Détection automatique de l'ironie dans les tweets en français

Jihen Karoui, Farah Benamara Zitoune, Véronique Moriceau, Nathalie Aussenac-Gilles, Lamia Hadrich Belguith

Détection automatique de l'ironie dans les tweets en français

Jihen Karoui[1, 3] Farah Benamara Zitoune[1] Véronique Moriceau[2] Nathalie Aussenac-Gilles[1]
Lamia Hadrich Belguith[3]
(1) IRIT, CNRS, Université Paul Sabatier, 118 Route de Narbonne, F-31062 TOULOUSE CEDEX 9
(2) LIMSI, CNRS, Université Paris Sud, Rue John von Neumann, 91403 ORSAY CEDEX
(3) MIRACL, Pôle technologique de Sfax, Route de Tunis Km 10 B.P. 242, 3021 SFAX
jihen.karoui@irit.fr, benamara@irit.fr, moriceau@limsi.fr, Nathalie.Aussenac-Gilles@irit.fr,
l.belguith@fsegs.rnu.tn

Résumé. Cet article présente une méthode par apprentissage supervisé pour la détection de l'ironie dans les tweets en français. Un classifieur binaire utilise des traits de l'état de l'art dont les performances sont reconnues, ainsi que de nouveaux traits issus de notre étude de corpus. En particulier, nous nous sommes intéressés à la négation et aux oppositions explicites/implicites entre des expressions d'opinion ayant des polarités différentes. Les résultats obtenus sont encourageants.

Abstract.

Automatic Irony Detection in French tweets.
This paper presents a supervised learning method for irony detection in tweets in French. A binary classifier uses both state of the art features whose efficiency has been empirically proved and new groups of features observed in our corpus. We focused on negation and explicit/implicit oppositions of opinions with different polarities. Results are encouraging.

Mots-clés : Analyse d'opinion, détection de l'ironie, apprentissage supervisé.

Keywords: Opinion analysis, irony detection, supervised learning.

1 Introduction

L'extraction d'opinion dans les textes s'est beaucoup développée pendant la dernière décennie (Liu, 2012), et surtout depuis l'expansion du web social qui permet aux internautes d'émettre des opinions, des émotions ou des évaluations (critiques, etc.). Il existe plusieurs approches pour l'extraction d'opinions allant de représentations par sac de mots à des modèles plus complexes qui traitent de phénomènes dépendant du contexte ou du niveau discursif. Bien que les systèmes actuels obtiennent des résultats relativement bons sur des tâches de classification objective/subjective, l'analyse de polarité (positif ou négatif) doit encore être améliorée pour pouvoir prendre en compte notamment des formes figuratives telles que l'ironie.

L'ironie est un phénomène linguistique complexe largement étudié en philosophie et en linguistique (Grice *et al.*, 1975; Sperber & Wilson, 1981; Utsumi, 1996). Même si les théories différent sur la définition, elles s'accordent sur le fait que l'ironie implique une incongruité entre ce qui est dit et la réalité. Par exemple, dans les tweets ironiques, l'incongruité consiste souvent en l'opposition d'au moins deux propositions P_1 et P_2 qui s'opposent. Elles peuvent être dans le même énoncé (c'est-à-dire explicitement lexicalisées), ou bien l'une est présente et l'autre est implicite. Nous avons défini deux types d'opposition. L'*opposition explicite* peut impliquer une contradiction entre des mots de P_1 et des mots de P_2 qui ont des polarités opposées comme dans (1), ou qui n'ont pas de relation sémantique comme dans (2). L'opposition explicite peut aussi provenir d'un contraste positif/negatif explicite entre une proposition subjective P_1 et une situation P_2 qui décrit une activité ou un état indésirable. L'ironie est alors inférée grâce aux connaissances partagées ou aux normes sociales et culturelles : par exemple, (3) suppose que tout le monde s'attend à ce qu'un téléphone sonne assez fort pour être entendu.

(1) J'**adore** quand mon téléphone **tombe en panne** quand j'en ai besoin.

(2) **The Voice** est plus important que **Fukushima** ce soir.

(3) J'**adore** quand mon téléphone **baisse le son automatiquement**.

L'*opposition implicite* quant à elle, se produit quand il y a une opposition entre une proposition lexicalisée P_1 décrivant un événement ou un état et un contexte pragmatique externe à l'énoncé qui souvent nie P_1 ou son existence. La proposition P_1 peut être soit subjective (cf. (4)) ou objective. Par exemple, dans (5), l'ironie vient du fait que les lecteurs savent que la proposition P_1 (en italique) n'est pas vraie au moment de l'émission du tweet.

(4) #Hollande est vraiment un bon diplomate #Algérie.

(5) #Valls a appris la mise sur écoute de #Sarkozy en lisant le journal. *Heureusement qu'il n'est pas Ministre de l'Intérieur*.

Pour détecter l'ironie dans les oppositions explicites et implicites, la plupart des approches existantes utilisent un classifieur binaire (i.e. apprendre si un texte est ironique ou non) avec une variété de traits allant de simples traits de surface (ponctuations, émoticônes, etc.) à des traits comme la polarité, la synonymie ou le contexte émotionnel (Reyes & Rosso, 2014; Barbieri & Saggion, 2014). Une analyse de notre corpus de tweets en français (nous le présentons dans la section suivante) montre que plus de 62 % des tweets contiennent des opérateurs de négation ("ne...pas") ou des quantifieurs de négation ("jamais", "personne"). Ainsi, la négation nous semble être un indice important dans les énoncés ironiques. Nous faisons donc l'hypothèse que la présence d'une négation ou d'une opposition implicite ou explicite peut aider à la détection automatique de l'ironie.

Dans les sections suivantes, nous présentons notre corpus de tweets puis nous présentons le classifieur binaire utilisé pour la détection de l'ironie dans les tweets. Enfin, nous présentons les expériences menées et les résultats. Finalement, nous concluons en présentant quelques pistes de travaux futurs.

2 Corpus

Un de nos objectifs est de tester si la négation est un bon indice pour détecter les tweets ironiques. Pour cela, nous avons constitué un corpus de tweets ironiques et non ironiques contenant ou non des mots de négation tels que *ne, n', pas, non, ni, sans, plus, jamais, rien, aucun(e), personne*. Nous considérons comme ironiques les tweets contenant les hashtags #*ironie* ou #*sarcasme*, les autres sont considérés comme non ironiques.

Pour collecter les tweets, nous avons dans un premier temps sélectionné un ensemble de thèmes discutés dans les médias au printemps 2014. Nous avons choisi 184 thèmes répartis en 9 catégories (politique, sport, musique, etc.). Pour chaque thème, nous avons sélectionné un ensemble de mots-clés avec et sans hashtag, par exemple : politique (Sarkozy, Hollande, UMP, ...), santé (cancer, grippe), sport (#Zlatan, #FIFAworldcup, ...), médias sociaux (#Facebook, Skype, MSN), artistes (Rihanna, Beyoncé, ...), télévision (TheVoice, XFactor), pays ou villes (Cordée du Nord, Brésil, ...), Printemps Arabe (Marzouki, Ben Ali, ...) et d'autres thèmes plus génériques (pollution, racisme). Nous avons ensuite sélectionné des tweets ironiques contenant les mots-clés, le hashtag #*ironie* ou #*sarcasme* et un mot de négation ainsi que des tweets ne contenant pas de négation. De la même manière, nous avons aussi sélectionné des tweets non ironiques (*i.e.* ne contenant pas #*ironie* or #*sarcasme*). Une fois les tweets collectés, nous avons supprimé les doublons, les retweets et les tweets contenant des liens vers du contenu extérieur. Pour les expériences décrites par la suite, les hashtags #*ironie* et #*sarcasme* sont supprimés des tweets. Pour identifier automatiquement les vrais usages de négation (par exemple, *pas* peut être un nom, *une personne* n'est pas une négation), nous avons utilisé l'analyseur syntaxique MElt [1] ainsi que des règles manuelles pour corriger les sorties de l'analyseur si nécessaire.

Au total, nous avons un ensemble de 6742 tweets. Pour mesurer l'effet de la négation sur la tâche de détection de l'ironie, nous avons constitué 3 corpus : les tweets avec négation (*NegOnly)*, les tweets sans négation (*NoNeg*), et un corpus regroupant l'ensemble des tweets (*All*). Le tableau 1 montre la répartition des tweets.

Pour s'assurer que les hashtags indiquant l'ironie sont fiables, deux annotateurs ont annoté 3 sous-ensembles : 50 tweets ironiques et 50 non ironiques pour chacun des corpus *All*, *NoNeg* et *NegOnly*. L'accord inter-annotateur (kappa de Cohen) par rapport à la référence (*i.e.* par rapport aux hashtags) est $\kappa = 0.78$ pour *All*, $\kappa = 0.73$ pour *NoNeg* et $\kappa = 0.43$ pour *NegOnly*. Ces scores montrent que les hashtags #*ironie* and #*sarcasme* sont relativement fiables mais que la présence d'une négation est une cause d'ambiguïté pour la détection de l'ironie par des humains.

1. http://alpage.inria.fr/statgram/frdep/fr_stat_dep_malt.html

Corpus	Ironique	Non ironique	TOTAL
NegOnly	470	3761	**4231**
NoNeg	1075	1436	**2511**
All	1545	5197	**6742**

TABLE 1 – Répartition des tweets dans le corpus.

3 Un classifieur pour la détection de l'ironie

3.1 Traits utilisés

Un tweet est représenté par un vecteur composé de 6 groupes de traits que nous présentons ici. Certains d'entre eux ont été utilisés avec succès pour la détection de l'ironie (dans ce cas, nous citons les références), d'autres sont nouveaux.

Traits de surface : ce sont principalement les traits utilisés dans l'état de l'art. Le premier est la longueur du tweet en nombre de mots (Tsur *et al.*, 2010). Les autres sont tous binaires et indiquent la présence ou non de : ponctuation (Kreuz & Caucci, 2007; Gonzalez-Ibanez *et al.*, 2011), mots en lettres majuscules (Tsur *et al.*, 2010; Reyes *et al.*, 2013), interjections (Gonzalez-Ibanez *et al.*, 2011; Buschmeier *et al.*, 2014), émoticônes (Gonzalez-Ibanez *et al.*, 2011; Buschmeier *et al.*, 2014), citation (Tsur *et al.*, 2010; Reyes *et al.*, 2013), argot (Burfoot & Baldwin, 2009), mots d'opposition tels que "mais" et "bien que" (Utsumi, 2004), séquence de points d'exclamation ou d'interrogation (Carvalho *et al.*, 2009), combinaison de points d'exclamation et d'interrogation (Buschmeier *et al.*, 2014). Nous avons ajouté un nouveau trait qui indique la présence de connecteurs discursifs qui ne déclenchent pas d'opposition ("ainsi, donc, ...") car nous faisons l'hypothèse que les tweets non ironiques sont susceptibles d'être plus verbeux. Pour implémenter ces traits, nous avons utilisé trois lexiques : un pour les connecteurs discursifs (Roze *et al.*, 2012), un pour l'argot (389 entrées), construit manuellement à partir de diverses sources trouvées sur le web [2] et le lexique CASOAR (Benamara *et al.*, 2014) pour les interjections (236 entrées) et les émoticônes (595 entrées).

Traits de sentiment : ce sont les traits qui indiquent la présence de mots ou d'expressions d'opinion positive ou négative (Reyes & Rosso, 2011, 2012), leur nombre (Barbieri & Saggion, 2014). Nous avons ajouté 3 nouveaux traits : la présence de mots ou expressions de surprise ou d'étonnement, la présence et le nombre d'opinions neutres. Pour obtenir ces traits, nous avons utilisé deux lexiques :

– CASOAR (Benamara *et al.*, 2014), un lexique pour le français de 2732 mots ou expressions d'opinion catégorisés en 4 catégories sémantiques (REPORTAGE, JUGEMENT, SENTIMENT-APPRÉCIATION et CONSEIL comme définies dans (Asher *et al.*, 2009)), ainsi que 184 entrées correspondant aux adverbes de doute, affirmation, intensifieur, et adverbes de négation,
– EMOTAIX [3], un lexique émotionnel et affectif disponible publiquement de 4921 entrées regroupées en 9 catégories : malveillance, mal-être, anxiété, bienveillance, bien-être, sang-froid, surprise, impassibilité, émotion non spécifique. Il contient 1308 entrées positives, 3078 négatives et 535 neutres.

Traits pour les modifieurs de sentiment : ils regroupent deux nouveaux traits qui indiquent si un tweet contient un mot d'opinion dans la portée d'une modalité ou d'un adverbe d'intensité. Les **traits pour les modifieurs** vérifient aussi si un tweet contient : un intensifieur (Liebrecht *et al.*, 2013; Barbieri & Saggion, 2014), une modalité, un mot de négation ou un verbe de discours rapporté.

Traits d'opposition : ils sont nouveaux par rapport à ceux traditionnellement utilisés. Ils indiquent la présence d'opposition explicite grâce à des patrons lexico-syntaxiques spécifiques. Ces traits ont été partiellement inspirés de (Riloff *et al.*, 2013) qui a proposé une méthode par bootstrapping pour détecter les tweets sarcastiques correspondant à une opposition entre un sentiment/opinion positif et une situation négative. Nous avons donc étendu ce patron afin de traiter d'autres types d'opposition. Par exemple, nos patrons indiquent si un tweet contient (a) une opposition de sentiment/opinion, ou (b) une opposition explicite positive/négative entre une proposition subjective et une proposition objective. Soit P_+ (resp. P_-) une proposition subjective contenant une expression positive (resp. negative), soit P_{obj} une proposition objective ne contenant pas d'expression d'opinion (P_{obj} peut contenir une négation ou non), et soit Neg un opérateur qui change la polarité des mots subjectifs dans P_+ (resp. P_-). Les patrons pour (a) sont de la forme $[Neg(P_+)].[P'_+]$, $[Neg(P_-)].[P'_-]$, $[P_+].[Neg(P'_+)]$, $[P_-].[Neg(P'_-)]$, $[P_-].[P'_+]$, et $[P_+].[P'_-]$;

2. http://www.linternaute.com/dictionnaire/fr/usage/argot/1/
3. http://www.tropes.fr/download/EMOTAIX_2012_FR_V1_0.zip

ceux pour (b) sont de la forme : $[Neg(P_+)].[P'_{obj}]$, $[Neg(P_-)].[P'_{obj}]$, $[P_{obj}].[Neg(P'_+)]$, $[P_{obj}].[Neg(P'_-)]$, $[P_+].[P'_{obj}]$, $[P_-].[P'_{obj}]$, $[P'_{obj}].[P_+]$, et $[P'_{obj}].[P_-]$.
Nous considérons qu'un mot d'opinion est dans la portée d'une négation, s'ils sont séparés par au maximum deux tokens (puisque les tweets sont des messages courts limités à 140 caractères).

Traits de contexte : le contexte d'énonciation est important pour comprendre l'ironie d'un énoncé. Ces traits indiquent donc la présence/absence d'éléments de contexte tels que les pronoms personnels, les mots-clés d'un thème donné et les entités nommées donnés par l'analyseur syntaxique. Par exemple, l'ironie dans le tweet *Elle nous avait manqué !* est difficile à détecter car il ne contient pas d'élément contextuel.

3.2 Expériences and résultats

Nous avons testé plusieurs classifieurs sous Weka et avons obtenu les meilleurs résultats avec SMO. Comme nous avons 3 corpus (*NegOnly*, *NoNeg* et *All*), nous avons entraîné 3 classifieurs, un par corpus, notés $C_{NegOnly}$, C_{NoNeg}, et C_{All}. Comme le nombre d'instances ironiques dans *NegOnly* est relativement petit (470 tweets), le classifieur $C_{NegOnly}$ a été entraîné sur un sous-ensemble équilibré de 940 tweets avec une validation croisée sur 10 échantillons. Pour C_{NoNeg} et C_{All}, nous avons utilisé 80% du corpus pour l'apprentissage et 20% pour le test, avec une distribution égale entre les instances ironiques (notées `IR`) et non ironiques (notées `NIR`)[4]. Le nombre de tweets non ironiques étant plus grand que le nombre d'ironiques (cf. Tableau 1), nous avons entraîné les classifieurs en fixant l'ensemble de tweets ironiques tout en faisant varier l'ensemble de tweets non ironiques. Les résultats pour ces différentes combinaisons sont relativement similaires. Les résultats présentés ici ont été obtenus en entraînant C_{NoNeg} sur 1720 tweets et en testant sur 430 tweets. C_{All} a été entraîné sur 2472 tweets (1432 contenant une négation –404 `IR` et 1028 `NIR`) et testé sur 618 tweets (360 contenant une négation –66 `IR` et 294 `NIR`).

Pour chaque classifieur, nous avons étudié l'apport de chaque groupe de traits (cf. Section 3.1) au processus d'apprentissage. Nous avons appliqué à chaque ensemble d'apprentissage un algorithme de sélection de traits (Chi2 et GainRatio), puis avons mesuré l'impact des groupes de traits les plus pertinents sur la tâche de détection de l'ironie. Pour toutes les expériences, nous avons utilisé les traits de surface comme baseline. Pour C_{NoNeg} et $C_{NegOnly}$, le trait qui indique la présence d'une négation a été désactivé. Les résultats en terme d'exactitude sont présentés dans le tableau 2.

	NegOnly	*NoNeg*	*All*
Baseline (traits de surface)	**73.08**	63.25	55.50
Meilleurs traits de surface	73.08	64.65	56.31
Meilleurs traits de sentiment	57.02	**67.90**	58.25
Modifieurs de sentiment	53.51	56.51	51.94
Modifieurs	53.72	55.81	**86.89**
Opposition	55.31	63.02	79.77
Contexte interne	55.53	53.25	53.55

TABLE 2 – Résultats des 3 expériences en terme d'exactitude.

Comparé aux autres traits, la baseline obtient de bons résultats sur *NegOnly* alors que les résultats sont beaucoup moins bons que les 2 autres corpus. Pour *NoNeg*, les meilleurs résultats sont obtenus en utilisant les traits {longueur du tweet, interjections, connecteurs discursifs, ponctuations, citations} alors que pour *All*, la meilleure combinaison correspond à {présence de ponctuation, mots en lettres majuscules}. Les principales conclusions que l'on peut tirer du tableau 2 sont : (1) Dans *NegOnly*, les traits sémantiques pris séparément (sentiment, modifieurs, opposition, etc.) ne sont pas suffisants pour classer les tweets `NIR` et `IR`. On note en particulier que les résultats de la baseline pour la classe `NIR` sont meilleurs que ceux pour `IR` (respectivement 77,60 et 66,40 en F-mesure). (2) Les traits de sentiment sont les plus fiables pour *NoNeg* en utilisant le trait de surprise/étonnement associé aux traits de fréquence des mots d'opinion. Ici aussi, `NIR` obtient 12,7 points de plus que `IR` avec une F-mesure de 73,30. (3) Les traits pour les modifieurs et oppositions sont les meilleurs pour *All*. Comme pour les autres corpus, on remarque que les prédictions du classifieur sont meilleures pour la classe `NIR` que pour `IR` mais avec un écart moindre (2,2 en utilisant les modifieurs et 7,4 en utilisant les oppositions).

4. Pour C_{NoNeg} et C_{All}, nous avons testé une validation croisée sur 10 échantillons avec une distribution équilibrée entre les instances ironiques et non ironiques mais les résultats sont beaucoup moins bons

Le tableau 3 détaille les résultats globaux quand les classifieurs sont entraînés sur tous les traits pertinents de groupe. Les résultats sont donnés en termes de précision (P), rappel (R), F-mesure (F, macro-moyenne) et exactitu résultats sont meilleurs pour *All* que pour *NegOnly* et *NoNeg*. Ces résultats sont obtenus en utilisant les 3 traits face {mots en lettres majuscules, connecteurs d'opposition, longueur du tweet}, les modifieurs {présence d'inter et négations} et les traits d'opposition {présence d'opposition explicite et implicite}. La meilleure combinais *NegOnly* est composée de 2 traits de surface {mots en lettres majuscules, citation} et du trait d'opposition. Fina si on ne considère pas les tweets contenant des négations (i.e. *NoNeg*), les performances tombent à 69,30% d tude. La meilleure combinaison est la suivante : traits de surface {ponctuation, mots en lettres majuscules, inter citation, connecteurs discursifs, connecteurs d'opposition, longueur du tweet}, sentiment {(présence de mots d' positifs/négatids/neutres) et modifieurs de sentiment {mots d'opinion modifiés par un intensifieur ou une modalité pouvons ainsi tirer 4 conclusions : (1) Les traits de surface sont essentiels pour la détection de l'ironie, surtout p tweets sans négation, (2) La négation est un trait important pour cette tâche mais ne suffit pas : en effet, parm tweets mal classés par C_{All}, 60% contiennent des négations (37 `IR` et 9 `NIR`), (3) Pour les tweets contenant une n les traits d'opposition sont les plus efficaces, (4) Les mots d'opinion sont plus susceptibles d'être utilisés dans le sans négation.

	Ironique (IR)			**Non ironique (NIR)**		
	P	R	F	P	R	F
$C_{NegOnly}$	0.889	0.56	0.687	0.679	0.933	0.785
C_{NoNeg}	0.711	0.651	0.68	0.678	0.735	0.705
C_{All}	0.93	0.816	0.869	0.836	0.939	0.884

	Résultats (meilleure combinaison)	
	F-score (macro-moyenne)	Exactitude
$C_{NegOnly}$	73.60	74.46
C_{NoNeg}	69.25	69.30
C_{All}	**87.65**	**87.70**

TABLE 3 – Résultats pour les meilleures combinaisons de traits.

Pour les 3 classifieurs, une analyse d'erreur montre que les erreurs de classification sont principalement dues à 4 facteurs : la présence de comparaison, l'absence de contexte, l'humour ou de mauvais hashtags *#ironie* ou *#sarcasme*. La comparaison est une forme d'ironie par laquelle on attribue des caractéristiques à un élément en le comparant à un élément complètement différent (*e.g. "Benzema en équipe de France c'est comme le dimanche. Il sert à rien"*). Ce type d'ironie utilise souvent des marqueurs de comparaison. Nous ne traitons pas ce phénomène pour le moment mais une approche par similarité sémantique pourrait être utilisée (Veale & Hao, 2010). L'absence de contexte est responsable de la majorité des erreurs. En effet, l'interprétation des tweets mal classés nécessite des connaissances contextuelles extérieures aux tweets. Cette absence de contexte peut se manifester de plusieurs façons : (1) Le thème du tweet n'est pas mentionné (*e.g.* "*Elle nous avait manqué !*" ou bien l'ironie doit être inférée des hashtags (e.g. #poissondavril) ; (2) L'ironie porte sur une situation spécifique, par exemple un épisode d'une série télé ou une situation géographique ; (3) de fausses assertions comme dans "*Ne vous inquietez pas. Le Sénégal sera champion du monde de Football*" ; (4) Des oppositions qui impliquent une contradiction entre 2 mots qui ne sont pas sémantiquement reliés (*e.g. "ONU"* et *"organization terroriste"*, *"Tchad"* et *"élection démocratique"*). Ce cas est plus fréquent dans les tweets sans négation alors que les cas (2) et (3) le sont plus dans les tweets avec négation. Ces résultats sont très encourageants car les travaux qui se sont intéressés à cette même tâche ont atteint des scores de précision de 30% pour le néerlandais (Liebrecht *et al.*, 2013) et 79% (Reyes *et al.*, 2013) pour l'anglais par exemple.

4 Conclusion

Dans cet article, nous avons présenté une approche par apprentissage automatique pour la détection de l'ironie dans les tweets. Nous avons vu que les traits de surface traditionnellement utilisés pour cette tâche dans d'autres langues sont aussi efficaces pour le français. Nous avons introduit de nouveaux traits qui nous ont permis de tester deux hypothèses : la présence de négation et celle d'opposition explicite ou implicite peut aider à détecter l'ironie. Les résultats obtenus sont très encourageants. A court terme, nous prévoyons d'améliorer la classification des tweets pour lesquels le contexte est absent, en exploitant par exemple l'information extra-linguistique.

Remerciements

Ce travail a été financé par le projet ANR ASFALDA ANR-12-CORD-023.

Références

ASHER N., BENAMARA F. & MATHIEU Y. (2009). Appraisal of Opinion Expressions in Discourse. *Linguisticae Investigationes 32 :2*.

BARBIERI F. & SAGGION H. (2014). Modelling Irony in Twitter : Feature Analysis and Evaluation. In *Proceedings of Language Resources and Evaluation Conference (LREC)*, p. 4258–4264.

BENAMARA F., MORICEAU V. & MATHIEU Y. Y. (2014). *TALN-RECITAL 2014 Workshop DEFT 2014 : DÉfi Fouille de Textes (DEFT 2014 Workshop : Text Mining Challenge)*, chapter Catégorisation sémantique fine des expressions d'opinion pour la détection de consensus, p. 36–44. Association pour le Traitement Automatique des Langues.

BURFOOT C. & BALDWIN C. (2009). Automatic satire detection : Are you having a laugh ? In *Proceedings of the ACL-IJCNLP 2009 conference short papers*, p. 161–164 : Association for Computational Linguistics.

BUSCHMEIER K., CIMIANO P. & KLINGER R. (2014). An impact analysis of features in a classification approach to irony detection in product reviews. *ACL 2014*, p.42.

CARVALHO P., SARMENTO L., SILVA M. J. & OLIVEIRA E. D. (2009). Clues for detecting irony in user-generated contents : oh... ! ! it's so easy ;-). In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, p. 53–56 : ACM.

GONZALEZ-IBANEZ R., MURESAN S. & WACHOLDE N. (2011). Identifying sarcasm in Twitter : a closer look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies : short papers-Volume 2*, p. 581–586 : Association for Computational Linguistics.

GRICE H. P., COLE P. & MORGAN J. L. (1975). Syntax and semantics. *Logic and conversation*, **3**, 41–58.

KREUZ R. J. & CAUCCI G. M. (2007). Lexical influences on the perception of sarcasm. In *Proceedings of the Workshop on computational approaches to Figurative Language*, p. 1–4 : Association for Computational Linguistics.

LIEBRECHT C., KUNNEMAN F. & VAN DEN B. A. (2013). The perfect solution for detecting sarcasm in tweets# not. In *Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* : New Brunswick, NJ : ACL.

LIU B. (2012). Sentiment Analysis and Opinion Mining (Introduction and Survey). In M. . C. PUBLISHERS, Ed., *Synthesis Lectures on Human Language Technologies*.

REYES A. & ROSSO P. (2011). Mining subjective knowledge from customer reviews : a specific case of irony detection. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, p. 118–124 : Association for Computational Linguistics.

REYES A. & ROSSO P. (2012). Making objective decisions from subjective data : Detecting irony in customer reviews. *Decision Support Systems*, **53**(4), 754–760.

REYES A. & ROSSO P. (2014). On the difficulty of automatically detecting irony : beyond a simple case of negation. *Knowledge and Information Systems*, **40**(3), 595–614.

REYES A., ROSSO P. & VEALE T. (2013). A multidimensional approach for detecting irony in Twitter. *Language Resources and Evaluation*, **47**(1), 239–268.

RILOFF E., QADIR A., SURVE P., SILVA L. D., GILBERT N. & HUANG R. (2013). Sarcasm as contrast between a positive sentiment and negative situation. In *EMNLP*, p. 704–714.

ROZE C., DANLOS L. & MULLER P. (2012). Lexconn : A French lexicon of discourse connectives. *Discours, Multi-disciplinary Perspectives on Signalling Text Organisation*, **10**, (on line).

SPERBER D. & WILSON D. (1981). Irony and the use-mention distinction. *Radical pragmatics*, **49**, 295–318.

TSUR O., DAVIDOV D. & RAPPOPORT A. (2010). ICWSM-A Great Catchy Name : Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews. In *Proceedings of ICWSM*.

UTSUMI A. (1996). A unified theory of irony and its computational formalization. In *Proceedings of the 16th conference on Computational linguistics-Volume 2*, p. 962–967 : Association for Computational Linguistics.

UTSUMI A. (2004). Stylistic and contextual effects in irony processing. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, p. 1369–1374.

VEALE T. & HAO Y. (2010). Detecting ironic intent in creative comparisons. In *Proceedings of ECAI*, volume 215, p. 765–770.