

## INTRODUCTION

**Luc Grivel**

**Lavoisier** | *Les Cahiers du numérique*

**2011/2 - Vol. 7**  
**pages 9 à 14**

**ISSN 1622-1494**

Article disponible en ligne à l'adresse:

-----  
<http://www.cairn.info/revue-les-cahiers-du-numerique-2011-2-page-9.htm>  
-----

Pour citer cet article :

-----  
Grivel Luc, « Introduction »,  
*Les Cahiers du numérique*, 2011/2 Vol. 7, p. 9-14.  
-----

Distribution électronique Cairn.info pour Lavoisier.

© Lavoisier. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

# INTRODUCTION

LUC GRIVEL

Le développement de nouveaux espaces d'expression sur internet (les réseaux sociaux, les blogs personnels, les forums, etc.) génère de nouvelles pratiques sociales et donc de nouveaux champs d'études socio-économiques, politiques, scientifiques. L'enjeu est de détecter le plus tôt possible les changements de perception relatifs à un produit, un service, des faits sociétaux concernant la famille, l'éducation, la religion, le travail, la santé, la politique, etc. Dans ce cadre, une opinion peut se définir comme un sentiment, un jugement, un avis (une évaluation) émis par une source (un individu ou un groupe social) à propos d'un objet (une cible). Pour caractériser une opinion, il faudrait donc identifier ces trois éléments principaux : le sentiment, le jugement ou l'avis exprimé, la source de l'opinion, la cible de l'opinion.

Analyser les expressions d'opinions sur internet est une tâche qui impose de résoudre un certain nombre de problèmes difficiles, de natures différentes.

Sur un plan épistémologique, se confrontent autour de la notion d'opinion de nombreux antagonismes (fait/valeur, subjectivité/objectivité, opinion publique/opinion privée, individu/groupe, comportement individuel/comportement collectif, etc.) auxquels on peut rattacher un certain nombre de questions essentielles pour l'analyse d'opinions :

- Comment distinguer ce qui relève de l'opinion ou du jugement personnel (sentiment que l'on a à propos de quelque chose), de ce qui relève de l'information factuelle (un fait, un événement) ou de la connaissance ?
- Peut-on séparer a priori les faits des opinions ?
- Existe-t-il un langage subjectif ? Peut-on le caractériser ?
- Comment caractériser la cible (marque, personne, entreprise, etc.), la source, les circonstances dans lesquelles l'opinion a été exprimée (quand et où ?)

Derrière ces questions se cache la difficulté de définir l'opinion. Il n'existe pas de définition formelle consensuelle, mais des approches empiriques, telles :

- l'opinion, c'est le passage de texte, la phrase ou la partie de phrase à laquelle on peut rattacher une polarité (positive ou négative).
- l'opinion, c'est la phrase ou la partie de phrase qui permet de répondre à la question : que pense X de Y ?
- ...

Dresser un état de l'opinion, dégager des tendances pose également des questions d'ordre sociologique et méthodologique.

Dans le cadre d'un sondage d'opinion classique (enquête), les problèmes méthodologiques posés par les questions fermées sont bien identifiés. En demandant aux personnes interrogées de répondre à des questions fermées, l'enquêteur pose des questions auxquelles une partie des enquêtés n'a jamais réfléchi et leur demande de choisir parmi des réponses préétablies, formellement identiques, mais qui pour les enquêtés pourront avoir des significations très différentes.

L'analyse d'expressions d'opinions sur internet (la fouille d'opinion) pourrait être une façon de surmonter ces limitations, mais introduit de nouvelles questions méthodologiques concernant le cadre d'interprétation :

- Sur quels critères regrouper, agréger des opinions ?
- Par quelles procédures construire du sens, synthétiser des résultats ?

Nous avons souhaité dans ce numéro spécial que puissent se croiser les regards et les approches de chercheurs issus de différentes disciplines (les sciences de l'information et de la communication, la linguistique, l'informatique, la sociologie) pour répondre aux problèmes posés par le traitement de la textualité et aux questions épistémologiques et méthodologiques soulevées plus haut. Le comité éditorial a sélectionné 7 contributions.

*De la fouille de données à la fabrique de l'opinion* de Egle Eensoo-Ramdani, Evelyne Bourion, Monique Slodzian, Mathieu Valette, propose une réflexion sur les enjeux épistémologiques et l'histoire de la fouille d'opinions, avec une discussion approfondie sur les principaux concepts fondateurs du champ de la fouille d'opinions (fait/valeur, subjectivité/objectivité, opinion publique/opinion privée, individu/groupe, comportement individuel/comportement collectif, etc.). Ce n'est pas seulement un article *savant*, bien documenté, avec une bibliographie riche. En s'appuyant sur des exemples issus de différents projets qu'ils ont menés, les auteurs proposent des pistes pour l'amélioration des performances de l'analyse de la subjectivité, en s'intéressant à l'incidence de l'émetteur sur le message et ses interprétations possibles par les récepteurs du message. « L'émetteur, en élaborant son texte, choisit, inconsciemment ou non, un ensemble de contraintes de production destinées à donner au lecteur un cadre d'interprétation » qu'il faut prendre en compte pour la détection des valeurs véhiculées par un texte. En résumé, un article utile, notamment pour les praticiens de la fouille d'opinions !

*Vers une caractérisation automatique de critères pour l'opinion-mining* de Benjamin Duthil, Francois Troussel, Gérard Dray, Jacky Montmain et Pascal Poncelet, s'inscrit dans la continuité de la réflexion menée par l'article précédent sur la prise en compte du genre dans la fouille d'opinions, puisqu'il s'agit ici d'analyser des critiques dans deux domaines (le cinéma et la restauration). L'objectif est d'identifier des fragments de textes en relation avec une thématique donnée pour permettre une analyse d'opinions personnalisable (selon les critères de l'utilisateur), analogue à celles produites par les systèmes d'analyse et de comparaison des opinions de consommateurs sur des produits. L'utilisateur associe un certain nombre de mots à chaque critère qu'il a défini. La technique consiste à identifier des mots corrélés (statistiquement) à ces mots initiaux (qualifiés de germes dans l'article). L'analyse d'opinion se limite ensuite à identifier la polarité (positif/négatif) pour chacun des critères en s'appuyant sur un outil du domaine public senticNet. L'apport principal de cet article est de proposer une méthode de construction de corpus d'apprentissage en constituant des classes de documents rassemblant des exemples de documents du domaine pour chaque mot germe, et des anti-classes de documents

rassemblant des exemples de documents du domaine ne contenant aucun des mots germes.

L'article *Pénalisation des mots fréquents pour la classification de sentiments* de Abdelhalim Rafrafi, Vincent Guigue, Patrick Gallinari s'intéresse aux prétraitements qui peuvent être réalisés pour améliorer la classification de sentiments par des techniques d'apprentissage supervisé. Plus précisément, les auteurs font l'hypothèse que les mots fréquents dans un texte caractérisent plus le thème que l'émotion. Cependant le codage classiquement utilisé en classification supervisé (le TFIDF) pour pénaliser les mots fréquents n'est, selon eux, pas adapté pour la classification de sentiments. Les auteurs proposent une technique statistique de sélection de mots caractérisant l'émotion qui peut être mise en œuvre en cours d'apprentissage. Ils l'ont testé sur cinq bases de données couramment utilisées pour l'évaluation et la comparaison d'algorithmes de classification de sentiments. Les résultats montrent une amélioration des performances de la précision par rapport aux techniques de codage classiques. Un article qui sera apprécié par les spécialistes d'analyse de données pour la rigueur des formulations et modèles mathématiques employés, les analyses détaillées des résultats.

Les forums, les blogs, réseaux sociaux, sont des lieux rassemblant des personnes qui partagent des objets d'intérêt sur lesquels ils s'expriment. Le choix d'un lieu et d'un objet n'est évidemment pas neutre vis-à-vis de la méthodologie à employer dans le cadre d'une étude des sentiments. Les textes qui suivent traitent du problème de la prise en compte du contexte énonciatif.

Dans l'article présenté par Magali Bigey, *Littérature et sentiments : assumer ses choix de lecture sur le web*, l'objet partagé est la lecture sérielle (le roman populaire et sériel). Le lieu est un forum sur les romances. Dans ce type de lieu, s'exprimer dans un langage correct est un facteur de crédibilité des participants, ce qui justifie le choix d'une méthode d'analyse des propos tenus qui soit basée sur des outils d'analyse linguistique, en l'occurrence un analyseur morpho-syntaxique, Nooj, des dictionnaires spécifiques (des verbes de sentiment basés sur les verbes du lexique de M. Gross dans le but de réaliser un concordancier réunissant tous les verbes de sentiment d'un corpus en contexte. L'étude réalisée met en évidence les représentations de l'objet roman et de son lectorat. Les représentations sociales doivent être également prises en compte, et notamment les pressions qui s'exercent sur les participants qui, ici, peuvent se retrouver accusés d'effectuer de mauvais choix de lecture ou attaqués sur leur crédibilité.

*Les réseaux sociaux de santé*, présenté par Sébastien Broca et Raphaël Koster discute des propriétés que peuvent avoir des opinions exprimées dans un forum. L'article est positionné dans le champ sociologique et questionne la

nature des réseaux communautaires, la diversité des représentations et des sentiments d'appartenance, la co-construction de savoirs profanes. Le forum étudié concerne l'autisme (Yahoo ! Autisme), une liste de discussion nationale nécessitant une inscription préalable, et rassemblant presque exclusivement des personnes confrontées à l'autisme et intervenant sous leur vraie identité. Les échanges sur cette liste de discussion ont été analysés sous la forme d'une enquête qualitative (entretiens).

*Une expérience d'utilisation d'e-portfolio en formation continue* d'Hédia Mhiri Sellami présente une analyse des opinions émises par un échantillon de population au sein du corps enseignant concernant l'usage d'un portfolio numérique (edu-portfolio.org) dans le cadre d'un master professionnel intitulé « Nouvelles Technologies Educatives » (NTE) Pour évaluer cette expérience, des enseignants ont répondu à un questionnaire composé de questions fermées et ouvertes. La démarche du questionnaire et les méthodes d'analyse employées (ici, quantitatives), sont justifiées et décrites en détail : objectifs, type de questionnaire, mode de recueil, période/durée de recueil, nombre de répondants, impact sur les marges d'erreurs, méthodes d'analyses, enjeux de l'interprétation des résultats, difficultés rencontrées, solution ou approche adoptée pour y répondre, etc.

*Reconnaissance des émotions dans la musique* présenté par Jean Debaecker, traite d'un objet de recherche récent dont les retombées économiques sont potentiellement très importantes et les objectifs très ambitieux : extraire des données suffisamment significatives d'un objet sonore, une chanson par exemple, ce qui permettrait de développer des fonctions de recherche de musique par émotion ou par humeur (triste, apaisant, excitant, etc.). Le protocole expérimental mis en place (évaluation d'une chanson par un grand nombre de personnes (265) sur un site web) permet d'évaluer le caractère intersubjectif de sujets exprimant un jugement de goût et de valider une hypothèse de travail : l'impossibilité d'établir une connaissance a priori de l'émotion lors de la rencontre d'un objet esthétique et d'un individu. C'est un travail préliminaire à des travaux d'envergure plus importante, avec une expérimentation plus dynamique, sur une base musicale importante, en présence de groupes témoins.

Coordonner un numéro en essayant de considérer conjointement les points de vue de plusieurs disciplines n'a pas été une tâche facile. J'exprime mes sincères remerciements aux auteurs des articles pour leurs contributions de qualité, et aux membres du comité scientifique sans qui je n'aurais pu mener à bien cette aventure.