
Le microblogage pour la microanalyse des sentiments et des opinions

Alexander Pak — Patrick Paroubek

LIMSI - CNRS, Bât. 508 Université Paris XI BP 133 - 91403 ORSAY Cedex - France
alexpak@limsi.fr; pap@limsi.fr

RÉSUMÉ. Le microblogage est une tendance récente dans l'Internet d'aujourd'hui. Les utilisateurs expriment leurs opinions par le biais des plates-formes de microblogage comme Twitter. Dans nos travaux, nous utilisons Twitter comme une source de données multilingues pour collecter un corpus de textes exprimant des sentiments annotés. Avec le lexique issu du corpus collecté, nous construisons un classificateur des sentiments que nous appliquons à trois types de tâches : la classification des sentiments dans des textes courts, la désambiguïsation des adjectifs de sentiments ambigus et la construction de lexiques affectifs pour d'autres langues. Nous les appelons des tâches de microanalyse des sentiments car elles opèrent sur des textes courts ou des portions de textes. Les évaluations expérimentales utilisant l'ensemble des données annotées à la main, la participation à la campagne d'évaluation SemEval 2010 et la corrélation avec le lexique affectif ANEW montrent que notre méthode fonctionne bien même si nous n'utilisons pas d'outils spécifiques à la langue ou de ressources construites par des humains.

ABSTRACT. Microblogging is a recent trend in today's Internet. Users express their opinions using microblogging platforms such as Twitter. In our research, we use Twitter as a multilingual data source to collect a corpus of sentiment labeled texts. Using the lexicon extracted from our corpus, we build a sentiment classifier which we apply to three kinds of tasks: classification of sentiments in short texts, disambiguation of sentiment ambiguous adjectives, and construction of affective lexicons in different languages. We call them "micro sentiment analysis" tasks as they operate on small texts or spans of texts. Experimental evaluations using hand-annotated dataset, participation in the SemEval 2010 evaluation campaign and correlation with ANEW affective lexicon prove that our method performs well, even if we do not use language specific tools and human-built resources.

MOTS-CLÉS : analyse des sentiments, microblogage, Twitter, SemEval, ANEW, approche multilingue.

KEYWORDS: sentiment analysis, microblogging, Twitter, SemEval, ANEW, multilingual approach.

1. Introduction

Allant de pair avec un développement de l'utilisation d'Internet comme un moyen de communication populaire, l'analyse de sentiments et la fouille d'opinion sont devenues les sujets de recherche de beaucoup de scientifiques. Sur la Toile, il y a de plus en plus de sites qui donnent aux internautes les moyens d'exprimer leur opinion ou de s'enquérir de celle de leurs semblables, comme par exemple les sites de critiques de films, les guides gastronomiques, de shopping ou les réseaux sociaux. Les derniers apparus sur cette scène sont les sites de « microblogage ». Ils permettent aux utilisateurs de poster de petits messages (certains sites autorisent l'attachement aux messages d'images issues d'autres médias). En raison de sa simplicité, il y a de plus en plus d'utilisateurs qui ont ajouté le microblogage comme un outil principal de communication au même titre que le courrier électronique ou les messages instantanés. Les grandes compagnies utilisent le microblogage comme moyen de publication de leurs nouvelles et les célébrités pour communiquer avec leurs fans. Jansen *et al.* (2009) l'ont identifié à *du bouche-à-oreille en ligne* concernant les marques commerciales.

Puisque le public des services de microblogage croît de jour en jour, il est devenu possible d'utiliser des données venant de ces sites pour faire de l'analyse de sentiments ou de la fouille d'opinion. Par exemple, les sociétés fabriquant des produits manufacturés peuvent être intéressées par les réponses aux questions suivantes :

- Que pensent les clients de leurs produits (services, compagnie, etc.) ?
- Est-ce que leurs produits ont plutôt une bonne ou une mauvaise image auprès du public ?
- Quels sont les produits que les gens souhaiteraient se voir offrir ?

Les partis politiques peuvent être intéressés de savoir comment leur programme est perçu. Les organisations sociales peuvent souhaiter obtenir des retours d'opinion sur les débats en cours. Toutes ces informations peuvent être collectées facilement et rapidement à partir des sites de microblogage puisque leurs utilisateurs y publient au jour le jour leurs opinions sur de nombreux aspects de leur vie courante.

Dans nos travaux, nous nous intéressons au microblogage, et plus particulièrement à l'un des principaux acteurs de ce type de média, Twitter¹, qui est à ce jour une des plates-formes les plus populaires. Le tableau 1 présente des exemples de messages qui ont circulé sur Twitter. Nous pensons que ce type de document peut être une source de données utile dans l'analyse de sentiments et la fouille d'opinion pour les raisons suivantes :

- Twitter est utilisé par un très large public pour exprimer son opinion sur des sujets variés, c'est donc une source intéressante pour connaître les sentiments et les opinions que le public entretient à l'égard de différentes questions ;
- il contient une énorme quantité de messages qui est en augmentation journalière. Le corpus que l'on y collecte peut donc être aussi grand qu'on le souhaite ;

1. <http://twitter.com>

funkeybrewster : @redeyechicago I think Obama's visit might've sealed the victory for Chicago. Hopefully the games mean good things for the city.
vcurve : I like how Google celebrates little things like this : Google.co.jp honors Confucius Birthday Japan Probe
mattfellows : Hai world. I hate faulty hardware on remote systems where politics prevents you from moving software to less faulty systems.
brrooklyn : I love the sound my iPod makes when I shake to shuffle it. Boo bee boo

Tableau 1. Exemples de messages publiés dans Twitter et contenant une expression d'opinion

– le public de Twitter a des origines socioculturelles très variées, de l'utilisateur lambda, aux célébrités, en passant par les représentants des grandes compagnies, les figures politiques² ; on y trouve même des chefs d'État. Il est donc possible d'obtenir facilement des textes issus de différents milieux sociaux et de groupes d'intérêts variés ;

– des internautes de nombreux pays utilisent Twitter dans leur langue d'origine³. Bien que les utilisateurs en provenance des États-Unis soient les plus nombreux, il est néanmoins possible de rassembler des données pour différentes langues en quantités suffisantes. Dans nos travaux, nous avons collecté des corpus en anglais, espagnol, français et chinois.

Dans cet article, nous utilisons Twitter comme une source de corpus multilingue permettant d'assembler des données utiles pour l'analyse de sentiments et la fouille d'opinion. Nous montrons qu'il est possible d'obtenir des listes de termes annotés avec des étiquettes représentant les sentiments exprimés dans le corpus, de manière automatique en utilisant les « émoticônes » (*emoticons*) comme un étiquetage bruité des valeurs de sentiments. Nous avons ensuite utilisé ces listes de marqueurs de sentiments dans trois types de tâches :

- 1) la classification de textes courts en fonction des sentiments qui y sont exprimés ;
- 2) la désambiguïsation d'adjectifs de sentiments ambigus ;
- 3) la construction d'un lexique affectif pour différentes langues.

Concernant notre approche, nous employons les termes de « microanalyse de sentiments » car nous effectuons une analyse de sentiments sur des textes courts pour aboutir à un champ d'analyse qui se restreint à un seul mot.

Notre article est organisé de la façon suivante : nous faisons d'abord un rapide tour d'horizon des recherches qui ont été effectuées sur le microblogage ainsi que l'analyse de sentiments et la fouille d'opinion dans la section 2. Nous présentons notre

2. <http://www.sysomos.com/insidetwitter/politics>

3. <http://www.sysomos.com/insidetwitter/#countries>

approche dans la section 3. Dans la section 4, nous appliquons notre méthode aux tâches de microanalyse de sentiments que nous avons définies. Enfin, nous présentons une évaluation expérimentale de notre travail dans la section 5, avant de conclure dans la section 6.

2. Travaux antérieurs

2.1. *Analyse de sentiments par rapport à fouille d'opinion*

Avec la création de communautés d'internautes autour des « blogs » et des réseaux sociaux, l'analyse de sentiments et la fouille d'opinion sont devenues des sujets de recherche à la mode. Un survol très large des travaux existant en la matière est proposé dans (Pang et Lee, 2008). Les auteurs y décrivent les approches et les techniques mises en œuvre dans la recherche d'information orientée vers la fouille d'opinion.

Bien que l'analyse de sentiments et la fouille d'opinion soient souvent considérées ensemble dans les publications scientifiques, certains auteurs considérant même les deux termes comme des synonymes, nous pensons qu'il est nécessaire de faire une distinction entre les concepts d'« opinion » et de « sentiment », c'est pourquoi nous en proposons les définitions suivantes :

– une **opinion** est l'expression d'une position personnelle ou d'une croyance à propos d'un événement particulier, un objet ou une personne (la cible de l'expression d'opinion) ;

– un **sentiment** est la communication d'un état émotionnel de l'auteur, dont l'origine peut être un événement, un objet ou une personne (l'origine du sentiment).

Pour illustrer ces différences, nous montrons des exemples des deux types dans le tableau 2 en nous limitant à des expressions de polarité positive ou négative afin de rester concis. Cependant, nous pensons que les expressions d'opinion peuvent avoir des structures beaucoup plus complexes que celles montrées en exemple.

Des exemples précédents nous en tirons les remarques suivantes afin de préciser nos définitions.

1) Une expression d'opinion a toujours une cible, tandis qu'un sentiment peut ne pas avoir d'origine clairement exprimée.

2) Une opinion et un sentiment peuvent être associés dans une expression, mais il est possible d'avoir des expressions n'exprimant qu'une seule opinion ou qu'un seul sentiment.

3) Une expression peut faire référence en même temps à une opinion et à un sentiment de polarités différentes.

Pour formaliser notre modèle, nous représentons une expression de sentiment comme un triplet (auteur, polarité, origine), où l'auteur possède un état émotionnel dont la polarité est positive ou négative et dont l'origine est optionnelle. Une expres-

Énoncé	Sentiment	Opinion	Cible de l'o.	Origine du s.
I am happy, cooking and listening to music on my iPod	positive	–	–	–
I am happy with the quality of my new iPod	positive	–	–	iPod
I think iPod is the best mp3 player	–	positive	iPod	–
I am happy that I have bought an iPod, it is the best mp3 player	positive	positive	iPod	purchase of iPod
I am happy that I have bought an iPod, although its price is too high	positive	négative	price of iPod	purchase of iPod
I read the new iPod press-release	–	–	–	–
So sad because my iPod is broken	négative	–	–	iPod is broken

Tableau 2. Exemples d'énoncés montrant la différence entre des sentiments et des opinions

sion d'opinion est un triplet (source, prédicat, cible), où la source exprime une assertion à propos de la cible. La représentation d'une opinion possède la même structure que celle d'un sentiment, mais le prédicat remplace la polarité et la cible est requise tandis que l'origine est optionnelle. Dans nos exemples, l'énoncé « *I think iPod is the best mp3 player* » (Je pense que l'iPod est meilleur baladeur mp3), le prédicat de l'opinion est « <CIBLE> *is the best mp3 player* » (<CIBLE> est le meilleur baladeur mp3), la cible est « iPod » et la source l'auteur/locuteur de la phrase. Dans l'énoncé « *I am happy that I have bought an iPod, although its price is too high* » (Je suis content d'avoir acheté un iPod, bien que son prix soit trop élevé.), le prédicat est « *the price of <the Target> is too high* » (le prix de <CIBLE> est trop élevé) et la cible est « iPod ». Notre représentation est similaire au quadruplet d'opinion (*Topic, Holder, Claim, Sentiment*) (sujet, source, prédicat, sentiment) de (Kim et Hovy, 2004), où le sujet correspond à la cible de notre modèle, toutefois dans notre modèle nous séparons les opinions des sentiments en deux classes distinctes.

Tandis que les sentiments sont, d'après nous, toujours polarisés, nous pensons que ce n'est pas le cas pour les opinions. Une expression d'opinion peut souvent être simplifiée voire remplacée par une expression polaire, mais dans l'exemple « *I believe the price of iPod will be \$399* » (Je pense que le prix de l'iPode sera de 399 \$), il n'y a pas

de polarité exprimée. Cependant, il se peut que le contexte implique une interprétation positive ou négative, par exemple « *Apple's products are always overpriced. I believe the price of iPod will be \$399.* » (Les produits d'Apple sont surcotés. Je pense que le prix de l'iPod sera de 399 \$.). Dans (Balahur *et al.*, 2010) les auteurs pensent qu'il est nécessaire de distinguer trois points de vue différents, qui peuvent aboutir à une interprétation différente pour une même expression d'opinion : l'auteur, le lecteur et le texte lui-même.

Les expressions d'appréciation/dépréciation nécessitent un traitement particulier car elles mêlent expression de sentiment et expression d'opinion. Par exemple, l'énoncé *I like iPod* (J'aime l'iPod.) exprime à la fois un sentiment et une opinion positive, car le sentiment positif en faveur de l'origine (la cause du sentiment) implique que l'auteur associe un prédicat positif à la cible de l'opinion. Dans notre recherche, nous avons limité la portée de nos investigations aux sentiments et aux expressions polaires d'opinion telles que appréciation/dépréciation qui sont réparties en deux classes (positive et négative). Nous supposons de plus que l'auteur/la source est le locuteur ou le rédacteur de l'énoncé considéré.

2.2. *Microblogage et émoticônes*

Dans (Yang *et al.*, 2007), ce sont des « weblogs » qui sont utilisés pour construire un corpus à des fins d'analyse de sentiments en se reposant sur les icônes d'émotion affectées aux messages postés sur les blogs comme indicateurs de l'humeur de l'auteur du message. Dans cet article, les chercheurs appliquent les algorithmes d'apprentissage de machines à vecteurs de support (SVM) et de champs aléatoires conditionnels (CRF) pour classer des phrases en fonction des sentiments qu'elles expriment. Ils ont ensuite testé plusieurs stratégies pour déduire le sentiment majoritaire au niveau du document. Dans ces expériences, la stratégie qui l'a emporté a consisté à considérer le sentiment exprimé dans la dernière phrase comme étant le sentiment majoritairement porté par l'ensemble du document.

J. Read a étudié les émoticônes telles que « :-) » et « :-(» pour construire un ensemble de données d'apprentissage pour la classification à base de sentiments dans (Read, 2005). L'auteur a assemblé des textes contenant des émoticônes en provenance des groupes de nouvelles Usenet. Les données ont été divisées en deux classes : positive (regroupant les textes contenant une émoticône joyeuse) et négative (associant les textes avec des émoticônes de tristesse ou de colère). Des systèmes de classement à base de SVM ou d'algorithme bayésien naïf entraînés sur ces données ont obtenu jusqu'à 70 % d'exactitude.

Une approche similaire à (Read, 2005) est utilisée dans (Go *et al.*, 2009), où les auteurs utilisent Twitter pour rassembler des données d'apprentissage afin d'effectuer des recherches documentaires fondées sur les sentiments. Ils construisent un corpus bipolaire en se fondant sur les émoticônes avant d'appliquer différents systèmes de classement. Les meilleurs résultats obtenus dans cette étude l'ont été avec une méth-

ode bayésienne naïve associée à la mesure d'information mutuelle pour la sélection des traits. Les auteurs rapportent des performances en exactitude mesurées sur leurs jeux de test allant jusqu'à 81 %. Cependant, la même méthode a obtenu de mauvaises performances lorsque les auteurs sont passés à une répartition en trois classes, par l'ajout d'une classe neutre pour les documents ne contenant aucune expression de sentiment.

2.3. Les lexiques affectifs

Pour l'anglais, il existe un certain nombre de lexiques affectifs pour l'analyse de sentiments, tels que ANEW (Bradley et Lang, 1999), WordNet-Affect (Strapparava et A, 2004), ou encore Balanced Affective Word List (Siegle, 1994), SentiWordNet (Baccianella *et al.*, 2010). Des versions d'ANEW adaptées à d'autres langues ont été produites, par exemple pour l'espagnol (Redondo *et al.*, 2007) et l'allemand (Vö *et al.*, 2009), mais il n'existe pas de données comparables pour le français car l'adaptation d'une base de données comme ANEW au français nécessiterait un effort humain conséquent.

Dans (Mathieu, 2006), l'auteur a assemblé un lexique sémantique électronique pour les verbes du français exprimant des sentiments, des émotions ou des états psychologiques. Le lexique contient six cents verbes répartis sur trente-trois classes sémantiques. Toutes les classes descendent de trois classes primitives associées respectivement à une polarité : positive, négative ou neutre. Ce lexique contient en outre des liens entre les classes représentant des relations de sens, d'intensité et d'antinomie. L'auteur propose avec le lexique le logiciel FEELING pour détecter les émotions dans des phrases. Bien que ce lexique représente à ce jour un des travaux les plus aboutis pour le français dans le domaine de l'analyse de sentiments, son extension à d'autres verbes ou aux adjectifs, nécessiterait un effort humain important pour un expert du domaine.

En ce qui concerne les adjectifs, dans (Harb *et al.*, 2008) les auteurs proposent d'utiliser une méthode automatique pour construire avec l'aide du moteur de recherche Google sur les blogs (*Google Blog search engine*) des listes d'adjectifs indicateurs de polarité positive et de polarité négative, et ce pour n'importe quel domaine. Le moteur de recherche est utilisé pour construire un corpus de textes exprimant des émotions en prenant comme point de départ des adjectifs « graines » pour les classes positives (bon, gentil, excellent, etc.) et négatives (mauvais, méchant, pauvre, etc.). Les auteurs ont ensuite extrait les adjectifs et déterminé leur polarité en examinant les mots présents dans les contextes des adjectifs graines. Dans ces travaux, les auteurs ne s'intéressent qu'aux adjectifs, alors que dans nos travaux nous construisons une liste de termes qui contient aussi des verbes et des noms.

3. Notre approche

3.1. Collecte du corpus

Pour toutes les expériences décrites dans cet article, nous avons utilisé un corpus de messages de Twitter récupéré avec l'interface de programmation fournie par Twitter⁴. Lors de la collecte du corpus, nous avons automatiquement réparti les messages sur trois classes : positive, négative, et neutre (Pak et Paroubek, 2010b). Pour obtenir des messages à connotation positive ou négative, nous avons utilisé les émoticônes comme représentant un étiquetage bruité indicatif du sentiment exprimé dans le message. Puisque aucun message ne peut dépasser une longueur de 140 caractères, d'après les règles des plates-formes de microblogage, ceux-ci sont pour la plupart composés d'une seule phrase. C'est pourquoi nous faisons l'hypothèse qu'une émoticône à l'intérieur d'un message représente la teneur émotionnelle du message dans son ensemble et que tous les mots du message sont en rapport avec cette émotion. Nous avons utilisé les émoticônes exprimant la joie (comme « :-) ») pour la classe des messages positifs et les émoticônes représentant la tristesse (telles que « :(») pour la classe des messages négatifs. Lorsqu'un message contenait les deux types d'émoticônes, nous l'avons ignoré. Afin de constituer un ensemble de messages neutre, nous avons identifié une liste de 44 journaux qui publient leurs titres sur leurs comptes Twitter (comme par exemple le *New York Times*, *Washington Post*, etc.).

Les étapes suivantes de normalisation des textes ont été effectuées sur le corpus ainsi collecté afin de produire les n-grammes.

1) Filtrage : nous avons supprimé les liens d'URL (e.g. <http://example.com>), nous avons aussi supprimé les noms d'utilisateurs Twitter (e.g. @alex, grâce au symbole @ marqueur d'identité utilisateur), les mots spéciaux propres à Twitter (comme « RT »⁵), et les émoticônes.

2) Segmentation : elle a été effectuée selon les espaces et les signes de ponctuation. Pour l'anglais, nous avons considéré les formes contractées suivantes comme formant un seul mot : « *don't, I'll, she'd* ».

3) Suppression des formes fréquentes : pour l'anglais, nous avons supprimé les articles : « *a, an, the* ».

Un traitement particulier a été effectué pour les particules de négations puisque celles-ci jouent un rôle spécial dans l'expression des sentiments et des opinions (Wilson *et al.*, 2005). Nous regroupons la particule de négation avec le mot qui la précède et avec le mot qui la suit, en la dupliquant. Par exemple, une phrase comme « *I do not like fish* » engendrera les trois bigrammes suivants : « *I do+not, do+not like, not+like fish* ». Pour l'anglais nous avons considéré les particules « no » et « not » ; pour le français les particules « ne » et « pas » et pour le chinois, les particules suivantes :

4. <http://dev.twitter.com/doc/get/search>

5. Une abréviation pour « *retweet* », qui signifie la citation ou la republication d'un message.

- 1) 不 – n'est pas + nom
- 2) 未 – auxiliaire présent/futur ne + verbe
- 3) 莫 (別) – ne fais pas (impératif)
- 4) 無 (沒有) – n'a pas

3.2. Analyse de corpus

Le tableau 3 présente les caractéristiques du jeu de données Twitter.

Caractères par message	70,14
Caractères par mot	6,84
Mots par message	10,26

Tableau 3. *Caractéristiques du jeu de données Twitter*

Pour l'étiquetage morphosyntaxique, nous avons utilisé le TreeTagger (Schmid, 1994). Nous nous sommes intéressés aux différences de répartition des étiquettes syntaxiques entre les différents sous-ensembles de textes (positif, négatif, neutre). Afin d'effectuer une comparaison deux à deux des distributions d'étiquettes, nous avons calculé les quantités suivantes pour chaque étiquette et pour les deux sous-ensembles de messages à connotation positive et négative :

$$P_{1,2}^T = \frac{N_1^T - N_2^T}{N_1^T + N_2^T} \quad [1]$$

où N_1^T et N_2^T sont les nombres d'occurrences d'étiquettes T dans le premier et le second ensemble respectivement. Voici la correspondance entre les différentes valeurs P^T et les configurations d'étiquettes :

- si P^T est positive et proche de 1, alors l'ensemble 1 contient beaucoup plus d'occurrences de l'étiquette considérée que l'ensemble 2 ;
- si P^T est négative et proche de -1 , alors l'ensemble 1 contient bien moins d'occurrences de l'étiquette considérée que l'ensemble 2 ;
- si P^T est proche de 0, alors les deux ensembles contiennent approximativement le même nombre d'occurrences de l'étiquette.

3.2.1. Subjectif par rapport à objectif

La figure 1 montre les valeurs de P^T pour toutes les étiquettes avec l'ensemble 1 qui est l'ensemble des documents subjectifs (union des sous-ensembles positifs et négatifs) et avec l'ensemble 2 qui est l'ensemble contenant les documents qualifiés d'objectifs (documents à polarité neutre). Sur le graphe nous observons que les étiquettes morphosyntaxiques ne sont pas distribuées de manière homogène entre les deux ensembles et donc qu'elles peuvent être utilisées comme indicateurs de polarité.

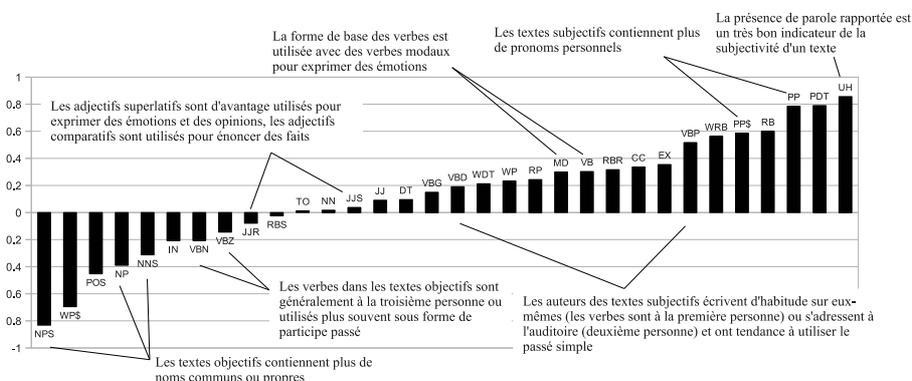


Figure 1. P^T Valeurs indicatives d'objectivité ou de subjectivité

Par exemple, les énoncés UH peuvent être considérés comme des indicateurs forts de subjectivité. Dans ce qui suit, nous détaillons nos observations.

Nous observons que les textes objectifs ont tendance à contenir plus de noms communs et de noms propres (NPS, NP, NNS), tandis que les auteurs de textes subjectifs utilisent plus souvent les pronoms personnels (PP, PPS).

Les auteurs de textes subjectifs parlent souvent d'eux-mêmes (première personne) ou s'adressent au public (deuxième personne) (VBP), tandis que les verbes des textes objectifs sont souvent à la troisième personne (VBZ). En ce qui concerne le temps, les textes subjectifs contiennent plus souvent des verbes au passé simple (VBD) au lieu de participes passés (VBN). En outre, on trouve plus de formes verbales de base (infinitif sans préposition) (VB) dans les textes subjectifs, ce qui peut s'expliquer par l'emploi fréquent dans ces textes de verbes modaux (MD).

Sur le graphe, nous voyons que les adjectifs superlatifs (JJS) sont utilisés plus souvent pour exprimer des émotions et des opinions, tandis que les adjectifs comparatifs (JJR) sont utilisés pour énoncer des faits et pour fournir de l'information. Les adverbes (RB) sont essentiellement utilisés dans les textes subjectifs pour ajouter une coloration émotionnelle aux verbes.

3.2.2. Positif par rapport à négatif

La figure 2 montre les valeurs de P^T pour les ensembles négatifs et positifs.

Comme nous le voyons sur le graphe, un ensemble positif possède un nombre plus important de pronoms WH comme « whose » (WH\$), ce qui est inattendu. Cependant si nous regardons le corpus de plus près, nous nous apercevons que les utilisateurs de Twitter ont tendance à utiliser la forme « whose » comme un raccourci pour « who is ».

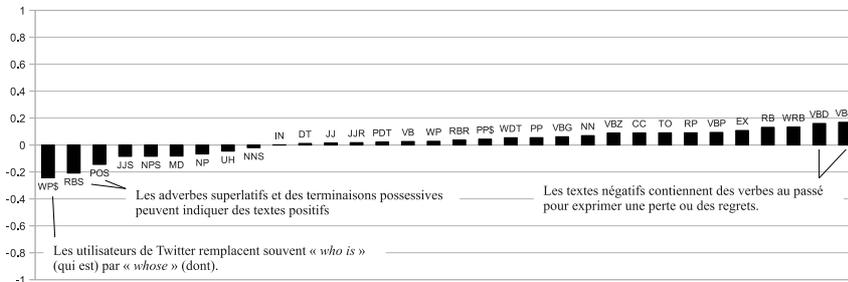


Figure 2. P^T valeurs pour la distinction positif par rapport à négatif

Par exemple :

dinner & jack o'lantern spectacular tonight ! :) *whose* ready for some pumpkins ??

Un autre indicateur de texte positif est la présence d'adverbes superlatifs (RBS), tels que « *most* » et « *best* ». Les textes positifs sont aussi caractérisés par l'emploi de terminaisons possessives (POS).

Au contraire de l'ensemble des textes positifs, l'ensemble des documents négatifs contient plus souvent des verbes aux temps du passé (VBN, VBD), car de nombreux auteurs expriment leurs sentiments négatifs à propos d'une perte ou d'une déception. Voici quelques exemples de verbes fréquemment rencontrés exprimant cette idée : « *missed* », « *bored* », « *gone* », « *lost* », « *stuck* », « *taken* ».

Nous avons comparé les distributions des étiquettes morphosyntaxiques de deux parties d'un même ensemble. La proximité des distributions obtenues nous permet de conclure à la relative homogénéité du corpus.

3.3. Système de classement

Nous avons construit un système de classement de textes en fonction des sentiments exprimés par les adjectifs qu'ils contiennent selon une approche multinomiale bayésienne naïve, en faisant l'hypothèse que la polarité exprimée est homogène dans le message. Ce choix est justifié par la brièveté des messages. Afin de tester cette hypothèse, nous avons classé manuellement des textes courts d'un corpus d'essai selon deux classes, positive et négative, puis nous avons comparé la polarité associée au texte avec celle associée à l'adjectif qu'il contenait. Nous avons observé que dans 94 cas sur 100, la polarité de l'adjectif est la même que celle du texte.

Nous avons expérimenté en utilisant des modèles à base d'unigrammes, de bigrammes et de trigrammes. Pang *et al.* (2002) mentionnent que les modèles à base d'unigrammes obtiennent de meilleures performances lorsqu'ils ont effectué le classement de critiques de films. Mais Dave *et al.* (2003) ont obtenu un résultat opposé en constatant que les bigrammes et les trigrammes donnaient de meilleurs résultats pour la classification de critiques de produits. On pourrait s'attendre à ce que des n-grammes d'ordre supérieur, comme les trigrammes, permettent de mieux capturer les constructions associées aux expressions de sentiments, alors que les modèles à base d'unigrammes devraient permettre d'obtenir une meilleure couverture des données. Au moyen d'un algorithme de vote majoritaire, nous avons combiné trois systèmes de classement construits sur des modèles de n-grammes d'ordres différents (unigrammes, bigrammes et trigrammes) en faisant, dans les calculs, l'hypothèse simplificatrice que les n-grammes étaient indépendants.

4. La microanalyse de sentiments

4.1. Classification en sentiments pour les textes courts

Nous avons entraîné deux systèmes de classement en sentiments sur un corpus de messages Twitter que nous avons ensuite testés sur un corpus annoté manuellement avec trois classes : positive, négative et neutre. Les deux systèmes de classement utilisent des traits différents : présence/absence de n-grammes et distribution des étiquettes morphosyntaxiques. Le système de classement fondé sur les étiquettes morphosyntaxiques estime la probabilité de la présence d'une étiquette dans différents ensembles de textes et utilise cette estimation pour calculer la probabilité *a posteriori*. Bien que les étiquettes morphosyntaxiques soient dépendantes des n-grammes, nous faisons encore une fois ici l'hypothèse d'indépendance entre les n-grammes et les étiquettes morphosyntaxiques pour la simplicité des calculs. Le calcul final de la probabilité d'un sentiment pour un message donné se ramène alors à un simple produit de probabilités conditionnelles.

De manière à augmenter l'exactitude de la classification, nous nous sommes débarrassés des n-grammes ambigus, c'est-à-dire des n-grammes qui ne sont ni clairement indicateurs d'un sentiment particulier, ni marqueurs d'objectivité. Ces n-grammes apparaissent de manière homogène dans tous les jeux de données. À cette fin, nous avons utilisé une mesure de « contraste » qui est calculée pour chaque n-gramme, en notant g les n-grammes, s les étiquettes de sentiments et N le nombre total de sentiments :

$$\text{contraste}(g) = 1 - \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\min(P(g|s_i), P(g|s_j))}{\max(P(g|s_i), P(g|s_j))} \quad [2]$$

Dans notre cas, nous avons trois sentiments : s_{pos} , s_{neg} , s_{neut} . Par exemple, si pour un n-gramme nous avons :

$$P(g|s_{pos}) > P(g|s_{neg}) > P(g|s_{neut})$$

N-grammes	N-grammes
so sad	time of
miss my	place to
so sorry	of people
love your	music at
i'm sorry	states of
sad i	from thursday
i hate	chicago is
lost my	watch out
have great	week with
i miss	takes more
gonna miss	by both
wishing i	your seat
miss him	moment of

Tableau 4. Bigrammes avec des valeurs de contraste élevées (gauche) et basses (droite)

L'équation deviendra :

$$contrast(g) = 1 - \frac{1}{3} \left[\frac{P(g|s_{neg})}{P(g|s_{pos})} + \frac{P(g|s_{neut})}{P(g|s_{pos})} + \frac{P(g|s_{neut})}{P(g|s_{neg})} \right] \quad [3]$$

Cette mesure heuristique prend ses valeurs entre 0 et 1. Une faible valeur est indicatrice d'un contraste faible pour le n-gramme correspondant, impliquant le rejet de ce dernier. De la même manière qu'avec l'entropie, nous pouvons contrôler les performances du système de classement en ajustant le seuil de contraste en deçà duquel un n-gramme est rejeté. Le tableau 4 donne des exemples de n-grammes qui possèdent des valeurs élevées de contraste et d'autres ayant de très faibles valeurs de contraste.

4.2. Désambiguïisation des adjectifs ambigus de sentiments

Nous avons participé à la campagne d'évaluation SemEval 2010 (Wu *et al.*, 2010) (Pak et Paroubek, 2010c) sur la désambiguïisation d'adjectifs de sentiments en chinois. Le corpus de test de SemEval est constitué de textes courts en chinois qui contiennent des adjectifs ambigus de sentiments. Le but de la tâche d'évaluation est donc de préciser pour chacune de ces occurrences d'adjectif quel est le sentiment exprimé en fonction du contexte dans lequel elle s'inscrit. Les adjectifs considérés sont les suivants : 大 grand, 小 petit, 多 nombreux, 少 peu, 高 haut, 低 bas, 厚 épais, 薄 fin, 深 profond, superficiel, 重 lourd, léger, 巨大 énorme, 重大 sérieux. Cette tâche a été effectuée de manière non supervisée, aucune donnée d'apprentissage n'a été fournie mais l'utilisation de ressources externes était encouragée. Le tableau 5 contient un exemple des textes du jeu de données d'essai de la campagne d'évaluation.

Textes	Polarité
<i>Yang Weize said that labor costs is Low in Suzhou, talent quality is very high, with a better business conditions</i> (Yang Weize a dit que le coût de la main d'œuvre est bas dans le Suzhou, la compétence est très élevée, avec une meilleure situation entrepreneuriale)	P
<i>It is reported that the accident occurred, water fog big, visibility is very low. Two ships collided, the tonnage of the smaller ferry sank immediately, while others took advantage of a ferry boat fled the scene in fog</i> (On rapporte que lorsque l'accident a eu lieu, beaucoup d'eau et de brouillard, la visibilité est très basse . Deux bateaux sont entrés en collision, le tonnage du plus petit ferry a coulé immédiatement tandis que d'autres profitèrent d'un ferry pour fuir la scène dans le brouillard)	N
<i>Hong Kong officials said that the Mainland "individual visit" travelers crime is very low Hong Kong SAR Government deficit 81.1 billion</i> (Les officiels de Hong Kong rapportent que l'activité criminelle des visiteurs individuels du continent est très basse Le gouvernement SAR de Hong Kong a un déficit de 81.1 milliard)	P
<i>(International) Iraqi Foreign Minister said that the morale of the Iraqi people is high</i> ((International) Le ministre des affaires étrangères irakien a dit que le moral du peuple irakien est bon)	P
<i>TV quality of series « New Silk Road » is high</i> (La qualité des séries TV « New Silk Road » est élevée .)	P
<i>the click rate of youtube video is high</i> (la fréquence des clicks sur youtube video est élevée)	P
<i>hpv3917tu : why the cpu temperature is high ?</i> (hpv3917tu : pourquoi est-ce que la température du cpu est élevée)	N

Tableau 5. Exemples de textes de notre base de données, traduits et accompagnés d'une mise en gras l'adjectif cible. La polarité de l'adjectif est donnée en deuxième colonne.

C'est donc dans ce cadre d'évaluation que nous avons testé notre approche pour identifier les sentiments exprimés par des messages courts, en faisant l'hypothèse que l'adjectif cible avait la même polarité que le contexte dans lequel il se trouvait. Lors de nos expériences, nous avons fait varier la taille de la fenêtre d'observation centrée sur l'adjectif et nous avons mesuré pour chaque taille de fenêtre la variation induite sur l'exactitude du système de classement. Nous avons testé deux approches différentes pour notre algorithme de classement en sentiments.

1) Nous avons collecté un corpus de messages chinois provenant de Twitter et nous l'avons utilisé pour entraîner un système de classement afin d'annoter les données SemEval.

2) Nous avons utilisé un système de traduction automatique pour traduire les données SemEval du chinois vers l'anglais et nous avons ensuite annoté les textes traduits au moyen d'un système de classement en sentiments entraîné sur un corpus anglais issu de Twitter.

Notre seconde approche été motivée par le fait que la quantité de messages en anglais disponibles à partir de Twitter est beaucoup plus importante que celle des messages en chinois ; nous voulions voir ainsi si la perte de performance résultant de la traduction automatique ne pouvait pas être compensée par le gain produit par l'augmentation de la taille du corpus d'entraînement. Les systèmes de traduction automatique que nous avons utilisés sont Google Translate ⁶ et Yahoo Babelfish ⁷.

C'est avec le système Google Translate que nous avons obtenu les meilleurs résultats, comme nous l'indiquons dans la section 5.2. Des travaux antérieurs (Kettunen, 2009) rapportent également que Google Translate est le système de traduction automatique qui a permis d'obtenir les meilleurs résultats pour la tâche SemEval de desambiguïsation des adjectifs ambigus de sentiments.

4.3. Construction de lexiques affectifs dans différentes langues

Anglais	Français	Score moyen de polarité
happy	heureux	8,21
good	bien	7,47
bored	ennuyé	2,95
angry	en colère	2,85
sad	triste	1,61

Tableau 6. Exemples de mots issus de la base ANEW avec leur traduction en français

Les lexiques annotés sont souvent utilisés pour l'analyse de sentiments ou pour classer des textes en fonction des sentiments qu'ils expriment. Ces lexiques contiennent, pour chaque mot, une estimation des sentiments référencés, exprimée soit par une valeur numérique (par exemple, sur une échelle de valeur entre -1 et $+1$, représentant un continuum allant d'une opinion très négative à une opinion très positive), soit par une étiquette symbolique (par exemple, « peur », « colère », etc.). ANEW est une liste de mots anglais construite par Bradley et Lang (1999). Elle a souvent été utilisée pour des recherches sur la fouille d'opinion et l'analyse de sentiments.

Le lexique ANEW est une liste de 1 034 mots, avec pour chacun la moyenne et l'écart type des scores de « plaisir », « intérêt » et « dominance » relevés par des anno-

6. <http://code.google.com/apis/ajaxlanguage/documentation/>

7. <http://babelfish.yahoo.com/>

tateurs humains utilisant une échelle de valeur à neuf niveaux⁸. Un extrait du lexique ANEW est donné dans le tableau 6. Dans nos travaux (Pak et Paroubek, 2010a), nous nous sommes intéressés uniquement à l'estimation du score moyen de « valence » (c'est-à-dire de « plaisir »), puisque ce dernier peut être utilisé pour la classification en sentiments.

D'après notre hypothèse de travail, un mot doit avoir un score de valence élevé, s'il apparaît fréquemment dans la classe des textes à connotation positive et s'il n'apparaît que très rarement dans la classe des textes à connotation négative. *A contrario*, un mot qui apparaît de manière significative, plus souvent dans l'ensemble des textes à connotation négative, se verra attribuer un score de valence bas.

Pour chaque mot w du lexique ANEW, étant donné le décompte de ses occurrences dans le sous-ensemble de textes positifs $N(w, M_{pos})$ et dans le sous-ensemble de textes négatifs $N(w, M_{neg})$, nous estimons sa polarité de la manière suivante :

$$polarite^*(w) = \frac{9 \cdot N(w, M_{pos}) + N(w, M_{neg})}{N(w, M_{pos}) + N(w, M_{neg})} \quad [4]$$

où 9 est le score maximal de polarité dans ANEW. Ce qui est égal à :

$$polarite^*(w) = 8 \cdot \frac{N(w, M_{pos})}{N(w, M_{pos}) + N(w, M_{neg})} + 1 = 8 \cdot P(M_{pos}|w) + 1 \quad [5]$$

où $P(M_{pos}|w)$ est la probabilité qu'un mot w soit associé à un sentiment positif, 8 est un facteur d'échelle et 1 est une valeur de translation.

Pour montrer que notre méthode est indépendante de la langue, nous avons collecté un corpus en espagnol de messages Twitter et utilisé l'adaptation du lexique ANEW en espagnol (Redondo *et al.*, 2007) comme validation. Pour cette expérience, nous avons rassemblé 250 000 messages pour chaque classe (positive et négative). Nous avons traduit manuellement le lexique ANEW en français (1 034 mots) et utilisé une collection de 250 000 messages Twitter en français pour notre méthode d'estimation de polarité.

5. Résultats expérimentaux

5.1. Validation manuelle

Nous avons testé notre système de classement sur un corpus de messages Twitter annoté manuellement. Le corpus de référence utilisé est le même que celui qui a servi pour les expériences décrites dans (Go *et al.*, 2009); ses caractéristiques sont présentées dans le tableau 7.

8. En fait le lexique ANEW est constitué de trois jeux de données : les scores moyens pour les annotateurs, les scores moyens pour les annotatrices et les scores pour tous les annotateurs quel que soit leur sexe. Dans nos travaux, nous n'avons considéré que le troisième jeu de données.

Sentiment	Nombre d'exemples
Positif	108
Négatif	75
Neutre	33
Total	216

Tableau 7. *Caractéristiques du corpus de référence annoté manuellement*

Nous avons calculé l'exactitude de notre système de classement sur la base de données complètes, c'est-à-dire :

$$exactitude = \frac{N(\text{classifications correctes})}{N(\text{toutes classifications})} \quad [6]$$

Nous utilisons aussi la mesure de « décision » (Adda *et al.*, 1998) qui indique quelle proportion des données a été classée par le système :

$$decision = \frac{N(\text{documents retournés})}{N(\text{tous documents})} \quad [7]$$

Si nous augmentons la valeur du seuil de contraste (éq. 2), le nombre de messages que nous sommes capables de classer (décision) diminue, car certains des messages ne contiennent plus assez de n-grammes ayant une valeur de contraste supérieure au seuil. Cependant, l'exactitude s'améliore.

D'abord nous avons testé l'impact de l'ordre des n-grammes sur les performances du système de classement. Les résultats de cette comparaison sont présentés dans la figure 3. La meilleure performance a été obtenue avec des bigrammes et la plus mauvaise avec des trigrammes. Ceci peut s'expliquer par le fait que les bigrammes sont le meilleur compromis entre la couverture (les unigrammes sont plutôt indiqués, car plus nombreux) et la discrimination (les trigrammes sont mieux adaptés car capables d'identifier des expressions complexes de plusieurs mots).

Ensuite, nous avons examiné l'impact de la prise en compte de la négation (duplication des particules de négation et attachement au mot qui précède et qui suit) lors de la construction des n-grammes. Les résultats présentés dans la figure 4 montrent que notre manière particulière de prendre en compte les particules de négation améliore l'exactitude.

Nous avons aussi regardé quel était l'impact de la taille des données d'apprentissage sur les performances du système. La mesure de performance que nous considérons ici est la F -mesure (Manning et Schütze, 1999) :

$$F = (1 + \beta^2) \frac{\text{précision} \cdot \text{rappel}}{\beta^2 \cdot \text{précision} + \text{rappel}} \quad [8]$$

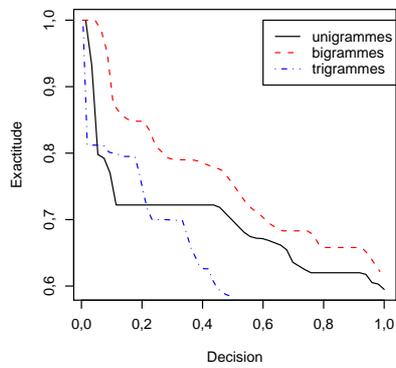


Figure 3. Comparaison des mesures d'exactitude de classification avec les unigrammes, les bigrammes et les trigrammes

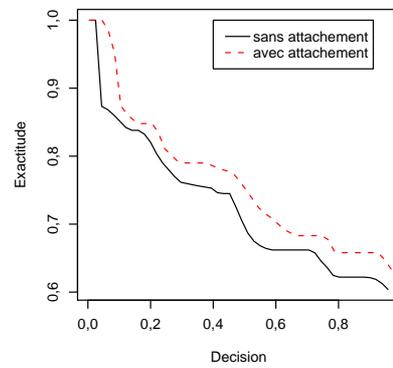


Figure 4. L'impact apporté par la prise en compte des négations

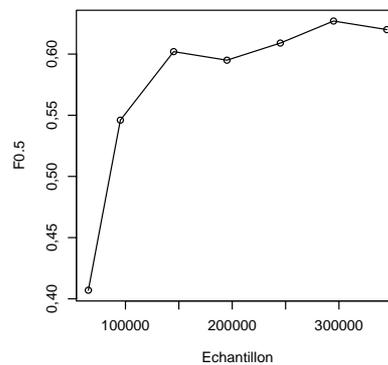


Figure 5. L'impact de l'augmentation de la taille des données d'apprentissage sur la $F_{0,5}$ -mesure

Dans notre évaluation, nous avons remplacé la précision par l'exactitude (6) et le rappel par la décision (7), car nous sommes dans un cas de classement non binaire :

$$F = (1 + \beta^2) \frac{\text{exactitude} \cdot \text{décision}}{\beta^2 \cdot \text{exactitude} + \text{décision}} \quad [9]$$

Les résultats sont présentés dans la figure 5. Comme le graphe le montre, en augmentant la taille des données d'apprentissage, nous améliorons les performances du

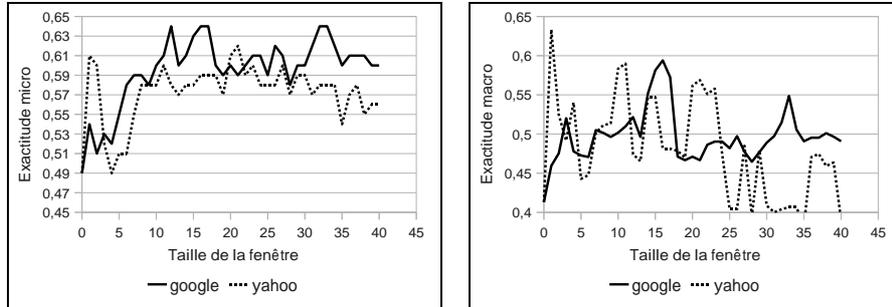


Figure 6. *Micro (gauche) et macro (droite) exactitude avec Google Translate et Yahoo Babelfish*

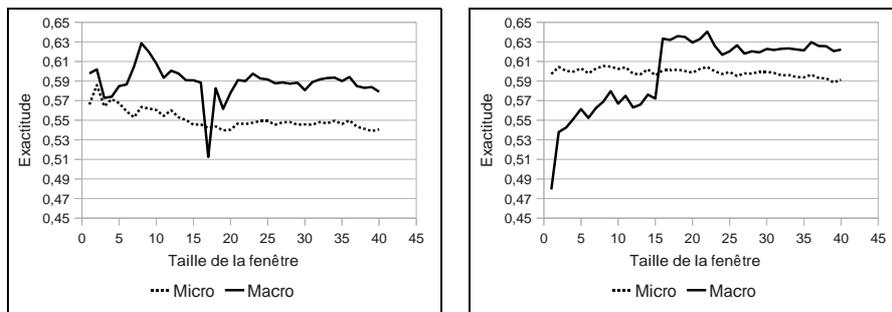


Figure 7. *Micro et macro exactitude pour la première (gauche) et la seconde (droite) approche*

système jusqu'à un certain point, où la taille des données est suffisante et au-delà desquelles il faut recourir à d'autres artifices (par exemple sélection de traits supplémentaires) pour continuer à avoir une amélioration.

5.2. SemEval évaluation

La tâche SemEval fournissait deux jeux de données : un jeu d'essai contenant 100 phrases en chinois et un jeu de test avec 2 917 énoncés. Les annotations pour le jeu d'essai ont été fournies avant l'évaluation et celles du jeu de test, bien évidemment après. Nous avons utilisé le jeu d'essai pour choisir le système de traduction automatique que nous allions utiliser et pour définir la taille de la fenêtre de contexte optimale. Dans ce qui suit, nous indiquons aussi les résultats que nous avons obtenus sur les données d'essai, bien que celles-ci ne constituent pas vraiment des données

d'apprentissage à cause de leur taille réduite. L'évaluation considèrait à la fois la micro et la macro exactitude :

$$Exact_{micro} = \frac{\sum_{i=1}^N m_i}{\sum_{i=1}^N n_i} \quad [10]$$

$$Exact_{macro} = \frac{\sum_{i=1}^N \frac{m_i}{n_i}}{N} \quad [11]$$

où N est le nombre de mots cibles, n_i est le nombre de toutes les occurrences d'un mot dans les données de test et m_i est le nombre de ses instances qui ont été correctement étiquetées.

Nous avons d'abord comparé les performances de notre méthode en utilisant Google Translate et Yahoo Babelfish appliqués sur les données d'essai. Les résultats en micro et macro exactitude sont donnés dans le graphe 6. L'axe des abscisses représente la taille de la fenêtre de contexte autour du mot cible. L'axe des ordonnées donne les valeurs d'exactitude. Sur ce graphe, nous voyons que Google Translate permet d'obtenir de meilleurs résultats, c'est pourquoi nous l'avons choisi pour traiter les données de test.

Ensuite, nous avons étudié l'impact de la taille de la fenêtre de contexte sur la micro et la macro exactitudes, présenté dans la figure 7.

La seconde approche que nous avons essayée donne de meilleurs résultats. Nous avons été capables d'obtenir 64 % de macro et 61 % de micro exactitudes en utilisant cette dernière, alors que nous n'avons obtenu que 63 % de macro et 61 % de micro exactitudes avec la première approche. Le meilleur système (Yang et Liu, 2010) de la campagne SemEval a obtenu 94,20 % et 92,23 %, respectivement en micro et en macro exactitudes. Il a été créé à partir de ressources constituées à la main, comme des lexiques ou des systèmes de classement à base de règles.

Une autre observation qu'il est possible de faire à partir des graphes de performance est que le traitement du chinois peut se contenter d'une taille de fenêtre de contexte plus petite. Avec la première approche, une taille de fenêtre de contexte de 8 mots nous a permis d'obtenir la meilleure macro exactitude. Pour la seconde approche, la meilleure performance en exactitude a été obtenue avec une fenêtre de contexte de 22 mots.

5.3. *Corrélation avec ANEW*

Nous avons estimé les valeurs de polarité pour les mots du lexique ANEW pour l'anglais, l'espagnol et le français. Pour valider nos estimations nous avons calculé l'erreur quadratique moyenne :

$$EQM = \frac{1}{|W|} \sum_{w \in W} (\text{polarite}(w) - \text{polarite}^*(w))^2 \quad [12]$$

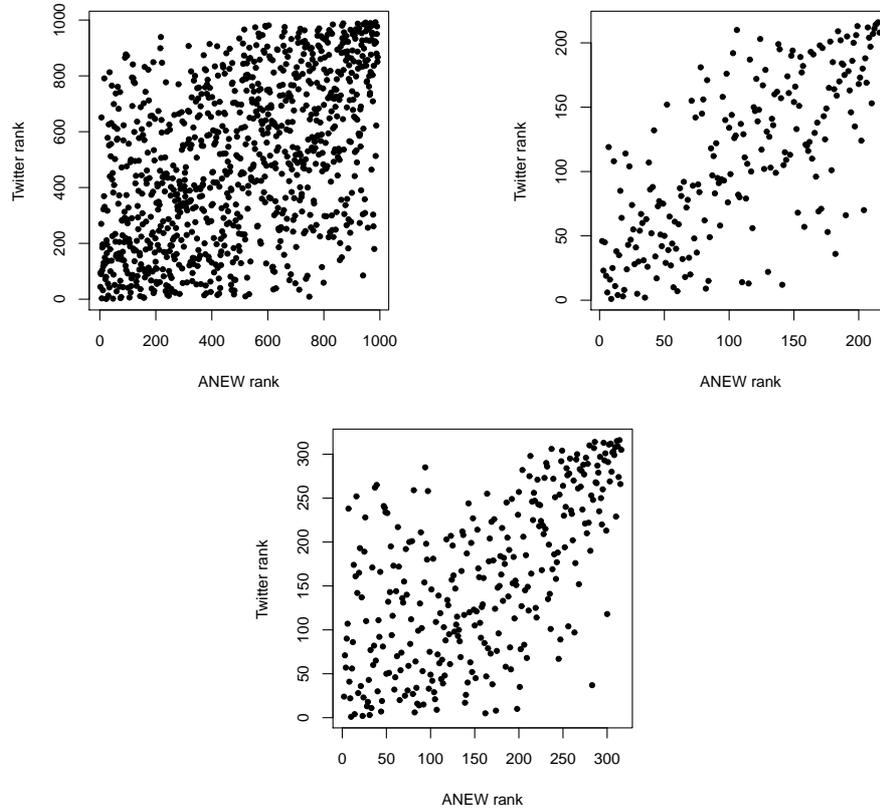


Figure 8. *Corrélation entre les estimations de rang de polarité entre ANEW et Twitter pour tous les mots (gauche), pour les adjectifs (centre) et pour les mots fréquents (droite)*

où W est la liste de tous les mots et $polarite(w)$ est le score de valence du lexique ANEW.

Une autre manière de valider nos estimations de score de valence est de vérifier la corrélation entre les rangs d'un même mot : le rang d'après les valeurs de polarité du lexique ANEW et le rang d'après notre estimation de polarité. Deux scores ont été calculés pour mesurer la corrélation : le coefficient τ de Kendall (Kendall, 1938) et le coefficient de rang de Spearman (Maritz, 1981).

Le coefficient τ de Kendall est défini par :

$$\tau = \frac{n_c - n_d}{\frac{1}{n}(n-1)} \quad [13]$$

où n_c est le nombre de paires de concordances, n_d est le nombre de paires discordantes

et n est le nombre d'éléments du jeu de données. Le coefficient τ de Kendall prend ses valeurs entre -1 et 1 :

- $\tau = -1$ indique un désaccord total entre deux rangs ;
- $\tau = 1$ indique un accord parfait.

Le coefficient de rang de Spearman est défini comme :

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad [14]$$

où $d_i = x_i - y_i$ est la distance entre les deux rangs d'un même élément entre les deux classements considérés et n est le nombre d'éléments. Le coefficient de rang de Spearman prend aussi ses valeurs entre -1 et 1 , avec la même interprétation des valeurs que pour le coefficient τ de Kendall.

Nous avons testé trois listes différentes de mots :

- tous les mots du lexique ANEW (*Tous les mots* dans le tableau 8) ;
- seulement les adjectifs de la liste ANEW (*Adjectifs* dans le tableau 8) ;
- les mots du lexique ANEW qui apparaissent au moins 100 fois dans notre corpus (*Mots fréquents* dans le tableau 8).

Nous avons représenté le classement des mots pour l'anglais dans la figure 8 où chaque point représente un mot, l'abscisse correspondante représente le rang selon le score de valence du lexique ANEW, l'ordonnée correspond à un classement d'après notre estimation du score de valence. Les résultats pour l'anglais, l'espagnol et le français sont présentés dans le tableau 8. Le coefficient τ de Kendall et le coefficient de rang de Spearman montrent une bonne corrélation entre les deux classements. Nous notons que la valeur de l'erreur moyenne quadratique est élevée. Cependant elle peut être facilement abaissée si l'on emploie des techniques d'apprentissage automatique ou une analyse en régression. Nous considérons que la mesure pertinente pour l'évaluation de notre estimation de score de valence est celle des coefficients de corrélation. Le coefficient τ de Kendall et le coefficient de rang de Spearman ont indiqué une bonne corrélation pour les adjectifs et les mots fréquents, avec l'anglais, l'espagnol et le français. Il nous paraît donc justifié de construire des lexiques du genre d'ANEW en sélectionnant les unigrammes les plus fréquents de notre corpus. Nous avons ainsi obtenu une liste de 3 383 mots avec une estimation de leur valeur de polarité. Le tableau 9 donne des exemples des mots de cette liste.

6. Conclusions et perspectives

De nos jours, le microblogage est devenu un des principaux modes de communication sur Internet. Dans des travaux de recherche récents, le microblogage a été identifié à *du bouche-à-oreille en ligne* à propos des marques commerciales. Les quantités importantes de messages disponibles sur les sites de microblogage le rendent très intéressant comme source de données pour l'analyse de sentiments et la fouille d'opinion.

Score	Kendall τ	Spearman	EQM
Tous les mots (anglais)	0,359	0,510	3,055
Adjectifs	0,550	0,736	2,046
Mots fréquents	0,454	0,626	2,861
Tous les mots (espagnol)	0,327	0,468	5,068
Adjectifs	0,434	0,610	4,251
Mots fréquents	0,441	0,626	3,287
Tous les mots (français)	0,287	0,376	6,184
Adjectifs	0,384	0,489	4,836
Mots fréquents	0,448	0,543	3,174

Tableau 8. Coefficients de corrélation et EQM pour l'anglais, l'espagnol et le français

Terme	Polarité estimée	Terme	Polarité estimée
félicitations	8,19	déçue	1,4
bienvenue	8,18	dégouté	1,41
joyeux	8,12	pleut	1,58
héhé	7,74	déprime	1,58
plaisir	7,62	craint	1,62
merci	7,58	malheureusement	1,62
bonsoir	7,56	déteste	1,7
grâce	7,53	pleurer	1,73
bonjour	7,35	m'ennuie	1,76
bonne	7,25	perd	1,83
jolie	7,19	manques	1,89
rêves	7,19	dommage	1,88
rire	7,18	manquer	1,87
parfait	7,12	pauvre	1,85
anniversaire	7,1	désolée	2,02
hâte	7,08	marre	2,05
nouvel	7,05	devoir	2,09
voulez	6,92	impossible	2,1
bonheur	6,84	rater	2,16
génial	6,82	malade	2,18

Tableau 9. Exemple des mots ayant une valeur de polarité estimée élevée (gauche) et faible (droite)

Dans nos travaux, nous avons présenté une méthode pour la collecte automatique de corpus utilisables comme données d'entraînement pour des systèmes de classe-

ment en sentiments ou en opinions. Nous avons présenté plusieurs expériences de mise en application de notre approche. Nous considérons, que ce que nous faisons, est de la microanalyse de sentiments, puisque nous travaillons sur des textes ou des extraits de textes très courts. Notre intention n'était pas de tester des méthodes d'analyse linguistique sophistiquées, mais nous avons voulu conserver une indépendance par rapport à la langue traitée, permettant ainsi d'adapter facilement notre méthode à d'autres langues que celles présentées dans cet article, pourvu que des corpus existent. L'approche que nous avons développée repose essentiellement sur des systèmes de classement bayésiens naïfs multinomiaux, des modèles n-grammes et un étiquetage morphosyntaxique de base. Notre méthode est aussi entièrement automatique et ne nécessite aucune intervention humaine dans la préparation ou le traitement des données.

Nous avons évalué la performance de notre approche en comparant ses résultats avec un corpus de référence annoté manuellement et un lexique de référence annoté manuellement ; nous avons aussi participé à la campagne d'évaluation SemEval 2010 sur l'analyse de sentiments. Dans cette campagne, notre système n'a pas obtenu d'aussi bonnes performances que les systèmes à base de règles construites manuellement et utilisant des lexiques, mais compte tenu de nos hypothèses de travail, notre système a obtenu de bonnes performances sur la tâche considérée. En outre, il est toujours possible d'améliorer la performance de notre approche en utilisant des ressources complémentaires manuellement construites et spécifiques à la langue cible.

Pour l'avenir, nous avons l'intention de passer de la microanalyse de sentiments à la macroanalyse de sentiments, portant sur des paragraphes étendus ou des documents de taille conséquente, tout en incorporant d'autres langues. Il serait intéressant également d'étudier la façon 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.

Remerciements

Ces travaux ont reçu le soutien financier du projet DOXA du pôle de compétitivité CAP-DIGITAL.

7. Bibliographie

- Adda G., Mariani J., Lecomte J., Paroubek P., Rajman M., « The GRACE French Part-of-Speech Tagging Evaluation Task », vol. I, Granada, p. 433-441, May, 1998.
- Baccianella S., Esuli A., Sebastiani F., « SentiWordNet 3.0 : An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining », *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may, 2010.
- Balahur A., Steinberger R., Kabadjov M., Zavarella V., van der Goot E., Halkia M., Pouliquen B., Belyaeva J., « Sentiment Analysis in the News », *Proceedings of the Seventh conference*

- on *International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, May, 2010.
- Bradley M. M., Lang P. J., « Affective norms for English words (ANEW). Gainesville, FL. The NIMH Center for the Study of Emotion and Attention », *University of Florida*, 1999.
- Dave K., Lawrence S., Pennock D. M., « Mining the peanut gallery : opinion extraction and semantic classification of product reviews », *WWW '03 : Proceedings of the 12th international conference on World Wide Web*, ACM, New York, NY, USA, p. 519-528, 2003.
- Go A., Huang L., Bhayani R., « Twitter Sentiment Analysis », 2009, Final Projects from CS224N for Spring 2008/2009 at The Stanford Natural Language Processing Group.
- Harb A., Dray G., Plantié M., Poncelet P., Roche M., Troussel F., « Détection d'Opinion : Apprenons les bons Adjectifs ! », *INFORSID'08 : INformatique des Organisations et Systèmes d'Information et de Décision - Atelier FODOP'08*, p. 59-66, 2008.
- Jansen B. J., Zhang M., Sobel K., Chowdury A., « Micro-blogging as online word of mouth branding », *CHI EA '09 : Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, ACM, New York, NY, USA, p. 3859-3864, 2009.
- Kendall M. G., « A New Measure of Rank Correlation », *Biometrika*, vol. 30, no. 1/2, p. 81-93, 1938.
- Kettunen K., « Packing it all up in search for a language independent MT quality measure tool », *Proceedings of the 4th Language & Technology Conference (LTC'09)*, Poznań, Poland, November, 2009.
- Kim S.-M., Hovy E., « Determining the sentiment of opinions », *COLING '04 : Proceedings of the 20th international conference on Computational Linguistics*, Association for Computational Linguistics, Morristown, NJ, USA, p. 1367, 2004.
- Manning C. D., Schütze H., *Foundations of statistical natural language processing*, MIT Press, Cambridge, MA, USA, 1999.
- Maritz J., *Distribution-Free Statistical Methods*, Chapman & Hall, 1981.
- Mathieu Y., « A Computational Semantic Lexicon of French Verbs of Emotion », *Computing Attitude and Affect in Text : Theory and Applications*, vol. 20, p. 109-124, 2006.
- Pak A., Paroubek P., « Constructing an Affective Lexicon of French using Twitter », *TALN 2010, ATALA*, Montreal, Canada, july, 2010a.
- Pak A., Paroubek P., « Twitter as a Corpus for Sentiment Analysis and Opinion Mining », *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may, 2010b.
- Pak A., Paroubek P., « Twitter based system : Using Twitter for Disambiguating Sentiment Ambiguous Adjectives », *SemEval 2010 : Proceedings of International Workshop of Semantic Evaluations*, 2010c.
- Pang B., Lee L., « Opinion Mining and Sentiment Analysis », *Found. Trends Inf. Retr.*, vol. 2, p. 1-135, January, 2008.
- Pang B., Lee L., Vaithyanathan S., « Thumbs up ? Sentiment classification using machine learning techniques », *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, p. 79-86, 2002.
- Read J., « Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentiment Classification », *ACL*, The Association for Computer Linguistics, 2005.

- Redondo J., Fraga I., Padrón I., Comesaña M., « The Spanish adaptation of ANEW (affective norms for English words). », *Behavior research methods*, vol. 39, no. 3, p. 600-605, August, 2007.
- Schmid H., « Probabilistic Part-of-Speech Tagging Using Decision Trees », *Proceedings of the International Conference on New Methods in Language Processing*, p. 44-49, 1994.
- Siegle G., « <http://www.sci.sdsu.edu/CAL/wordlist/origwordlist.html> », 1994.
- Strapparava C. V., A., « WordNet-Affect : an affective extension of WordNet », *In Proceedings of the 4th International Conference on Language Resources and Evaluation , LREC*, 2004.
- Võ M. L.-H., Conrad M., Kuchinke L., K. Urton, Hofmann M. J., Jacobs A. M., « The Berlin Affective Word List Reloaded (BAWL-R) », *Behavior research methods*, vol. 41, no. 2, p. 534-538, May, 2009.
- Wilson T., Wiebe J., Hoffmann P., « Recognizing contextual polarity in phrase-level sentiment analysis », *HLT '05 : Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Morristown, NJ, USA, p. 347-354, 2005.
- Wu Y., Jin P., Wen M., Yu S., « SemEval-2010 Task 18 : Disambiguating Sentiment Ambiguous Adjectives », *SemEval 2010 : Proceedings of International Workshop of Semantic Evaluations*, 2010.
- Yang C., Lin K. H.-Y., Chen H.-H., « Emotion Classification Using Web Blog Corpora », *WI '07 : Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, IEEE Computer Society, Washington, DC, USA, p. 275-278, 2007.
- Yang S.-C., Liu M.-J., « YSC-DSAA : An Approach to Disambiguate Sentiment Ambiguous Adjectives Based On SAAOL », *SemEval 2010 : Proceedings of International Workshop of Semantic Evaluations*, 2010.