

Génération aléatoire de phrases d'après une grammaire générative

Introduction

On a cherché à écrire une grammaire générative capable de générer des phrases françaises syntaxiquement correctes. On a réalisé un ensemble de scripts en python3 capable d'exploiter cette grammaire pour produire des phrases aléatoires.

Vue d'ensemble des scripts

`generate_sentence.py` est le point d'entrée, il gère les arguments de la ligne de commande et la lecture des fichiers.

`sentence_generator.py` est le générateur.

`argparse.py` est le module utilisé pour traiter les arguments de la ligne de commande, on l'inclut dans l'archive car il n'est pas intégré à python3.1.

Générateur

Les fonctionnalités du générateur sont limitées : toutes les possibilités sont équiprobables, les répétitions ne sont pas détectées, etc.

Système de variables

Un système de variables est implémenté afin de permettre un minimum de factorisation, dans notre cas utilisé pour les accords en genre et en nombre.

Les deux points sont utilisés pour la transmission de variable. Par exemple `SN:G,V` tire au sort un genre et un nombre si ce n'est pas déjà fait (les possibilités sont définies dans le même fichier) puis transmet les valeurs à une des formes de `SN`.

Le dollar est utilisé pour le remplacement d'une variable par sa valeur (une valeur est choisie aléatoirement si la variable n'en a pas encore). Par exemple `NOM_$G_$N` peut être remplacé par `NOM_F_P`, c'est à dire un nom féminin pluriel.

Valeurs spéciales

Deux valeurs spéciales sont utilisées par le générateur : `_Sentence` est le symbole de départ des non-terminaux ; `_Replacements` est une liste de remplacements à effectuer après la génération (par exemple « de les » devient « des »).

Le reste des symboles est totalement indépendant du générateur, y compris les noms et valeurs des variables.

Grammaire

La grammaire générative créée est elle aussi limitée. Les structures de phrase ne contiennent qu'un seul verbe, donc pas de concordance des temps à gérer. Les pronoms personnels ne sont pas utilisés, tout est à la troisième personne afin de s'épargner la conjugaison, la distinction singulier/pluriel est par contre prise en compte. Des adjectifs et des adverbes sont incorporés.

Trois formes de phrase sont intégrées. La première est la forme affirmative classique composée d'un

syntagme nominal et d'un verbal. Les deux autres sont des formes interrogatives: celles commençants par un pronom interrogatif suivi d'un syntagme nominal comme « Est-ce que... ? »; et celles où le PI est suivi du nom comme « Quel-le-s... ? ».

Notre grammaire contient une règle récursive, celle du syntagme nominal qui peut s'appeler lui-même via le complément du nom. On note que sans détection des répétitions ni contrôle de profondeur, la récursivité crée régulièrement des phrases trop longues et/ou répétitives.

Conclusion

Il est possible de générer aléatoirement des phrases simples et syntaxiquement correctes, mais la difficulté augmente rapidement si l'on essaye de créer une grammaire capable de générer n'importe quelle phrase d'une langue.

Annexe

Exemple de phrase affirmative simple

_Sentence → SN:G,N SV:G,N .
SN → DET_\$G_\$N NOM_\$G_\$N ADJ_\$G_\$N
DET_M_P → ces
NOM_M_P → tigres
ADJ_M_P → gris
SV → Vi_\$G_\$N
Vi_M_P → craqueront
Ces tigres gris craqueront.

Exemple de phrase interrogative simple

Sentence → Quel\$G_\$N NOM_\$G_\$N SV:G,N ?
Quel_M_P → quels
NOM_M_P → éléphants
SV → Vt_\$G_\$N CV
Vt_M_P → dérangèrent
CV → SN
SN → DET_\$G_\$N NOM_\$G_\$N
DET_M_P → les
NOM_M_P → vers de terre
Quels éléphants dérangèrent les vers de terre ?

Exemple de récursion excessive

_Sentence → PI SN:G,N SV:G,N ?
PI → est-ce que
SN → DET_\$G_\$N NOM_\$G_\$N
DET_F_S → cette
NOM_F_S → poule
SV → Vt_\$G_\$N ADV CV
Vt_F_S → poursuit
ADV → férocement
CV → SN
SN → DET_\$G_\$N NOM_\$G_\$N ADJ_\$G_\$N CN
DET_M_P → ces
NOM_M_P → éléphants
ADJ_M_P → lents
CN → Préposition_CN SN
Préposition_CN → de
SN → DET_\$G_\$N NOM_\$G_\$N ADJ_\$G_\$N CN
DET_F_P → les
NOM_F_P → girafes
ADJ_F_P → lentes
CN → Préposition_CN SN
Préposition_CN → de
SN → DET_\$G_\$N NOM_\$G_\$N ADJ_\$G_\$N CN
DET_F_P → ces
NOM_F_P → girafes
ADJ_F_P → rapides
CN → Préposition_CN SN
Préposition_CN → de
SN → DET_\$G_\$N NOM_\$G_\$N ADJ_\$G_\$N CN
DET_M_P → les
NOM_M_P → vers de terre
ADJ_M_P → rouges
CN → Préposition_CN SN
Préposition_CN → de
SN → DET_\$G_\$N NOM_\$G_\$N
DET_F_P → ces
NOM_F_P → chèvres

Est-ce que cette poule poursuit férocement ces éléphants lents des girafes lentes de ces girafes rapides des vers de terre rouges de ces chèvres ?