HAL open science

Catégorisation des évaluations dans un corpus de blogs multi-domaine

Matthieu Vernier, Laura Monceaux, Béatrice Daille, Estelle Dubreil

To cite this version:

Matthieu Vernier, Laura Monceaux, Béatrice Daille, Estelle Dubreil. Catégorisation des évaluations dans un corpus de blogs multi-domaine. Revue des Nouvelles Technologies de l'Information, 2009, pp.45–70. hal-00405407

HAL Id: hal-00405407

https://hal.science/hal-00405407v1

Submitted on 20 Jul 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Catégorisation des évaluations dans un corpus de blogs multi-domaine

Matthieu Vernier*, Laura Monceaux*, Béatrice Daille*, Estelle Dubreil*

*LINA - CNRS UMR 6241
UFR de Sciences et Techniques, 2, rue de la Houssinière
BP 92208, 44322 Nantes Cedex 03, France
{*Prénom.Nom*}@univ-nantes.fr,
http://www.lina.univ-nantes.fr

Résumé. Dans le cadre de la fouille d'opinion, nous proposons une méthode automatique pour la détection et la catégorisation des évaluations localement exprimées dans un corpus de blogs multi-domaine. Cette méthode s'appuie sur deux théories linguistiques modélisant le processus d'évaluation dans le langage naturel et sur des ressources lexicales. Nous présentons deux composants logiciels qui s'intègrent à la plateforme *UIMA* [1] et qui permettent : l'extraction automatique de structures symboliques spécifiques à l'expression évaluative et la catégorisation des évaluations à partir des structures symboliques apprises. L'outil de catégorisation vise en particulier à analyser la signification axiologique, la modalité et la configuration énonciative d'une évaluation. L'objectif à terme est d'associer les évaluations ainsi catégorisées à leur sujet.

1 Introduction

L'image est un aspect fondamental dans notre société pour les personnes et les entreprises pour lesquelles l'opinion du public est importante. Celles-ci ont besoin de se tenir au courant de l'évolution de leur image et des sujets qui intéressent la population pour s'adapter à leurs attentes et améliorer leur réactivité. Ces aspects impliquent particulièrement l'industrie des nouvelles technologies, la politique, la publicité, les médias ou la finance pour lesquels l'étude de l'image représente un enjeu et un pouvoir économique majeur.

Les blogs se caractérisent par l'usage évaluatif qui en est fait, au sens où les internautes l'utilisent pour exprimer librement et partager leurs opinions sur leurs centres d'intérêts. De plus, la quantité de blogs et leur réactivité rendent possible des analyses en quasi temps réel et l'observation de variations au fil des jours, des semaines ou des mois. Les blogs prennent une importance de plus en plus grande ; les sociologues se penchent sur ce phénomène, d'autres medias en parlent et font référence à des informations provenant de blogs, que ce soit pour leur caractère provocateur, original, novateur, etc. De plus, ils sont en perpetuelle évolution, tant du point de vue qualitatif que quantitatif. Le nombre de blogs est passé de 8 à 72 millions

[1] « Unstructured Information Management Architecture » - voir Actes de l'atelier *Towards Enhanced Interoperability for Large HLT Systems: UIMA for NLP (LREC'08)* (2008).

entre mars 2004 et mars 2006 [2]. Figure emblématique du web 2.0, les blogs proposent à l'internaute une diversité d'outils associée à une simplicité d'utilisation, et sur le plan sociologique, la possibilité d'interactions actives des internautes entre eux, encourageant la formation de réseaux sociaux. Pour ces raisons, les blogs constituent une ressource linguistique idéale pour une analyse d'image réactive, multi-domaines et à grande échelle.

A l'heure du développement de la recherche d'information, l'enjeu premier réside donc dans la création de programmes informatiques capables de détecter automatiquement les évaluations émises par les internautes à propos d'un sujet donné. C'est ce que nous qualifions dans le cadre de ce travail d'accès au couple concept cible - évaluation, où le concept correspond par exemple à un homme politique, une marque, un produit, un événement (une grève, une guerre, une élection, etc), une idée (le développement durable, le droit à l'avortement, etc), et l'évaluation, à l'émotion, le jugement, l'appréciation, la subjectivité et l'intensité, émis à propos de ce concept. Selon (Dave et al., 2003), l'outil de *fouille d'opinion* le plus abouti produirait, pour un concept cible donné, une liste de caractéristiques évaluatives. L'idée était alors essentiellement de pouvoir guider le processus de décision d'un utilisateur lorsqu'il souhaite acheter un produit, en lui proposant un aperçu ou un résumé des avis d'autres consommateurs. Parallèlement, l'expression *Analyse des sentiments* [3] est utilisée dans un très grand nombre de travaux sur la classification de critiques, ce qui vaut à ce terme d'être fréquemment associé à cette tâche. En réalité ces deux expressions dénotent le même champ d'étude dans lequel nous nous positionnons.

D'un point de vue terminologique, nous préférons adopter le terme d'*évaluation* et de *langage évaluatif* pour faire référence aux différentes formes linguistiques possibles pour exprimer une opinion, une appréciation, un jugement, une émotion, un accord, un désaccord, etc. Nous justifions cet emploi en adoptant le point de vue des théories linguistiques anglo-saxonnes (Martin et White, 2005) et françaises (Charaudeau, 1992), (Galatanu, 2000) sur l'évaluation, ainsi qu'aux travaux de fouille d'opinions qui s'appuient sur ces théories (Wiebe et Mihalcea, 2006), (Whitelaw et al., 2005), (Read et al., 2007), (Ferrari et al., 2008). Dans l'optique d'une extraction d'information robuste, orientée vers l'analyse automatique des évaluations en usage dans la blogosphère, il faut d'abord pouvoir qualifier la diversité du phénomène et des différentes formes d'expressions possibles de l'évaluation. Pour plusieurs applications, identifier simplement la polarité générale des documents ou des phrases peut ne pas être suffisant : « il n'est pas rare de trouver deux opinions ou plus dans une seule phrase, ou de trouver une phrase contenant aussi bien des opinions que des informations factuelles » (Wiebe et Mihalcea (2006)). Nous nous intéressons à délimiter précisément ces expressions et les termes utilisés pour évaluer. Nous cherchons également à savoir quels sont les concepts qui sont évalués particulièrement subjectivement et intensément et à quel degré un locuteur s'implique dans l'évaluation qu'il exprime ou s'il n'est que le relais de l'opinion générale ?

De part sa nature et ses usages, les blogs imposent de nouveaux défis et de nouvelles problématiques de recherche. Les approches classiques de classification de critiques par apprentissage automatique supervisée sont efficaces sur des textes d'opinions portant sur un sujet fixé. Les applications possibles peuvent être d'ordre **marketing** afin de guider les utilisateurs dans leur consommation. Il peut ainsi s'agir de catégoriser des critiques de films, de livres, de

[2] http ://www.journaldunet.com/cc/03_internetmonde/intermonde_blog.shtml

[3] « Sentiment analysis » : expression apparue via les travaux de classification de critiques de (Turney, 2002) et (Pang et al., 2002).

produits technologiques (lecteurs MP3, ordinateurs portables, caméras, etc), de voitures, des album musicaux ou encore des destinations de voyages touristiques selon la polarité positive, négative ou neutre de l'ensemble de l'unité textuelle considérée. Les applications peuvent également être **financières** ou **politiques**. Ainsi, il peut s'agir de classer des textes de réactions positives ou négatives suite à une dépêche boursière (Devitt et Ahmad, 2007) afin d'améliorer l'anticipation d'événements économiques. D'autres travaux ont le projet de classer les opinions des politiciens afin d'améliorer l'accès et la qualité d'informations des votants : apportent-ils leur soutien ou s'opposent-ils à telle idée politique ou à un projet de loi (Bansal et al., 2008), (Thomas et al., 2006) ? L'aspect mono-thématique des corpus analysés, associé à la disposition de corpus d'entraînements déjà catégorisé sur le web, rend possible l'utilisation de méthodes d'apprentissage classiques (SVM, Naïves Bayes, etc).

Toutefois, avec ces approches, comment traiter les billets de blogs qui évoquent (et évaluent) une multitude de concepts différents ? Comment traiter les commentaires très courts d'un blog ? Comment disposer d'un corpus d'entraînement et comment décider de la polarité générale d'un blog s'il évalue plusieurs concepts différemment ? Afin de répondre aux problèmatiques de notre tâche applicative, nous adoptons une approche symbolique basée sur l'apprentissage de structures spécifiquement évaluatives. En amont de notre travail informatique, nous nous appuyons sur la grammaire de la langue proposée par (Charaudeau, 1992) pour modéliser les principales caractéristiques des phénomènes évaluatifs et nous empruntons à (Galatanu, 2000) la terminologie linguistique pour définir correctement les notions de subjectivité et d'axiologie.

La suite de cet article est divisé en cinq sections : dans la section 2, nous tentons de synthétiser les différents axes de recherche dans le champ particulièrement vaste de la fouille d'opinion afin de resituer et justifier l'utilité de notre méthode ; la section 3 présente les deux modèles linguistiques du langage évaluatif et de la subjectivité sur lesquels nous fondons notre approche informatique et sur lesquels nous fixons la terminologie de notre étude. Dans la section 4, nous présentons une étude manuelle du langage de l'évaluation menant à la création d'un corpus annoté manuellement (Dubreil et al., 2008) et d'un lexique de l'évaluation. Nous expliquons par la suite (section 5) les différents modules informatiques intégrés dans une chaîne de traitements *UIMA* permettant l'extraction supervisée de structures symboliques évaluatives, puis la catégorisation des évaluations à partir des structures apprises. Parallèlement, nous testons l'efficacité de notre outil sur un corpus test, en cherchant à quantifier et qualifier les évaluations exprimées par les internautes, à travers deux requêtes potentiellement données par un utilisateur : que pensent les internautes de *Sarah Palin* et que pensent-ils de la *nourriture japonaise* ?

2 Axes de recherche en fouille d'opinion

Les travaux de fouille d'opinion se caractérisent selon trois aspects : les objectifs de recherches ou d'applications qu'ils visent, la nature des corpus qu'ils analysent et les méthodes informatiques qu'ils mettent en oeuvre. Ces trois aspects sont généralement interdépendants et permettent de distinguer trois groupes de travaux dans ce domaine : les approches de classification de textes, les approches de recherche d'informations sur les concepts évalués dans une unité textuelle. La troisième catégorie de travaux consiste à élaborer manuellement, semi-automatiquement ou automatiquement des ressources lexicales et sémantiques sur l'évaluation

(Esuli et Sebastiani, 2006). Elle se positionne comme un apport aux deux types d'approches précédentes.

Dans cette section, nous nous intéressons en particulier à synthétiser les nombreux travaux sur la fouille d'opinion en classification de textes et en recherche d'informations en les distingant en particulier sur l'aspect numérique ou symbolique de leur méthode informatique. Nous entendons par approche numérique, les techniques d'acquisitions d'informations basées sur l'aspect fréquentiel des données considérées. Ce type de méthode représente très majoritairement les travaux du domaine de la fouille d'opinion. De façon plus marginale, les méthodes symboliques s'intéressent à représenter la structure grammaticale ou sémantique des données analysées pour améliorer la qualité des informations extraites.

2.1 Catégorisation de textes

Les travaux pionniers en fouille d'opinion répondent à un besoin applicatif en essor avec l'émergence du web, des sites commerciaux et des premiers types de réseaux sociaux : la classification de documents (plus particulièrement des critiques ou des textes d'opinions) selon la polarité générale exprimée dans l'ensemble du texte par l'auteur. Il s'agit de classer des textes dont le sujet principal est unique : critique de film, de voiture, de livre, d'un album musical, dépêche boursière, texte de relecture d'article scientifique, débat politique, etc. L'objectif applicatif est essentiellement d'aider le processus décisionnel d'un utilisateur, en synthétisant les informations dont il pourrait avoir besoin, lorsqu'il souhaite acheter, voter, ou prendre une décision qui impliquerait le besoin de connaître l'opinion d'une ou plusieurs autres personnes.

Les approches discutées ci-dessous partagent toute l'idée de représenter une unité textuelle (un document, un paragraphe, une phrase) par un ensemble prédéfini de caractéristiques linguistiques (ou *traits*), puis d'utiliser la fréquence de ces traits pour décider de la catégorie d'un texte. Les méthodes de classification d'opinions héritent des méthodes d'apprentissage automatique et de la fouille de données en adaptant ces traits au contexte de l'analyse d'opinions. Ces traits sont variables selon les travaux et la première problématique concerne le choix des meilleurs traits possibles pour décrire une unité textuelle.

2.1.1 Caractéristiques linguistiques d'une unité textuelle

Pour décrire une unité textuelle dans une optique de fouille d'opinion, les traits considérés traditionnellement appartiennent à cinq niveaux d'abstraction linguistique :

- **Lexical** : traditionnellement, un texte peut être représenté par un vecteur de mots simples, en tenant éventuellement compte de la fréquence des termes ou de leur rapport tf-idf au sein du corpus considéré. (Pang et al., 2002) montrent toutefois que considérer la simple présence ou absence d'un mot plutôt que sa fréquence est plus efficace sur certains corpus, en particulier des corpus de critiques de films. Dans la même idée, les hapax ou les mots qui n'apparaissent qu'une seule fois dans un corpus semblent être des indicateurs très précis de subjectivité (Wiebe et Mihalcea, 2006). De façon voisine, (Yang et al., 2006) s'intéressent aux termes rares du corpus, non listés dans les dictionnaires préexistants, pour extraire les néologismes qui sont statistiquement souvent associés à une évaluation. L'utilisation d'unigrammes ou de n-grammes (Dave et al., 2003) (Vernier et al., 2007) comme vecteur de traits est sujet à débat. Les résultats sont variables selon

les corpus analysés, il semble que ce choix résulte davantage de la nature des corpus et de la richesse de leurs structures lexicales.

- **Morpho-syntaxique** : l'information sur la catégorie grammaticale d'un mot est également couramment exploité. Il s'agit généralement d'une façon un peu brute de désambiguïser certains mots. Les adjectifs sont particulièrement employés (Mullen et Collier, 2004) (Whitelaw et al., 2005) (Harb et al., 2008) de part la grande corrélation entre leurs présences et le fait qu'un texte soit subjectif. Mais les noms et les verbes sont également d'importants indicateurs d'évaluation. (Hatzivassiloglou et Wiebe, 2000) ont spécialement étudié l'extraction des noms (*inquiétude, espoir, etc*), par amorce, en observant les noms qui co-occurent fréquemment avec des marqueurs subjectifs.
- **Relation syntaxique** : des tentatives de représentation des relations syntaxiques ont également été testées pour améliorer les résultats qui étaient notamment faussés par l'utilisation de tournures négatives ou diminutives (Kudo et Matsumoto, 2004). Dans un corpus monothématique évaluant des téléphones portables, le terme *difficile*, généralement négatif, n'a pas la même signification dans certains contextes : *difficile de raccrocher*. Ces expressions collocationnelles sont un des phénomènes linguistiques particulièrement observées dans les études sur la subjectivité et l'évaluation (Riloff et Wiebe, 2003) (Dubreil, 2006).
- **Sémantique** : bien qu'ils ne s'inscrivent pas tous explicitement dans le cadre de la classification de textes, des recherches sont actuellement menées sur la façon de caractériser des expressions évaluatives complexes en considérant des classes sémantiques propres à l'évaluation : la modalité évaluative (affect, appréciation, jugement), l'orientation axiologique (positif, négatif), l'intensité ou l'engagement du locuteur (Whitelaw et al., 2005). Les expressions ainsi détectées peuvent servir de traits dans un processus de classification numérique à l'image de (Vernier et al., 2007).
- **Enonciatif** : la position d'un terme dans une unité textuelle (au début, au milieu ou à la fin d'un texte) peut potentiellement avoir un effet sur la façon dont ce terme affecte le sens général du texte. Cette information est représentée dans le système de traits de (Kim et Hovy, 2006) et (Vernier et al., 2007).

Les outils informatiques de classification prenant en compte la structure sémantique, voire énonciative, des évaluations restent toutefois assez rares.

2.1.2 Méthodes

Les méthodes supervisées de classification de textes étudiées par (Sebastiani, 2002) sont souvent appliquées dans les travaux de fouille d'opinion. (Pang et al., 2002) ont par exemple comparé les classifieurs Naive Bayes, de Séparateur à Vaste Marge (SVM) et de maximum entropie pour la classification de critiques de film à partir d'un important corpus d'entraînement. D'autres comparaisons de ce type sur la performance des techniques standard d'apprentissages automatiques ont été menées avec différents types de traits caractérisants un texte (Airoldi et al., 2006) (Dave et al., 2003) (Gamon, 2004).

L'efficacité des travaux de classification des textes d'opinions est également particulièrement influencée par le domaine dans lequel ils s'appliquent (Aue et Gamon, 2005) (Blitzer et al.). La principale raison est qu'une même phrase peut avoir un sens différent dans un autre domaine ; l'expression *N'hesitez pas à lire le livre* est positive si l'on est en train d'évaluer un livre, et peut être négative si l'on parle d'un film. Certains mots ont également un sens ambigü

selon le domaine (*un film* ***imprévisible*** / *une voiture* ***imprévisible***) car ils activent des stéreotypes linguistiques différents (voir paragraphe 3.). La différence de vocabulaire entre domaine rend difficile la possibilité d'appliquer un classifieur entraîné sur les données d'un domaine vers un autre domaine. (Yang et al., 2006) proposent toutefois une approche simple pour le transfert de domaine : les traits qui sont des bons indicateurs de subjectivité dans au moins deux domaines différents sont considérés comme des traits non dépendants d'un domaine.

2.1.3 Limites

Les méthodes de classification sont donc essentiellement numériques. L'efficacité des traits selon les corpus est difficilement interprétable, mis à part en comparant les méthodes sur une même tâche. Le besoin d'un large corpus d'entraînement (initialement catégorisé) diminue la portabilité des méthodes et restreint leur efficacité à des types de corpus monothématiques spécifiques où le concept évalué est déterminé. Une majorité d'applications possibles en fouille d'opinions sont limités par ces aspects et sont à l'origine du deuxième type de problématique que nous présentons dans le paragraphe suivant.

2.2 Recherche d'informations sur un concept

La deuxième problématique, plus récente, est de savoir précisément ce qui, à l'intérieur d'un texte, est positif et ce qui est négatif. D'un point de vue applicatif, la classification de texte permet à une entreprise d'estimer l'avis général exprimé à propos d'un produit, mais celle-ci veut désormais savoir précisément qu'elles sont les aspects négatifs et les aspects positifs de son produit. L'idée retenue par une majorité de travaux dans cette problématique est de considérer qu'un concept (un produit, une personne, un événement, etc) est composé de plusieurs concepts associés[4] que l'on peut évaluer séparément (par exemple, un appareil photo s'évalue sur son prix, la qualité de ses images, sa taille, son aspect esthétique, etc.).

La tâche se décompose essentiellement en trois étapes : identifier et extraire les termes correspondants à des concepts associés dans un texte ; déterminer s'ils sont évalués positivement ou négativement en détectant les évaluations à proximité ; et regrouper les termes qui sont synonymes et représentent le même concept. L'objectif est notamment d'être capable de produire des résumés détaillés pour un concept cible donné.

2.2.1 Extraction des concepts associés

Les approches numériques se restreignent pour la plupart à l'analyse de critiques et donc à des corpus mono-thématiques. (Yi et al., 2003) considèrent des heuristiques linguistiques pour extraire des noms et des groupes nominaux représentant les concepts potentiellement évalués. L'intuition est de considérer les groupes nominaux les plus fréquents dans un texte ou dans un corpus mono-thématique comme étant des concepts du domaine.

De façon symbolique, (Popescu et Etzioni, 2005) considèrent les concepts associés reliés avec le concept central du domaine en cherchant à identifier les connexions linguistiques méronymiques[5]. Les connecteurs qui permettent de retrouver les concepts associés à un *scanner* sont par exemple : *du scanner, scanner a, etc*, ils sont extraits en explorant le web.

[4] Les travaux anglophones utilisent le terme *features* pour parler de ces concepts associés.
[5] X est une partie de Y.

2.2.2 Extraction du couple concept évalué-évaluation

Pour identifier les expressions évaluatives liées à un concept, une heuristique consiste à extraire les adjectifs qui apparaissent dans la même phrase que l'un de ces concepts (Hu et Liu, 2004), puis à utiliser ces adjectifs dans un algorithme de décision pour déterminer si le concept est évalué positivement ou négativement. Dans le cas des corpus multi-domaines thématiques, une solution proposée consiste à extraire de manière supervisée, lors d'une première passe, les unités textuelles (phrases, paragraphes) correspondant au concept considéré et se replacer ainsi dans les conditions d'une classification de texte mono-thématique (Hurst et Nigam, 2004). La délimitation et la catégorisation précise des expressions évaluatives restent un aspect peu développé dans les méthodes de recherche d'informations.

Nous nous positionnons davantage dans cette seconde problématique qui consiste à rechercher des informations sur un concept en analysant toutes les évaluations portant directement sur lui et sur ses concepts associés. Nous nous confrontons à une limite des méthodes existantes qui se focalisent fréquemment sur un seul domaine pour réaliser cette tâche. Il s'agit pour nous d'identifier toutes les évaluations contenues dans un corpus de blogs, puis de les relier aux bons concepts. Pour catégoriser et qualifier sémantiquement ces évaluations de façon la plus précise possible, nous nous appuyons sur des théories linguistiques du langage évaluatif que nous présentons dans la section suivante.

3 Langage évaluatif

Nous entendons par évaluation, la *rupture de l'indifférence par laquelle nous mettons toutes les choses sur le même plan et considérons toutes les actions comme équivalentes* (Lavelle, 1950). Cette rupture d'indifférence peut s'exprimer à travers un large spectre d'actes allant de la réaction physiologique (rougissement, tremblement, etc) ou affective (peur, désir, culpabilité, etc) à un acte de croyance, d'appréciation, ou de jugement, voire de décision institutionnelle (texte de loi, verdict d'un juge, etc) (Ferrari et Legallois, 2006).

L'axiologie[6] recouvre la zone sémantique qui renvoie à l'idée de préférence et de rupture de l'indifférence. Elle est associée à une notion de polarité positive/négative et comporte les évaluations référant aux champs d'expérience humaine : esthétique (beau/laid), pragmatique (utile/inutile, important/dérisoire, efficace/inefficace), cognitif ou intellectuel (intéressant/inintéressant), éthique ou morale (bien/mal, bon/mauvais), hédonique-affectif (agréable/désagréable, plaisir/souffrance). Exprimée dans le langage, l'évaluation est au cœur d'un certain nombre de champs linguistiques :

- la **sémantique** : (Kerbrat-Orecchioni, 1997) décrivent certaines unités lexicales comme étant intrinséquement évaluatives, elles possèdent une orientation axiologique positive ou négative interne. Par exemple, un *crime*[7] contient, par définition, une orientation négative d'un point de vue éthique et morale. Par opposition, le mot *tuer*[8] n'est pas orienté axiologiquement de façon interne.
- la **pragmatique** : (Anscombre et Ducrot, 1983) placent l'évaluation au coeur du dispositif argumentatif. Chaque mot a un potentiel argumentatif, et donc un potentiel évaluatif,

[6] Etymologiquement : axios = « ce qui vaut », « ce qui ne vaut pas ».
[7] Infraction grave à la morale et punissable par la loi. - Trésor de la langue française.
[8] Action de donner la mort. - Trésor de la langue française.

selon le contexte, selon une culture dominante, selon des stéreotypes partagés par le locuteur et ses interlocuteurs. D'un point de vue pragmatique, *tuer* possède un potentiel axiologique selon le stéreotype mis en jeu : *il l'a tué de sang-froid*, *il l'a tué pour se défendre* ou *il tue pour manger*. De la même façon, *imprévisible* est un adjectif à axiologie bivalente selon le stéreotype culturel exprimé : *une voiture imprévisible*, *un film imprévisible*.

– Traditionnellement, dans la lignée des travaux de l'école française d'analyse du discours (Charaudeau, 1983)(Maingueneau, 1991), l'évaluation est également observée à travers les marques linguistiques de l'**énonciation**, définie comme le phénomène qui témoigne de la façon dont le locuteur s'approprie le langage et l'organise en discours. Dans ce processus d'appropriation, le locuteur est amené à se positionner, en regard de son interlocuteur, en regard du monde qui l'entoure et en regard de son énoncé (Benveniste, 1974).

La **modalisation** se définit comme l'inscription dans le discours, par une marque linguistique, de la prise d'attitude du locuteur à l'égard du contenu de son énoncé. Elle est *susceptible de faire apparaître, au-delà de la littéralité du texte, les systèmes de valeurs et de représentations du locuteur* (Galatanu, 2000). Notre approche de l'évaluation est fondée sur cet usage discursif, à la fois sémantique, pragmatique et énonciatif, du langage évaluatif et sur la caractérisation de ses modalités.

3.1 Modalité de l'évaluation

La classification de (Charaudeau, 1992) propose cinq catégories (constituée de vingt-trois sous-catégories) de modalités du langage pouvant intervenir dans l'expression d'une évaluation : l'**opinion**, l'**accord ou le désaccord**, l'**acceptation ou le refus**, le **jugement**, l'**appréciation**. Pour chacune de ces modalités, Charaudeau définit ce qu'elles révèlent dans l'attitude du locuteur et donne plusieurs exemples de marqueurs linguistiques.

L'**opinion**, présuppose un fait (ou une information) à propos duquel le locuteur explicite la place que celui-ci (ou celle-ci) occupe dans son univers de croyance. Le locuteur évalue donc la vérité de son propos et révèle du même coup ce qu'est son point de vue. Ce point de vue est d'ordre *intellectif*, c'est à dire que, tout en gardant le propos à distance, le locuteur exprime une attitude de croyance plus ou moins certaine qui relève de la raison (*je suis convaincu, je pense, je crois, je doute*).

L'**accord** ou le **désaccord**, présuppose qu'il a été adressé au locuteur une demande de dire s'il adhère ou non à la vérité d'un propos tenu par un autre (que cette demande ait été faite réellement ou non). Le locuteur répond en exprimant qu'il adhère ou non au propos tenu. Du même coup, il contribue à la validation (positive ou négative) de la vérité de celui-ci. Exemples : *bien entendu*, *je ne suis certainement pas d'accord*.

L'**acceptation** ou le **refus**, présuppose qu'une demande d'accomplissement d'un acte a été adressée au locuteur. Il a la possibilité de répondre favorablement ou défavorablement à cette demande de faire. Exemples : *J'accepte de te suivre au bout du monde*, *Je refuse de lui prêter de l'argent*.

Pour le **jugement**, Le locuteur pose dans son énoncé, une action réalisée. Il postule que l'interlocuteur est responsable de cet acte et juge si cet acte est bon ou mauvais en déclarant son approbation ou réprobation en qualifiant l'interlocuteur. Le locuteur se donne l'autorité morale de celui qui peut juger. Cette modalité est également axiologique et concerne en particulier des

valeurs appartenant aux champs d'experience de l'éthique, de la pragmatique, de l'intellectuel. Exemples : *Je vous félicite pour votre travail !*, *Ton attitude n'est pas correct*.

L'**appréciation**, présuppose un fait à propos duquel le locuteur dit quel est son sentiment. Le locuteur évalue donc, non plus la vérité du propos, mais sa valeur, en révélant ses propres sentiments selon un champ d'expérience lié à l'affect, à l'hédonique et à l'esthétique. Cette évaluation est donc d'ordre axiologique. Exemples : *Je suis content que vous soyez là*, *Quelle malchance !*

Cette classification possède de nombreuses similitudes avec les travaux anglophones sur la théorie de l'évaluation (Appraisal Theory (Martin et White, 2005)) tout en tenant compte des spécificités du français. Parallèlement, le modèle proposé par (Charaudeau, 1992) est complémentaire avec certains aspects du modèle sur l'argumentation de (Galatanu, 2000) que nous abordons ci-dessous.

3.2 Objectivation et subjectivation du discours

Le modèle de (Galatanu, 2000) introduit une relation hiérarchique entre les modalités de l'évaluation fondée sur une *échelle de subjectivité*[9]. L'auteur précise que lorsqu'un locuteur organise son énoncé en discours, il peut choisir d'objectiver ou de subjectiver son discours en activant certaines modalités (Exemples 1-2) ou par la configuration énonciative des modalités qu'il active (Exemples 3-4).

1. *Je pense qu'elle est jolie* → **Je pense**[Opinion], **elle est jolie**[Appréciation]
2. *Oui, elle est jolie*→ **Oui**[Accord], **elle est jolie**[Appréciation]
3. *Elle est jolie* → **elle est jolie**[Appréciation implicite]
4. *Je l'adore* (car elle est jolie) → **je l'adore**[Appréciation explicite]

Dans l'exemple 1, l'activation de la modalité d'opinion par l'utilisation du verbe modal *penser* rend l'énoncé plus objectif. Dans son énonciation, le locuteur prend garde à replacer explicitement l'appréciation qui suit dans son propre univers de croyance. Dans l'exemple 2, le locuteur replace son appréciation dans un univers de croyance déjà établi par le marqueur d'accord (*Oui*).

Dans les exemples 3 et 4, le locuteur n'active pas d'univers de croyance pour exprimer son appréciation. Cependant, dans 4, le locuteur configure plus explicitement la modalité d'appréciation, avec le verbe modal *adorer* et avec l'utilisation du pronom *je*, que dans 3. Pour (Galatanu, 2000), *l'absence de modalisation est signifiante* : dans 3, le locuteur adopte une configuration d'énonciation implicite pour exprimer son appréciation et masque son identité énonciative (absence du pronom *je*) ce qui tend à donner une forme plus objective à l'évaluation. Les choix dans la modalisation sont parfois stratégiques, parfois involontaires mais ils rendent compte précisément de la prise d'attitude du locuteur.

Nous renvoyons à (Galatanu, 2000) et (Charaudeau, 1992) pour des considérations plus précises sur ces modèles. Dans la section suivante, nous présentons en quoi ces aspects théoriques du langage évaluatif nous sont utiles pour étudier l'évaluation dans le contexte des blogs.

[9] Par ordre croissant d'objectivation : Appréciation, Jugement, Accord-Desaccord et Opinion.

4 Etude des évaluations sur les blogs

Dans l'optique d'une détection et d'une catégorisation automatique des évaluations, nous nous intéressons dans un premier temps à décrire le langage évaluatif dans le contexte des blogs par une première phase d'analyse manuelle. Un corpus de 250 billets de blogs distincts constitue le support unique de cette analyse. Il est constitué de documents au format XML extraits de la plateforme d'hébergement de blogs OverBlogs[10]. Chaque document contient un billet principal, éventuellement un ou plusieurs commentaires associés et des métadonnées (date de publication du billet ou du commentaire, url, pseudonyme de l'auteur, etc).

4.1 Corpus

4.1.1 Extraction du corpus

Pour les besoins du travail informatique détaillé dans la section 5, nous distinguons deux sous-corpus : un corpus d'entraînement et un corpus de test.

Corpus d'entraînement Nous constituons un premier corpus contenant 200 documents extraits à partir de deux types de requêtes :

- via l'utilisation de 30 thématiques internes d'OverBlog (*actualité*, *politique*, *santé*, etc),
- via l'utilisation de 10 mots-clés (*Wii*, *Poutine*, *Nucléaire*, etc).

L'objectif de cette extraction est de disposer d'un corpus qui soit le plus large possible en termes de thématiques abordées, afin de ne pas influencer vers un domaine particulier la constitution du lexique et des structures évaluatives. D'autres contraintes ont été fixées comme la longueur des documents (environ 700 mots), la diversité des auteurs ou encore l'élimination de blogs trop descriptifs comme les programmes télévisés, la succession de liens vers des vidéos.

Corpus de test Selon les mêmes contraintes de longueur et de pertinence, une seconde extraction a donné lieu à la sélection de 50 documents, nous l'appelons *corpus de test*. Ces documents sont extraits à partir de l'utilisation des mots-clés : *Sarah Palin* et *Sushi*. Parmi ces documents, 25 contiennent l'expression *Sarah Palin*, 25 contiennent le mot *sushi*.

4.1.2 Analyse du corpus

L'annotation du corpus a été réalisée manuellement par une linguiste, après une phase commune d'annotation de plusieurs billets par 4 annotateurs ayant mené à l'élaboration d'une DTD XML précise et d'un manuel détaillé d'annotations. Son expertise consiste à annoter tous les segments textuels exprimant une évaluation et tous les concepts évalués selon la DTD XML déterminée. Cette DTD considère en particulier : les cinq modalités évaluatives (Opinion, Accord-Désaccord, Acceptation-Refus, Jugement Appréciation) et les trois configurations énonciatives (explicites, implicites, exclamatives) décrites par Charaudeau, ainsi que l'aspect axiologique (positif/négatif) des évaluations établi par Galatanu. Nous renvoyons à (Dubreil et al., 2008) pour une description plus précise des annotations utilisées et de la méthodologie d'annotations.

[10] http://www.over-blog.com

Corpus d'entraînement L'annotation du corpus d'entraînement permet ainsi d'obtenir 4945 exemples d'évaluations. La répartition des modalités d'évaluations, des configurations d'énonciations des appréciations, et de l'axiologie favorable ou défavorable sont présentées dans les tableaux 1, 2 et 3 pour chaque corpus.

	Opinion	appréciation	Jugement	Accord/Des.	Acceptation/Refus
Entrainement	270	4163	143	351	18
Test (sushi)	28	475	-	34	-
Test (Sarah Palin)	32	376	-	10	-

TAB. 1 – *Répartition des modalités d'évaluations*

Ressource	Implicite	Explicite	Exclamative
Entrainement	3314	309	540
Test (sushi)	299	55	121
Test (Palin)	315	20	41

TAB. 2 – *Config. énonciative des appréciations*

Ressource	Favorable	Défavorable
Entrainement	2507	1656
Test (sushi)	355	120
Test (Palin)	171	205

TAB. 3 – *Axiologie des évaluations*

Corpus de test Dans le corpus de test, nous concentrons pour le moment notre étude aux modalités d'opinion, d'accord/désaccord et appréciation (nous regroupons les modalités d'appréciation et de jugement dans une seule catégorie). 846 instances d'évaluations ont ainsi été annotées (voir tableaux 1, 2 et 3). Dans ces tableaux, nous distinguons les textes extraits par mots clés. Il est toutefois important de préciser que ces textes n'évoquent et n'évaluent pas spécifiquement les *sushis* et *Sarah Palin*, d'autres concepts sont évalués.

À partir des annotations du corpus d'entraînement, et seulement de celui-ci, nous regroupons les entités lexicales fréquentes ou spécifiques au langage évaluatif dans un lexique.

4.2 Lexique de l'évaluation

Dans ce lexique, il s'agit en particulier de regrouper les mots ou les expressions :

- qui marquent spécifiquement une modalité de l'évaluation (*adorer*, *utile* comme Appréciation ; *penser*, *douter* comme Opinion ; *D'accord*, *Pourtant* comme . Accord-Desaccord ...),
- qui marquent spécifiquement une configuration énonciative (*adorer*, *apprécier* comme explicite ; *Wahou*, *Ouf* comme exclamative ...),
- qui ont une signification intrinsèquement axiologique (*adorer*, *utile* comme positive ; *détester*, *ennuyant* comme négative ; *oser*, *terrible* comme ambigüe).

La création du lexique de l'évaluation a consisté à exploiter les données annotées dans le corpus *d'entraînement* afin d'extraire des termes (des unigrammes, des expressions ou des syntagmes) candidats pour ce lexique. Chaque entrée du lexique de l'évaluation est issue d'une

instance d'évaluation annotée dans le corpus en y intégrant des informations structurées sur la valeur évaluative prise par cette entrée. Cette structuration répond à la nécessité d'une utilisation du lexique dans un cadre applicatif tout en assurant une cohérence linguistique indépendante de ce cadre. Le format de données XML a été retenu dans cette perspective. Le corpus annoté est traité, automatiquement et manuellement, afin de donner lieu à un ensemble de candidats à l'entrée dans le lexique final (Stern, 2008).

La recherche d'entrées potentielles ne s'est cependant pas limitée aux entrées simples ; le parcours manuel du fichier a également permis de repérer un certain nombre de formes linguistiques figées à différents degrés. Une catégorie complexe a alors été associée à l'entrée, avec l'idée de rendre compte de sa compositionnalité ultérieurement, au niveau de l'intégration des entrées au lexique. Exemples : *chapeau bas !*, *descente aux enfers*. Au final, le lexique se compose de 971 entrées dont essentiellement des adjectifs (493), des verbes (192) et des noms (166).

L'annotation des évaluations dans le corpus d'entraînement et le lexique de l'évaluation constituent deux types de ressources à la base de notre approche de catégorisation des évaluations expliquée dans la section suivante. Nous complétons nos ressources par deux ressources lexicales légères constituées de marqueurs d'intensité (*véritablement*, *incroyablement*, *terriblement*, etc) et de marqueurs de négations (*ne*, *pas*, *rien*, *jamais*, etc).

5 Catégorisation des évaluations par approche symbolique

Notre méthode de fouille d'opinion a pour objectif d'apprendre automatiquement des structures linguistiques spécifiques à l'expression d'une évaluation, puis à exploiter la généricité des structures apprises pour détecter les évaluations localement exprimées dans un texte. Les différentes utilisations de ces structures dans le corpus d'entraînement permet de leur associer des règles pour catégoriser la modalité, la configuration d'énonciation et l'axiologie des évaluations. Nous utilisons *UIMA*, une plateforme dédiée à la structuration de documents et au TAL. Elle permet, dans une même chaîne de traitement, de créer des annotations sur un objet linguistique et de les échanger entre plusieurs composants dans un format normalisé (XMI). Nous présentons les deux principaux moteurs d'analyses UIMA qui permettent : l'extraction de structures évaluatives et la catégorisation des évaluations dans un texte.

5.1 Extraction de structures évaluatives

Le composant d'extraction de structures évaluatives reçoit en entrée le flux de texte d'un document à analyser. Ce flux de texte a été préalablement annoté automatiquement par un étiqueteur grammatical (TreeTagger) et par la projection des différentes ressources lexicales (voir 4.2). Dès lors, le flux de texte est vu, par le composant, comme une succession de symboles à plusieurs niveaux linguistiques.

5.1.1 Symboles

Afin de prendre en compte les différents aspects des théories sur l'évaluation présentées dans la section 3, nous considérons quatre types de symboles :

– marqueurs d'**évaluation** : les termes issus du lexique de l'évaluation,

– marqueurs de **négation** : termes issus du lexique de la négation,
– marqueurs d'**intensité** : termes issus du lexique de l'intensité,
– les autres **mots** : termes qui n'existent dans aucun de ces lexiques.

Une fois l'étiquetage grammatical et la projection lexicale réalisés, chaque symbole possède une structure de traits à trois niveaux d'abstractions linguistiques : lexical (forme et lemme), grammatical (pos) et sémantique (type du symbole, modalité, configuration énonciative et axiologie) (voir fig.1 structure de traits des symboles *des*, *plus* et *utile*). Nous nous intéressons en particulier aux suites de symboles qui ont été annotées manuellement comme évaluation.

5.1.2 Apprentissage des structures évaluatives

L'apprentissage des structures évaluatives se déroule en deux étapes : l'extraction des évaluations qui ont été annotées manuellement dans le corpus d'entraînement, puis la généralisation symbolique de celles-ci afin de produire des règles de catégorisation.

Extraction des structures évaluatives Cette étape consiste à extraire les 4945 exemples d'évaluations annotées dans le corpus d'entraînement et à générer pour chacune d'elle la chaîne évaluative correspondante (voir exemple fig.1).

$$\begin{bmatrix} \text{lex} & \begin{bmatrix} \text{forme} & \text{'des'} \\ \text{lem} & \text{'du'} \end{bmatrix} \\ \text{gram} & \begin{bmatrix} \text{pos} & \text{'det'} \end{bmatrix} \\ \text{sem} & \begin{bmatrix} \text{type} & \text{'mot'} \\ \text{modal} & \text{' '} \\ \text{config} & \text{' '} \\ \text{axiol} & \text{' '} \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \text{lex} & \begin{bmatrix} \text{forme} & \text{'plus'} \\ \text{lem} & \text{'plus'} \end{bmatrix} \\ \text{gram} & \begin{bmatrix} \text{pos} & \text{'adv'} \end{bmatrix} \\ \text{sem} & \begin{bmatrix} \text{type} & \text{'mot'} \\ \text{modal} & \text{' '} \\ \text{config} & \text{' '} \\ \text{axiol} & \text{' '} \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \text{lex} & \begin{bmatrix} \text{forme} & \text{'utile'} \\ \text{lem} & \text{'utile'} \end{bmatrix} \\ \text{gram} & \begin{bmatrix} \text{pos} & \text{'adj'} \end{bmatrix} \\ \text{sem} & \begin{bmatrix} \text{type} & \text{'evaluation'} \\ \text{modal} & \text{'appreciation'} \\ \text{config} & \text{'implicite'} \\ \text{axiol} & \text{'positif'} \end{bmatrix} \end{bmatrix}$$

FIG. 1 – *Chaîne évaluative obtenue pour l'évaluation des plus utile.*

Généralisation Pour chaque chaîne évaluative extraite, il s'agit de générer sa structure symbolique en opérant un gain de généricité par le biais des règles suivantes (voir exemple fig. 2) :

– Généralisation de la valeur des traits **axiol**, **forme** et de **lemme** (X sur la fig. 2) pour tous les symboles de **type** *évaluation* et de **modalité** *appréciation*,

  En effet, certaines chaînes évaluatives annotées ayant des structures proches peuvent être regroupées puis généralisées pour couvrir d'autres chaînes non présentes dans notre corpus d'apprentissage. Par exemple, *des plus inintéressant*, *des plus utile* possèdent la même structure *des plus X* où X est un symbole évaluatif de modalité *appréciation* (voir fig. 2). Cette nouvelle structure symbolique permettra de détecter de nouvelles chaînes évaluatives comme *des plus inutile*, *des plus intéressant ...*

– Ajout de l'opérateur standard ***** sur les symboles de **type** *intensité* et généralisation de la valeur des traits **forme** et de **lemme** de ces symboles,

  Par exemple, *véritablement des plus utiles* et *des plus utiles* sont couverts par la même structure évaluative *Y* des plus X* où Y est un adverbe d'intensité et X un symbole évaluatif de modalité *appréciation*.

– Ajout de l'opérateur standard **+** (une ou plusieurs fois) pour les symboles de **config** *explicite* et de **pos** *pronom* et généralisation de la valeur des traits **forme** et de **lemme** de ces symboles.

  Par exemple, nous considérons *j'aime* et *moi j'aime* comme la même structure évaluative.

$$\begin{bmatrix} \text{lex} & \begin{bmatrix} \text{forme} & \text{'des'} \\ \text{lem} & \text{'du'} \end{bmatrix} \\ \text{gram} & \begin{bmatrix} \text{pos} & \text{'det'} \end{bmatrix} \\ \text{sem} & \begin{bmatrix} \text{type} & \text{'mot'} \\ \text{modal} & \text{' '} \\ \text{config} & \text{' '} \\ \text{axiol} & \text{' '} \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \text{lex} & \begin{bmatrix} \text{forme} & \text{'plus'} \\ \text{lem} & \text{'plus'} \end{bmatrix} \\ \text{gram} & \begin{bmatrix} \text{pos} & \text{'adv'} \end{bmatrix} \\ \text{sem} & \begin{bmatrix} \text{type} & \text{'mot'} \\ \text{modal} & \text{' '} \\ \text{config} & \text{' '} \\ \text{axiol} & \text{' '} \end{bmatrix} \end{bmatrix} \quad \begin{bmatrix} \text{lex} & \begin{bmatrix} \text{forme} & \text{X} \\ \text{lem} & \text{X} \end{bmatrix} \\ \text{gram} & \begin{bmatrix} \text{pos} & \text{'adj'} \end{bmatrix} \\ \text{sem} & \begin{bmatrix} \text{type} & \text{'evaluation'} \\ \text{modal} & \text{'appreciation'} \\ \text{config} & \text{'implicite'} \\ \text{axiol} & \text{X} \end{bmatrix} \end{bmatrix}$$

FIG. 2 – *Structure évaluative obtenue par la généralisation de la chaîne évaluative des plus utile.*

Règles de catégorisation associées aux structures Pour chaque structure symbolique générée, le composant produit un ensemble de méta-données qui serviront lors de la catégorisation. La figure 3 représente les méta-données associées à la structure générée par *des plus utile*.

Pour calculer la valeur du trait *tournure*, le composant considère que :
– les structures qui conservent la polarité du symbole axiologique utilisé sont directes,
– les structures qui inversent la polarité du symbole axiologique utilisé sont inversives,
– les structures qui n'activent pas de symbole axiologique sont figées.

Nous entendons par *tournure*, le fait que certaines structures peuvent être
– **directes** : *pour notre plus grand* ***plaisir*** positive avec *plaisir* positif
– **inversives** : *loin d'être* ***génial*** négative avec *génial* positif
– **figées** : *faire taire la critique*, *une goutte d'eau dans un océan*.

Dans le cas des structures figées, une méta-donnée supplémentaire est ajoutée pour déterminer la polarité générale de la structure. Par exemple, *en mettre plein la vue* est positif, *descente aux enfers* est négatif. Les structures et leurs méta-données sont conservées dans une ressource au format XML. Elles serviront de règles lors de la catégorisation. Le nombre de structures obtenues à partir du corpus d'entraînement est résumé dans le tableau 4 en les répartissant par modalité.

- un identifiant (id),
- le nombre d'occurences de cette structure annotées manuellement comme une évaluation (type - evaluation) et non annotées comme une évaluation (type - nonEvaluation) mais présentes dans le corpus d'entraînement,
- le nombre d'occurences de cette structure pour chaque modalité (modal) et pour chaque configuration d'énonciation (config),
- la *tournure* de la structure (tournure).

$$\begin{bmatrix} \text{id} & \\ \text{type} & \begin{bmatrix} \text{evaluation} & \text{'6'} \\ \text{nonEvaluation} & \text{'0'} \end{bmatrix} \\ \text{modal} & \begin{bmatrix} \text{appreciation} & \text{'6'} \\ \text{opinion} & \text{'0'} \\ \text{acc-des} & \text{'0'} \end{bmatrix} \\ \text{config} & \begin{bmatrix} \text{explicite} & \text{'0'} \\ \text{implicite} & \text{'6'} \\ \text{exclamatif} & \text{'0'} \end{bmatrix} \\ \text{tournure} & \begin{bmatrix} \text{directe} & \text{'6'} \\ \text{inversive} & \text{'0'} \\ \text{figee} & \text{'0'} \end{bmatrix} \end{bmatrix}$$

FIG. 3 – *Méta-données associées.*

Modalité	**Opinion**	**Appréciation**			**Acc/Désaccord**		
		exp. figée	inversif	direct	accord	désaccord	rectificatif
nb structures	49	1769	253	721	11	8	19

TAB. 4 – *Répartition des structures évaluatives apprises.*

5.2 Catégorisation des évaluations

Le deuxième moteur d'analyse UIMA que nous décrivons consiste, à partir des règles apprises, à détecter et catégoriser automatiquement les évaluations du corpus.

5.2.1 Détection des évaluations

Principe La tâche de détection consiste à parcourir itérativement chacun des documents du corpus et à analyser le flux de texte pour annoter les segments évaluatifs sans les catégoriser. Nous considérons que ces segments évaluatifs sont d'un niveau intra-phrastique.

Algorithme **Pour** chaque phrase du corpus **Faire** :
- Transformation de la phrase en chaîne symbolique (n symboles dans la phrase),
- Recherche des chaînes évaluatives présentes dans la phrase :
  - i = 1 (i étant la position du symbole courant)
  - **Tant que** i <= n **Faire** :
    - Recherche d'unification d'une chaîne symbolique à partir du symbole courant (en position i) avec les structures évaluatives apprises (la plus longue possible)
    - **Si** unification possible entre i et j **Alors** annotation de la chaîne et i = j+1
    - **Sinon** i = i+1 (on regarde le symbole suivant) **FinSi**
  - **Fin Tant que**

FinPour

Par ailleurs, nous attribuons un coefficient de confiance à la tâche de détection à l'aide des méta-données de S_i.

$$\alpha_{detection}(S_i) = \frac{Eval(S_i)}{Eval(S_i)+NonEval(S_i)}$$

$Eval(S_i)$ correspond à la valeur de la métadonnée *evaluation* de la structure S_i. $NonEval(S_i)$ est la valeur de la métadonnée *nonEvaluation* (fig. 4) de la structure S_i obtenue lors de la phase d'entraînement.

Résultats Pour maximiser la précision de notre méthode, nous rejettons les structures symboliques dont le coefficient est inférieur à 0,80. Cette borne est obtenue lors de la phase d'entraînement, elle permet d'obtenir la meilleure précision de détection (89,8%) sur le corpus d'entraînement sans affecter le rappel de façon disproportionné. Ainsi, lors de cette phase d'entraînement, l'outil détecte 70,2% des évaluations annotées (mesure du rappel), mais il faut garder à l'esprit que les différentes ressources lexicales ont été développées spécifiquement sur ce corpus d'entraînement.

De ce fait, nous testons la détection des évaluations sur le corpus test, pour lequel nous n'avons aucune connaissance lexicale en amont. Les résultats obtenus sont résumés dans le tableau 5. Nous considérons comme bien détectées, les évaluations correctement délimitées automatiquement, et celles dont une des deux bornes est erronée d'un ou deux mots d'écarts par rapport à la délimitation manuelle[11]. Les structures symboliques apprises semblent être des indicateurs assez précis pour détecter les évaluations, nous obtenons une précision similaire à celui de la phase d'entraînement.

Tâche	**Précision**	**Rappel**
DETECTION	88,4% (478/541)	50,1% (478/955)

TAB. 5 – *Résultats de la détection des évaluations sur le corpus test.*

En revanche et de manière assez prévisible, le rappel chute. Cette chute s'explique par plusieurs facteurs :

- un facteur **morphologique** : la nature des blogs et le langage sms, font que de nombreuses variations orthographiques sont possibles pour certains mots. Quelques mots évaluatifs échappent donc à notre outil pour ces raisons (ex : *j'adÔoore*).
- un facteur **lexical** : beaucoup d'adjectifs évaluatifs n'ont pas été rencontrés lors de l'entraînement, l'outil ne possède aucune information sémantique sur eux et ne les annotent donc pas. Ces adjectifs sont assez spécifiques aux sujets évalués (ex : *télégénique*, *énergique*, *puritaine*, *discrédité*), il serait donc particulièrement intéressant d'être capable de les détecter. Toutefois, nous ne pouvons pas considérer tous les adjectifs d'un texte comme étant évaluatifs (*un produit scalaire*, *une voiture rouge*), la mesure de précision chuterait fortement. Il faut pouvoir détecter les adjectifs qui apparaissent dans un contexte subjectif.
- un facteur **discursif** : en particulier dans les textes sur Sarah Palin, les internautes choisissent d'évaluer à partir de nouvelles structures qui correspondent la majorité du temps à des stéréotypes sociaux, à des expressions figées ou à des figures de styles particuliè-

[11] Par exemple, nous considérons correcte la détection automatique suivante : *la Wii risque vite de* ***devenir introuvable***, alors que l'annotation manuelle est : *la Wii risque vite de devenir* ***introuvable***.

rement nombreuses et que nous n'avions pas rencontrées durant la phase d'entraînement (ex : *sa fibre écolo*, *pitbull aux lèvres rouges*, *fardeau*, *bon débarras*, etc).

Dans la section 6, nous envisageons des perspectives à travers ces trois facteurs pour améliorer le rappel de la détection des évaluations. Par ailleurs, nous pouvons déjà noter que les opinions et les accords/désaccords sont détectées plus facilement en raison d'une moins grande variation lexicale et discursive pour ces types d'évaluations. Nous précisons ce constat dans le paragraphe suivant.

5.2.2 Catégorisation de la modalité des évaluations

Principe La catégorisation de la modalité consiste à décider automatiquement si l'évaluation détectée préalablement est : une opinion, une appréciation ou un accord/désaccord. Nous tirons profit de la phase d'entraînement supervisée pour calculer un coefficient de confiance pour chaque modalité selon la formule suivante :

$$\alpha_{catModalite}(S_i, M_j) = \frac{Eval(S_i,M_j)}{Eval(S_i,M_j)+NonEval(S_i,M_j)}.$$

$Eval(S_i, M_j)$ correspond au nombre de fois pour lequel la structure S_i a été annotée comme une évaluation de modalité M_j. La structure S_i est celle qui a servi lors de la détection de l'évaluation. La modalité qui possède le coefficient le plus élevé est choisie. Le composant de catégorisation de la modalité se charge de ce calcul et complète la structure de traits de l'évaluation détectée en précisant sa modalité.

Résultats Nous observons qu'il y a très peu d'ambiguïté entre modalités (TAB. 6)[12].

Modalité	**Précision**	**Rappel**
Opinion	88,0% (44/50)	100,0% (44/44)
Appreciation	100,0% (393/393)	99,0% (393/397)
Accord/Désaccord	100,0% (35/35)	94,6% (35/37)

TAB. 6 – *Résultats de la catégorisation des modalités sur les évaluations correctement détectées du corpus test.*

Les structures et les entités lexicales qui composent ces différents types d'évaluations sont bien distincts. De ce fait, les résultats obtenus sont presque maximales et viennent corroborer les définitions théoriques données par Charaudeau. La catégorisation automatique de la modalité est toutefois légèrement ambigüe entre les opinions et les accords/désaccords. La distinction sémantique entre *se positionner dans son propre univers de croyance* et *se positionner dans un univers de croyance pré-existant* est assez fine et provoque quelques erreurs de catégorisation. Certaines structures sont utilisées dans les deux modalités.

Nous pouvons également observer que :

– sur les 60 opinions annotées manuellement, 44 ont été correctement reconnues et catégorisées (rappel : 73,3%),

[12]Nous précisons que les résultats concernant cette tâche ne tiennent pas compte des évaluations non détectées, pour lesquelles la tâche de catégorisation de la modalité est d'avance un échec.

– sur les 851 appréciations annotées manuellement, 393 ont été correctement reconnues et catégorisées (rappel : 46,2%),
– sur les 44 accords/désaccords annotées manuellement, 35 ont été correctement reconnues et catégorisées (rappel : 79,5%).

La tâche critique concerne donc la détection des évaluations dont la modalité est l'appréciation. Nous détaillons ce constat dans le paragraphe suivant en nous intéressant à la catégorisation de la configuration d'énonciation des appréciations.

5.2.3 Catégorisation de la configuration d'énonciation des appréciations

Principe La catégorisation de la configuration d'énonciation consiste à décider automatiquement si une appréciation détectée est : implicite, explicite ou exclamative. Nous rappellons qu'il s'agit d'estimer le degré de subjectivité avec lequel s'exprime le locuteur en observant comment il positionne son identité énonciative par rapport à son propos. De la même façon que précédement, nous nous servons de la phase d'entraînement supervisée pour calculer un coefficient de confiance pour chaque configuration selon la formule suivante :

$$\alpha_{catConfigEnon}(S_i, appreciation, CE_j) = \frac{Eval(S_i, appreciation, CE_j)}{Eval(S_i, appreciation, CE_j) + NonEval(S_i, appreciation, CE_j)}.$$

$Eval(S_i, appreciation, CE_j)$ correspond au nombre de fois pour lequel la structure S_i a été annotée comme appréciation avec la configuration énonciative CE_j. La structure S_i est celle qui a servi lors de la détection de l'appréciation. La configuration d'énonciation qui possède le coefficient le plus élevé est choisie. Le composant de catégorisation se charge de ce calcul et complète la structure de traits de l'évaluation détectée en précisant la configuration énonciative de l'appréciation. Lorsque la phrase analysée est exclamative, nous appliquons une heuristique supplémentaire : le dernière appréciation de la phrase est exclamative quelque soit les coefficients de confiance de chaque configuration d'énonciation.

Résultats La tâche de catégorisation de la configuration d'énonciation des appréciations fournit de bons résultats (TAB. 7).

Configuration énonciative	Précision	Rappel
Explicite	89,6% (60/67)	96,8% (60/62)
Implicite	97,6% (290/297)	97,9% (290/296)
Exclamative	100,0% (29/29)	82,9% (29/35)

TAB. 7 – *Résultats obtenus sur la catégorisation de la configuration énonciative sur les évaluations correctement détectées du corpus test.*

De la même façon que pour les modalités, les structures symboliques sont particulièrement différentes entre les appréciations implicites et explicites. Les appréciations explicites sont en particulier reconnaissables grâce à l'utilisation de pronoms (je, nous, mon, etc) et de verbes d'appréciations (aimer, détester, apprécier, etc). Lorsque ces symboles ne sont pas présents, il s'agit très souvent d'une appréciation implicite. Les résultats obtenus indique que l'ambiguïté se situe plutôt sur la catégorisation des appréciations exclamatives. Il n'est pas rare de trouver plusieurs appréciations dans une même phrase exclamative ; dans ce cas, il est assez difficile

de savoir si l'exclamation porte sur toutes les appréciations de la phrase ou seulement sur la dernière.

A travers les résultats obtenus, nous observons les faits suivants :

- sur les 614 appréciations implicites annotées manuellement, seulement 290 ont été correctement reconnues et catégorisées (rappel : 47,2%),
- sur les 162 appréciations exclamatives annotées manuellement, seulement 29 ont été correctement reconnues et catégorisées (rappel : 17,9%),
- sur les 75 appréciations explicites annotées manuellement, 60 ont été correctement reconnues et catégorisées (rappel : 80,0%).

Nous mesurons d'avantage la difficulté à détecter les appréciations implicites avec le lexique de l'évaluation à notre disposition. Sans apprentissage préalable sur le corpus test, l'outil est toutefois capable de reconnaître la moitié des appréciations implicites. Le faible rappel pour la détection des appréciations exclamatives s'explique par l'utilisation fréquente du langage SMS ou de formes d'expressions orales (*waahhh ! !*, *oufff !*, etc) dont les retranscriptions écrites sont très variables. Enfin, les appréciations explicites sont, par nature, les plus aisément détectables.

5.2.4 Catégorisation axiologique des appréciations

Principe La catégorisation de l'axiologie consiste à décider si une appréciation est positive, négative ou ambigüe. Nous laissons en effet la possibilité à l'outil de décider qu'une évaluation est axiologiquement ambigüe. Dans le lexique de l'évaluation, certaines entités lexicales évaluatives sont potentiellement polarisées différemment selon le domaine, le contexte ou le stéréotype culturel mis en jeu (ex : *Elle **ose** les choses les plus **folles**, Une omelette sucrée **surprenante**, la visite est **terrible***). Nous décidons pour le moment de ne pas chercher à désambiguiser ces évaluations mais nous notons tout de même leur occurence. Pour décider de la polarité axiologique d'une appréciation, nous nous appuyons sur les tournures des règles apprises. Nous rappelons que celles-ci peuvent être *inversives*, *directes* ou *figées*. Une évaluation détectée à l'aide d'une règle :

- à structure directe, aura la même polarité axiologique que le symbole axiologisé,
- à structure inverse, aura une polarité axiologique inverse au symbole axiologisé,
- à structure figée, aura la polarité axiologique indiquée dans les méta-données de la structure apprise.

Le composant de catégorisation axiologique tient compte de ces règles pour compléter la structure de traits de chaque appréciation en ajoutant une information sur l'axiologie.

Résultats La tâche de catégorisation axiologique fournit des résultats encourageants (TAB. 8).

Suite à la phase d'entraînement, nous notons qu'il n'y a presque pas de structures symboliques qui soient à la fois une tournure inversive et une tournure directe. Dès lors, il semble logique qu'il y ait très peu d'ambiguïté pour catégoriser l'axiologie d'une expression évaluative locale. Les résultats montrent néanmoins que le rappel des appréciations défavorables est plus faible. Autrement dit, les appréciations défavorables ont tendance à être catégorisées comme favorables. Nous expliquons ces erreurs à travers deux facteurs :

- un facteur **syntaxique** : certaines structures inversives n'ont pas été apprises durant l'entraînement, notamment lorsque les marques de négation sont éloignées de l'évaluation.

Axiologie	**Précision**	**Rappel**
Favorable	93,3% (279/299)	97,6% (279/286)
Défavorable	96,5% (83/86)	77,6% (83/107)
Ambigüe	8 occurences	

TAB. 8 – *Résultats obtenus pour la catégorisation axiologique sur les évaluations correctement détectées du corpus test.*

- un facteur **discursif** : de nombreux cas d'évaluations ironiques occurent au sujet de Sarah Palin (*voilà un bel exemple*, *toujours aussi passionnant*) et des sushis (*merci de m'avoir foutu la gerbe*). Ici, la sémantique lexicale (*bel*, *passionnant*, *merci*) induit en erreur la catégorisation axiologique.

Il est d'ailleurs intéressant de noter que les structures inversives et ironiques sont plus souvent activées pour évaluer défavorablement que favorablement. Nous constatons également que :

- sur les 526 appréciations favorables annotées manuellement, 279 ont été correctement reconnues et catégorisées (rappel : 53,0%),
- sur les 325 appréciations défavorables annotées manuellement, 83 ont été correctement reconnues et catégorisées (rappel : 25,5%).

En plus des deux facteurs précédemment évoqués, nous pouvons en ajouter un troisième par l'utilisation de figures de styles métaphoriques qui perturbe la détection des appréciations défavorables. Les appréciations portant sur Sarah Palin, en particulier celles qui sont défavorables, utilisent très souvent des figures de styles métaphoriques (ex : *se faire lyncher*, *une clown*, *bouc émissaire*, *pitbull aux lèvres rouges*), ce qui a pour effet de limiter leur détection automatique. Différents travaux ont déjà fait état de l'importance des figures de styles métaphoriques dans le discours évaluatif de part la façon dont elles témoignent des représentations mentales du locuteur et donc de sa subjectivité. Toutefois, nous préférons pour le moment accentuer nos efforts sur la reconnaissance des évaluations non métaphoriques.

6 Perspectives & conclusion

La poursuite de nos travaux et l'amélioration de l'outil se scindent en deux axes :

- l'amélioration automatique de la couverture lexicale pour une meilleure détection,
- la détection du couple sujet évalué - évaluation afin de filtrer les évaluations ne portant pas sur les sujets cibles.

La figure 4 présente les résultats obtenus par la catégorisation axiologique automatique sur le corpus test. Comparativement à la figure 5 (les résultats théoriques obtenus par la catégorisation manuelle), nous observons que la qualité de la catégorisation souffre d'un déséquilibre dans la proportion des évaluations positives et négatives détectées. Nous pensons qu'une meilleure couverture lexicale permettra de rétablir cette proportion. Pour réaliser ce premier objectif, nous envisageons une méthode non supervisée pour apprendre de nouveaux termes et de nouvelles expressions figées utilisés pour évaluer subjectivement. L'association fréquente d'un terme ou d'une expression avec les structures subjectives apprises est un indice permettant d'induire leur rôle évaluatif. En s'inspirant de Hatzivassiloglou et McKeown (1997), il

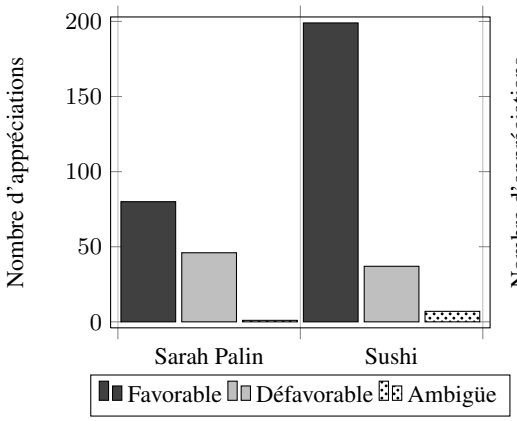

FIG. 4 – *Cat. automatique de l'axiologie.*

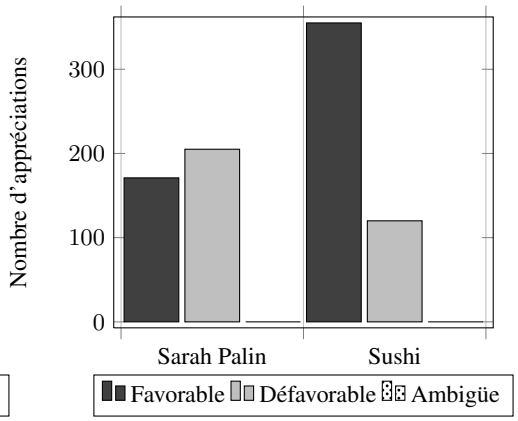

FIG. 5 – *Cat. manuelle de l'axiologie.*

est également possible de déterminer automatiquement la polarité axiologique des termes et expressions ainsi extraites.

Le second objectif tient compte du fait que les évaluations détectées ne concernent pas toutes *Sarah Palin* ou les *Sushis*. Beaucoup d'autres concepts sont évalués : Barack Obama, la politique, des canulards téléphoniques pour piéger Sarah Palin, des restaurants, des recettes, des voyages, etc. La piste de réflexion suivie actuellement consiste à calculer une proximité entre les évaluations détectées et les groupes nominaux ou les pronoms anaphoriques les plus proches dans le texte selon une approche similaire à celles utilisées pour la résolution d'anaphores. Dans un second temps, il s'agira de filtrer les évaluations dont les sujets candidats ne concernent pas directement le sujet ciblé par la requête (ici, Sarah Palin et les sushis).

Par ailleurs, nous nous intéressons à mesurer le degré de subjectivité avec lequel un locuteur s'exprime et évalue un concept cible. Les figures 6 et 7 montrent les résultats de la catégorisation automatique et manuelle des modalités évaluatives utilisées pour évaluer dans les textes sur Sarah Palin et sur les sushis. Les résultats expérimentaux ont une légère tendance à confirmer l'observation manuelle suivante : comparativement à Sarah Palin, le discours évaluatif sur les sushis est plus fréquemment subjectivé, par l'utilisation des modalités d'appréciations explicites (*j'adore*, *je déteste*), et exclamatives (*miam*, *beurk !*. *c'est trop bon ! ! !*). En d'autres termes, le locuteur ne masque pas son identité énonciative (*je*) pour évaluer les sushis et ses évaluations sont énoncées de manière explicite et exclamative. Par opposition, le discours sur Sarah Palin induit souvent une stratégie argumentative. Il s'agit, pour le locuteur, de convaincre ses lecteurs que son opinion sur Sarah Palin est la bonne car plus objective. Ainsi, il utilise plutôt des appréciations de forme implicite (1), plutôt que de forme explicite (2) :

1. *Sarah Palin est* + **adjectif évaluatif**
2. **J'aime/Je déteste** *Sarah Palin*

Le rôle des évaluations de modalité d'opinion ou d'accord/désaccord (*Je pense, C'est véritablement, Je doute, etc.)* possède également un intérêt que nous souhaitons développer dans nos travaux futurs. De fait, ces évaluations sont souvent associées à une modalité d'appréciation et permettent de situer l'appréciation dans un univers de croyance plus ou moins fort.

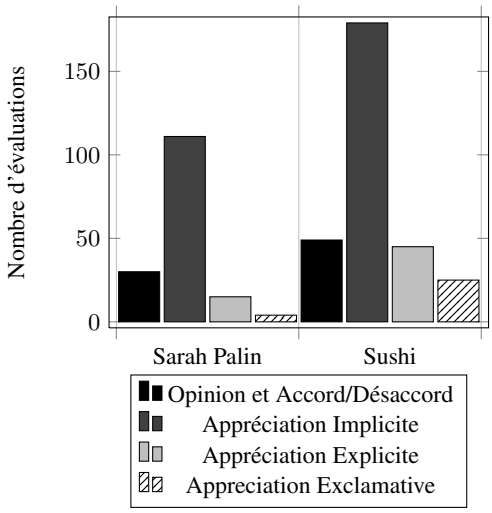

FIG. 6 – *Cat. automatique de la modalité.*

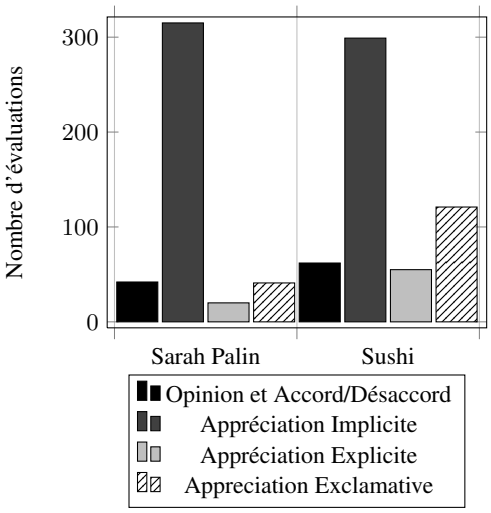

FIG. 7 – *Cat. manuelle de la modalité.*

Par exemple, l'appréciation *être une réussite* n'est pas la même dans les univers de croyance suivant :

1. **C'est véritablement** une réussite.
2. **Oui**, c'est une réussite.
3. **Je doute** que ce soit une réussite.
4. **Ce n'est pas véritablement** une réussite.

En conclusion, nous avons présenté une méthode de détection et de catégorisation des évaluations localement exprimées dans un corpus de blogs multi-domaines. Cette méthode s'appuie sur deux théories linguistiques modélisant le langage évaluatif et l'expression de la subjectivité dans le discours. Il fait suite à une expertise linguistique ayant mené à la création de deux types de ressources : un corpus contenant environ 5000 annotations d'évaluations et un lexique de l'évaluation composé d'environ 1000 entrées lexicales. La méthode proposée s'articule en deux étapes : une extraction supervisée de structures symboliques à trois niveaux d'abstraction linguistique (lexicale, grammaticale et sémantique) et la catégorisation des évaluations dans un corpus test à partir des structures apprises. Cette catégorisation porte sur trois aspects : la modalité, la configuration énonciative et l'axiologie des évaluations. Les résultats obtenus à partir de deux requêtes applicatives montrent des résultats encourageants et invitent à poursuivre la démarche générale de notre travail. En effet, sur des billets de domaines différents et d'auteurs différents de notre corpus d'entrainement, les structures évaluatives apprises permettent de détecter un certain nombre d'évaluations présentes dans ces nouveaux billets. Ces travaux futurs consistent à extraire les couples sujet évalué - évaluation évoqué par les internautes et à enrichir nos ressources par une méthode d'apprentissage non supervisé afin de détecter et catégoriser d'avantage d'évaluations et augmenter ainsi l'efficacité de notre outil.

Remerciements Ces travaux s'inscrivent dans le projet BLOGOSCOPIE, soutenu par le programme Technologies Logicielles 2006 de l'Agence Nationale de la Recherche, et réalisé en collaboration avec Syllabs, Sinequa et Over-Blog.

Références

Airoldi, E. M., X. Bai, et R. Padman (2006). Markov blankets and meta-heuristic search : Sentiment extraction from unstructured text. *Lecture Notes in Computer Science 3932 (Advances in Web Mining and Web Usage Analysis)*, 167–187.

Anscombre, J. et O. Ducrot (1983). *L'argumentation dans la langue*. Pierre Mardag.

Aue, A. et M. Gamon (2005). Customizing sentiment classifiers to new domains : A case study. In *Submitted to Recent Advances in Natural Language Processing (RANLP)*.

Bansal, M., C. Cardie, et L. Lee (2008). The power of negative thinking : Exploiting label disagreement in the min-cut classification framework. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, pp. 13–16. Poster paper.

Benveniste, E. (1974). Problèmes de linguistique générale. Gallimard.

Blitzer, J., M. Dredze, et F. Pereira. Biographies, Bollywood, boom-boxes and blenders : Domain adaptation for sentiment classification. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, pp. 440–447.

Charaudeau, P. (1983). Langage et discours. Hachette.

Charaudeau, P. (1992). *Grammaire du sens et de l'expression*. Hachette Education, COMMUNICATION, PARA UNIVERSITAIRE.

Dave, K., S. Lawrence, et D. M. Pennock (2003). Mining the peanut gallery : Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th World Wide Web Conference (WWW 2003)*, pp. 519–528.

Devitt, A. et K. Ahmad (2007). Sentiment polarity identification in financial news : A cohesion-based approach. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL 2007)*, Prague, Czech Republic, pp. 984–991.

Dubreil, E. (2006). *La dimension argumentative des collocations textuelles en corpus électronique spécialisé au domaine du TAL(N)*. Ph. D. thesis, Sciences du Langage : Connaissance Langages Cultures.

Dubreil, E., M. Vernier, L. Monceaux, et B. Daille (2008). Annotating opinion – evaluation of blogs. In *Workshop of LREC 2008 Conference, Sentiment Analysis : Metaphor, Ontology and Terminology (EMOT-08)*, pp. 124.

Esuli, A. et F. Sebastiani (2006). SentiWordNet : A publicly available lexical resource for opinion mining. In *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)*, pp. 417–422.

Ferrari, S. et D. Legallois (2006). Vers une grammaire de l'évaluation des objets culturels. In *Actes du colloque international discours et documents (ISDD 2006), Schedae, fasicule 1*, pp. 57–68. Presse Universitaire de Caen.

Ferrari, S., Y. Mathet, T. Charnois, et D. Legallois (2008). Analyse d'opinion : discours évaluatif et classification de documents. In *Actes de l'atelier FODOP'08 (fouille de données d'opinions) (INFORSID'08)*, pp. 23–36.

Galatanu, O. (2000). Signification, sens, formation. In *Education et Formation, Biennales de l'éducation, (sous la direction de Jean-Marie Barbier, d'Olga Galatanu)*, Paris. PUF.

Gamon, M. (2004). Sentiment classification on customer feedback data : noisy data, large feature vectors, and the role of linguistic analysis. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004)*, pp. 841–847.

Harb, A., G. Dray, M. Plantié, P. Poncelet, et M. R. F. Trousset (2008). Détection d'opinion : Apprenons les bons adjectifs ! In *actes de l'atelier FODOP'08 (fouille de données d'opinions) (INFORSID'08)*, pp. 59–66.

Hatzivassiloglou, V. et K. McKeown (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the Joint ACL/EACL Conference*, pp. 174–181.

Hatzivassiloglou, V. et J. Wiebe (2000). Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, pp. 299–305.

Hu, M. et B. Liu (2004). Mining opinion features in customer reviews. In *Proceedings of the 19th National Conference on Artificial Intelligence*, pp. 755–760. AAAI Press / The MIT Press.

Hurst, M. et K. Nigam (2004). Retrieving topical sentiments from online document collections. In *Document Recognition and Retrieval XI*, pp. 27–34.

Kerbrat-Orecchioni, C. (1997). L'Énonciation, de la subjectivité dans le langage. Colin (réédition 2002).

Kim, S.-M. et E. Hovy (2006). Automatic identification of pro and con reasons in online reviews. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pp. 483–490.

Kudo, T. et Y. Matsumoto (2004). A boosting algorithm for classification of semi-structured text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, pp. 301–308.

Lavelle, L. (1950). *Traité des valeurs*, Volume tome 1. PUF.

Maingueneau, D. (1991). L'analyse du discours. Hachette.

Martin, J. et P. White (2005). *The Language of Evaluation, Appraisal in English*. Palgrave Macmillan.

Mullen, T. et N. Collier (2004). Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of the 9th Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)*, pp. 412–418. Poster paper.

Pang, B., L. Lee, et S. Vaithyanathan (2002). Thumbs up ? sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pp. 79–86.

Popescu, A.-M. et O. Etzioni (2005). Extracting product features and opinions from reviews. In *Proceedings of the Human Language Technology Conference and the Conference on Em-*

pirical Methods in Natural Language Processing (HLT/EMNLP 2005), pp. 339–346.

Read, J., D. Hope, et J. Carroll (2007). Annotating expressions of appraisal in english. In *Proceedings of the Linguistic Annotation Workshop (LAW 2007)*, pp. 93–100. Association for Computational Linguistics.

Riloff, E. et J. Wiebe (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2003)*, pp. 105–112. Association for Computational Linguistics.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys 34*(1), 1–47.

Stern, R. (2008). Constitution d'un lexique de sentiment. Mémoire de Master de Recherche Linguistique-Informatique, Université Paris 7.

Thomas, M., B. Pang, et L. Lee (2006). Get out the vote : Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, pp. 327–335.

Turney, P. (2002). Thumbs up or thumbs down ? semantic orientation applied to unsupervised classification of reviews. pp. 417–424.

Vernier, M., Y. Mathet, F. Rioult, T. Charnois, S. Ferrari, et D. Legallois (2007). Classification de textes d'opinions : une approche mixte n-grammes et sémantique. In *actes de l'atelier DEFT'07 (défi fouille de textes) (AFIA'07)*, pp. 95–109.

Whitelaw, C., N. Garg, et S. Argamon (2005). Using appraisal groups for sentiment analysis. In *Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM)*, pp. 625–631. ACM.

Wiebe, J. et R. Mihalcea (2006). Word sense and subjectivity. *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (ACL'06)*, p.1065–1072.

Yang, K., N. Yu, A. Valerio, et H. Zhang (2006). WIDIT in TREC-2006 Blog track. In *Proceedings of the 15th Text REtrieval Conference (TREC 2006)*.

Yi, J., T. Nasukawa, R. Bunescu, et W. Niblack (2003). Sentiment analyzer : Extracting sentiments about a given topic using natural language processing techniques. In *ICDM '03 : Proceedings of the 3rd IEEE International Conference on Data Mining*, pp. 427.

Summary

Our works deal with appraisal language analysis, we propose a automatical method for appraisal detection and categorization in a multi-domain weblog corpus. Our approach takes into consideration two theories modelling appraisal expression in natural language and uses lexical ressources. We present two *UIMA* analyze engines for : automatical extraction of symbolic structures especially used for appraisal language and appraisal categorization using symbolic structures learned. In particular, the categorization tool aim to analyze axiological signification, attitude and enunciation configuration of an evaluation. We discuss about improvements we plan and about applications perspectives of this tool.