(|a|, |b|) & \text{si } \min(|a|, |b|) = 0 \\ \text{lev}(a-1, b-1) & \text{si } a[0] = b[0] \\ 1 + \min \begin{cases} \text{lev}(a-1, b) \\ \text{lev}(a, b-1) \\ \text{lev}(a-1, b-1) \end{cases} & \text{sinon} \end{cases}$$

Complexité :  $O((n+1)*(m+1))$

Autres possibilités :

Nombre de correction

# Fréquence d'utilisation

Ajout d'une fréquence d'utilisation des mots au dictionnaire  
Dictionnaire adapté à l'utilisateur

Résultats :

Le jeune grçon rames

Distance de levenshtein :	le jeune garçon rame
Fréquence Eschyle :	le jeune garçon ramerait
Fréquence Usuelle :	le jeune garçon rame
Reverso :	le jeune garçon ramène

Reverso : «Le correcteur français le plus précis au monde »

~~Mairci pour vottre attanssion~~

Merci pour votre attention

# Dictionnaire

0,de,de,D,e,i

0,du,du,D,m,s

0,la,le,D,f,s

0,un,un,D,m,s

0,sont,être,V,i,\_,\_,\_,\_,\_,a,lp,3p

0,être,être,V,i,\_,\_,\_,\_,\_,a,Y

0,ont,avoir,V,i,t,\_,\_,\_,\_,a,lp,3p

0,été,être,V,i,\_,\_,\_,\_,\_,a,Q,e,i,

0,la,la,N,m,i

0,est,est,N,m,s

0,a,a,N,m,i