

Analyse de discours évaluatif, modèle linguistique et applications

Stéphane Ferrari^{*,***}, Thierry Charnois^{*,***}
Yann Mathet^{*,***}, François Rioult^{*,***}
Dominique Legallois^{**},***

*GREYC - CNRS UMR 6072

**CRISCO EA 4255

***Université de Caen Basse-Normandie
{Prenom.Nom}@unicaen.fr

Résumé. Notre étude porte sur le discours évaluatif. L'approche que nous suivons est celle d'une analyse symbolique en vue d'un traitement sémantique de l'expression d'opinions.

Un travail préliminaire d'observation sur corpus a permis une première modélisation linguistique qui révèle la complexité du phénomène étudié. Nous en décrivons une mise en oeuvre destinée à constituer un outil d'aide à l'observation pour l'expert linguiste et permettre un retour sur le modèle.

Nous présentons et discutons ensuite deux approches informatiques testées pour la classification de documents d'opinions dans le cadre du défi DEFT07. La première s'appuie sur des n-grammes de mots, la deuxième reprend en partie le modèle linguistique, complété par un processus de classification fondé sur des règles d'association généralisées.

1 Introduction

Dans cet article, nous proposons un état de l'art synthétique pour mieux situer les différentes méthodes que nous avons développées pour l'analyse d'opinion. Nous présentons ensuite l'étude linguistique qui sous-tend notre approche, menée sur un corpus de 443 critiques de livres déposées par les internautes sur les sites fnac et amazon. Des formes récurrentes ainsi que des champs sémantiques privilégiés apparaissent qui caractérisent l'expression d'opinions, de jugements d'évaluation. Ces régularités en langue renseignent non seulement sur la polarité des sentiments exprimés, mais aussi sur d'autres propriétés comme leur intensité ou leur prise en charge par le locuteur. Une première mise en oeuvre a partiellement validé ce modèle tout en mettant en évidence des limites dues à l'implémentation. Elle produit une qualification partielle de l'expression locale d'opinions au sein des textes en repérant différents motifs issus de l'étude linguistique, essentiellement à une échelle infra-phrastique.

La campagne d'évaluation DEFT07 (<http://deft07.limsi.fr>) a été l'occasion d'une exploitation de cette approche pour la classification de textes d'opinion. Cette édition 2007 du