



HAL
open science

CANÉPHORE : un corpus français pour la fouille d'opinion ciblée

Joseph Lark, Emmanuel Morin, Sebastián Peña Saldarriaga

► **To cite this version:**

Joseph Lark, Emmanuel Morin, Sebastián Peña Saldarriaga. CANÉPHORE : un corpus français pour la fouille d'opinion ciblée. 22e conférence sur le Traitement Automatique des Langues Naturelles (TALN), Jun 2015, Caen, France. hal-01169293

HAL Id: hal-01169293

<https://hal.science/hal-01169293v1>

Submitted on 29 Jun 2015

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.

CANÉPHORE : un corpus français pour la fouille d'opinion ciblée

Joseph Lark^{1,2} Emmanuel Morin¹ Sebastián Peña Saldarriaga²

(1) LINA, UMR CNRS 4261, 2 chemin de la Houssinière, Nantes, France

(2) Dictanova, 2 chemin de la Houssinière, Nantes, France

joseph@dictanova.com, emmanuel.morin@univ-nantes.fr, sebastian@dictanova.com

Résumé. La fouille d'opinion ciblée (*aspect-based sentiment analysis*) fait l'objet ces dernières années d'un intérêt particulier, visible dans les sujets des récentes campagnes d'évaluation comme *SemEval* 2014 et 2015 ou bien DEFT 2015. Cependant les corpus annotés et publiquement disponibles permettant l'évaluation de cette tâche sont rares. Dans ce travail nous présentons en premier lieu un corpus français librement accessible de 10 000 tweets manuellement annotés. Nous accompagnons ce corpus de résultats de référence pour l'extraction de marqueurs d'opinion non supervisée. Nous présentons ensuite une méthode améliorant les résultats de cette extraction, en suivant une approche semi-supervisée.

Abstract.

CANÉPHORE : a French corpus for aspect-based sentiment analysis evaluation

Aspect-based sentiment analysis knows a renewed interest these last years, according to recent opinion mining evaluation series (*SemEval* 2014 and 2015, DEFT 2015). However, publicly available evaluation resources are scarce. This work firstly introduces a publicly available annotated French Twitter corpus for sentiment analysis evaluation on aspect, subject and opinion word levels (10 000 documents). We present baseline results on this corpus for the task of opinion word extraction and then show that these results can be improved with simple semi-supervised methods.

Mots-clés : Fouille d'opinion, web social, corpus annoté, extraction d'information semi-supervisée.

Keywords: Opinion mining, social web, annotated corpus, semi-supervised information extraction.

1 Introduction

Ce travail s'inscrit dans le cadre de la fouille d'opinion sur le web social en français, et plus particulièrement la fouille d'opinion ciblée (*aspect-based sentiment analysis*). Nous entendons par fouille d'opinion ciblée l'analyse des arguments avancés par les internautes au niveau des sujets ou de leurs aspects, par opposition à l'analyse d'un document dans son ensemble (Pang & Lee, 2004; Peña Saldarriaga *et al.*, 2013). Ce domaine de recherche fait dernièrement l'objet d'un fort intérêt d'après les sujets de récents ateliers tels que *SemEval* 2014 (*Semantic Evaluation*) (Pontiki *et al.*, 2014) ainsi que DEFT 2015 (Défi Fouille de Texte). D'un point de vue applicatif, ce type de fouille est une promesse très intéressante.

En dehors des corpus annotés fournis pour les campagnes d'évaluation citées précédemment, peu de ressources sont disponibles pour l'évaluation de ces systèmes, et donc pour leur bon développement. Nous proposons ici un corpus annoté d'interactions sur Twitter, CANÉPHORE¹ (Corpus Annoté pour l'Évaluation de Fouille d'Opinion pour la Recherche), disponible publiquement. L'annotation réalisée manuellement permet d'évaluer un système de fouille d'opinion au niveau des sujets abordés, de leurs aspects (partie ou caractéristique jugée du sujet), ou des marqueurs d'opinion, c'est-à-dire des mots portant une orientation sémantique (ou polarité) et qualifiant une cible. Nous associons à ce corpus les résultats d'une méthode de référence non supervisée pour l'extraction de ces marqueurs. Dans un deuxième temps nous expérimentons l'identification faiblement supervisée des patrons morpho-syntaxiques caractéristiques des marqueurs d'opinion, à partir d'un lexique affectif restreint. Les résultats obtenus confortent cette approche du problème de l'adaptation au domaine.

Après une brève mise en contexte de nos travaux (*cf.* section 2), nous décrivons dans cet article le processus d'annotation ainsi que le contenu du corpus d'évaluation proposé (*cf.* section 3). Enfin nous montrons les résultats de nos expériences en extraction de marqueurs d'opinion (*cf.* section 4), dont nous commentons les points forts et les limites (*cf.* section 5).

1. Le corpus est disponible à <https://github.com/ressources-tal/canephore>. Les tweets doivent être acquis à partir de leur identifiant, qui est associé à un fichier unique au format d'annotation de Brat (<http://brat.nlplab.org/standoff.html>).

2 Travaux connexes

Le corpus que nous proposons permet d'évaluer dans un cadre formel un système de fouille d'opinion ciblée. Après avoir présenté en quoi cela diffère d'un corpus pour la fouille d'opinion au niveau document, nous définissons ce qu'est ce cadre formel, puis nous dressons un bref état de l'art sur les problèmes soulevés dans le domaine de la fouille d'opinion ciblée.

2.1 Corpus pour la fouille d'opinion

En ce qui concerne la classification de documents en polarité positive ou négative, plusieurs méthodes efficaces permettent d'obtenir automatiquement un corpus d'évaluation à partir du web social. Ainsi parmi les travaux français, Harb *et al.* (2008) effectuent des requêtes incluant des marqueurs d'opinion sur le moteur de recherche de Google Blog, et Pak & Paroubek (2010) interrogent l'API de Twitter avec des émoticônes positives ou négatives. Une autre méthode consiste à utiliser les notes des internautes pour déduire la polarité d'une critique, c'est le cas par exemple de Boubel & Bestgen (2011) pour des critiques de films, ou de Vincent & Winterstein (2013) sur des critiques d'hôtels, de films et de romans.

Malheureusement il n'existe à ce jour aucune méthode permettant de construire de la même façon un corpus d'évaluation pour la fouille d'opinion ciblée – une annotation manuelle est donc nécessaire. À notre connaissance les données disponibles en anglais sont celles produites lors de la campagne d'évaluation *SemEval* 2014, proposant des critiques annotées sur des restaurants et ordinateurs portables. En français, nous n'avons connaissance d'aucun corpus libre de ce type.

2.2 Fouille d'opinion ciblée

À notre sens, le travail fondateur de la fouille d'opinion ciblée est celui de Hu & Liu (2004), dans lequel les auteurs recherchent les adjectifs cooccurents des sujets abordés parmi des critiques sur le web social afin d'extraire les opinions portées sur les caractéristiques d'appareils numériques. Le choix des adjectifs s'explique par la corrélation forte entre la subjectivité du discours et la présence d'adjectifs, comme ont pu le montrer Hatzivassiloglou & Wiebe (2000). L'orientation sémantique (ou polarité) des adjectifs retrouvés est soit directement attribuée si l'adjectif est présent dans un lexique affectif, soit inférée en recherchant un synonyme ou un antonyme dans la ressource WordNet. Liu (2010) propose par la suite une définition formelle d'une opinion ciblée, modélisée par un quintuplet (e, a, s, h, t) représentant respectivement la cible de l'opinion (*entity*), l'aspect visé (*aspect*), le marqueur d'opinion (*sentiment*), l'émetteur de l'opinion (*holder*) et le moment d'émission de cette opinion (*time*). Le problème de la fouille d'opinion ciblée est alors décomposé en sous-tâches visant à détecter chaque élément de ce tuple. Nous ne considérons dans cet article qu'un triplet (e, a, s) soit (*cible, aspect, marqueur*). Les sous-tâches dans notre cas sont donc l'identification des sujets, des aspects et des marqueurs d'opinion ainsi que la détection de la polarité de ces marqueurs.

Pour ce qui est de l'identification des sujets et aspects cibles, Brun *et al.* (2014) ont recours à la création d'un arbre syntaxique, indiquant le lien entre un marqueur et une cible. Cependant Kiritchenko *et al.* (2014) montrent que des résultats satisfaisants peuvent être atteints par apprentissage sur une fenêtre de n mots incluant un marqueur d'opinion et sa cible. Enfin, Qiu *et al.* (2009) proposent une méthode dite de double propagation par laquelle les sujets sont détectés à l'aide des marqueurs déjà identifiés et *vice versa*, par le moyen d'un modèle de champ aléatoire conditionnel (CRF).

L'identification des marqueurs d'opinion et l'inférence de leur polarité peuvent être réalisées conjointement, c'est notamment le cas si un lexique affectif est utilisé. Plusieurs travaux visant à construire ce type de ressource ont été entrepris, du fait de la difficulté de retrouver l'orientation sémantique d'un mot sans ce point de repère. Parmi ceux-ci nous pouvons citer le lexique de subjectivité MPQA (Wiebe *et al.*, 2005) ainsi que le lexique SentiWordNet (Baccianella *et al.*, 2010). Plusieurs projets de traduction de SentiWordNet ont vu le jour, cependant aucun équivalent français n'est encore disponible. En revanche, Pak & Paroubek (2010) décrivent la construction d'un tel lexique à partir de Twitter.

Il est important de rappeler que ces lexiques affectifs indépendants du domaine peuvent toutefois induire des ambiguïtés, puisque de nombreux mots présentent une orientation sémantique voire une valence subjective différente selon leur contexte d'apparition. Garcia-Fernandez & Ferret (2012) réalisent une étude sur les différentes stratégies applicables pour la construction de ressources spécifiques au domaine, permettant de réduire ces ambiguïtés. Parmi ces stratégies nous notons l'adaptation d'un lexique affectif générique à un domaine spécifique (Jijkoun *et al.*, 2010) et l'adaptation de ressources pour la fouille d'opinion d'un domaine spécifique vers un autre (Marchand, 2013).

3 Présentation du corpus CANÉPHORE

Notre objectif est d'évaluer progressivement un système de fouille d'opinion ciblée, depuis la détection des sujets abordés jusqu'à la qualification en polarité binaire des marqueurs qualifiant les aspects de ces sujets. Nous expliquons ici en quoi le corpus annoté que nous proposons est adapté à cette évaluation.

3.1 Description du corpus

Le corpus provient d'un ensemble de tweets échangés pendant l'événement "Miss France" en 2012. Les doublons, les retweets, les citations ainsi que les tweets considérés trop courts (moins de 3 mots) ont été retirés afin d'éviter le biais que peuvent apporter les répétitions. Ces suppressions réduisent le corpus à 10 000 tweets, soit environ 127 000 mots.

Marqueurs d'opinion CANÉPHORE est analogue à un corpus de critiques comparatives, tels que ceux proposés pour les campagnes d'évaluation récentes, au sens où les internautes s'expriment sur les différents aspects de quelques entités. D'après notre analyse sur plusieurs corpus de ce type, une conséquence de cette configuration est que les internautes emploient fréquemment les mêmes mots porteurs d'opinion pour qualifier la plupart des sujets abordés. Toutefois ce constat peut être nuancé dans le cas de ce corpus puisque bon nombre de ces mots, que nous appelons marqueurs d'opinion, sont uniques (tableau 2) car provenant d'un vocabulaire argotique, et bien souvent incorrectement orthographiés.

Éléments annotés du corpus	#	Marqueurs d'opinion	#
Tweets contenant une annotation	5372	Positifs	687
Sujets (toutes variantes)	708	Négatifs	1290
Marqueurs	1967	Mots (unigrammes)	1075
Aspects	292	Dont uniques	740
Triplets cible-aspect-marqueur	955	Expressions (n-grammes, n > 1)	892
Couples cible-marqueur	5220	Dont uniques	800

Tableau 1: Informations sur les éléments du corpus annotés

Tableau 2: Informations sur les marqueurs d'opinion

Sujets Les principaux sujets du corpus sont bien évidemment les "Miss", dont les caractéristiques sont jugées par les internautes. Cependant l'identification des cibles précises des opinions émises reste un défi car dans bien des cas plusieurs sujets sont jugés dans un même tweet. De plus, les erreurs orthographiques et l'emploi fréquent d'anaphores nominales (par exemple "Alsace" ou "Miss Rousse" pour "Miss Alsace") complexifient la consolidation des opinions qualifiant chaque sujet. Nous avons ainsi retrouvé 708 variantes ou erreurs orthographiques pour ces sujets (tableau 1), alors qu'il n'est question que d'une centaine de cibles uniques au cours de ces discussions.

3.2 Annotation du corpus

Nous décrivons la méthode utilisée pour annoter les éléments du corpus indiquant une opinion, ainsi que les difficultés que nous avons rencontrées lors de cette étape.

Protocole L'annotation du corpus a été réalisée grâce à l'outil libre Brat², avec lequel il est possible d'étiqueter pour chaque tweet les entités (sujets, aspects ou marqueurs) et les relations entre ces entités. Nous modélisons donc par ces annotations les tuples (*sujet, aspect, marqueur*) définis précédemment. Pour chaque tweet exprimant une opinion directe sur un sujet explicitement mentionné, nous relevons les segments de texte les plus courts permettant à un humain sans connaissance du sujet de disposer d'une information non ambiguë. Si l'opinion est indirecte, ou fait référence à un sujet par une anaphore pronominale (sans mention du sujet dans le tweet), aucune annotation n'est retenue. Les informations ajoutées à chaque tweet sont donc : le sujet jugé, l'aspect du sujet le cas échéant, le marqueur d'opinion, sa polarité et éventuellement la marque de négation inversant cette polarité. À ces informations s'ajoutent les relations possibles entre les entités : un lien "est un aspect de" entre un aspect et un sujet, un lien "est une opinion positive (ou négative) sur" entre un marqueur d'opinion et un sujet ou un aspect et enfin un lien "inverse" entre une marque de négation et un marqueur d'opinion. L'exemple d'annotation en figure 1 présente ces différents éléments dans un cas non ambigu.

2. <http://brat.nlplab.org/>

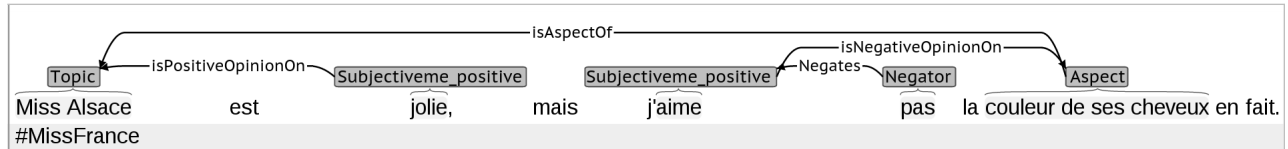


FIGURE 1: Capture d'écran d'annotation à l'aide de Brat, dans un cas non ambigu

Difficultés Les annotations effectuées décrivent des opinions directes, qu'elles soient explicites ou implicites. Notre objectif est que l'information annotée désigne ce qu'un humain lisant le tweet pourrait comprendre sans ambiguïté. Cependant, une opinion peut être exprimée au moyen d'une construction complexe ou de connaissances implicites, ce qui peut impliquer des difficultés d'annotation. La figure 2 répertorie quelques exemples parmi ces difficultés. Le cas de la construction comparative (figure 2.a) est fréquent. Nous avons choisi dans ce cas de considérer le pronom relatif "que" comme l'élément inversant la polarité du marqueur d'opinion, car cela représente un pivot indissociable d'une comparaison entre deux sujets. Un autre annotation pourrait considérer la construction "plus...que" (ou "moins...que") en entier, cependant l'outil que nous avons utilisé ne permet pas d'annoter des empanns de texte non connexes.

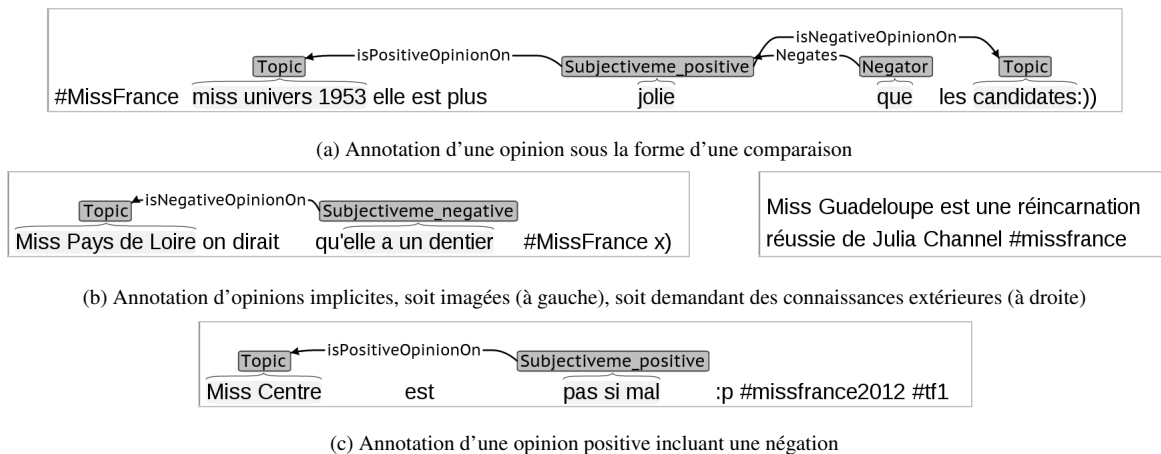


FIGURE 2: Captures d'écran d'annotations, dans le cas d'opinions énoncées de façon ambiguë ou complexe

Une autre forme d'opinion complexe est celle reposant sur des connaissances externes, ce que nous appelons opinion implicite. En réalité, toute forme d'opinion repose sur des connaissances extérieures à son expression, et l'utilisation même de ressources linguistiques pour la fouille d'opinion pose la question de la limite de cet apport. En effet, de la même façon que nous disposons dans un lexique de la polarité "positif" pour le mot "belle", nous pourrions disposer d'une telle polarité *a priori* pour l'expression "réincarnation réussie de Julia Channel" (figure 2.b, à droite). La spécificité des expressions de ce type nous encourage cependant à les considérer comme trop ambiguës – aucune annotation n'est alors effectuée. Ce n'est pas le cas des expressions imagées (figure 2.b, à gauche), dont la polarité peut être inférée dans leur contexte, ou des expressions idiomatiques, dont le sens est connu. Enfin, nous avons fait le choix d'annoter comme un tout certaines expressions composées courantes, comme "pas si mal" (figure 2.c), "pas mal" ou "pas trop mal". Compte tenu de la granularité de l'analyse visée, et du grand nombre de ces cas particuliers, nous avons préféré effectuer plusieurs passes de vérification, en particulier pour valider la cohérence entre les informations, à une annotation multiple. Nous n'avons par conséquent pas d'accord inter-annotateur pour ce corpus.

3.3 Extraction des marqueurs d'opinion du corpus

Le système que nous souhaitons évaluer à l'aide de ce corpus doit en définitive produire un récapitulatif des éléments positifs ou négatifs exprimés par les internautes à propos des sujets discutés. Cette tâche complexe est décomposée en plusieurs sous-tâches, dont la détection des sujets et de leurs aspects, l'extraction des marqueurs d'opinion et l'inférence de la polarité de ces marqueurs. La première des tâches à réaliser pour la désambiguïsation des opinions émises est l'identification des marqueurs d'opinion. La subjectivité et la polarité de ces mots sont bien souvent spécifiques au contexte

d'apparition, c'est pourquoi nous pensons qu'une approche peu supervisée est adaptée pour ce problème. Nous montrons dans ce travail les résultats de l'extraction des marqueurs (mots simples uniquement) sans tenir compte de leur polarité.

Sujet	Précision	Rappel	F1
Miss Alsace	30,75	33,16	31,91
Miss Bretagne	19,07	19,07	19,07
Miss Réunion	40,00	37,36	38,64
Miss Languedoc	56,82	55,56	56,18
Miss Martinique	35,23	35,23	35,23
Miss Guyane	46,60	47,06	46,83
Miss Provence	28,89	36,62	32,30

Tableau 3: Évaluation de la méthode de référence pour l'extraction des marqueurs d'opinion (%)

Sujet	Précision	Rappel	F1
Miss Alsace	29,10	9,87	14,74
Miss Bretagne	55,56	25,77	35,21
Miss Réunion	45,10	12,64	19,74
Miss Languedoc	12,90	2,96	4,82
Miss Martinique	59,00	22,73	32,79
Miss Guyane	25,71	8,82	13,14
Miss Provence	40,00	8,45	13,95

Tableau 4: Évaluation de l'extraction par projection du lexique affectif restreint (%)

La méthode de référence dont nous montrons les performances en tableau 1 est celle proposée par Hu & Liu (2004), qui consiste à extraire les adjectifs présents dans la même phrase qu'un sujet identifié. Nous avons choisi sept sujets parmi les plus discutés dans le corpus afin d'évaluer cette extraction, ce qui correspond à notre cadre d'application. L'étiquetage grammatical a été réalisé en suivant le travail de Dejean *et al.* (2010), car nous utilisons UIMA³ pour cette analyse. Nous observons que cette méthode fournit des résultats très variables, et dans le meilleur des cas assez moyens. Plusieurs facteurs peuvent expliquer cela. Premièrement la précision ne peut être maximale dans la mesure où nous ne réalisons pas de détection du lien entre un marqueur et sa cible. Deuxièmement, le rappel est limité puisque seuls les adjectifs sont extraits, alors que les marqueurs d'opinion peuvent être des verbes ou des substantifs. Enfin des erreurs d'étiquetage grammatical peuvent expliquer ces performances, car nous savons notre prétraitement imparfait sur ce corpus de tweets.

4 Expériences

Au vu des performances moyennes de la méthode de référence sur notre corpus, nous avons souhaité mener quelques expériences sur l'extraction de marqueurs d'opinion peu supervisée, dont nous montrons ici les résultats.

4.1 Méthodes proposées

Les méthodes proposées réalisent l'extraction en deux étapes : nous identifions dans un premier temps des patrons morpho-syntaxiques – c'est-à-dire les séquences d'étiquettes grammaticales ou de lemmes de n mots consécutifs – caractérisant la présence d'un marqueur d'opinion à partir d'un lexique de mots d'opinion restreint (que nous décrivons par la suite). Ces patrons sont ensuite projetés sur le corpus pour découvrir de nouveaux marqueurs.

Patrons fixes Notre première méthode consiste à identifier les patrons syntaxiques fixes les plus fréquents entourant un marqueur connu. Nous retenons les patrons syntaxiques dans une fenêtre centrée sur le marqueur de 7 mots au maximum et dont la fréquence figure parmi le top 3. Cela représente en tout quatre patrons : *nom-verbe-candidat adjectif* (24 occurrences), *pronom-verbe-candidat verbe* (24 occurrences), *verbe-adverbe-candidat adjectif* (22 occurrences) et *déterminant-adverbe-candidat adjectif* (20 occurrences).

Modèle SVM Notre seconde méthode consiste à réaliser un apprentissage par machine à vecteurs de support⁴ (SVM) sur les patrons morpho-syntaxiques, en reprenant une méthode que nous avons expérimentée précédemment (Lark *et al.*, 2014). Les traits de classification pour ces patrons sont les étiquettes grammaticales et les lemmes des mots entourant un mot candidat dans une phrase, en tenant compte de leur position. Les exemples positifs pour ce modèle sont les patrons entourant les marqueurs du lexique. Le modèle est ensuite utilisé pour classifier tous les adjectifs et verbes du corpus, considérés comme exemples négatifs lors de la phase d'apprentissage s'ils sont inconnus du lexique.

4.2 Lexique de marqueurs stables

L'extraction est réalisée de manière semi-supervisée au sens où aucun exemple du corpus n'est fourni, cependant nous l'amorçons à partir d'un lexique affectif stable, et indépendant du domaine. Nous entendons par stable le fait que les mots

3. Unstructured Information Management, <https://uima.apache.org/>

4. Nous utilisons la librairie LIBLINEAR (<http://liblinear.bwaldvogel.de/>)

indexés dans cette ressource ont été choisis pour leur faible variation de sens et de polarité en fonction de la thématique abordée ou de leur contexte d'apparition. Afin d'obtenir un tel lexique nous avons retenu les éléments les plus fréquents dans un large corpus du web social depuis un lexique affectif indépendant du domaine dont nous disposons⁵. Le lexique final, vérifié manuellement, contient 189 marqueurs d'opinion, dont 70 positifs et 119 négatifs.

4.3 Résultats

Nous reprenons le cadre d'expérimentation de la méthode de référence pour évaluer l'extraction des marqueurs par les méthodes proposées, dont les résultats sont indiqués dans le tableau 5. Afin de mettre en perspective ces résultats, nous montrons les performances d'une simple projection de notre lexique restreint sur le corpus, en tableau 4. Enfin nous évaluons l'ensemble des marqueurs d'opinion obtenus par l'une et l'autre des deux méthodes proposées (colonne de droite dans le tableau 5). Les marqueurs extraits dans les deux cas n'étant pas les mêmes, nous montrons que le rappel peut être amélioré en utilisant cette combinaison.

Sujet	Patrons fixes			SVM			Union		
	Précision	Rappel	F1	Précision	Rappel	F1	Précision	Rappel	F1
Miss Alsace	96,47	20,76	34,17	83,08	13,67	23,48	88,89	26,33	40,62
Miss Bretagne	80,95	8,76	15,81	80,36	23,20	36,00	80,56	29,90	43,61
Miss Réunion	98,08	28,02	43,59	96,77	16,48	28,17	96,83	33,52	49,80
Miss Languedoc	94,59	51,85	66,99	85,19	17,04	28,40	90,12	54,07	67,59
Miss Martinique	92,31	13,64	23,76	95,45	23,86	38,18	93,55	32,95	48,74
Miss Guyane	95,00	37,25	53,52	77,27	16,67	27,42	89,80	43,14	58,28
Miss Provence	88,89	11,27	20,00	83,33	7,04	12,99	84,62	15,49	26,19

Tableau 5: Évaluation de l'extraction par occurrence (%)

4.4 Discussion

Les résultats que nous obtenons par ces méthodes simples sont globalement satisfaisants en matière de précision. Cela corrobore notre approche par amorçage dans la mesure où des marqueurs inconnus du lexique sont retrouvés, tandis que les éléments faisant partie de notre ressource mais non pertinents pour ce corpus sont filtrés. Toutefois, c'est sur la notion de couverture que les méthodes testées trouvent leurs limites. D'une part le choix des patrons les plus présents peut éliminer des candidats les mots peu fréquents. D'autre part, d'après notre analyse des patrons retenus, l'agrégation des marqueurs candidats dont le patron morpho-syntaxique est similaire à ceux déjà reconnus ne permet pas d'acquérir les marqueurs dont le patron est très différent, quand bien même leur fréquence serait élevée.

5 Conclusions et perspectives

Cet article présente la diffusion d'un corpus français annoté pour l'évaluation de méthodes de fouille d'opinion ciblée. Une telle ressource n'existait pas à notre connaissance, et nous espérons que d'autres seront créées selon le même type d'annotation que nous décrivons ici. La diffusion de ce corpus offre la possibilité à chacun de l'exploiter, mais également de l'enrichir. Certaines informations ne sont pour le moment pas présentes car elles ne correspondent pas à notre cadre d'analyse : l'opinion générale de chaque tweet n'est pour le moment pas annotée, et les marqueurs d'opinion sont simplement catégorisés selon une polarité binaire et non en classes d'opinion fines. Enfin, pour que ce corpus puisse être exploité par tout un chacun, il serait pertinent d'évaluer son annotation dans un cadre plus large que celui de ce travail.

Nous évaluons sur ce corpus des méthodes faiblement dépendantes du domaine et de la langue. Les résultats indiquent qu'elles permettent d'extraire avec précision les éléments clés qualifiant l'opinion émise sur les sujets abordés. Cette extraction non supervisée montre cependant certaines limites quant à la capacité de détection des marqueurs peu fréquents ou peu similaires aux marqueurs déjà reconnus. Afin de détecter ces éléments nous prévoyons d'étudier l'extraction par double propagation au moyen d'une classification par séquence comme ont pu le tester Qiu *et al.* (2009).

5. Ressource établie à partir du lexique Apopsis, disponible sur demande : <http://taln.lina.univ-nantes.fr/apopsis/>

Références

- BACCIANELLA S., ESULI A. & SEBASTIANI F. (2010). Sentiwordnet 3.0 : An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of LREC 2010*, Valletta, Malta.
- BOUBEL N. & BESTGEN Y. (2011). Une procédure pour identifier les modifieurs de la valence affective d'un mot dans des textes. In *Actes de TALN 2011*, p. 137–142, Montpellier, France.
- BRUN C., POPA D. N. & ROUX C. (2014). XRCE : Hybrid classification for aspect-based sentiment analysis. In *Proceedings of SemEval 2014*, p. 838–842, Dublin, Ireland.
- DEJEAN C., FORTUN M., MASSOT C., POTTIER V., POULARD F. & VERNIER M. (2010). Un étiqueteur de rôles grammaticaux libre pour le français intégré à Apache UIMA. In *Actes de TALN 2010*, Montréal, Canada.
- GARCIA-FERNANDEZ A. & FERRET O. (2012). Etude de différentes stratégies d'adaptation à un nouveau domaine en fouille d'opinion. In *Proceedings of JEP-TALN-RECITAL 2012*, p. 391–398, Grenoble, France.
- HARB A., DRAY G., PLANTIÉ M., PONCELET P., ROCHE M. & TROUSSET F. (2008). Détection d'opinion : Apprenons les bons adjectifs ! In *Actes de INFORSID 2008 - Atelier FODOP*, p. 59–66, Fontainebleau, France.
- HATZIVASSILOGLOU V. & WIEBE J. M. (2000). Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of COLING 2000*, p. 299–305, Saarbrücken, Germany.
- HU M. & LIU B. (2004). Mining and summarizing customer reviews. In *Proceedings of KDD 2004*, p. 168–177, Seattle, WA, USA.
- JIJKOUN V., DE RIJKE M. & WEERKAMP W. (2010). Generating focused topic-specific sentiment lexicons. In *Proceedings of ACL'10*, p. 585–594, Stroudsburg, PA, USA.
- KIRITCHENKO S., ZHU X., CHERRY C. & MOHAMMAD S. (2014). Nrc-canada-2014 : Detecting aspects and sentiment in customer reviews. In *Proceedings of SemEval 2014*, p. 437–442, Dublin, Ireland.
- LARK J., PEÑA SALDARRIAGA S. & MORIN E. (2014). Consumer concern extraction in social web reviews. In *Proceedings of Digital Intelligence 2014*, Nantes, France.
- LIU B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*.
- MARCHAND M. (2013). Fouille d'opinion : ces mots qui changent de polarité selon le domaine. In *Actes de CORIA 2013*, p. 347–352, Neuchâtel, Switzerland.
- PAK A. & PAROUBEK P. (2010). Construction d'un lexique affectif pour le français à partir de twitter. In *Actes de TALN 2010*, Montréal, Canada.
- PANG B. & LEE L. (2004). A sentimental education : Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of ACL'04*, Barcelona, Spain.
- PEÑA SALDARRIAGA S., VINTACHE D. & DAILLE B. (2013). Démonstrateur Apopsis pour l'analyse des tweets. In *Actes de TALN 2013*, p. 807–808, Les Sables d'Olonne, France.
- PONTIKI M., GALANIS D., PAVLOPOULOS J., PAPAGEORGIOU H., ANDROUTSOPOULOS I. & MANANDHAR S. (2014). Semeval-2014 task 4 : Aspect based sentiment analysis. In *Proceedings of SemEval 2014*, Dublin, Ireland.
- QIU G., LIU B., BU J. & CHEN C. (2009). Expanding domain sentiment lexicon through double propagation. In *Proceedings of IJCAI 2009*, p. 1199–1204, Pasadena, CA, USA.
- VINCENT M. & WINTERSTEIN G. (2013). Construction et exploitation d'un corpus français pour l'analyse de sentiment. In *Actes de TALN 2013*, p. 764–771, Les Sables d'Olonne, France.
- WIEBE J., WILSON T. & CARDIE C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation, Volume 39*, p. 165–210.