



HAL
open science

Préface

Marc El Bèze, Agata Jackiewicz, Susan Hunston

► **To cite this version:**

Marc El Bèze, Agata Jackiewicz, Susan Hunston. Préface. Revue TAL : traitement automatique des langues, 2010, Opinions, sentiments et jugements d'évaluation, 51 (3). hal-01878913

HAL Id: hal-01878913

<https://hal.science/hal-01878913v1>

Submitted on 21 Sep 2018

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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Préface

« *L'opinion est chose instable* »
Aristote, *Seconds analytiques*, I, 33, 89a5.

1. Problématique

L'acte d'évaluation est enraciné dans l'existence même des êtres humains. Il est courant de s'exprimer pour valoriser autrui, soi-même et plus généralement le monde qui nous entoure. Si cet exercice fait partie de notre quotidien, c'est que valoriser est une façon privilégiée d'étayer nos choix, guider et orienter nos actions, intervenir sur ce qui régit les rapports sociaux.

L'analyse automatique de ces actes, qu'ils soient oraux ou écrits, n'est pas chose aisée, en raison notamment de leur caractère privé, culturel, instinctif, versatile, bref, subjectif. La difficulté croît, et c'est souvent le cas, lorsque les expressions franches et directes sont délaissées au profit de formulations voilées, de citations ou d'allusions produites dans le but de louer, encenser, manipuler, déstabiliser ou assassiner, avec finesse et raffinement. Dans cet exercice de style, le contexte, qui joue toujours un rôle, devient essentiel. Les relations entre les unités linguistiques employées et les intentions qu'elles sous-tendent sont si peu transparentes qu'il est quasiment impossible d'en codifier les règles.

Sans parvenir ni même chercher à « décrypter le fond de la pensée » des locuteurs, il est devenu possible aujourd'hui d'extraire et de catégoriser selon leur polarité les opinions ou les sentiments exprimés explicitement dans des textes. Comme en témoignent les cinq contributions réunies dans ce numéro spécial, malgré les difficultés rencontrées, des progrès notables ont été accomplis dans le traitement automatique des messages porteurs d'opinions.

Le fait qu'ils interviennent massivement au travers des nouvelles pratiques sociales de communication (notamment depuis ces dernières années) a ouvert de nouvelles perspectives et entraîné un engouement pour la résolution des problèmes posés. Mais l'intérêt pour ces problématiques s'est accru aussi et surtout parce que dans plusieurs domaines – politiques, sociaux, économiques et scientifiques, pour n'en citer que quelques-uns –, elles correspondent à des enjeux majeurs.

2. Le contexte politique, économique, social et culturel

2.1. De la démocratie directe au Web 2.0

Le développement (amplifié depuis l'apparition du Web 2.0) des échanges, des interactions entre utilisateurs, des espaces de débats aboutit à une production abondante de textes dans lesquels la subjectivité occupe une très grande place. L'importance prise par cette dimension dans les messages s'explique aussi par l'apparition de nouveaux espaces d'expression. La multiplication des blogs de journalistes, qui en est un exemple particulièrement significatif, vient se surajouter aux pratiques plus traditionnelles des dépêches, critiques, « lettres d'information » et autres communiqués accessibles ou diffusés sur le Web depuis ses origines. Dans ce contexte, on peut observer une démultiplication des lieux d'échanges où au fil de réponses, citations, reprises mutuelles, et autres débats croisés, peuvent être confrontés les points de vue des uns et des autres.

En fait, le Web social ne fait qu'accompagner l'évolution des mentalités et des usages. Sur le plan politique, l'autonomisation croissante des individus dans les sociétés modernes renforce leur sentiment que de nouvelles formes de démocraties directes sont nécessaires et qu'elles sont aussi légitimes que celles en vigueur dans les systèmes parlementaires. Les sources/formes de la légitimité sont en train de se multiplier avec des caractéristiques qu'il serait imprudent de négliger. Échappant aux impératifs de temps et de lieu, pour le pire ou le meilleur, la consultation et le débat restent ouverts en permanence et sont *a priori* accessibles à tous.

Au fond, c'est l'addition de la révolution technologique et du désir populaire de participation et de transparence qui concourt à la multiplication des lieux d'expression des opinions. Sondages, élections, référendums, débats collectifs, blogs, forums... ne sont que des modalités différentes d'expression d'un même phénomène : l'opinion publique.

2.2. Opinions sur... opinion

Dans la littérature, on trouve une production abondante à propos de la notion d'opinion, associant et croisant des points de vue philosophique, historique, politique et sociologique. Y sont essentiellement traités les enjeux actuels liés à la montée de l'opinion publique dans la sphère politique, à sa vitalité démocratique, mais aussi à son immaturité et les risques qu'elle engendre (Cautrès et Reynié, 2000), (Sauvy, 1997). Les titres d'ouvrages sont évocateurs : « *Les fers de l'opinion* » (Bénéton, 2000), « *L'opinion publique et la science. À chacun son ignorance* » (Bensaude-Vincent, 2000), « *La revanche de l'opinion* » (Cayrol, 2007), « *La Reine du monde* » (Julliard, 2008). L'opinion y apparaît tantôt comme une force aveugle, tantôt comme une conscience, un arbitre, voire une sorte de tribunal redouté et redoutable même s'il n'est doté d'aucun pouvoir juridique.

Encensée ou méprisée, robuste ou malléable, écoutée ou bâillonnée, « *Reine du monde* » ou maîtresse inconstante, l'opinion est aujourd'hui investie de valeurs fortes. Les arguments invoqués en faveur ou à l'encontre de l'opinion présentent à nos yeux un intérêt dans la mesure où ils font apparaître les propriétés que l'on tend à attribuer à cette notion, si difficile à cerner.

On note que la disqualification de l'opinion peut être double : intellectuelle et éthique. Ainsi, sous un certain angle, l'opinion serait le siège de croyances, d'illusions, voire d'ignorance. Une forme de savoir inférieur, un parent pauvre de la science, un obstacle à dépasser. Une attitude motivée par des intérêts privés, fondée sur un réalisme immédiat, teintée de préjugés, source de réponses faciles. Mais, pour défendre l'opinion, on ne manque pas, non plus, d'arguments percutants, d'ordre pragmatique notamment. En effet, on peut y voir une forme de connaissance en prise avec le terrain, les circonstances et le vécu quotidien. Une attitude nourrie par la prudence de l'utilisateur et du citoyen. Une forme d'engagement nécessaire au fonctionnement de la démocratie.

Ce deuxième point de vue, positif, accorde un crédit important aux profanes et à leurs compétences. Les « citoyens ordinaires » ne se contenteraient pas d'appliquer les cadres interprétatifs véhiculés par les experts, les éducateurs ou les médias, mais ils mobiliseraient leurs propres expériences et leurs compétences critiques et discursives, pour négocier le sens des enjeux politiques et sociétaux. Ainsi, on peut trouver, dans un ouvrage paru récemment (Flichy, 2010) des éléments proclamant le « sacre de l'amateur ». Une lecture critique de cet ouvrage s'impose pour appréhender plus finement l'opposition entre d'une part les professionnels experts et d'autre part les amateurs autodidactes. Nous ne cherchons pas à positionner ce numéro dans un champ qui n'est pas le sien, mais nous voulons signaler au lecteur que les éclairages apportés par les sociologues sur cette activité particulière ne peuvent être ignorés.

2.3. Impact économique et politique

La compréhension fine et réactive des mouvements et des perceptions de l'opinion représente un enjeu politique et économique majeur.

Il est devenu indispensable de connaître au plus vite les opinions et sentiments exprimés par les citoyens. Gouvernements, syndicats, partis politiques, collectivités territoriales s'inquiètent de l'opinion et cherchent à la cerner pour agir en conséquence. Le besoin touche en particulier les administrations en charge de la défense et de la sécurité. Les collectivités territoriales, à leur niveau, cherchent de plus en plus à mettre en place des stratégies de communication efficaces qui englobent de plus en plus le Web social. Les projets locaux qu'elles proposent amènent les internautes à réagir en ligne : ils expriment leurs accords, leurs désaccords avec telle ou telle initiative. Cette préoccupation est aussi celle des grandes entreprises soucieuses de leur image, spécialement dans un contexte de

gestion des crises. Mais, elle tend à être partagée par un nombre croissant d'acteurs de la sphère économique, notamment du fait de l'expansion fulgurante du commerce en ligne. Il s'agit d'accéder aux jugements relatifs à un produit ou un service particulier, afin de mieux anticiper les besoins et mieux cibler l'impact du marketing visant des catégories de consommateurs clairement identifiées.

La problématique de l'opinion s'étend à l'échelle internationale. En témoignent, par exemple, le projet américain CERATOPS¹ visant à observer à travers la presse la perception mondiale de tel ou tel fait politique ou militaire, et les études européennes EVS (*European Values Surveys*)², menées depuis 1981, sur les valeurs des Européens et évolution de ces valeurs : « *How do Europeans think about life, family, work, religion, sex, politics, and society?* »

Dans tous ces univers, on cherche à détecter les opinions de plus en plus tôt, quand elles commencent à se former et à se cristalliser. En gestion de crise, les entreprises tentent de comprendre, en temps quasi réel, la perception de l'opinion, publique, spécialisée ou cultivée. Aller « pêcher » les opinions directement à la source, dans les discours, là où elles s'expriment spontanément permet de s'affranchir de la médiation – sécurisante mais aussi source de biais et d'appauvrissement – des sondages. Les sondages recueillent les préférences, mais escamotent les discours et les réactions corporelles et émotives qui les accompagnent. En effet, les dispositifs d'enquête ou de questionnement standardisé sont conçus pour objectiver les opinions, lesquelles, de ce fait, se trouvent décontextualisées et désindexées des situations réelles.

Les retombées espérées sont multiples. Dans le peu d'espace réservé à cette préface, on se limitera à ne donner qu'un seul exemple, celui de l'analyse de l'opinion publique :

« *Ces avancées sur le front des méthodes permettent ainsi d'envisager des dispositifs d'enquête hybrides donnant une plus large place aux questions ouvertes, voire à des séquences mêlant questions fermées et ouvertes [...]. Ils rendent ainsi possible la constitution de véritables corpus "d'énoncés d'opinion publique".* » (Brugidou, 2005).

2.4. Enjeux scientifiques et techniques

Tous ces besoins ont ouvert de nouvelles perspectives de recherche, tant dans les milieux universitaires qu'industriels. Un vaste champ d'études, attesté par une multitude de publications, de projets et d'événements s'en réclamant, s'est rapidement constitué.

1. [http : //www.cs.pitt.edu/mpqa/ceratops/](http://www.cs.pitt.edu/mpqa/ceratops/)

2. [http : //www.europeanvaluesstudy.eu/](http://www.europeanvaluesstudy.eu/)

Un changement notable de paradigme est en train de s'opérer du point de vue épistémologique, lié à l'introduction du sujet connaissant/énonçant/opinant au cœur des dispositifs informatiques. En TAL, l'essentiel des efforts de recherche a porté jusqu'à un passé récent sur le repérage et l'extraction des « faits », des informations « positives » (ou supposées telles) portées par les textes. Il s'agissait de s'affranchir de la subjectivité pour garantir la vérité. Or, on s'accorde aujourd'hui sur le fait qu'il convient de s'intéresser non plus seulement à « ce qui est » mais à « ce qui en est dit, et par qui », incluant des notions telles que les « sources du savoir » ou les « positionnements » des auteurs/contributeurs. Ce mouvement de désobjectivisation, bien que relativement récent, est en train d'évoluer de manière assez rapide.

De plus en plus, les internautes sont appelés à participer à un travail collaboratif d'élaboration des connaissances. Wikipédia en est un exemple particulier. Pour contribuer à l'un ou l'autre de ces projets « encyclopédiques », il ne faut pas forcément être un expert connu et reconnu. À cause de cela, il est important, aujourd'hui plus qu'hier, que les utilisateurs de ces sites soient avertis de la façon dont les savoirs se construisent. D'un côté, il est nécessaire que les lecteurs arrivent à distinguer dans les textes qu'ils lisent la part de rhétorique chargée de les pousser à agir ou à penser d'une façon précise. De l'autre côté, il faudrait éviter que cette vigilance ne se transforme en méfiance voire en un cynisme qui finirait par les dissuader de prendre part à quelque action que ce soit.

3. La recherche universitaire et industrielle

Ces dernières années ont été marquées par un intérêt croissant pour le traitement informatisé des opinions, des attitudes, des sentiments ou des jugements d'évaluation exprimés dans des documents textuels, issus par exemple d'éditions en ligne des médias traditionnels ou de supports tels que les blogs ou les forums. Un nouveau champ de recherche s'est rapidement constitué autour d'équipes déjà nombreuses visant, avec des objectifs divers, l'étude d'opinions émises sur des entités variées (produits, comportements...) à l'aide de méthodes et de techniques non moins diversifiées. En témoignent, entre autres, l'apparition de nombreux ateliers et les récentes campagnes d'évaluation en fouille d'opinion FODOP'08, DEFT'07, DEFT'09, NTCIR6 (2006-2007).

Les premiers travaux en matière de « fouille d'opinion » se sont concentrés sur une caractérisation des contenus relativement simple, en termes de « subjectif/objectif » et/ou de « polarité appréciative » (positive/négative/neutre). L'équation entre opinion, sentiment ou jugement d'évaluation s'étant rapidement avérée peu satisfaisante, ont été ensuite menées des études sur des aspects plus ciblés, tels que l'analyse émotionnelle des conversations. Progressivement, des recherches s'orientent, de manière plus ambitieuse et à plus long terme, vers une caractérisation plus riche, plus fine, plus informative, et donc plus fiable.

En dehors du développement de ces traitements automatiques, des recherches académiques ont été menées pour étudier ce qui dans le langage relève de la subjectivité, des opinions. En suivant des voies parfois parallèles, parfois divergentes voire même conflictuelles, ces travaux qui peuvent être qualifiés d'approches fondées sur les textes ou les corpus ont permis d'aboutir à une meilleure compréhension du phénomène étudié. Rapprocher les trajectoires suivies par ces familles d'approches ne peut que présenter des avantages comme ceux que peuvent procurer l'alliance entre la norme rigoureuse mais rigide, son assouplissement et son enrichissement par les exemples particuliers sur lesquels la catégorisation automatique peut s'appuyer pour atteindre les objectifs de performance qui lui sont fixés.

Parmi les approches fondées sur les textes, on peut citer les travaux de Martin et White (2005) qui proposent de distinguer entre trois classes « *Affect* », « *Judgement* » et « *Appreciation* ». Les expressions affectives sont employées pour exprimer une réaction personnelle (satisfaction, bonheur...). Les jugements attribuent des qualités (capacité, ténacité...) à des personnes en tant qu'acteurs sociaux. Les appréciations véhiculent des évaluations sur tout autre type d'entité (valeur, composition...). Cette distinction met en lumière la différence entre l'expression d'une évaluation en tant que réponse personnelle et l'expression de qualités inhérentes ou contextuellement associées à une personne, une action ou à un objet. Dans ces travaux également, les différents niveaux d'expression de l'évaluation ont été explorés allant de l'explicite à l'implicite. Les intentions peuvent être clairement affichées ou enfouies dans des formulations plus ou moins obliques et, dans le premier cas, fortement contextuelles, avec à l'extrême limite, une simple invite du lecteur à évaluer, plutôt que l'expression même d'une évaluation.

Les recherches fondées sur les corpus, se sont focalisées sur les unités de sens « *Unit of Meaning* » (Sinclair, 2004 ; Stubbs, 2001) ou les séquences sémantiques « *Semantic Sequences* » et le potentiel qu'elles ont de véhiculer une évaluation. Ces concepts ont pour objet de cerner la tendance de voir certains mots former par leur cooccurrence des fragments de phrases qui apparaissent fréquemment dans un contexte évaluatif consistant. Parmi les exemples bien connus donnés par (Sinclair 2004), citons celui des « *true feelings* » trahissant un état émotionnel d'anxiété ou de frustration et plus généralement employé dans un contexte de difficulté à s'exprimer.

En schématisant, les principaux axes de recherche concernent : (i) la modélisation linguistique et informatique ainsi que la gestion des données d'opinion (qu'est-ce qu'une « opinion », comment la représenter informatiquement ?) ; (ii) l'expression en langue et en discours (comment les opinions, sous leurs différentes facettes, sont-elles formulées ?) ; (iii) la construction, l'acquisition et la validation des ressources linguistiques ; (iv) les méthodes pour identifier, annoter et extraire automatiquement opinions et sentiments dans des documents textuels ou audiovisuels ; (v) la présentation synthétique de la diversité des opinions.

4. La présentation du numéro

Si toutes les questions ayant trait en TAL à l'analyse de subjectivité ne sont pas couvertes par le présent volume, les contributions qui y sont rassemblées permettent de donner une bonne idée de quelques problématiques retenues, et laissent apercevoir, au-delà de ce qui a déjà été entrepris, la variété et l'étendue de l'effort restant à accomplir.

L'article *De la classification d'opinion à la recommandation : l'apport des textes communautaires*, par Damien Poirier, Françoise Fessant et Isabelle Tellier, propose d'exploiter et de classer des textes de blogs pour un système de recommandation de contenu. Le traitement des opinions (exprimées assez souvent en cyberlangue dans des messages très courts) se ramène ici à une tâche de classification supervisée qui vient alimenter un composant de recommandation fondé, quant à lui, sur un filtrage collaboratif. C'est là que réside l'originalité de la proposition des auteurs, proposition qui s'avère de plus être une alternative efficace au filtrage fondé sur la thématique. Envisagée spécifiquement dans l'optique de la recommandation, la notion d'opinion renvoie aux attitudes qui traduisent une préférence (« aime »/« n'aime pas »). L'article s'attarde sur les particularités des textes communautaires, lesquelles expliquent ou justifient certains choix de traitements (faits ou à faire) : l'importance de la ponctuation, la coprésence de particules de négation et de mots à polarité négative, l'emploi significatif de mots ayant la même racine, la corrélation entre temps verbaux et tonalité axiologique globale des textes.

L'article *Résumé automatique de textes d'opinion* de Aurélien Bossard, Michel Génereux et Thierry Poibeau permet de découvrir comment le problème du résumé automatique de textes d'opinion peut être résolu avec les outils et techniques existants. Les auteurs s'intéressent au problème du résumé automatique des opinions exprimées sur la Toile et notamment dans des blogs. Ils proposent un système de résumé multidocument CBSEAS, associant l'analyse de la redondance, le repérage des informations nouvelles et le calcul de la polarité des opinions. Le traitement des opinions, visant une distinction simple entre contenus positifs et négatifs, s'appuie sur la stratégie du report de la polarité calculée au niveau du billet sur chaque phrase de celui-ci. Grâce à ces méthodes, de bons résultats ont été obtenus dans le cadre de la campagne d'évaluation TAC 2008 où il s'agissait de produire un résumé d'opinion polarisée sur un sujet donné. Pour améliorer leur système, les auteurs envisagent quelques pistes, dont tout particulièrement la simplification de phrases et une agrégation plus élaborée des opinions émanant de différentes sources.

Alexander Pak et Patrick Paroubek dans *Le microblogage pour la microanalyse des sentiments et des opinions* utilisent de manière conjointe des mots simples et des n-grammes pour détecter des opinions et des sentiments dans des messages issus de Twitter. La méthode proposée évite toute théorie linguistique et offre un moyen pour constituer des corpus de messages de Twitter dans plusieurs langues. L'étude s'appuie sur une définition précise et distinctive des opinions et des sentiments. Les

premières sont une expression des positions ou des croyances personnelles portant sur une cible. Les seconds renvoient à des états émotionnels dont l'origine n'est pas nécessairement connue. Une expression linguistique peut faire référence à une opinion et à un sentiment de polarités différentes. Comme le font ici les auteurs, il est important de prendre en compte la dimension multilingue du problème. La méthode est adaptée, c'est sa force et sa faiblesse, au traitement de messages très courts, estimés homogènes tant du point de vue thématique que sur le plan axiologique. Une hypothèse forte et implicite pose une équivalence entre des fragments de lexiques évaluatifs appartenant à des langues différentes. On appréciera tout particulièrement le fait que les auteurs envisagent d'étudier dans le futur la manière dont les « sentiments sont exprimés avec plus ou moins d'intensité, de façon plus ou moins marquée selon les langues, et quel est l'impact sur ces usages de la dimension culturelle ».

La contribution de Rémi Lavalley, Chloé Clavel et Patrice Bellot intitulée *Extraction probabiliste de chaînes de mots relatives à une opinion* s'inscrit dans la lignée des méthodes numériques appliquées au TALN et converge avec des recherches actuelles sur les agglomérats de mots en linguistique de corpus. L'article propose d'extraire des séquences de mots (lemmes) dans des sous-corpus porteurs d'une opinion donnée (parmi deux ou trois opinions), puis utilise ces séquences pour catégoriser des textes dans l'une de ces trois opinions. La méthode d'extraction de collocations est fondée sur un processus automatique itératif. L'idée sous-jacente est que grâce à ce dispositif il est possible d'identifier pour chacune des classes les expressions linguistiques composites, relevant d'une « phraséologie évaluative » à l'œuvre dans les textes analysés. Le gain recherché est de pouvoir s'affranchir d'une écriture manuelle, nécessairement coûteuse, de patrons linguistiques. On remarque que la notion même de l'opinion est de cette façon induite ou construite dynamiquement en fonction du corpus applicatif choisi. Si cette approche ne permet pas encore de surpasser les résultats obtenus par un petit nombre d'autres méthodes, elle présente plusieurs avantages dont les plus évidents sont l'indépendance vis-à-vis du domaine d'application et *a fortiori* de la langue, et surtout la possibilité de montrer à l'utilisateur en les ordonnant les séquences qui ont le plus contribué à la prise de décision.

Dans *Analyse des sentiments et transcription automatique : modélisation du déroulement de conversations téléphoniques*, Frederik Cailliau et Ariane Cavet décrivent la démarche qu'ils ont retenue pour traiter les transcriptions automatiques de conversations téléphoniques enregistrées dans un centre d'appels. Les auteurs s'appuient sur des recherches en analyse d'opinions pour visualiser la dimension émotionnelle des échanges. L'hypothèse retenue est que les expressions linguistiques employées dans ces tours de parole peuvent être caractérisées d'un triple point de vue : (i) type d'attitude (appréciation, acceptation, accord...), (ii) la polarité (positive, négative), (iii) la charge émotionnelle (marquée, non marquée). Le système notionnel mis en œuvre dans l'étude est relativement élaboré. Il constitue une adaptation du modèle de (Charaudeau, 1992), notamment par ajout de

la notion de « surprise ». Ayant mesuré l'impact des performances du système de reconnaissance de la parole sur la détection de ces entités les auteurs sont conduits à proposer un cumul des poids d'intensité accordés à chaque entité détectée en espérant que cette somme viendra compenser les détections erronées ainsi que les omissions. Si leur objectif premier est atteint, il reste encore des pas à accomplir pour pouvoir détecter automatiquement avec fiabilité l'évolution de la charge émotionnelle telle qu'elle peut se déployer au fil d'un échange.

5. Bilan

Le présent volume donne un bon aperçu d'un état partiel des lieux. On ne doit pas y chercher un bilan reflétant l'intégralité du champ des recherches concernées. À plusieurs titres, la situation est contrastée et nous devons nous garder de toute généralisation ou extrapolation, faites uniquement à partir des contributions incluses dans ce numéro.

Les articles rassemblés dans ce numéro thématique abordent les traitements automatiques des opinions, des sentiments et des jugements d'évaluation essentiellement d'un point de vue applicatif. La tâche commune que partagent presque toutes les applications présentées est de catégoriser automatiquement des objets textuels en positifs et négatifs, parfois neutres. Les auteurs assument ce « modeste » objectif, lequel n'est souvent qu'une étape dans une problématique plus vaste et plus complexe : résumé de textes d'opinion, recommandation nourrie d'appréciations, modélisation du déroulement affectif des transactions téléphoniques, etc.

Il est important de noter que les méthodes décrites sont pensées, et cela est parfaitement justifié, en fonction des caractéristiques des textes/discours à traiter.

En simplifiant, deux familles de textes sont représentées (i) les textes issus de blogs et de *tweets*, (ii) les conversations orales téléphoniques transcrites. Les textes traités sont généralement courts. Cela permet, faisant l'hypothèse d'une univocité thématique et axiologique du message, de leur attribuer une polarité unique. Il faut dire aussi que les avis exprimés dans ces textes sont des plus explicites. Les discours contournés et allusifs se logent rarement dans des espaces de taille limitée. On dispose pour certains d'entre eux d'appréciations chiffrées censées traduire l'opinion exprimée verbalement. Certains textes comprennent des éléments iconographiques (*smileys* ou émoticônes, séquences de signes de ponctuation) destinés à représenter l'état affectif de leur auteur. Ces éléments peuvent alors être exploités comme un repère supplémentaire. Par ailleurs, souvent les textes sont bruités (transcription de l'oral) ou ne répondent pas aux normes d'écriture des grammaires standard (fautes, raccourcis sms, *smileys*, emplois particuliers de signes de ponctuation). Il est évident que cette composante influe tant sur les traitements réalisés que sur la qualité du résultat final.

On constate que les exigences des auteurs quant à (i) la définition des objets traités (opinions, sentiments...), (ii) la précision de la terminologie, ou (iii) la qualité des lexiques utilisés (prise ou non en compte de la polysémie...) sont fortement contrastées. Si, pour certains, la définition même des contenus subjectifs et, à plus forte raison, la catégorisation plus fine posent problèmes, pour d'autres une approche intuitive semble suffire. Dans le premier cas, on se trouve face à la difficulté de trouver des définitions consensuelles et théoriquement plausibles. Dans le second, c'est la non-reproductibilité de la démarche qui pose problèmes. Car l'on ignore sur quels critères sont choisis, construits et annotés les exemples/corpus d'opinion utilisés (dans des situations d'apprentissage supervisé, par exemple).

Les contributions réunies dans ce volume montrent que les méthodes classiques du TAL peuvent être employées avec des résultats satisfaisants pour le traitement des contenus subjectifs. Les auteurs soulignent le fait que l'information extraite avec ces moyens est déjà précieuse pour certains domaines d'application. La robustesse des approches pouvant traiter plusieurs langues et plusieurs registres (oral/écrit) est nettement privilégiée. L'équation mise en avant est donc celle de l'efficacité conjuguée avec l'économie des moyens.

6. Bibliographie

- Bensaude-Vincent B., *L'opinion publique et la science. A chacun son ignorance*. Institut d'édition Sanofi-Synthelabo, 2000.
- Bénéton P., *Les fers de l'opinion*, PUF, Béhémouth, 2000.
- Brugidou M., « Vers une représentation discursive de l'opinion publique. Les problèmes posés par la constitution et l'analyse de corpus de questions ouvertes de sondage », *Corpus*, n°4, 2005.
- Cautrès B., Reynié D., (dir), *L'opinion européenne*. Presses Sciences Po, 2000.
- Cayrol R., *La revanche de l'opinion. Médias, sondages, Internet*. Editions Jacob-Duvernet. 2007.
- Charaudeau P., *Grammaire du sens et de l'expression*, Paris, Hachette Éducation, 1992.
- Flichy P., *Le sacre de l'amateur*, Le Seuil 2010.
- Julliard J., *La Reine du monde*, Flammarion, Café Voltaire, 2008.
- Martin J.R., White P.R.R., *The Language of Evaluation : Appraisal in English*, London & New York, Palgrave Macmillan, 2005.
- Sauvy A., *L'opinion publique*, PUF, 1997.
- Sinclair J.M., *Trust the Text : Language, Corpus and Discourse*. London : Routledge, 2004.
- Stubbs M., *Words and Phrases : Corpus Studies of Lexical Semantics*, Oxford, Blackwell, 2001.

Remerciements

Nous tenons à exprimer notre gratitude envers le comité de lecture de la revue TAL, une mention particulière pour Béatrice Daille et l'aide qu'elle nous a apportée. Nos plus vifs et chaleureux remerciements vont aux membres du comité spécifique formé pour ce numéro :

Nicholas Asher, IRIT, Université Paul-Sabatier, Toulouse
Monika Bednarek, Université de Sidney, Australie
Farah Benamara Zitoune, IRIT, Université Paul-Sabatier, Toulouse
Yves Bestgen, CECL, Université catholique de Louvain, Belgique
Thierry Charnois, GREYC, Université de Caen
Chloé Clavel, EDF, Paris
Stéphane Ferrari, GREYC, Université de Caen
Catherine Gouttas, Thales
Mark Lee, Université de Birmingham, Royaume-Uni
Dominique Legallois, CRISCO, Université de Caen
Laura Monceaux, LINA, Université de Nantes
Maria Teresa Pazienza, Université de Rome Tor Vergata, Italie
Mathieu Roche, LIRMM, Université de Montpellier 2
Horacio Rodriguez, Université Polytechnique de Catalogne, Barcelone, Espagne
Horacio Saggion, Université de Sheffield, Royaume-Uni
Geoff Thompson, Université de Liverpool, Royaume-Uni
Juan Manuel Torres-Moreno, LIA, Université d'Avignon

Agata Jackiewicz
STIH, Université de Paris-Sorbonne
Agata.Jackiewicz@paris-sorbonne.fr

Susan Hunston
Centre for Corpus Research, Université de Birmingham
s.e.hunston@bham.ac.uk

Marc El-Bèze
LIA, Université d'Avignon
marc.elbeze@univ-avignon.fr