

Synthèse de texte avec le logiciel Syntox

Lionel Clément

Université Bordeaux 1 – LaBRI (UMR 5800) – CLEE-ERSS (UMR 5263)

lionel.clement@labri.fr

RÉSUMÉ

Le logiciel **Syntox**, dont une interface utilisateur en ligne se trouve à cette *URL* : <http://www.syntox.net>, est une mise en application d'un modèle basé sur les grammaires attribuées, dans le cadre de la synthèse de texte. L'outil est une plateforme d'expérimentation dont l'ergonomie est simple. **Syntox** est capable de traiter des lexiques et des grammaires volumineux sur des textes ambigus à partir de la description explicite de phénomènes linguistiques.

ABSTRACT

Automated generation of text with Syntox

Syntox, which includes an online user interface at URL <http://www.syntox.net>, is an implementation of a model based on attribute grammars, in the context of automated generation of text. The software is intended as a platform for experimentation with an ergonomic interface. **Syntox** is usable with large vocabularies and grammars to produce ambiguous texts from an explicit description of linguistic phenomena.

MOTS-CLÉS : Synthèse de texte, Grammaire attribuée, Syntaxe.

KEYWORDS: Text generation, Attribute Grammars, Syntax.

Introduction

Le programme Syntox demande une grammaire, un lexique (ou une connexion à un serveur qui délivre ce lexique), et une structure de traits censée correspondre à une représentation de la syntaxe d'un texte. Le logiciel produit une forêt partagée. L'ensemble des textes engendrés par cette forêt est ensuite affiché à l'écran avec des éléments permettant la mise au point des grammaires (forêts partielles, messages d'erreurs, étapes de synthèse).

Synthèse de texte à l'aide de grammaires attribuées

Contrairement à ce qui se fait dans les modèles habituels «d'unification», comme PATR-II (Shieber, 1984), il n'est jamais question de marquer une contrainte d'égalité $A = B$ entre deux éléments dont l'un entre dans la structure compositionnelle de l'autre. Nous rejetons systématiquement cette approche qui pose un problème de complexité algorithmique (dont l'indécidabilité dans le cas général) pour l'analyse ou la synthèse lorsque A et B sont dans des rapports pluri-univoques (homonymie, polysémie, synonymie). A cette approche qui se résume à résoudre par unification

un système d'équations et qui donne son nom aux grammaires «d'unification», nous opposons un modèle où A et B sont dans une relation de dépendance, plus précisément de copie.

Le modèle utilisé par Syntox est celui, bien connu et assez ancien des grammaires attribuées (Knuth, 1968). En conformité avec celui-ci, Syntox impose la non circularité des calculs sur les attributs pour la raison évoquée immédiatement. Ceci sera essentiel pour garantir la décidabilité des algorithmes qui mettent en correspondance les textes et les représentations profondes de ces textes (ici la synthèse de texte à partir d'une représentation de la syntaxe).

Pour nous approcher des formalismes utilisés en linguistique, notamment des grammaires d'unification, les structures de traits, censés contenir les représentations profondes des textes, ont été choisies comme attributs de la grammaire. L'analyse syntagmatique est supportée par la grammaire hors-contexte sous-jacente de la grammaire attribuée.

La grammaire décline pour chaque syntagme, deux attributs, l'un synthétisé, que je note \uparrow , et l'autre hérité que je note \uparrow . Le premier est calculé en fonction des attributs des termes qui composent le syntagme. Le second est calculé en fonction des attributs du terme qui le contient. D'évidence les propriétés lexicales seront synthétisées, y compris celles qui ne sont pas locales, les propriétés globales¹ seront héritées.

Écrire une grammaire jouet du français

Grâce à quelques exemples simples, on montre l'intérêt de manipuler les deux types d'attributs (synthétisé et hérités). Les phénomènes lexicaux peuvent être complexes. Je montre comment faire avec Syntox pour qu'un prédicat nominal s'articule avec un verbe support ou encore comment un prédicat impose des constantes dans le lemme ou les catégories grammaticales de ses compléments.

Sur l'exemple des adjectifs épithètes, je montre comment utiliser les listes d'ajouts et comment récupérer les attributs synthétisés de façon récursive pour traiter l'accord.

Au delà de la phrase En conclusion, je propose quelques exemples aux limites de l'outil avec des textes qui mettent en oeuvre une ou plusieurs propositions. En aucune manière, Syntox ne permet de traiter la résolution des anaphores ou l'ellipse. Il conviendra alors de savoir articuler Syntox dans un système de génération de texte pour ces problèmes en particulier.

Références

- KNUTH, D. E. (1968). Semantics of context-free languages. *Mathematical Systems Theory*, 2(2):127–145.
- SHIEBER, S. M. (1984). The design of a computer language for linguistic information. In *Proceedings of the Tenth International Conference on Computational Linguistics*, pages 362–366, Stanford University, Stanford, California.

1. Par propriétés globales, j'entends par exemple celles qui ne sont pas sujettes à la variation.