

# Application de la théorie de l'Appraisal à l'analyse d'opinion

Pierre Gardin

Université de Caen-Basse-Normandie

Contact : pierre.gardin@etu.info.unicaen.fr

---

## Résumé

Cet article présente un travail de recherche en cours dans le domaine de l'analyse d'opinions. Il commence par introduire cette discipline, ses principales branches et ses applications, ainsi que les différentes approches existantes pour la classification d'opinions. Puis, il expose l'Appraisal, une théorie récente qui permet d'analyser l'évaluation et le positionnement dialogique, et les travaux inspirés par cette théorie dans le domaine du traitement automatique des langues. Enfin, il introduit une approche pour l'extraction d'opinions basée sur la théorie de l'Appraisal, dans laquelle des groupes adjectivaux sont extraits dans l'optique de réaliser des tâches comme la classification.

**Mots-clés :** Analyse d'opinions, analyse de textes subjectifs, classification de sentiments, théorie de l'Appraisal.

## Abstract

This paper presents an ongoing research study in the field of opinion analysis. First, it briefly introduces the discipline, its main subdomains, applications and the differences with traditional, fact-based textual analysis. Traditional approaches to sentiment classification are presented. Then, it describes Appraisal theory, a new theoretical framework for dealing with evaluation and stance, and the works it has inspired. Finally, it shows an application of Appraisal theory to extraction of opinion-holding adjectival groups.

**Keywords:** Sentiment analysis, subjective text analysis, sentiment classification, Appraisal theory.

---

## 1. Introduction

### 1.1. Contexte

L'analyse automatique de l'expression de l'évaluation dans le discours (ou *sentiment analysis*) se différencie en plusieurs points de l'analyse textuelle classique, celle-ci consistant à traiter des textes supposés objectifs et visant principalement à la classification thématique de documents et à l'extraction d'informations ciblées. Tout d'abord, les différentes caractéristiques d'une opinion sont relativement cohérentes quelque soit le type de document [8]. Ensuite, la tâche de classification ne concerne plus un ensemble ouvert de classes, mais un ensemble limité à quelques valeurs (généralement les polarités : positive, négative ou neutre) [8]. Cette discipline s'est développée fortement depuis une dizaine d'années en raisons de ses nombreuses applications potentielles. La veille industrielle en est la plus évidente, mais il faut également citer la possibilité d'améliorer la recherche d'informations sur des produits de consommation, ou la prise en compte des caractéristiques des opinions pour des tâches qui relèvent de la recherche d'information classique [8].

### 1.2. Problème

Le travail réalisé a pour but d'effectuer une extraction des opinions présentes dans des textes, pour obtenir un ensemble de caractéristiques qui peuvent servir à la classification de textes d'opinion selon leur polarité. Il s'agit d'un problème assez difficile, comme le prouvent les résultats du

concours de classification de textes d'opinion DEFT'07 [4]. Les opinions possèdent des attributs (polarité, intensité, source, cible...) dont le calcul nécessite un grand nombre de ressources externes comme des lexiques sémantiques ou des bases de données d'entités nommées. L'extraction de ces caractéristiques est subordonnée à un étiquetage morpho-syntaxique et une analyse syntaxique préalables, ce qui constitue un facteur d'erreur supplémentaire. Pour finir, certains types de documents, tel que le corpus d'interventions parlementaires de la tâche DEFT'07, résistent même, dans une certaine mesure, à l'évaluation humaine.

## 2. Approches existantes

Les approches existantes de classification de textes d'opinions – en particulier celles présentées à DEFT'07 – décomposent toutes le problème en deux sous-tâches : la sélection des traits pour construire une représentation simplifiée du texte et la construction du classifieur proprement dite, qui fait souvent appel aux techniques classiques de la fouille de données [4]. Ces approches peuvent se répartir selon le degré de théorie linguistique auquel elles font appel. Dans tous les cas, des prétraitements linguistiques simples sont effectués : lemmatisation, regroupement par synonymes. Certaines approches utilisent les méthodes classiques de la recherche d'information – l'analyse sémantique latente, la similarité cosinus entre des vecteurs de classes, la réduction du nombre de traits à l'aide de critères comme l'information mutuelle ou l'indice Tf\*Idf... D'autres approches utilisent les techniques plus spécifiques à la fouille d'opinion et effectuent une analyse plus fine des textes, notamment en cherchant à déterminer l'orientation sémantique de certains termes. Ces techniques font une utilisation plus intensive de lexiques sémantiques externes tels que SentiWordNet [3], comportant des informations sur la polarité des termes.

## 3. La théorie de l'Appraisal

L'Appraisal [7] est une théorie sur la façon dont sont exprimées les évaluations (ce terme inclut les opinions et les réactions émotionnelles). Elle se représente sous la forme d'un graphique appelé *système*, qui modélise les différents choix du locuteur dans les fonctions qu'il peut exprimer (significations) et dans la façon de les exprimer (formulations). Les trois sous-systèmes de l'Appraisal sont l'Attitude, l'Engagement et la Gradation. L'Attitude est le système qui permet de décrire les réactions émotionnelles (Affect) et les différentes opinions selon qu'elles portent sur des comportements (Jugement) ou sur d'autres entités (Appréciation). L'Engagement est le système qui concerne le degré d'acceptation par le locuteur de sa proposition, ce qui inclut le discours rapporté. La Gradation est le système qui rend compte de l'intensité des évaluations exprimées, ainsi que du degré avec lequel les propositions sont modulées par l'Engagement. Ses deux sous-systèmes sont la Force, qui correspond à une Gradation en fonction de l'intensité ou de la quantité, et le Focus, qui modélise le degré de prototypicalité.

La théorie de l'Appraisal n'a pas encore, à notre connaissance, connu d'application en Traitement Automatique des Langues en français, mais un certain nombre de travaux anglophones visaient à détecter des spams [11, 13] et à classer des articles scientifiques selon la discipline concernée [11]. Elle a été utilisée pour la classification de critiques de films [12] qui a permis d'améliorer les résultats par rapport à la méthode par sacs de mots. Une expérience a prouvé que l'accord inter-annotateurs des différents systèmes de l'Appraisal est assez faible [9], même quand l'annotation est peu précise. Enfin, une procédure a été développée pour l'extraction des opinions dans le cadre de l'Appraisal [2].

## 4. Approche proposée

Les travaux réalisés ont pour but d'extraire des groupes adjectivaux (GAdj) exprimant l'Attitude, pouvant inclure des adverbes modificateurs, et de calculer leurs caractéristiques. Les adjectifs sont en effet la principale forme que peut prendre l'Attitude, et aussi la façon la plus naturelle d'exprimer une évaluation [2, 12]. Les GAdj possèdent les caractéristiques suivantes, exprimées sous forme de structure de traits :

- Polarité : indique la dépréciation ou l'appréciation dont est porteuse une évaluation.
- Force : mesure de l'intensité ou de la quantité (« faible », « moyenne » ou « forte »).

- Focus : degré de prototypicalité, pouvant prendre les mêmes valeurs que la Force.
  - Système : sous-système d'Attitude (type d'émotion ou opinion).
- Les valeurs possibles de l'attribut Système sont issues de la théorie de l'Appraisal.

#### 4.1. Création des lexiques

Un lexique comprenant 325 adjectifs a été constitué manuellement en traduisant les exemples tirés de [7] et à l'aide de dictionnaires de synonymes. Chaque adjectif est décrit par une lexie et des attributs : la Force, la polarité et le système, et on considère que chaque adjectif possède intrinsèquement un Focus neutre. La Force d'un adjectif a été déterminée manuellement. L'existence d'adjectifs équivalents mais de Force différente a permis de placer ces adjectifs dans chacune des catégories. Les adjectifs formés avec un préfixe privatif (par ex. « irréprochable ») ont été considérés comme hyperboliques et donc de Force élevée, ceux possédant beaucoup d'homographes comme de Force faible en raison du risque de confusion. Les termes relevant du registre familier possèdent une Force plus élevée que leurs synonymes du registre courant. Un autre lexique contient 37 adverbes qui peuvent modifier les syntagmes placés après lui : soit un adjectif seul, soit un adjectif précédé d'un ou plusieurs adverbes, soit les deux. Certains possèdent des valeurs axiologiques intrinsèques (par ex. « trop ») qui remplacent celles des adjectifs qu'ils modifient. Un dernier lexique est composé de sept verbes d'état.

#### 4.2. Modèle de calcul

Le traitement est réalisé à l'aide de grammaires DCG au sein de la plate-forme LinguaStream [1] qui permet d'enchaîner des modules d'analyse qui effectuent des annotations. Les prétraitements consistent en une segmentation en mots et un étiquetage de ceux-ci par leur catégorie grammaticale avec le TreeTagger [10]. L'analyse s'effectue en trois temps. D'abord, les groupes adjectivaux sont repérés en partant des adjectifs et en ajoutant successivement les adverbes placés avant eux, ces derniers modifiant les attributs du groupe auxquels ils sont rattachés. Quel que soit le type de syntagme (adjectif seul ou GAdj) auquel il s'applique, un adverbe peut augmenter ou diminuer la Force et le Focus et inverser la polarité. Ensuite, les groupes nominaux (GN) sont délimités à l'aide d'une grammaire utilisant les annotations déjà effectuées pour les GAdj. Pour finir, deux grammaires permettent de rattacher un GAdj porteur d'une opinion au GN correspondant à l'entité qui fait l'expérience d'une émotion ou porte une opinion, selon les trois patrons : épithète, attribut, apposition. Les évaluations annotées sont ensuite visualisées sous forme d'un fichier XML.

### 5. Tests

Nous avons mené quelques expériences préliminaires afin de déterminer si le modèle de traitement pouvait être exploité efficacement pour la classification d'opinions.

#### 5.1. Corpus

Nous avons utilisé deux corpus au format XML : le premier (repris de [5]) est constitué d'articles de type « portraits » (sans les nécrologies) du journal « Le Monde » couvrant la période juillet 2002–décembre 2002, le second est un corpus de critiques de produits culturels provenant du site Web « avoir-alire ». Nous avons réalisé les tests sur les 24 premiers textes de chaque corpus.

#### 5.2. Résultats

Le nombre moyen d'évaluations extraites par article est significativement plus élevé pour le premier corpus (2,04) que pour le second (1,13). Dans les deux cas, le nombre d'épithètes est très majoritaire par rapport aux autres patrons et l'Appréciation est plus fréquente que les deux autres types d'évaluation. Dans les portraits, il y a une forte présence de discours rapporté qui permet de supposer que le point de vue de l'auteur ne se manifeste pas directement.

En ce qui concerne le second corpus, il manque une ontologie du domaine qui permettrait d'analyser les évaluations selon leur lien avec le thème de l'article. Les évaluations détectées le sont presque toutes correctement mais les limites du modèle sont clairement visibles. Par exemple, le GN « l'unique intérêt du film » est analysé comme positif en raison de la valeur positive accordée à l'adjectif « unique ». De plus, pour un tiers des articles, aucune évaluation n'a été trouvée, ce qui est dû à la fois au silence dans les lexiques des verbes d'état et d'adjectifs et à certaines constructions complexes – comme les adverbes de commentaires phrastiques situés au milieu de

la proposition. Afin d'utiliser les évaluations détectées dans le cadre d'une campagne de classification, il faudrait, en plus d'une ontologie d'un domaine, une résolution des coréférences pour interpréter la cible des évaluations.

## 6. Conclusion

Dans cet article, nous avons décrit la discipline récente que constitue la fouille d'opinions et plus précisément la classification de textes d'opinion selon la polarité. Nous avons également présenté un modèle basé sur la théorie de l'Appraisal, une nouvelle théorie sur l'expression de l'évaluation dans les textes. Nous avons décrit comment calculer les attributs de groupes adjectivaux en se fondant sur les propriétés intrinsèques des mots.

Différentes pistes sont envisagées afin d'arriver à appliquer la théorie de l'Appraisal à la classification d'opinions. Pour commencer, la détection du discours rapporté se doit d'être traitée pour assigner des sources aux évaluations. D'autres réalisations de l'Attitude pourraient être prises en compte, en particulier les substantifs et les processus mentaux qui sont les plus simples à modéliser. Enfin, une méthode générique pour la constitution d'opinion est envisageable à partir de WordNet [6]. Le manque de ressources lexicales en français constitue néanmoins un problème important.

## Bibliographie

1. Frédéric Bilhaut and Antoine Widlöcher. La plate-forme LinguaStream. In *Journée ATALA «Architectures logicielles pour articuler les traitements sur corpus»*, 2005.
2. Kenneth Bloom, Navendu Garg, and Shlomo Argamon. Extracting appraisal expressions. In *Proceedings of NAACL HLT*, pages 308–315, 2007.
3. Andrea Esuli and Fabrizio Sebastiani. SENTIWORDNET : A publicly available lexical resource for opinion mining. In *Proceedings of LREC-06, 5th Conference on Language Resources and Evaluation*, pages 417–422, Genova, IT, 2006.
4. Cyril Grouin, Jean-Baptiste Berthelin, Sarra El Ayari, Thomas Heitz, Martine Hurault-Plantet, Michèle Jardino, Zohra Khalis, and Michel Lastes. Présentation de DEFT'07 (Défi Fouille de Textes). In *Proceedings of the DEFT'07 workshop, Plate-forme AFIA*, pages 9–17, Grenoble, France, July 2007.
5. Agata Jackiewicz, Thierry Charnois, and Stéphane Ferrari. Jugements d'évaluation et constituants périphériques. In Adeline Nazarenko, editor, *Actes de la 16e conférence sur le traitement automatique des langues naturelles*, Senlis, France, June 2009.
6. Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten De Rijke. Using WordNet to measure semantic orientation of adjectives. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004, volume 4*, pages 1115–1118, 2004.
7. James R. Martin and Peter R. R. White. *The language of evaluation : Appraisal in English*. Palgrave Macmillan, Hampshire, England & New York, U.S.A., 2005.
8. Bo Pang and Lillian Lee. *Opinion mining and sentiment analysis*. Now Publishers, 2008.
9. Jonathon Read, David Hope, and John Carroll. Annotating expressions of appraisal in english. In *Proceedings of the Linguistic Annotation Workshop*, pages 93–100, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
10. Helmut Schmid. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*, volume 12. Manchester, UK, 1994.
11. Casey Whitelaw and Shlomo Argamon. Systemic functional features in stylistic text classification. In *Proceedings of AAAI Fall Symposium on Style and Meaning in Language, Art, Music, and Design*, 2004.
12. Casey Whitelaw, Navendu Garg, and Shlomo Argamon. Using appraisal groups for sentiment analysis. In *Proceedings of CIKM-05, 14th ACM International Conference on Information and Knowledge Management*, pages 625–631, Bremen, Germany, 2005. ACM, New York, NY, USA.
13. Casey Whitelaw and Jon Patrick. Selecting systemic features for text classification. In *Proceedings of Australasian Lang. Tech. Workshop*, pages 93–100, 2004.