

FipsColor : grammaire en couleur interactive pour l'apprentissage du français

Jean-Philippe Goldman, Kamel Nebhi, Christopher Laenzlinger

(1) LATL - Laboratoire d'Analyse et de Technologie du Langage,
Département de Linguistique, Faculté des Lettres,
Université de Genève

Jean-Philippe.Goldman@unige.ch,
Kamel.Nebhi@unige.ch,
Christopher.Laenzlinger@unige.ch

Résumé L'analyseur multilingue FiPS permet de transformer une phrase en une structure syntaxique riche et accompagnée d'informations lexicales, grammaticales et thématiques. On décrit ici une application qui adapte les structures en constituants de l'analyseur FiPS à une nomenclature grammaticale permettant la représentation en couleur. Cette application interactive est disponible en ligne (<http://latl.unige.ch/fipscolor>) peut être utilisée librement par les enseignants et élèves de primaire.

Abstract The FiPS parser analyzes a sentence into a syntactic structure reflecting lexical, grammatical and thematic information. The present paper describes the adaptation of the structures in terms of constituents as existent in FiPS to a grammatical annotation, as well as its coloured representation. This online interactive application (available at <http://latl.unige.ch/fipscolor>) can be freely used by teachers and pupils of primary education.

Mots-clés : analyse syntaxique, grammaire générative, services web, tei

Keywords: chart parser, generative grammar, web services, tei

1 Motivation

Dans cet article, nous présentons une application web directement destinée au monde de l'enseignement et basée sur l'analyseur syntaxique FiPS. FipsColor reprend les principes pédagogiques de la grammaire en couleur (Plan d'études et « Maîtrise du français » - Corome Suisse Romande) selon lesquels toute catégorie grammaticale et fonction de constituant sera représentée par une couleur. L'application va permettre à l'apprenant de différencier les catégories grammaticales de mots (représentées par des mots coloriés) d'une part, et les fonctions syntaxiques (représentées par des groupes de mots soulignés), d'autre part. Le système de couleur est un moyen mnémotechnique efficace qui permet également de mettre l'accent sur certains points problématiques de la langue, comme par exemple les ambiguïtés lexicales. La phrase suivante illustre ce cas : « Paul ferme la porte et déclare haut et ferme qu'il va gérer la ferme. »

Dans cet exemple, « ferme » est verbe, adverbe puis nom commun. Les différentes colorations permettent à l'élève de distinguer clairement les catégories grammaticales de mots. Outre les ambiguïtés de la langue, le sous-lignage des fonctions syntaxiques a pour but de mieux présenter les règles d'accord et d'appréhender au mieux les structures les plus complexes (proposition relative, proposition conjonctive, etc.). Dans ce sens, FipsColor est un support didactique intéressant en milieu scolaire.

2 L'analyseur syntaxique FiPS

L'analyseur syntaxique FiPS (Laenzlinger et Wehrli 1991, Wehrli 1997), développé depuis plusieurs années au LATL, est un outil linguistique capable d'associer à chaque phrase d'un texte une structure syntaxique accompagnée d'informations lexicales, grammaticales et sémantiques (« thématiques »)¹. Les applications de l'analyseur sont multiples: traduction automatique (ou aide à la traduction) (Wehrli 2003), synthèse et reconnaissance de la parole (Gaudinat et al. 1999 et Goldman et al. 2001), indexation et recherche 'intelligente' d'informations, extraction terminologique (Seretan 2008) et apprentissage des langues (L'Haire & Vandeventer-Faltin 2003). FiPS a été développé sur la base de la théorie *Principes & Paramètres* de la Grammaire Générative (Chomsky 1995 : chap. 1, Haegeman 1994, Laenzlinger 2003). La structure en constituants assignée aux phrases repose sur un schéma X-barre réduit à deux niveaux: [XP L X R]. XP est une projection maximale de la tête X, alors que L (Spécificateurs) et R (Compléments) sont des listes (éventuellement vides) de projections maximales correspondant respectivement aux sous-constituants gauches et droits de la tête X. X est une variable correspondant aux catégories Adv (adverbe), A (adjectif), N (nom), D (déterminant), V (verbe), P (préposition), C (conjonction), T (temps). Une phrase complète a donc la structure suivante :²

[TP [AdvP Hier] [DP le [NP garçon]] a [VP recueilli [DP un [NP [AP petit] chat [AP noir] [AP affamé]]]]]

La **stratégie d'analyse** est de type gauche à droite avec traitement parallèle des alternatives, combinant une approche incrémentale, essentiellement ascendante avec un filtre descendant. Selon cette stratégie dite du 'coin droit', l'algorithme est dirigé par les données (data-driven), c'est-à-dire on cherche à attacher un nouvel élément au coin droit d'un constituant dans le contexte gauche déjà existant. Ce dernier spécifie un ensemble de nœuds actifs auxquels le nouvel élément est susceptible de s'attacher. Trois mécanismes fondamentaux sont utilisés par l'analyseur : la projection, la combinaison et le déplacement.

L'**implémentation objet** de FiPS consiste à concevoir les objets linguistiques, tels que les structures lexicales et les projections syntaxiques, comme des structures abstraites dont l'implémentation peut varier

¹ Par contraste, les analyseurs 'superficiels' ne cherchent pas à construire une représentation globale, ni a fortiori une forme logique, mais restent à un niveau de représentation morpho-syntaxique, avec un regroupement des constituants minimaux (groupes nominaux, groupes prépositionnels, etc.).

² Le groupe nominal est analysé comme un syntagme déterminant (DP) contenant un syntagme nominal (NP).

d'une langue à l'autre. Ces variations sont traitées par l'extension de type en ce qui concerne les structures de données et par la redéfinition des méthodes pour ce qui est des processus de traitement de ces données. Le niveau le plus abstrait dans la hiérarchie des objets décrit les propriétés fondamentales qui sont vérifiées dans toutes les langues. Ceci s'apparente d'une certaine manière au concept chomskyen de 'grammaire universelle'. Cette approche syntaxique formelle et les avantages de l'implémentation objet permettent un temps de traitement rapide (environ un million de mots par heure), une souplesse et une facilité de développement.

3 Adaptation des étiquettes syntagmatiques et choix d'une sortie

Parmi les deux niveaux de représentations considérés : le niveau lexical concerne la **catégorie grammaticale des mots** (ou part-of-speech): *Nom Verbe, Adjectif, Préposition, Déterminant, Adverbe, Conjonction, Pronom*. La correspondance entre les étiquettes de FiPS et celles de la sortie FipsColor est univoque. Il est à signaler que la richesse de l'analyse permet d'associer à chaque mot des informations supplémentaires, par exemple le genre et le nombre pour les noms et adjectifs, ou encore le temps, la personne et le nombre pour les verbes. Le niveau syntagmatique concerne les groupes ou syntagmes. La catégorie de leur tête détermine généralement leur type : groupe prépositionnel, nominal, adjectival, adverbial ou verbal. Mais c'est la **fonction des constituants** qui est mis en valeur dans FipsColor. Seuls les constituants et fonctions majeurs de la phrase sont représentés. Sujet, Complément d'Objet Direct et Indirect, Attribut du Sujet, Complément Circonstanciel, et Prédicat. La figure ci-dessous résume les correspondances possibles entre constituants et fonctions et leurs équivalences dans l'analyse FiPS.

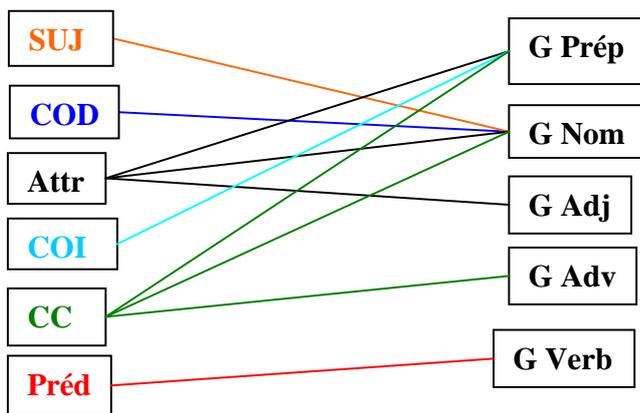


Figure 1 : Correspondances constituants/fonctions

Sujet	DP spécifieur de TP
Prédicat	Tête de VP Tête de TP
Cod	DP complément de VP ³
Coi	PP complément de VP
Attribut	DP PP AP complément de VP
Cc	AdvP spécifieur de TP VP CompVP PP spécifieur de TP CompVP DP spécifieur de TP CompVP

Tableau 1 : Correspondances FipsColor/Fips

[TP [AdvP Hier] [DP le [NP garçon]] a [VP recueilli [DP un [NP [AP petit] chat [AP noir] [AP affamé]]]]]

CC
Sujet
Prédicat
Cod

En sortie de l'analyseur, les structures syntaxiques arborescentes sont représentées au format XML avec la recommandation TEI (« Text Encoding Initiative »). La TEI permet de construire un schéma personnalisé à l'aide de différents modules dont l'utilisation forme une DTD particulière. Notre formalisme utilise la version P5 de la TEI et plus particulièrement le module consacré à l'analyse linguistique (« Simple Analytic Mechanisms »).

³ Certaines structures permettent le déplacement du COD. Les phrases (ou TP) sont, pour l'instant, exclues de la fonction COD.

4 Création du service web et mise en place de la web application

Notre sortie XML est produite par un module spécifique agissant tel un service web. L'application consomme ce service et transforme le résultat XML en HTML associé à diverses technologies web pour créer les interactions nécessaires. La figure ci-dessous est une vue de FipsColor : l'analyse montre que le pronom relatif « que » est bien interprété comme étant Cod du verbe « aimer ». Dans le second exemple, les pronoms clitiques sont aussi correctement analysés : « le » est bien Cod et « lui » est Coi. FipsColor rend compte des propositions enchâssées en utilisant un soulignage multiple. La structure arborescente inhérente au XML modélise au mieux cet enchâssement.

Figure 2 : FipsColor, fenêtre de résultat et menus

5 Conclusion et perspectives

Cet article propose un outil interactif d'aide à l'apprentissage du français qui permet à l'apprenant de mieux prendre conscience des concepts grammaticaux de base dans un cadre ludique. FipsColor est une réponse à la création de *services web* basée sur une application provenant du TAL et directement exploitable par la communauté pédagogique.

Références

- CHOMSKY, N. 1995. The Minimalist Program, Cambridge, Mass., MIT Press.
- GAUDINAT, A, GOLDMAN J-P. & WEHRLI E. 1999. "Syntax-Based Speech Recognition: How a Syntactic Parser Can Help a Recognition System". EuroSpeech Conference, Budapest, Hungary, 1999, volA, p.1587-1590
- GOLDMAN J-P, GAUDINAT A., NERIMA L., WEHRLI E. 2001. "FipsVox : a French TTS based on a syntactic parser". 4th Speech Synthesis Workshop. Edinburgh, 2001
- HAEGEMAN, L. 1994. Introduction to Government and Binding Theory, Oxford, Blackwell.
- LAENZLINGER, C. 2003. Initiation à la Grammaire Formelle du Français: Le Modèle Principes & Paramètres de la Grammaire Générative Transformationnelle. Peter Lang, Berne/Berlin.
- LAENZLINGER, C. ET E. WEHRLI, 1991. "FIPS : Un Analyseur interactif pour le français". TA Informations, 32:2, 35-49.
- L'HAIRE, S. & VANDEVENTER-FALTIN, A 2003. "Enor diagnosis in the FreeText project". CAUCO 20(3), T. Heift & M. Schulze (éds.). Special Issue Error Analysis and Error Correction in Computer-Assisted Language Learning
- SERETAN, V. 2008. "Collocation extraction based on syntactic parsing". Ph.D. thesis, Univ. of Geneva.
- WEHRLI, E. 1997. L'analyse syntaxique des langues naturelles: Problèmes et méthodes, Paris, Masson
- WEHRLI, E. 2003. "Translation of Words in Context". IXth MT Summit, New Orleans.