

Repérer les phrases évaluatives dans les articles de presse à partir d'indices et de stéréotypes d'écriture

Mathias Lambert

Université Paris IV-Sorbonne, Laboratoire STIH (LaLIC) - 28 rue Serpente, 75006 Paris
Mathias.Lambert@paris-sorbonne.fr

Résumé. Ce papier présente une méthode de recherche des phrases évaluatives dans les articles de presse économique et financière à partir de marques et d'indices stéréotypés, propres au style journalistique, apparaissant de manière concomitante à l'expression d'évaluation(s) dans les phrases. Ces marques et indices ont été dégagés par le biais d'une annotation manuelle. Ils ont ensuite été implémentés, en vue d'une phase-test d'annotation automatique, sous forme de grammaires DCG/GULP permettant, par filtrage, de matcher les phrases les contenant. Les résultats de notre première tentative d'annotation automatique sont présentés dans cet article. Enfin les perspectives offertes par cette méthode relativement peu coûteuse en ressources (à base d'indices non intrinsèquement évaluatifs) font l'objet d'une discussion.

Abstract. This paper presents a method to locate evaluative sentences in financial and economic newspapers, relying on marks and stereotyped signs. Peculiar to journalese, these are present concomitantly with the expression of evaluation(s) in sentences. These marks or signs have been found by means of a manual annotation. Then, in preparation for an automatic annotation phase, they have been implemented in the form of DCG/GULP grammars which, by filtering, allows to locate the sentences containing them. The results of our first automatic annotation attempt are shown in this article. Furthermore, the prospects offered by this method, which relies on non-intrinsically evaluative marks and therefore does not require long lists of lexical resources, are discussed.

Mots-clés : Opinion, évaluation, repérage de phrases évaluatives, presse économique et financière, style journalistique, indices/marques/stéréotypes d'écriture.

Keywords: Opinion, appraisal, detection of evaluative sentences, financial and economic newspapers, journalese, writing signs/marks/stereotypes.

1 Le repérage automatique d'évaluation : un thème de recherche stratégique d'actualité

Au cours des dix dernières années, le nombre de travaux, aussi bien à visées académiques qu'à visées industrielles, ayant pour thème la recherche informatisée des opinions, des évaluations, des attitudes, des sentiments ou des émotions exprimées dans les documents textuels a augmenté de manière significative. L'intérêt croissant que suscite ce thème atteste d'un réel besoin, notamment dans le domaine de la veille économique. Avec internet, la quantité de données disponibles (notamment dans le domaine de la presse) à traiter est aujourd'hui trop volumineuse pour un analyste humain seul. La demande d'outils informatiques de repérage ou d'aide au repérage des segments porteurs d'évaluation est donc très forte, car au cœur d'enjeux stratégiques et économiques importants.

La plupart des travaux actuels se limitent souvent à la simple détermination de polarités (positive, négative ou neutre) et n'arrivent à capter que partiellement les phénomènes visés qui recouvrent pourtant des significations riches, diversifiées et souvent complexes. Le but de notre recherche, qui s'inscrit plus largement au sein du projet ANR OntOpiTex¹, est d'avancer dans le domaine en cherchant à identifier, à agréger et à caractériser finement des segments textuels porteurs d'opinions, en fonction de plusieurs critères (valeur sémantique, source, intensité et force, type d'objet évalué...). Le présent article décrit une étude exploratoire visant à découvrir et à mettre en œuvre de nouvelles méthodes pour détecter les évaluations dans les textes. Nous avons cherché, à partir d'une annotation manuelle sur un corpus d'articles de la presse économique tirés du web, à mettre au jour des indices textuels et lexicaux récurrents voire stéréotypés apparaissant de manière concomitante à l'expression (ou au relais) d'évaluation(s). Ces indices ont ensuite été regroupés dans une dizaine de catégories, en fonction de critères syntaxiques et sémantiques. Une première implémentation, à l'aide de la plateforme SemioLabs², sous forme de grammaires locales d'unification a alors permis de tester, sur un nouveau corpus d'articles issus de la presse économique, la fiabilité des indices, et par la même la pertinence de notre méthode. Le principal intérêt de notre approche est que, contrairement aux autres travaux dans le domaine, elle est relativement peu coûteuse en ressources lexicales. Nous n'avons en effet pas eu à constituer de longues listes de ressources lexicales évaluatives (adjectifs, adverbes et/ou verbes).

1.1 Concept et domaine

Opinion, subjectivité, évaluation, attitude, jugement, appréciation, sentiment, émotion, affect, ..., la prolifération des termes pour nommer le concept semble, d'une part, révélatrice de la difficulté qu'ont les auteurs à l'appréhender et à le manipuler, et d'autre part, fortement liée à la diversité des travaux dans le domaine. De manière générale, les travaux en TAL relatifs au phénomène de l'évaluation sont en majorité ceux traitant de la fouille d'opinion et de l'analyse de sentiments. Ils s'inscrivent dans trois grandes catégories : (i) la constitution de ressources lexicales pour la fouille d'opinion ; (ii) la classification de textes et/ou de phrases (objectif vs subjectif et/ou positif vs négatif) ; (iii) l'analyse d'opinion dans les textes. Pour un état de l'art complet et détaillé de la question en TAL, voir (Pang et Lee, 2008).

Le travail présenté ici s'inscrit dans la deuxième catégorie : la classification de phrases (objectives vs subjectives). Pour être plus précis, il s'agit de repérage de phrases subjectives dans les textes. Pour faire référence à ces phrases, nous parlerons de *phrases évaluatives*. Cette appellation nous évitera de fait d'avoir à aborder la question délicate de la délimitation des segments évaluatifs. Quant à la problématique concernant ce qu'il faut considérer comme étant évaluatif ou non, nous la discuterons plus bas, dans la partie ayant trait à l'annotation.

1.2 Méthode

La méthode la plus triviale pour repérer des phrases évaluatives consiste à constituer manuellement (avec possibilité d'enrichissement semi-automatique) une liste de lexique évaluatif (principalement à base d'adjectifs et d'adverbes) et de projeter ce lexique sur les textes. [Bloom *et al*, 2007a] et [Bloom *et al*,

¹ Projet soutenu par l'ANR (2009 CORD 016) ; site : <https://ontopitex.greyc.fr/>

² Développée par la société Noopsis : www.noopsis.fr/

2007b] procèdent de cette manière³ pour extraire les phrases porteuses d'opinion, et les présentent dans un navigateur (Appraisal Navigator) qui offre à l'utilisateur la possibilité d'attribuer une étiquette aux phrases en fonction de catégories propres à la théorie de l'Appraisal de (Martin et White, 2005).

Certains auteurs, comme (Wiebe et al., 2002) et (Wiebe et al., 2005), en s'appuyant sur une liste d'éléments subjectifs (SE) récoltés à la suite d'une annotation manuelle, proposent de rechercher automatiquement dans les phrases des éléments potentiellement subjectifs ou PSE (hapax, collocations, adjectifs et verbes ayant des similarités distributionnelles avec des SE) et de décider ensuite si leur potentiel subjectif s'actualise ou pas en fonction de la densité de SE présents dans le cotexte. Ces travaux sont relativement proches de ceux de (Riloff et Wiebe 2003) qui utilisent une méthode de bootstrapping, où des phrases étiquetées évaluatives sont utilisées comme données d'apprentissage, pour produire automatiquement des modalités représentant des expressions subjectives et ainsi différencier les phrases objectives des phrases subjectives.

D'autres, comme (Yu et Hatzivassiloglou, 2003) qui s'intéresse à la classification de phrases et de propositions pour distinguer les opinions des faits dans un système de Question/Réponse, s'appuient essentiellement sur des techniques statistiques (approche par similarité, classifieur bayésien naïf, classifieur bayésien naïf multiple).

Notre méthode se distingue des précédentes dans la mesure où nous n'avons pas constitué de listes de ressources lexico-grammaticales évaluatives ni utilisé de procédés statistiques pour détecter les phrases porteuses d'évaluation(s). Notre idée est née d'une constatation suite à l'annotation manuelle. Etant donné que l'écriture journalistique se trouve soumise à de fortes contraintes (souci d'objectivisation, concision, clarté, ...), le journaliste a tendance à recourir à des tours spécifiques (formes morphosyntaxiques et/ou syntaxiques, marques rhétoriques ou emphatiques) récurrents lorsqu'il cherche à atténuer/dissimuler une évaluation. Ces formes, qui trahissent une certaine subjectivité, peuvent alors être considérées comme des indices ou des symptômes attestant de la présence d'évaluations.

2 Annotation, observations et hypothèse, expérience

2.1 Annotation

Cette tâche a été réalisée manuellement sur un corpus de 36 articles de presse économique collectés à partir de la base documentaire Factiva. Ces articles, répartis par groupes de 6, concernaient 6 entreprises/domaines (Apple, EDF-ENR, Goldman Sachs, Google, Total, les laboratoires pharmaceutiques). Pour distinguer ce qui est de l'ordre de l'évaluation de ce qui est du contenu factuel, nous nous sommes donné un critère simple : « si nous étions un investisseur, un analyste économique, ou tout autre type de partie prenante (au sens de 'stakeholder'), quels éléments seraient susceptibles de nous intéresser ? (« Is this relevant for my goals/needs? » (Bednarek, 2009, p.158)) ». Ci-dessous quelques exemples de phrases relevées :

- C'est un trophée que Christophe de Margerie peut se réjouir de perdre. (01total)
- « Les gens sont excédés par cette attente », explique Patrice Leclair, délégué syndical Force ouvrière. (04total)
- Mais ce rendez-vous incontournable des fans de la pomme, s'illustre, pour sa 25e édition, par l'absence d'Apple... (08apple)
- Le dieu de la finance semble s'être transformé en diable. (15glodmansachs)
- Pas facile de concilier morale et commerce, surtout en Chine... (21google)
- Certaines questions auraient dû être posées. (27labopharma)
- Gare aux désillusions en bourse ! (32EDFENR)

³ En s'appuyant sur la théorie Appraisal de (Martin et White, 2005) pour construire le lexique

2.2 Observations et hypothèse

A partir des phrases évaluatives relevées (plus de 300), les marques/indices ont été regroupés en 11 groupes/types se voulant syntaxiquement et/ou sémantiquement homogènes et cohérents :

1. Citation-Mise en relief (exemples d'indices : « », " ", d'après, selon) ;
2. Conjecture (exemples d'indices : verbes au conditionnel, verbes au futur simple, peut-être, sans doute, Reste (adv)⁴ à voir/savoir/attendre). En effet d'après (Wiebe et Wilson, 2002, p.112) : « Subjectivity in natural language refers to aspects of language used to express emotion, evaluation and **speculation** » ;
3. Constat (exemples d'indices : force est de, somme toute, en définitive, bref, finalement) ;
4. Opposition-Concession (exemples d'indices : mais, pourtant, néanmoins, toutefois, en revanche) ;
5. Construction attributive (exemples d'indices : verbes attributifs et assimilés). De fait, d'après (Wagner et Pinchon, 1962, p.147) : « l'attribut fait partie d'une phrase où l'on pose un jugement prédicatif. Il évoque une qualité qu'on reconnaît appartenir à une personne, à une chose, qu'on leur attribue. »
6. Phrase averbale (exemples d'indices : absence de verbe conjugué) ;
7. Qualité :(exemples d'indices : principal/premier/deuxième/second atout/avantage/point positif, ne manque (adv)⁵ pas de) ;
8. Rang (exemples d'indices : le géant/champion/leader (adj)* (de)?) ;
9. Subjectivité journalistique affleurante (exemples d'indices : !, ?, utilisation de 'on' hors de citations). En effet, d'après (Wiebe et Wilson, 2002, p.113) : « some expressions such as / are subjective in all contexts. ». Cette catégorie, un peu plus générale que les autres, a permis de regrouper certains phénomènes ;
10. Tournure emphatique (exemples d'indices : ce qui frappe le plus c'est, c'est... qui) ;
11. Volonté/Stratégie (exemples d'indices : verbes vouloir/souhaiter/préférer).

Sur ces 11 groupes, 8 (Citation, Conjecture, Constat, Opposition-Concession, Phrase attributive, Phrase averbale, Subjectivité journalistique affleurante, et Tournure emphatique) sont directement liés au style d'écriture. L'hypothèse qui découle de cette observation est donc la suivante : dans un corpus d'articles journalistiques, des marques/indices non intrinsèquement évaluatifs et propres au style d'écriture journalistique (i.e., des stéréotypes⁶) peuvent aider, par leur présence récurrente et concomitante aux évaluations, à repérer des phrases évaluatives.

2.3 Expérience

Pour tester cette hypothèse, une implémentation, sous forme de grammaires locales, des 8 catégories citées plus haut a été réalisée à l'aide de la plateforme SemioLabs. Cette dernière est une plateforme générique pour le développement d'applications TAL développée par la société Noopsis pour son usage interne. Noopsis a mis cet outil à notre disposition dans le cadre du projet Ontopitex. L'intérêt de la plateforme SemioLabs est qu'elle fonctionne de manière modulaire par articulation de composants de traitement plus ou moins indépendants parmi lesquels un tokeniseur et un tagger intégrés. Notre travail d'implémentation a donc pu être réduit à l'écriture de grammaires locales (DCG/GULP) permettant, par filtrage, de repérer les

⁴ Notations issues des expressions rationnelles : (X)* = zéro, une, deux ou plusieurs fois X

⁵ Notations issues des expressions rationnelles : (X)? = zéro ou une seule fois X

⁶ Le sens premier de ce terme peut être éclairant à ce sujet : méthode en imprimerie, au XIX^{ème} siècle, permettant la reproductibilité en masse d'un modèle fixe.

phrases contenant les marques/indices spécifiques dont nous avons parlé plus haut. Un total d'environ 200 règles (pour l'ensemble des 8 grammaires) a vu le jour. Comme SemioLabs offre la possibilité d'associer chaque grammaire à un module spécifique d'annotation pouvant fonctionner individuellement, nous avons ensuite pu observer le résultat de l'annotation pour chaque grammaire, i.e. pour chacune des catégories d'indices. Cette annotation automatique a été menée sur un nouveau corpus de 20 articles de presse économique récoltés à partir de Factiva et concernant Airbus-Boeing (10 textes) et les Cleantechs (10 textes). En parallèle, les 20 articles ont également été annotés manuellement par nos soins pour fournir une première base de comparaison.

3 Résultats et conclusion

La pertinence de chaque type de marques (i.e. dans quelle mesure chacun de ces types de marques apparaît de manière concomitante à l'expression d'évaluation(s) ?) a été évaluée. On peut parler d'un calcul de la précision P_i de chacun des types de marques i , selon la formule suivante :

$$P_i = \frac{\text{nombre de phrases évaluatives contenant au moins une marque de type } i \text{ correctement repérées par le système}}{\text{nombre total de phrases contenant au moins une marque de type } i \text{ repérées par le système}}$$

Le calcul du rappel (i.e. le nombre de phrases évaluatives contenant au moins une marque de type i correctement repérées par le système / le nombre total de phrases évaluatives contenant au moins une marque de type i présentes dans le corpus) a fait apparaître des taux entre 0,9 et 1. Deux interprétations coexistent : le silence est infime car i) les grammaires ont une excellente couverture et/ou ii) parce que l'évaluation n'est pas totalement objective (cf §2 du 3.2).

3.1 Détail des résultats pour la précision P_i

Pour chaque type de marques des tableaux⁷ récapitulatifs détaillés (un pour chacun des 20 articles du corpus) ont été constitués. Ci-dessus un tableau récapitulant la précision obtenue pour chacun des types d'indices/marques sur l'ensemble des 20 textes :

Type i de marques	P_i
Citation-Mise en relief	0,698
Conjecture	0,807
Constat	0,912
Opposition-Concession	0,951
Construction attributive #1 (verbe 'être' pris en compte)	0,592
Construction attributive #2 (verbe 'être non pris en compte)	0,879
Phrase averbale	0,950
Subjectivité journalistique affleurante	0,972
Tournure emphatique	0,967

Tableau 1 : Résultats du calcul de la précision pour chaque type de marques

A l'exception des marques de type Citation-Mise en relief et celles de type Construction attributive #1, on constate que, dans l'ensemble, le bruit est très faible, donc que les marques sont plutôt de bons, voire de très bons, jalons pour repérer les évaluations.

Concernant la catégorie Citation-Mise en relief, des phrases comme '*En plus de l'espace « Recherche et financement », Pollutec organise la convention d'affaires internationale B2Fair.*' (02CleanTech), contenant entre guillemets des spécifications de noms concourent à faire baisser la précision. Quant au résultat moyen concernant la catégorie Construction attributive #1, l'analyse des indices de type Construction attributive #2 montre que la copule 'être' est une source importante de bruit. Une phrase comme '*on est ici dans la région de Seattle.*' (01AirbusBoeing) a été relevée alors que, d'une part, il ne s'agit pas d'une phrase attributive et

⁷ Que par manque de place nous ne pouvons faire figurer dans cet article.

que, d'autre part, elle ne porte aucune évaluation. De même les phrases passives ('180 A320 ont été commandés par IndiGo, 60 autres par Virgin Atlantic.' (06AirbusBoeing)) ou avec des verbes dont les temps composés se forment avec l'auxiliaire 'être' ('L'événement annuel 'Materials Day' s'est tenu fin avril à l'Institut de Thermotechnique (sciences appliquées) de la KULeuven.' (06CleanTech)) ont occasionné du bruit.

3.2 Conclusion et perspectives

Loin d'être une évaluation dont on puisse tirer des résultats catégoriques, cette petite expérience de repérage automatique nous a permis i) d'effectuer quelques petits réglages au niveau des grammaires locales pour éviter des erreurs grossières, et ii) d'obtenir un premier retour, plutôt encourageant, quant à la pertinence de notre hypothèse. Globalement, s'appuyer sur des marques non-intrinsèquement évaluatives, récurrentes dans le discours journalistique, et apparaissant concomitamment à des évaluations semble une méthode prometteuse. Peu coûteuse en ressources et relativement simple à mettre en œuvre, cette méthode permettrait d'offrir une solution concrète à une problématique complètement d'actualité car en prise avec des enjeux sensibles et stratégiques.

La prochaine étape dans notre recherche devrait consister à nous procurer un corpus annoté par un analyste (ou tout autre partie prenante) afin de pouvoir mener une évaluation complète (précision et rappel) et pleinement objective (car basée sur des données d'annotation indépendantes).

Nous envisageons également par la suite de fournir, de manière automatique, pour chaque phrase repérée par le système, un indice de fiabilité basé sur la densité d'indices/marques présents dans la phrase. Cette idée repose sur l'hypothèse que plus une phrase contient d'indices/marques récurrentes dans le discours journalistique et apparaissant concomitamment à des évaluations, plus la probabilité qu'on ait à faire à une phrase évaluative est forte. La validité d'une telle hypothèse devra être démontrée par une étude concrète.

Références

- BEDNAREK M. (2009). Dimensions of evaluation: cognitive and linguistic perspectives. *Pragmatics & Cognition* 17/1: 146-175.
- BLOOM K., GARG N., ARGAMON S. (2007 a). Extracting Appraisal Expressions, *HLT/NAACL 2007*.
- BLOOM K., STEIN N., ARGAMON S. (2007 b). Appraisal Extraction for News Opinion Analysis, *NTCIR6 2007*.
- MARTIN J. R., WHITE P. R. R. (2005). *The Language of Evaluation : Appraisal in English*, New York and London : Palgrave MacMillan.
- PANG B., LEE L. (2008). Opinion Mining and Sentiment Analysis. In *Foundations and Trends in Information Retrieval*, Vol 2, pp 1–135.
- RILOFF E., WIEBE J. (2003). Learning Extraction Patterns for Subjective Expressions in *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP-2003)*, pp 105–112, Sapporo, Japan. ACL.
- WAGNER R.L., PINCHON J. (1962). *Grammaire du français classique et moderne*, Paris : Hachette Université.
- WIEBE J., WILSON T., CARDIE C. (2005). Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, vol. 39, issue 2-3, pp 165–210.
- WIEBE J., WILSON T. (2002). Learning to disambiguate potentially subjective expressions. In *Proceedings of CoNLL-2002*, Taipei, Taiwan.
- YU H., HATZIVASSILOPOULOU V. (2003). Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP 2003)*, Sapporo, Japan. pp 129–136. ACL.