Proposition méthodologique pour la détection automatique de Community Manager. Étude multilingue sur un corpus relatif à la Junk Food

Johan Ferguth, Aurélie Jouannet, Asma Zamiti, Yunhe Wu, Jia Li, Antonina Bondarenko, Damien Nouvel, Mathieu Valette

▶ **To cite this version:**

Johan Ferguth, Aurélie Jouannet, Asma Zamiti, Yunhe Wu, Jia Li, et al.. Proposition méthodologique pour la détection automatique de Community Manager. Étude multilingue sur un corpus relatif à la Junk Food. 22éme conférence sur le Traitement Automatique des Langues Naturelles, 2015, Caen, France. hal-01359439

HAL Id: hal-01359439
https://inalco.hal.science/hal-01359439

Submitted on 2 Sep 2016

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.

Proposition méthodologique pour la détection automatique de Community Manager. Étude multilingue sur un corpus relatif à la Junk Food

Johan Ferguth [1] Aurélie Jouannet [1] Asma Zamiti [1] Yunhe Wu [1] Jia Li [1]
Antonina Bondarenko [1] Damien Nouvel [1] Mathieu Valette [1]
(1) ERTIM, INALCO - 2 rue de Lille, 75007 Paris
jferguth@gmail.com, jouannet.aurelie@gmail.com, zamiti.asma@gmail.com,
yunhe.wu9@gmail.com, maggieleebonnie@gmail.com, tonyabondarenko@gmail.com,
damien.nouvel@inalco.fr, mvalette@inalco.fr

Résumé. Dans cet article, nous présentons une méthodologie pour l'identification de messages suspectés d'être produits par des Community Managers à des fins commerciales déguisées dans des documents du Web 2.0. Le champ d'application est la malbouffe (junkfood) et le corpus est multilingue (anglais, chinois, français). Nous exposons dans un premier temps la stratégie de constitution et d'annotation de nos corpus, en explicitant notamment notre guide d'annotation, puis nous développons la méthode adoptée, basée sur la combinaison d'une analyse textométrique et d'un apprentissage supervisé.

Abstract.
Methodological Proposal for Automatic Detection of Community Manager. Multilingual Study based on a Junk Food corpus
This article describes the methodology for identifying a certain kind of speech in internet forums. The detection of the speech of a Community Manager combines recent issues in the domain of Natural Language Processing, including opinion mining and sentiment analysis, with another more abstract problem. Going beyond detecting the polarity of a message, this project targets the underlying intentions and identity of the author of the message on the forum.

Mots-clés : Community Management, Textométrie, Multilinguisme, Fouille de texte.

Keywords : Community Management, Textometry, Multilingualism, Data Mining.

Introduction

Le métier de gestionnaire de communauté (*community manager* désormais CM) est né de l'essor des sites communautaires et des réseaux sociaux. Ces derniers ont bouleversé la relation hiérarchique, verticale et unilatérale qui était établie jusqu'alors entre les entreprises et les consommateurs. Ainsi, chaque individu est désormais un émetteur légitime. Ce nouveau mode de communication a obligé les entreprises à repenser entièrement leur stratégie pour placer les communautés au centre de leurs dispositifs. Dans ce contexte, le CM a notamment pour mission de développer la visibilité de l'entreprise et de faire fructifier son capital social (e-réputation, etc.). Si, habituellement, dans l'entreprise, l'existence du CM est officielle et son activité non dissimulée, il arrive que certains professionnels valorisent leurs produits ou enseignes en recourant à de faux commentaires positifs. L'objectif est de noyer un nombre important de vrais commentaires négatifs ou de pallier l'absence de commentaires.

Nous relatons dans cet article une expérience de classification automatique de commentaires de forums de discussions ayant trait à la restauration rapide (fast-food). L'objectif applicatif est d'identifier les messages potentiellement écrits par des CM sans que ceux-ci se soient explicitement présentés comme tels. Dans cette optique, nous allons adopter une approche hybride associant une méthode de linguistique de corpus pour la production de descripteurs sémantiques et une phase d'apprentissage supervisé pour la classification des messages. Dans un premier temps nous exposerons notre problématique en mettant le focus sur la difficulté d'identifier le texte d'un énonciateur dont les intentions sont dissimulées.

Nous avons développé quatre corpus de langues différentes : anglais, chinois, français et russe [1]. Pour des raisons éditoriales, nous nous focaliserons sur le corpus français pour la description précise de la méthodologie. Pour les autres langues, nous nous contenterons de présenter quantitativement les corpus et de donner les résultats de la classification.

Après un bref état de l'art, nous présenterons en détails notre corpus et la méthodologie de constitution en 2.1. Ensuite, nous détaillerons la méthodologie de sélection des descripteurs et leur caractérisation sémantique en 2.2. Nous indiquerons comment nous avons utilisé l'apprentissage automatique en 3. Enfin, nous conclurons avec un bilan du travail réalisé et présenterons quelques perspectives.

1 État de l'art

La problématique de la tromperie (*deception*) est corrélée à trois applications principales, relativement semblables : l'attribution d'auteurs, la détection de plagiat et la détection de messages à intention dissimulée. Pour les deux premières, les méthodes stylométriques (Juola 2012) ou par mesure statistique de la distance intertextuelle (Luong, éd. 2003) sont souvent utilisées, en plus des méthodes de similarité standard de la fouille de textes (Koppel *et al.* 2009, Afroz *et al.* 2012). D'autres utilisent des méthodes hybrides articulant une approche linguistique, comme par exemple (Rubin et Vashchilko 2012) qui associent une théorie linguistique d'inspiration cognitiviste (Rhetorical Structure Theory) et des méthodes d'apprentissage supervisé.

En fouille de texte, si le champ des méthodes utilisables est relativement bien balisé et correspond à l'état de l'art (dominé par les méthodes de classification), la constitution des jeux de données pour l'application de méthode supervisée est rendue difficile par la nature intrinsèquement cachée des corpus et fait, en soi, l'objet de recherche. (Gokhman *et al.* 2012) en font un intéressant état de l'art et distinguent deux approches. La première, que nous pourrions qualifier de descendante, consiste à demander à des participants de produire leur corpus dupé, volontairement ou par annotation rétrospective de textes. Cette pratique du panel volontaire est souvent utilisée dans le contexte de recherche en psychologie cognitive. Des approches ascendantes, plus en phase avec les pratiques de la fouille de textes, consistent à élaborer une heuristique pour l'identification de messages suspects. Par exemple, en mesurant sur un site recensant des avis de consommateurs (e.g. Amazon) les commentaires très similaires en fréquences anormalement élevées et/ou en un court laps de temps (Wu *et al.* 2010b) (Jindal et Liu 2008) ou en identifiant des anomalies dans les écarts de satisfaction sur un même produit (Wu *et al.* 2010b). (Gokhman *et al.* 2012) considèrent que ces approches n'offrent pas un véritable *gold standard* mais une approximation acceptable. Nous nous inspirons de cette méthodologie dans notre travail.

2 Corpus et méthodologie

2.1 Stratégies d'élaboration du jeu de données et présentation du corpus

La difficulté inhérente à ce type de tâches réside dans l'obtention d'un corpus annoté de textes écrits par des énonciateurs qui dissimulent leurs intentions. Pour pallier cette difficulté nous avons établi une catégorisation binaire : commentaires *suspectés* d'avoir été rédigés par un CM *vs* commentaires *non suspectés* d'avoir été rédigés par un CM. Nous avons sélectionné manuellement entre 300 et 700 commentaires (de dix mots minimum) suivant les langues de l'étude, principalement sur des forums de discussion [2]. Nous avons effectué une campagne d'annotation comprenant une validation croisée pour chaque langue (2 à 3 annotateurs par langue). Le guide d'annotation a été réalisé avec la participation d'un ancien professionnel du *community management*, qui nous a exposé les stratégies rédactionnelles les plus fréquentes. On pourra en lire le détail dans la liste des règles de sélection ci-après. En complément du corpus web 2.0, nous avons collecté un ensemble de textes émanant des sites web institutionnels des marques considérées dans le contexte applicatif de la *junk-food*, de façon à construire un corpus de référence qui sera intégré dans les tâches de classification. Nous avons fait l'hypothèse que ce corpus nous permettrait de caractériser un discours propre aux CM (nonobstant les variations liées aux genres textuels très différents) et d'établir une échelle de la suspicion.

- Règles de sélection des suspects
  - très (trop) bonne expression écrite,
  - stratégie de mauvaise écriture,

1. Les résultats obtenus sur le corpus russe ayant été jugés peu suffisants, nous ne les présenterons pas ici
2. Exemples pour le français : forum.hardware.fr, Tripadvisor, Doctissimo, www.jeuxvideo.com

- discours uniquement orienté vers le fait de convaincre,
- défense d'une marque contre un scandale,
- valorisation d'un produit et dénigrement d'un produit similaire ou valorise une enseigne plutôt qu'une autre (ex : *les frites de McDo sont pas terribles, je préfère celles de Quick*),
- insiste sur les quantités (ex : *Par exemple, chez Mc Do, je prends le filet-o-fish(un peu moins de 400 cal)*),
- vocabulaire choisi (ex : *Avec gros plan sur ces délicieux beignets de poulets qui vont faire trempette dans des sauces de toutes les couleurs. Miam !*),
- proposition de réductions (ex : *Et avec la carte étudiant, tu peux en avoir 3 de plus pour 1 €si je me souviens bien*),
- conseil (ex : *Essaye d'y aller a plusieurs et prendre un bucket vraiment bon rapport, pour avoir pour son argent*),
- déculpabilisation (ex : *Nous y allons sans aucun complexe ni scrupule !*),
- projection de la vie quotidienne mais générale (ex : *c'est à dire lors des voyages pour partir en vacances, ou lorsque nous faisons plusieurs magasins (ex : meubles, soldes)*),
- reprise claire d'un contre argument (ex : *je suis plutot bon client de Mcdo... sans y aller regulierement, j'y ai mes habitudes quand j'ai envie de "gras" avec un gout totalement chimique mais tellement bon... pas de fantaisie, j'y commande toujours a peu pres la meme chose (menu maxi best of deluxe/bacon/big mac + croque mcdo ou un autre gros sandwich) pas trop trop cher, ca cale convenablement*),
- très précis (chiffré) (ex : *McDonald's France accueille plus de 1,2 million de clients par jour, soit 440 millions par an ; cela représente 13 clients qui poussent la porte d'un restaurant Mc Donald en moyenne chaque seconde. Pas mal au pays de la gastronomie.*).

- Règles de sélection des non-suspects
  - présence de plusieurs noms de marques ou produits différents,
  - ne pas citer de nom de produit (ex : *j'aime bien aller au fastfood de temps en temps*),
  - parle de soi / sa famille (ex : *en y allant avec ma famille, ...*),
  - ne valorise pas une enseigne par rapport à une autre (ex : *frites, burger et milkshake : McDo. Quick pour les sauces*),
  - reprise du discours officiel pour le critiquer (comme des listes d'ingrédients et des calories, par ex : *Tout repas supérieur a ce repas "type" est a la fois hypercalorique (on mange facilement 1500kcal au mcdo, donc plus de la moitié des kcal nécessaire journalier !), hyperlipidique (on arrive a 50% lipides avec une part de frites...), et déséquilibré : même une fois par semaine c'est pas tres bon.*).

Le tableau 1 présente le nombre de messages ou de mots collectés pour les différents corpus. Pour le corpus chinois, le module JIEBA [3] a été utilisé pour segmenter les textes en mots. Nous constatons des disparités, autant du point de vue du nombre de mots dans les corpus que dans la proportion de messages suspects.

	Français	**Anglais**	**Chinois**
Corpus total (mots)	61 205	32 588	8 051
Non-suspects (messages)	249	251	516
Suspects (messages)	61 (20%)	25 (9%)	184 (26%)
Institutionnels (messages/mots)	4 / 30 917	26 / 13 604	3 / 1 850

TABLE 1 – Volume des corpus collectés

2.2 Méthodologie de constitution des descripteurs sémantiques

2.2.1 Sélection textométrique des descripteurs

Nous tentons de mettre en évidence les phénomènes textuels qui différencient les témoignages de nos deux catégories. Notre objectif est de trouver des descripteurs sémantiquement explicables et suffisamment robustes pour servir de caractéristiques aux méthodes d'apprentissage supervisé. Après le test de plusieurs logiciels implémentant les algorithmes de spécificités (Lafon, 1980) utilisés, notre choix s'est porté, pour l'analyse du corpus et l'extraction des critères, sur le logiciel Lexico 3 (Salem *et al.*, 2003), qui s'est avéré le plus robuste dans notre contexte multilingue.

3. https://github.com/fxsjy/jieba

Notre méthodologie s'inspire des travaux de Eensoo et Valette (2012, 369-370). Nous avons choisis les descripteurs en effectuant un calcul des spécificités (Lafon, 1980) sur les formes isolées pour chaque sous corpus (suspect/non-suspect). Nous avons ensuite examiné les concodances des candidats descripteurs pour nous assurer de leur pertinence sémantique et de leur monosémie (sauf pour le corpus chinois, les sinogrammes étant très polysémiques). Nous avons ensuite contrôlé visuellement leur répartition uniforme dans le sous corpus considéré en utilisant la fonction de partition (fréquence, spécificités, carte des sections).

En utilisant ce logiciel, nous avons comparé les spécificités de chaque classe à celle des deux autres afin de voir quels termes étaient les plus spécifiques aux suspects, aux non suspects et aux institutionnels. Nous avons ainsi inventorié plusieurs dizaines de critères sémantiques pour chaque classe, puis nous les avons caractérisés en fonction des composantes sémantiques suivant une lecture librement inspirée de (Rastier 2001) : composante *dialectique* (*i.e.* représentation du temps et du déroulement aspectuel, des rôles et des interactions entre acteurs), composante *dialogique* (positionnement énonciatif des acteurs) et composante *thématique*. Cette dernière inclut une composante *thymique* relative à l'affectivité du locuteur, l'expression de sa subjectivité.

2.2.2 Présentations des descripteurs

La classification de nos descripteurs nous a permis de constater des particularités propres à chacune des catégories. Pour le corpus français, on observe les tendances suivantes. Les messages suspects ont tendance à citer des ingrédients et ont ainsi un propos davantage centré sur des thématiques. Les non suspects, quant à eux, utilisent un contenu thématique plus varié ainsi que des éléments issus de la composante dialectique (emphase, argumentation).Leur positionnement énonciatif est aussi plus prononcé que chez les suspects, par l'usage des pronoms *je* et *tu* (composante dialogique). Enfin, les textes institutionnels, c'est-à-dire issus des sites web des marques, utilisent moins d'éléments caractéristiques, leur discours étant de fait plus formels que le discours provenant des messages de forums. L'utilisation des 1re et 2e personnes du pluriel illustre la relation entreprise-clientèle que l'on retrouve aussi dans leur contenu thématique avec des unités lexicales spécifiques au commerce.

Ci-dessous, on donne le détail de tous les descripteurs pour le français, à titre d'exemple.

- suspects
  - dialogique : (aucun)
  - thématique : *menu*, *mozzarella*, *chez*, *steak*, *fromage*, *pizzas*, *ingrédients*, *rumeur*, *nuggets*, *glaces*
    - thymique : *délicieux*, *simple*
  - dialectique : (aucun)
- non-suspects
  - dialogique : *je*, *tu*
  - thématique : *manger*, *plaisir*, *malbouffe*, *santé*, *semaine*, *gens*, *gout*, *nourriture*, *alimentation*, *animaux*, *plastique*, *problème*
    - thymique : *mauvais*, *merde*, *bon*
  - dialectique : *pense*, *mais*, *pas*, *moins*, *!*, *?*, *vraiment*, *très*
- institutionnels
  - dialogique : *vous*, *vos*, *nous*, *votre*, *notre*
  - thématique : *enfant*, *formation*, *préparation*, *allergène*, *produits*, *clients*, *services*
  - dialectique : (aucun)

3 Classification et évaluation des résultats

Nous utilisons Weka (Hall *et al.*, 2009) pour réaliser l'apprentissage automatique, pour lequel de nombreux algorithmes ont été expérimentés, parmi lesquels *Naive Bayes* (NB) et *Naive Bayes Multinomial* (NBM) se sont révélés les plus performants. Selon le corpus, nous disposons de deux ou trois classes. En anglais, la catégorisation doit distinguer entre les messages suspects (S), non-suspects (NS) et institutionnels. Le chinois ne considère que les messages suspects (S) et non-suspects (NS). Pour le français, la stratégie a consisté à fusionner [4] les classes des messages suspects et institutionnels (S+I) et de les distinguer des messages non-suspects (NS). Il serait de fait intéressant d'aller plus loin en examinant le

4. Cette fusion, basée sur l'hypothèse que les messages les plus suspects se rapprochent du discours institutionnel a permis d'améliorer les performances.

vocabulaire commun aux deux classes, de façon à affiner la stratégie de fusion (on pourrait ainsi considérer uniquement le vocabulaire spécifique aux deux classes). Pour cela, à nouveau, il faudrait obtenir un corpus clairement issu de *CMs* pour pouvoir établir un comparatif de qualité.

Le tableau 2 présente les résultats obtenus pour la classification des textes du corpus d'apprentissage et par validation croisée à dix plis, en les comparant à la stratégie consistant à affecter à tous les messages la classe majoritaire. Nous constatons que les expériences préliminaires, sans descripteurs textométriques et par utilisation du filtre StringToWordVector [5], obtiennent des résultats très proches des classes majoritaires : ceci montre clairement la difficulté de la tâche à laquelle nous sommes confrontés. L'utilisation des descripteurs textométriques permet d'améliorer les résultats en termes d'exactitude pour le français et le chinois.

	Français	**Anglais**	**Chinois**
Catégorisation	S + I / NS	S / I / NS	S / NS
Classe majoritaire	80.3%	97.1%	73.7%
Sans descripteurs			
Lemmatisation	non	non	non
Algorithme	NBM	NB	NB
Mots distincts	6000	1000	2854
Exactitude	80,2%	94,9%	74,6%
Avec descripteurs			
Nb descripteurs	51	63	80
Algorithme optimal	NBM	NB	NB
Exactitude	82,8%	91,5%	75,81%

TABLE 2 – Résultats obtenus sur corpus avec ou sans descripteurs sémantiques

Nous remarquons le gain en exactitude pour le français, dont le taux de bonne classification passe de 80,2% à 82,8%. Une analyse plus poussée des résultats sur ce corpus montre que, sur 65 messages suspects (dont 4 institutionnels, par fusion), seuls 3 sont correctement classés sans descripteurs, tandis qu'avec descripteurs, 18 le sont. Cela impacte par effet de bord la bonne classification des messages non-suspects pour lesquels, sur 249 messages, aucun n'était mal classé sans descripteurs, tandis qu'avec les descripteurs 7 deviennent étiquetés de manière erronée comme suspects. Globalement, l'apport est plus remarquable lorsque l'on calcule la f-mesure, qui passe de 72,3% à 79,6% grâce à l'utilisation des descripteurs textométriques.

Pour l'anglais, nos résultats, peu concluants, peuvent s'expliquer de plusieurs manières. D'une part, le grand déséquilibre des commentaires entre la catégorie *suspects* et les deux autres peut rendre la tâche plus difficile (mais nous pensons que ceci est représentatif de ce qui a été trouvé sur internet lors de la constitution du corpus : aussi bien dans les commentaires d'articles que dans les forums, les publications pouvant être catégorisées dans *suspects* selon le guide d'annotation semblent rares et hétérogènes). Il nous a également semblé, au vu des messages plus répandus sur les sites de partage d'avis, que la diffusion de faux commentaires relèverait plutôt d'initiatives personnelles de restaurants que d'une stratégie globale des marques, ce qui peut renforcer leur hétérogénéité. Compte tenu de ces résultats, notre intuition est que les différences culturelles et sociétales (lobbying, culture du fast-food, etc.) dans les pays anglophones, pourrait expliquer que les marques n'aient pas besoin d'intégrer ce type de pratiques dans leurs stratégies globales.

Le cas du corpus chinois présente des particularités : le langage est bien plus hétérogène entre les forums et les sites institutionnels. Ainsi, le registre du corpus institutionnel est beaucoup plus soutenu, ce qui pose des difficultés sur le plan du lexique et impacte fortement les apprentissages automatiques. Il a donc été décidé de ne pas inclure les sites institutionnels, qui n'amélioraient pas, voire dégradaient, les résultats. Les expériences portant alors sur deux classes font obtenir un score de 74,61% sur deux classes avec l'algorithme Naive Bayes. En y ajoutant les 60 descripteurs sémantiques repérés à l'aide de Lexico3, nous obtenons 75,81% soit un gain absolu de 1,2 points (malgré la présence de descripteurs sémantiquement ambigus - comme de nombreux mots en chinois).

Nous retenons, essentiellement, deux éléments de ces expérimentations. D'une part, l'utilisation de l'apprentissage automatique, sur une tâche aussi subjective et avec un volume aussi faible de données, est difficile à mettre en œuvre et demande un ajustement méticuleux des paramètres pour éviter le sur-apprentissage. Nous pensons être dans un cas qui se rapproche de la reconnaissance de *signaux faibles*. Nous remarquons que la textométrie aide manifestement à pallier

5. Filtre Weka qui convertit les attributs textuels en vecteurs de mots. Un paramètre permet d'ajuster le nombre de mots distincts sélectionnés (pour le chinois, il correspond au nombre total : 2854).

ces difficultés et permet d'obtenir des résultats qui, s'ils demanderaient à être mieux établis quantitativement, le sont très clairement du point de vue qualitatif : l'utilisation des descripteurs permet d'obtenir plus de messages classés suspects et évite de recourir à une stratégie trop proche de la classe majoritaire (*i.e.* classer tous les messages comme non suspects).

Conclusion

Dans cet article, nous avons tenté d'établir une classification en utilisant des descripteurs sémantiques pour une tâche de reconnaissance des messages dissimulés de gestionnaires de communautés dans un contexte multilingue. Nous avons adopté une méthode hybride associant la linguistique de corpus et la classification supervisée. Les résultats obtenus varient d'une langue à une autre et ne sont pas systématiquement concluants. Les difficultés rencontrées sont cependant inhérentes à la nature même de la problématique initiale, à savoir détecter des émetteurs qui ont pour but la dissimulation. Les résultats sont tout de même encourageants pour une première approche et nous permettent d'envisager des développements ultérieurs.

Références

AFROZ S., BRENNAN M. & GREENSTADT R. (2012). Detecting hoaxes, frauds, and deception in writing style online. In *Proceedings of the 33rd conference on IEEE Symposium on Security and Privacy*.

AKOGLU L., CHANDY R. & FALOUTSOS C. (2013). Opinion fraud detection in online reviews by network effects. In *ICWSM*.

ALMELA Á., VALENCIA-GARCÍA R. & CANTOS P. (2012). Seeing through deception : A computational approach to deceit detection in written communication. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, p. 15–22 : Association for Computational Linguistics.

BACHENKO J., FITZPATRICK E. & SCHONWETTER M. (2008). Verification and implementation of language-based deception indicators in civil and criminal narratives. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, p. 41–48 : Association for Computational Linguistics.

DOHSE K. (2013). Fabricating feedback : Blurring the line between brand management and bogus reviews. *Journal of Law, Technology and Policy, Forthcoming*.

EENSOO E. & VALETTE M. (2012). Sur l'application de méthodes textométriques à la construction de critères de classification en analyse des sentiments. In *TALN 2012*, volume 2, p. 367–374 : GETALP-LIG.

FENG S., BANERJEE R. & CHOI Y. (2012). Syntactic stylometry for deception detection. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics : Short Papers-Volume 2*, p. 171–175 : Association for Computational Linguistics.

FITZPATRICK E. & BACHENKO J. (2012). Building a data collection for deception research. In *Proceedings of the workshop on computational approaches to deception detection*, p. 31–38 : Association for Computational Linguistics.

FLEURY S. & ZIMINA M. (2014). Trameur : A framework for annotated text corpora exploration. *COLING 2014*, p.57̃.

GOKHMAN S., HANCOCK J., PRABHU P., OTT M. & CARDIE C. (2012). In search of a gold standard in studies of deception. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, p. 23–30 : Association for Computational Linguistics.

HALL M., FRANK E., HOLMES G., PFAHRINGER B., REUTEMANN P. & WITTEN I. H. (2009). The weka data mining software : an update. *ACM SIGKDD explorations newsletter*, **11**(1), 10–18.

HAMON T. (2012). Acquisition terminologique pour identifier les mots clés darticles scientifiques. *Actes du huitième DÉfi Fouille de Textes*, p.28̃.

HAUCH V., MASIP J., BLANDON-GITLIN I. & SPORER S. L. (2012). Linguistic cues to deception assessed by computer programs : a meta-analysis. In *Proceedings of the workshop on computational approaches to deception detection*, p. 1–4 : Association for Computational Linguistics.

JINDAL N. & LIU B. (2008). Opinion spam and analysis. In *WSDM '08 Proceedings of the 2008 International Conference on Web Search and Data Mining*, p. 219–230 : ACM.

JUOLA P. (2012). Detecting stylistic deception. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, p. 91–96 : Association for Computational Linguistics.

KOPPEL M., SCHLER J. & ARGAMON S. (2009). Computational methods in authorship attribution. In *Journal of the American Society for Information Science and Technology*, p. 9–26.

LAFON P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots*, **1**(1), 127–165.

LI D. & SANTOS JR E. (2012). Argument formation in the reasoning process : toward a generic model of deception detection. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, p. 63–71 : Association for Computational Linguistics.

LUONG X. (2003). La distance intertextuelle. *Revue Corpus*.

MO Q. & YANG K. (2014). Overview of web spammer detection. *Journal of Software 25(7)*.

RUBIN V. L. & VASHCHILKO T. (2012). Identification of truth and deception in text : Application of vector space model to rhetorical structure theory. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, p. 97–106 : Association for Computational Linguistics.

SALEM A., LAMALLE C., MARTINEZ W., FLEURY S., FRACCHIOLLA B., KUNCOVA A. & MAISONDIEU A. (2003). Lexico3–outils de statistique textuelle. manuel dutilisation. *Syled-CLA2T, Université de la Sorbonne nouvelle–Paris*, **3**.

SPORER S. L. (2012). Making the subjective objective ?: computer-assisted quantification of qualitative content cues to deception. In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, p. 78–85 : Association for Computational Linguistics.

VARTAPETIANCE A. & GILLAM L. (2012). I don't know where he is not : does deception research yet offer a basis for deception detectives ? In *Proceedings of the Workshop on Computational Approaches to Deception Detection*, p. 5–14 : Association for Computational Linguistics.

WU, GREENE, SMYTH & CUNNINGHAM (2010a). Distortion as a validation criterion in the identification of suspicious reviews. In *Proceedings of the First Workshop on Social Media Analytics*, p. 10–13 : ACM.

WU, GREENE, SMYTH & CUNNINGHAM (2010b). *Distortion as a validation criterion in the identification of suspicious reviews*. Rapport interne UCD-CSI-2010-04.