

Dépendances syntaxiques et méthodes de détection de passages pour une segmentation sur le locuteur et le thème

Loïc Maisonnasse, Caroline Tambellini

Laboratoire CLIPS IMAG – Université Joseph Fourier
385, rue de la bibliothèque - BP 53
38041 Grenoble cedex9
loic.maisonnasse@imag.fr
caroline.tambellini@imag.fr

Résumé. Nous montrons ici l'intérêt des dépendances syntaxiques et des méthodes de détection de passages pour la détection du locuteur dans les discours. Nous nous appuyons sur les méthodes développées dans le cadre de la campagne de fouilles de texte DEFT'05 dans le but de distinguer le discours de François Mitterrand de celui de Jacques Chirac. Nous évaluons l'utilisation des dépendances syntaxiques en tant qu'unité caractérisant les différences entre les deux discours. Ces unités, obtenues par traitement linguistique, constituent donc les descripteurs retenus pour la représentation des discours, et sur lesquels nous appliquons un apprentissage. Les deux présidents ayant des discours sur des thématiques différentes nous combinons cet apprentissage avec l'utilisation d'une méthode de segmentation thématique pour détecter les changements de locuteur.

1 Introduction

1.1 Principes

DEFT'05 a pour objectif la détection des phrases de F. Mitterrand dans des allocutions de J. Chirac. Pour ce faire, nous disposons d'un corpus composé d'allocutions de J. Chirac au sein desquelles des portions d'allocutions de F. Mitterrand ont été insérées.

La réussite de la tâche réside dans l'utilisation des éléments qui distinguent au mieux les deux discours. Pour cela, nous partons du principe que le vocabulaire des discours n'est pas le seul élément mettant en évidence la différence entre les discours des deux orateurs. Cette différence s'exprime également au niveau des tournures et des constructions de phrases propres à chaque auteur.

Globalement, nous montrons l'intérêt de l'utilisation d'éléments représentatifs de la syntaxe pour déterminer l'origine d'un discours : notre méthode utilise les dépendances syntaxiques pour la détection du locuteur.

Les discours de F. Mitterrand ayant une thématique différente des allocutions de J. Chirac, nous évaluons la complémentarité d'une détection des différences au niveau syntaxique avec une approche plus thématique. Pour cela, nous combinons notre première approche avec la piste des changements thématiques dans le but d'améliorer la détection des allocutions de F. Mitterrand.