

option LTAL

Linguistique et Traitements Automatiques des Langues

TP 4

Génération aléatoire de phrases d'après une grammaire générative

Analyse fonctionnelle :

Le programme à réaliser génère aléatoirement des phrases à partir d'une grammaire formelle (placée hors du générateur dans un fichier texte).

Vous êtes dans le **rôle du locuteur natif** qui juge si la phrase sortie fait partie de la langue; mais vous êtes aussi dans le **rôle du linguiste** qui recherche les règles syntagmatiques qui ne génèrent que des phrases acceptables par le locuteur natif, et qui pourraient générer toutes les phrases du français.

Chaque règle comporte un choix aléatoire: soit plusieurs structures possibles pour chaque élément du vocabulaire non terminal, soit plusieurs mots pour le vocabulaire terminal.

À chaque exécution, le programme sort la grammaire formelle utilisée et 5 phrases générées, avec, pour chaque phrase, les règles appliquées.

Grammaire de départ et exemple de formalisme :

(vous pouvez créer votre propre formalisme)

membres Droits = vocabulaire non terminal (les nœuds sauf les feuilles) :

P --> SN . | SN SV . (| = ou)

SN --> DET NOM | DET NOM ADJ

SV --> Vi | Vt SN

membres Droits = vocabulaire terminal (les feuilles) :

DET --> le | la | les | un | une | ce | ces | sa (mettez au moins 10 mots
pour obtenir des phrases variées)

NOM --> chat | renard | belette | souris | poule

ADJ --> rapide | fauve | rouges | lent | grise

Vi --> dormait | s'ennuyait | craquera

Vt --> poursuit | dérangea | adorait | a vu

exemple de phrase générée avec la grammaire ci-dessus, avec les règles appliquées:

P --> SN SV .

SN --> DET NOM ADJ

DET --> le

NOM --> renard

ADJ --> fauve

SV --> Vt SN

Vt --> poursuivait

SN --> DET NOM

DET --> une

NOM --> belette

le renard fauve poursuivait une belette .

Algorithme :

Deux algorithmes possibles (l'algorithme 2 est plus rapide à mettre en œuvre) :

1) en partant de la racine (P), parcourir l'arbre syntagmatique en profondeur d'abord :

à chaque nœud (membre G), si c'est une feuille, la sortir,
sinon choisir un des membres D au hasard
pour chaque constituant du membre D choisi
continuer le parcours

2) en partant de la racine (P), remplacer un membre G par un des membres D possibles choisis au hasard,
et recommencer tant que c'est possible

structure de données pour les règles : les membres Droits possibles d'un membre Gauche donné au moyen
d'un dictionnaire de clé "un membre Gauche" et de donnée : liste des membres Droits possibles (1 ou plus).

Préparation linguistique :

méthode: toujours construire ressources et algorithmes sur une observation des textes traités ou générés;
pour trouver vos règles, observez des textes, analysez des phrases en termes de syntagmes nominaux, verbaux,
ou prépositionnels (éventuellement récursifs);

Construction de la grammaire :

tentez de générer des phrases de structures analogues aux structures des phrases observées :
résoudre (partiellement) le problème des accords genre nombre (puis le contourner en ne mettant que des mots
du même genre et même nombre)
introduire des compléments du nom, du verbe
introduire l'adverbe de verbe, l'adverbe d'adjectif
rendez une règle récursive: SN --> DET NOM | DET NOM ADJ | DET NOM prép. SN
questions de surface (amalgames, élisions, majuscules de début de phrase) : vous pouvez les traiter par des
remplacements en fin de génération.

Travail à remettre :

Travail en binôme, à remettre en 1 exemplaire par binôme, au plus tard le **20 ou le 22 octobre** selon votre
groupe, en début du TP

Chaque binôme envoie un mail à <Julien.Gosme@info.unicaen.fr> avec le fichier du rapport, le programme
réalisé, les fichiers placés en entrée et ceux obtenus en sortie.

Le rapport devra comporter :

- une courte introduction présentant le sujet du TP;
- première partie : compte-rendu de votre observation des textes, et son résultat en termes de grammaires
formelles observées;
- une seconde partie décrivant la conception et l'implantation de votre générateur;
- troisième partie : vos grammaires génératives - au moins 2 versions de grammaire, l'une sans, l'autre avec
règle(s) récursive(s) - et vos observations sur vos résultats; entre autres, comparez les structures des phrases
générées avec les structures des phrases observées, et décrivez l'impact des règles récursives en génération;
- une conclusion rappelant l'ensemble des problèmes rencontrés et les solutions trouvées pour les résoudre;
- une annexe avec les pages de **résultats commentés** (et uniquement cela) : phrases générées, avec la
grammaire formelle pour chaque sortie;

(nombre de pages limité à 5, annexes non comprises; soin particulier pour la rédaction et l'orthographe)

•