

Fouille d'opinions

Béatrice Daille

Université de Nantes

béatrice.daille@univ-nantes.fr

3 octobre 2018

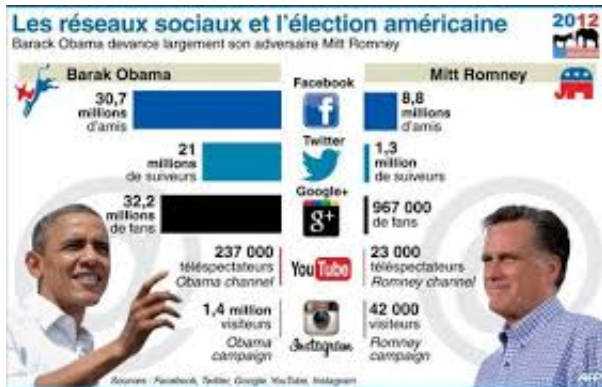
Introduction

Le Web est devenu une source d'information incontournable
diversité des contenus textuels porteurs d'opinions

- blogs
- commentaire
- réseaux sociaux
- tweets

Le développement d'outils pour extraire, synthétiser et comparer les opinions exprimées sur un sujet donné devient crucial.

Introduction



Introduction

- Les premiers travaux de recherche en extraction automatique d'opinion remontent à la fin des années 1990
- Plus de **26 000** publications recensées sur Google Scholar
- Nombreuses campagnes d'évaluation
 - TREC (Text REtrieval Conference) [Ounis et al., 2008]
 - DEFT (Défi Fouille de Textes) pour le français [2005-2017] (2007, 2009, 2015, 2017)
 - SemEval (Semantic Evaluation) [1998-2018] (2007, 2010, 2013, 2014, 2016, 2017)
 - Sentipolc (sentiment polarity classification) [2016]

Introduction

Un domaine interdisciplinaire

- linguistique, analyse du discours, pragmatique
- sociologie
- psychologie
- économie

Introduction

Le développement de systèmes d'analyse d'opinions n'est pas simple et nécessite de se confronter à plusieurs difficultés :

- Comment reconnaître les parties des textes qui renseignent l'utilisateur sur l'opinion qu'il recherche ?
- Comment évaluer la qualité des opinions qui en ressort : sont-elles plutôt positives, plutôt négatives ?
- Comment présenter le résultat de manière pertinente à l'utilisateur ?

Définition

Opinion est un terme générique utilisé pour désigner un ensemble d'expressions subjectives : sentiments, attitudes, points de vues, jugements, désirs, etc.

Une opinion est une expression subjective du langage qui :

- utilise un émetteur (une personne, une institution, etc.)
- pour juger ou évaluer un sujet (un objet, une personne, une action, un événement, etc.)
- en le positionnant sur une échelle polarisée
 - une norme sociale (comme un jugement esthétique)
 - une norme morale (comme la distinction entre le bien et le mal).

Opinion

En analyse d'opinions, il est important de distinguer entre le caractère subjectif ou objectif d'une expression.

Subjectif : opinion

J'ai adoré les plats servis dans ce restaurant.

Objectif : événement factuel

Le premier ministre a inauguré le nouvel hôpital.

Fouille d'opinion

Un quintuplet (Liu 2012) :

$$(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$$

avec :

e : entité

a : aspect de l'entité

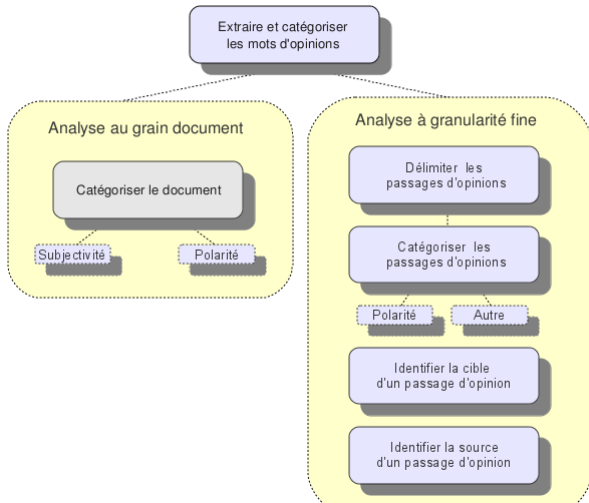
s : opinion exprimé sur l'aspect

h : émetteur de l'opinion

t : la date à laquelle l'opinion a été émise

Fouille d'opinion

Domaine de la fouille d'opinions



Fouille d'opinion au grain document

Soit un document d évaluant une entité, l'opinion générale émise sur l'entité est exprimée par le quintuplet :

$$(_, GENERAL, s, _, _)$$

avec :

e : entité connue

a : aspect général de l'entité

s : opinion

h : émetteur connu ou non pris en compte

t : date connue ou non prise en compte

Fouille d'opinion au grain document

Dans la plupart des systèmes, les méthodes et techniques employées dans chacune de ces tâches reposent sur quatre hypothèses majeures :

- Le document n'exprime qu'une opinion
- L'opinion ne porte que sur une seule entité.
- L'émetteur h est unique.
- Les phrases ou documents analysés sont indépendants les uns des autres.

Approches pour la détection d'opinion au niveau document

Un document sera classé comme émettant une opinion positive (resp. négative) s'il contient plusieurs unités lexicales d'opinion positive (resp. négative).

- **Lexique**
- **Apprentissage supervisé**

Lexiques

Dictionnaires

WordNet (wordnet anglais interfacé avec NLTK)

- ressource lexicale
- S'appuie sur des hypothèses psycholinguistiques selon lesquelles le lexique est organisé en réseau (Miller et al. 1999)
- Principes structurants (ex. green)
 - Unités lexicales organisées en synsets
 - Noms : hyperonymie (taxinomie), méronymie
 - Verbes : hyperonymie
 - Adjectifs et adverbes : antonymie
 - Tous : liens entre formes
- Disponibles dans plusieurs langues
 - Version anglaise :
<http://wordnetweb.princeton.edu/perl/webwn>
 - Version française : Wordnet Libre du Français (WOLF) :
<http://alpage.inria.fr/~sagot/>

Lexiques

Dictionnaires

SentiWordNet anglais (interfacé avec NLTK)

Baccianella et al. LREC 2010

- Repose sur Wordnet
 - Principe : ajouter à chaque synset un score positif, un score négatif ET un score d'objectivité compris entre 0 et 1
- `estimable(J,3)` “may be computed or estimated” Pos 0 Neg 0
Obj 1
- `estimable(J,1)` “deserving of respect or high regard” Pos .75
Neg 0 Obj .25

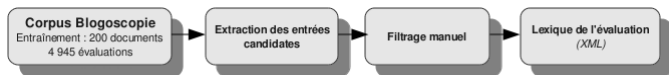
Lexiques

Dictionnaires

- autres lexiques anglais
 - MPQA Subjectivity Lexicon
 - Bing Liu Lexicon
- lexiques français
 - EMOTAIX - Lexique des émotions (4.600 entrées lexicales) - (Piolat et Bannour 2009)
 - CASOAR - 2 830 entrées lexicales dont 297 expressions composées de plusieurs mots
 - Blogoscopie

Lexique de l'évaluation

Lexique Blogoscopie pour le français



Quelques problèmes

- interprétation contextuelle
 - *petites envies*
 - *petite brune*
 - *petite retraite*
- terme neutre : *fruité*
- collocation : *vibrant* – contexte : *hommage vibrant*
- ambiguïté : *fou* – positif ou négatif
- négation : syntaxique *pas*, lexicale *enlever tout son charme*, comparatif *moins de plaisir*

Lexiques

Dictionnaires

Lexique Blogoscopie pour le français

lexique grammaire

- entrée lexicale ou forme
- catégorie pleine : adjectif, nom, verbe, adverbe
- évaluation type, sous-type
- forme : exclamation, configuration, négation
- contextes d'apparition

terrible, adjectif

- appréciation défavorable : *terrible* (solitude // Anorexiques)
- appréciation favorable : *Wah ça doit être terrible !* (Yaourt // recette)

Lexiques

Dictionnaires

Lexique Blogoscopie pour le français les chiffres

Catégorie	Nombre d'entrées	dont ambiguës
adjectif	493	26
nom	166	3
adverbe	60	9
verbe	192	15
syntagme verbal	24	0
phrase	14	0

Lexiques

Corpus

Lexiques construits à partir de corpus d'avis

Hypothèse : un mot ou un phrasème sera d'opinion positive (resp. négative) s'il apparaît dans plusieurs documents classés comme positifs (resp. négatifs).

Algorithme de Turney (2002)

Entrée : documents

Sortie : documents classés (positif vs. négatif)

- 1 Utilisation d'analyse morpho-syntaxique pour identifier les phrases
- 2 Estimation de l'orientation sémantique des groupes nominaux extraits
- 3 Affectation du document dans une classe (positive vs négative)

Lexiques

Corpus

Algorithme de Turney (2002)

- 1 Extraction de groupe nominaux de longueur 2 selon des patrons exprimés en termes de parties du discours

	1er mot	2ème mot	3ème mot (non extrait)
1	JJ	NN ou NNS	n'importe quoi
2	RB, RBR, RBS	JJ	pas NN pas NNS
3	JJ	JJ	pas NN pas NNS
4	NN ou NNS	JJ	pas NN pas NNS
5	RB, RBR, RBS	VB, VBD, VBN, VBG	n'importe quoi

Lexiques

Corpus

Algorithme de Turney (2002)

- calcul du score d'orientation d'opinion SO fondé sur l'information mutuelle gaussienne PMI

$$PM1(UL1, UL2) = \log_2 \frac{P(UL1, UL2)}{P(UL1)P(UL2)}$$

$$SO(GN) = PMI(GN, excellent) - PMI(GN, poor)$$

$$SO(GN) = \log_2 \frac{hits(GN NEAR excellent)hits(poor)}{hits(GN NEAR poor)hits(excellent)}$$

Lexiques

Corpus

Algorithme de Turney (2002)

3 Calcul de la moyenne des *SO* d'un document

- > 0 : positif
- < 0 : négatif

Problèmes :

- Les avis négatifs sont souvent moins fortement exprimés que les positifs
- Les adverbes peuvent inverser la polarité

Lexiques

Corpus

Algorithme de Turney (2002)
 Résultats obtenus sur des avis positifs

GN	Patron	SO
online service	JJ NN	2,8
online experience	JJ NN	2,3
direct deposit	JJ NN	1,3
local branch	JJ NN	0,42
...		
low fees	JJ NNS	0,33
true service	JJ NN	-0,73
other bank	JJ NN	-0,85
inconveniently located	RR VBD	-1,5
Moyenne		0,32

Lexiques

Corpus

Algorithme de Turney (2002)

Résultats obtenus sur des avis négatifs

GN	Patron	SO
direct deposits	JJ NNS	5,8
online web	JJ NN	1,9
very handy	RB JJ	1,4
...		
virtual monopoly	JJ NN	-2
lesser evil	RBR JJ	-2,4
other problems	JJ NNS	-2,8
low funds	JJ NNS	-6,8
Moyenne		-1,2

Lexiques

Corpus

Algorithme de Turney (2002)

Les adverbes peuvent inverser la polarité

- Négation sur unité lexicale négative
 - *The movie is bad* (négatif)
 - *The movie is not bad* (plutôt positif)
- Négation sur unité lexicale positive
 - *The movie is good* (positif)
 - *The movie is not good* (négatif)
 - *The movie is not so good* (négatif nuancé positif)
 - *The movie is not good enough* (négatif accentué)

Lexiques

Corpus

Algorithme de Turney

- 410 avis depuis le site Epinion
 - 170 (41 %) négatif
 - 240 (59 %) positif
- Résultats
 - classe majoritaire (méthode de référence) : 59 %
 - Turney : 74 %
- Remarques
 - uniquement des groupes nominaux de longueur 2
 - apprend du vocabulaire spécialisé à un domaine

Lexiques

Corpus

Améliorer la couverture d'un lexique de l'opinion

Enrichissement automatique par une approche faiblement supervisée

Nécessite

- un dictionnaire où les unités lexicales sont associées à leur polarité
- un ensemble de règles répondant à des hypothèses sémantiques



Sémantique : Classification objectif/subjectif

(Legalois 2005 ; Suhamy 2006)

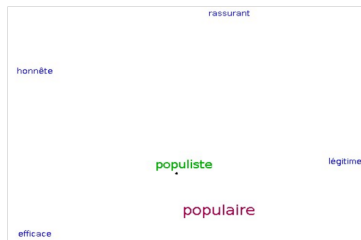
- **Principe A** : on intensifie rarement un adjectif objectif
 - C'est terriblement *législatif* ✘
 - Il est particulièrement *néo-zélandais* ✘
 - C'est terriblement *flatteur* ✔
 - Il est particulièrement *dynamique* ✔

Sémantique : Classification positif/négatif

(Hatzivassiloglou et McKeown 1997)

- **Principe b** : on oppose pas deux adjectifs de même polarité
 - Elle est *jolie* mais *belle* ✗
 - C'est *atroce* mais *douloureux* ✗
 - C'est *jolie* mais *inutile* ✓
 - C'est *atroce* mais *efficace* ✓

populiste



Morphologique : Classification positif/négatif

(Pupier 1998)

Hypothèses

- les mots ayant un suffixe en
 - *ard* sont probablement négatifs
 - *asse* sont probablement négatifs
 - *âtre* sont probablement négatifs
 - *eux* sont probablement négatifs
- les mots ayant un préfixe en
 - *mal* sont probablement négatifs
 - *im* sont probablement négatifs

Les résultats

Avant

- Lexique blogoscopie : 982 entrées
- précision 88,4
- rappel 50,1

Après

- Lexique : 3 974 entrées
- sémantique : 2 474 – 25,7 % évaluations en plus
- morphologique : 624 – 3 % évaluations en plus
- précision 79,5
- rappel 69,8

Apprentissage supervisé

Schéma de référence

2 phases : apprentissage et prédiction

- 1 Segmentation en mots
- 2 Calcul de traits
- 3 Application d'un classifieur
 - Naïve Bayes
 - Maximum Entropy (MaxEnt)
 - Support vector machines (SVM)
 - Logistic regression (LR)
 - ...

Apprentissage supervisé

Segmentation en mots

- annotation HTML ou XML
- annotation Twitter
- majuscules (conserver la casse)
- dates, numéros de téléphone
- émoticons

Apprentissage supervisé

Segmentation en mots

Des codes de nettoyage à base de regexp :

En : Happytokenizer

<http://sentiment.christopherpotts.net/tokenizing.html>

"RT @ #happyfuncoding : this is a typical Twitter tweet :-)"

rt, @, #happyfuncoding, :, this, is, a, typical, twitter, tweet, :-)

En : <https://github.com/brendano/tweetmotif>

Fr : <https://github.com/boudinfl/kea>

Apprentissage supervisé

Les traits

n-grammes de mots : unigrammes, bi-grammes, 3-grammes

"i'am going to chapel hill on sat. :)"

"i'm", 'going', 'to', 'chapel', 'hill', 'on', 'sat', '.', ':)', "i'm going",
'going to', 'to chapel',
'chapel hill', 'hill on', 'on sat', 'sat .', '. :)', "i'm going to", 'going
to chapel', 'to chapel hill',
'chapel hill on', 'hill on sat', 'on sat .', 'sat . :)'.

Apprentissage supervisé

Les traits

négation ajout du suffixe `_NOT` entre la négation et la première
marque de ponctuation

ponctuations : [", " ;", ". ", " !", " ?", "but", "|", "so"]

"I'm not happy"

"i'am", 'not', 'happy_`_NOT`', 'happy', "i'am not", 'not happy',

"i'am not happy"

Apprentissage supervisé

Les traits

dictionnaire twitter construction d'un dictionnaire pour les abréviations et mots d'argot.

Expression Twitter	Signification
:)	<i>veryhappy</i>
:)	<i>veryhappy</i>
b/c	<i>because</i>
Gr8	<i>Great</i>
OMG	<i>Oh my god</i>

"i'am going to chapel hill on sat. :)"

"i'am going to chapel hill on sat. *veryhappy*"

Apprentissage supervisé

Les traits

Zscore

$$Zscore(t_i, C_j) = \frac{tfr_{ij} - moyenne_i}{sd_i}$$

avec :

tfr : fréquence relative d'un terme t_i dans une classe C_j

sd_i : écart-type des mots dans la classe C_j

$moyenne_i$: moyenne des mots dans la classe C_j

Apprentissage supervisé

Les traits

	Positif	Zscore	Négatif	Zscore
Zscore	love	14,31	not	13,99
	good	14,01	fuck	12,97
	happy	12,3	don't	10,97
	great	11,1	shit	9,99
	excite	10,5	bad	8,4
	best	9,24	hate	8,29

Apprentissage supervisé

Les traits

Lexique d'opinions

- MPQA Subjectivity Lexicon
- Bing Liu Lexicon

Apprentissage supervisé

Naïve Bayes

Appliqué aux documents à affecter à une classe
Pour un document d et une classe c :

$$P(c | d) = \frac{P(d | c)P(c)}{P(d)}$$

Apprentissage supervisé

Naïve Bayes

Classifieur

$$\begin{aligned}c_{MAP} &= \operatorname{argmax}_{c \in C} P(c | d) \\ &= \operatorname{argmax}_{c \in C} \frac{P(d | c)P(c)}{P(d)} \\ &= \operatorname{argmax}_{c \in C} P(d | c)P(c)\end{aligned}$$

avec :

c_{MAP} : la classe la plus probable

Apprentissage supervisé

Naïve Bayes

Classifieur

$$\begin{aligned}c_{MAP} &= \operatorname{argmax}_{c \in C} P(d | c)P(c) \\ &= \operatorname{argmax}_{c \in C} P(x_1, x_2, \dots, x_n | c)P(c)\end{aligned}$$

avec :

x_1, x_2, \dots, x_n : document d est représenté par ses traits

Apprentissage supervisé

Naïve Bayes

Hypothèse d'indépendance des traits étant donné une classe

$$P(x_1, x_2, \dots, x_n | c) = P(x_1 | c)P(x_2 | c)P(x_3 | c) \dots P(x_n | c)$$

avec :

x_1, x_2, \dots, x_n : document d est représenté par ses traits

Apprentissage supervisé

Naïve Bayes

Classifieur Naïve Bayes multinomial

$$c_{MAP} = \operatorname{argmax}_{c \in C} P(x_1, x_2, \dots, x_n | c)$$
$$c_{NB} = \operatorname{argmax}_{c \in C} P(c_j) \prod_{x \in X} P(x | c)$$

Apprentissage supervisé

Naïve Bayes

Apprentissage du classifieur

- A partir d'un corpus d'apprentissage, extraire le vocabulaire

- Calculer $P(c_j)$

Pour chaque c_j dans C faire

$docs_j \leftarrow$ tous les documents de la classe c_j

$$P(c_j) \leftarrow \frac{|docs_j|}{|total\#documents|}$$

- Calculer $P(w_k | c_j)$

- $Text_j \leftarrow$ concaténation dans un document de $docs_j$

- Pour chaque mot w_k de vocabulaire

$n_k \leftarrow$ # d'occurrences de w_k dans $Text_j$

$$P(w_k | c_j) \leftarrow \frac{n_k + \alpha}{n + \alpha |vocabulary|}$$

Apprentissage supervisé

Naïve Bayes

Classifieur Naïve Bayes multinomial binarisé (traits à valeur booléenne)

Intuition

- pour la classification en opinion (et éventuellement d'autres domaines de classification)
- l'occurrence d'un mot est plus importante que sa fréquence
 - l'occurrence de *génial* est très parlante
 - le fait qu'il apparaisse 3 ou 4 fois n'apporte pas plus d'information
- Classifieur Naïve Bayes multinomial binarisé
- toutes les fréquences de mots dans un document sont fixées à 1

Apprentissage supervisé

Naïve Bayes

Classifieur Naïve Bayes multinomial binarisé
calculé sur un document d de l'ensemble de test

- on supprime tous les duplicats du document d
-

$$c_{NB} = \operatorname{argmax}_{c_j \in C} P(c_j) \prod_{i \in \text{positions}} P(w_i | c_j)$$

Binaire fonctionne mieux que les fréquences de mots pour l'opinion

Ironie et sarcasme

Définition

- Contrairement au langage littéral, le **langage figuratif** déforme le sens propre pour lui donner un sens appelé figuratif ou imagerie.
- L'ironie est une forme de langage figuratif défini comme une incongruité entre le sens littéral d'un énoncé et sa signification.
- L'ironie se chevauche avec une variété d'autres dispositifs figuratifs (satire, parodie et sarcasme).
- Le sarcasme est une ironie mordante. Le locuteur choisit de blesser et il le fait en présence de celui qu'il vise.
Rightarrow Le sarcasme souligne donc l'agressivité.

Ironie et sarcasme

Définition

- L'ironie est une forme de langage figuratif défini comme une incongruité entre le sens littéral d'un énoncé et sa signification.
- L'ironie se chevauche avec une variété d'autres dispositifs figuratifs (satire, parodie et sarcasme).
- Le sarcasme est une ironie mordante. Le locuteur choisit de blesser et il le fait en présence de celui qu'il vise.
Rightarrow Le sarcasme souligne donc l'agressivité.
- Difficile de distinguer ironie et sarcasme
 - **au niveau linguistique** : frontière floue entre ces deux notions
 - **au niveau informatique** : difficile de différencier ces deux notions dans un texte.

Ironie et sarcasme

Définition

- L'ironie est une forme de langage figuratif défini comme une incongruité entre le sens littéral d'un énoncé et sa signification.
- L'ironie se chevauche avec une variété d'autres dispositifs figuratifs (satire, parodie et sarcasme).
- Le sarcasme est une ironie mordante. Le locuteur choisit de blesser et il le fait en présence de celui qu'il vise.
Rightarrow Le sarcasme souligne donc l'agressivité.
- Difficile de distinguer ironie et sarcasme
 - **au niveau linguistique** : frontière floue entre ces deux notions
 - **au niveau informatif** : difficile de différencier ces deux notions dans un texte.
- ironie terme générique recouvrant ironie et sarcasme.

Langage figuratif

Spécificités

- L'interprétation du langage figuratif nécessite une connaissance du contexte de l'énonciation.
- Le contexte est disponible dans conversation orale ou dans un texte.

Paul : Il pleut aujourd'hui

Marie : Super !!! encore du beau temps

- difficile à identifier dans des petits textes comme les tweets.

De mieux en mieux. On avance #France @LeFigaroEmploi :
Chômage : le tour de passe-passe de Hollande"

Ironie et sarcasme

Détection

Apprentissage supervisé

les données : médias sociaux

Les hashtags pré-existants de l'ironie (#ironie, #sarcasme, etc.)
sont utilisés pour classer les tweets comme ironiques ou non.

Ironie et sarcasme

Détection

Apprentissage supervisé

les traits :

- marques de surface, indices sémantiques : interjections, emoticons, négation

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

Ironie et sarcasme

Détection

Apprentissage supervisé

les traits :

- contexte local : hashtag, indices lexicaux et sémantiques

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

Ironie et sarcasme

Détection

Apprentissage supervisé

les traits :

- contexte externe : connaissances à aller chercher dans des sources fiables

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

Ironie et sarcasme

Détection

L'incongruité dans les tweets ironiques consiste souvent en une contradiction entre au moins deux propositions (ou mots) P1 et P2.

Explicite (oxymore)

Ben non ! [Matraquer et crever des yeux] P1 , [ce n'est pas violent et ça respecte les droits] P2 !!! #ironie

Implicite (fausse affirmation)

La #NSA a mis sur écoutes un pays entier. Pas d'inquiétude pour la #Belgique : [ce n'est pas un pays entier.] P1 #ironie
→ P2 (connaissance externe) : la Belgique est un pays.

Ironie et sarcasme

Détection

Les marqueurs des trois catégories principales en vrac

- ponctuation
- majuscule
- verbes de discours rapporté (annoncer, dire, penser, etc.)
- émoticons
- interjections
- **négations**
- opinions
- pronoms personnels
- noms propres
- connecteurs de discours

Conclusion

Objectif

Analyse de l'opinion au niveau du document

Réalisations

- subjectif
- ironie
- polarité

Fouille d'opinion à granularité fine

J'ai une véritable passion pour J.K Rowling. Non seulement j'aime bien la femme écrivain et son univers, j'ai lu tous les Harry Potter et vu les films (nous sommes des incondtionnels at home) ...

Fouille d'opinion à granularité fine

Segments d'opinion

dimension subjective, notion de polarité, d'intensité, d'engagement du locuteur

J'ai une véritable passion pour J.K Rowling. Non seulement j'aime bien la femme écrivain et son univers, j'ai lu tous les Harry Potter et vu les films (nous sommes des inconditionnels at home)

...

Fouille d'opinion à granularité fine

Cible de l'opinion

J'ai une véritable passion pour J.K Rowling. Non seulement j'aime bien la femme écrivain et son univers, j'ai lu tous les Harry Potter et vu les films (nous sommes des incondtionnels at home) ...

Source de l'opinion

Quelle théorie linguistique ?

Appraisal (Martin and White 2005)

Modèle systémique de l'évaluation élaboré pour l'anglais

- l'attitude
- l'engagement
- la graduation

Modalités énonciatives (Charaudeau 1988, 1992)

Description de l'attitude de l'énonciateur pour le français.
Réduction théorique aux modalités d'expression d'une évaluation

- l'opinion
- l'accord/désaccord
- le jugement
- l'appréciation

Théorie de l'évaluation (Charaudeau 1988, 1992)

une typologie détaillée intuitive

Plus de 20 sous-catégories d'évaluations

- Opinion
 - Conviction – *je suis persuadé*
 - Supposition certitude forte – *je me doute*
 - Supposition certitude moyenne – *je crois*
 - Supposition certitude faible – *je doute*
 - Supposition pressentiment – *je sens*
- Appréciation
 - Explicite favorable – *je suis satisfait*
 - Explicite défavorable – *je suis triste*
 - Explicite exclamative favorable – *Géant !*
 - Explicite exclamative défavorable – *Flûte !*
 - Implicite favorable – *c'est super intéressant*
 - Implicite défavorable – *c'est mauvais*

Identifier la cible d'une évaluation

(Kesler et Nicolov 2009), (Jacob et Gurevych 2010)

Relier un passage évaluatif avec son objet

L'équipe de France *va mal* . Entre *honte* et *déception* .

Plus facile avec des entités nommées

J'adore jouer *à la Wii*

Identifier la cible d'une évaluation

Quelques problèmes

Différentes formes textuelles pour un objet

lors du fiasco des bleus en juin.

Elle a enfin été séduisante

Identifier la cible d'une évaluation

Quelques problèmes

Relation méronymiques entre objets

*Disons-le clairement : la défense française
n'a pas été très bonne , ce soir.*

*Le marquage sur les corners
reste encore approximatif .*

Identifier la cible d'une évaluation

Quelques problèmes

Plusieurs objets candidats autour de l'évaluation

*... pizzas commandées lors de
la dernière finale de coupe du monde avec
l'équipe de France , coursier courageux , pizzas aussi
banales que le match .*

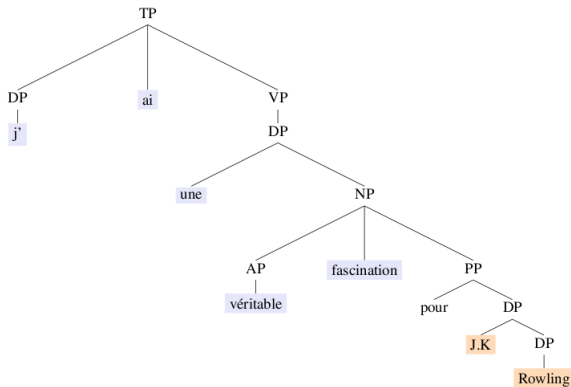
Identifier la cible d'une évaluation

Quelques problèmes

Proximité aléatoire entre l'évaluation et la cible

une réaction d'orgueil mais l'équipe de France devait se rassurer et rassurer son public ... C'était un devoir professionnel. Un ouf de soulagement .

Analyse des dépendances entre une évaluation et sa cible



Chemins des dépendances syntaxiques entre une évaluation et une cible

Chemin de dépendance(s)	Nb	Exemple
↑ AP	121	C'est un beau film
↓ DP	91	Cette question me taraude
↓ PP, ↓ DP	65	J'ai une véritable fascination pour J.K Rowling
↑ TP, ↑ CP	59	Un événement qui contribue à alourdir le débat
↑ VP, ↓ DP	45	Ce jeu est accessible pour les enfants de 12 ans
↑ AdvP, ↓ DP	29	Cette histoire de travailler plus [...] ne tient [...] pas la route

Méthodes intraphrastiques

Approche par apprentissage supervisé **MRankSVM**

Combinaison d'indices lexicaux et grammaticaux (Keesler et Nicolov 2009)

Approche syntaxique **MSyntaxe**

Chemins de dépendances le plus probable

Méthodes inter-phrastiques

Approche par proximité **MProximité**

La distance la plus courte entre le l'évaluation et sa cible en nombre de mots.

Méthode intra et inter-phrastique

Hypothèses

- Si la cible n'est pas trouvée dans la phrase, elle existe probablement dans les phrases voisines
- Plus un objet est la cible d'une évaluation, plus il est saillant, plus il est probable qu'il soit évalué plusieurs fois.

Méthode par mesure de saillance **MSaillance**

- 1 Classification des couples cible-opinion par **MRankSVM**
- 2 Calcul de la saillance

Résultats

Type Score	Intraphrastique Exactitude	Interphrastique Exactitude	Total Exactitude
MSyntaxe	68.8 % (407/592)	non applicable	40.7 % (407/1 000)
MRankSVM	71.5 % (423/592)	non applicable	42.3 % (423/1 000)
MProximite	53.5 % (317/592)	27.2 % (111/408)	42.8 % (428/1 000)
MSaillance	72.8 % (431/592)	60.0 % (245/408)	67.6 % (676/1 000)

Conclusion

Objectif

Analyse à granularité fine de l'opinion

Réalisations

-
-
-
-