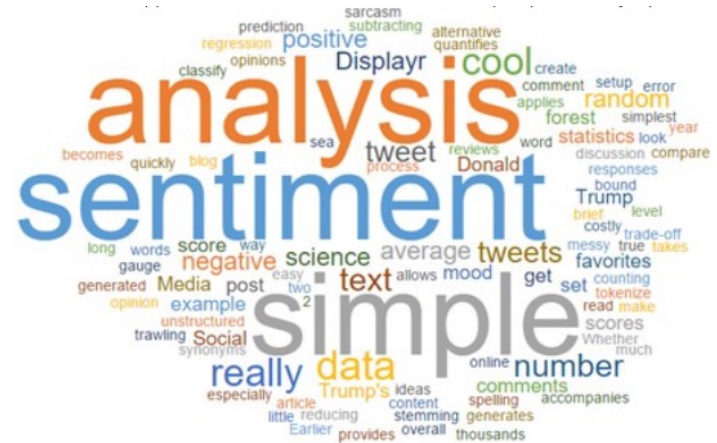


# DÉTECTION D'OPINION



# Historique

3

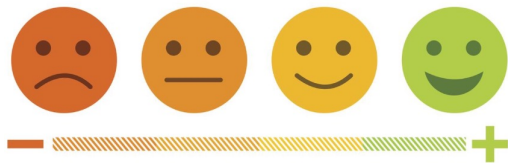
- Domaine multidisciplinaire
  - ▣ Psychologie (Typologie de Scherer 84), Sociologie, Economie
- Linguistique ou NLP :
  - ▣ Premiers travaux publiés dans les années 90
  - ▣ Polarité des adjectifs : *Predicting the Semantic Orientation of Adjectives* [Hatzivassiloglou and McKeown, EACL 1997]
  - ▣ Polarité + ou - de documents
    - Thumbs up? Sentiment Classification Using Machine Learning Techniques [Pang et al., EMNLP 2002]

# Sources d'opinion

4

## □ Le **web** est une mine d'or!

- ▣ Blogs
- ▣ Forum
- ▣ Avis de consommateurs sur site marchand
- ▣ Réseaux sociaux !!!



## □ Et ne pas oublier les entreprises!

# Influence...

5

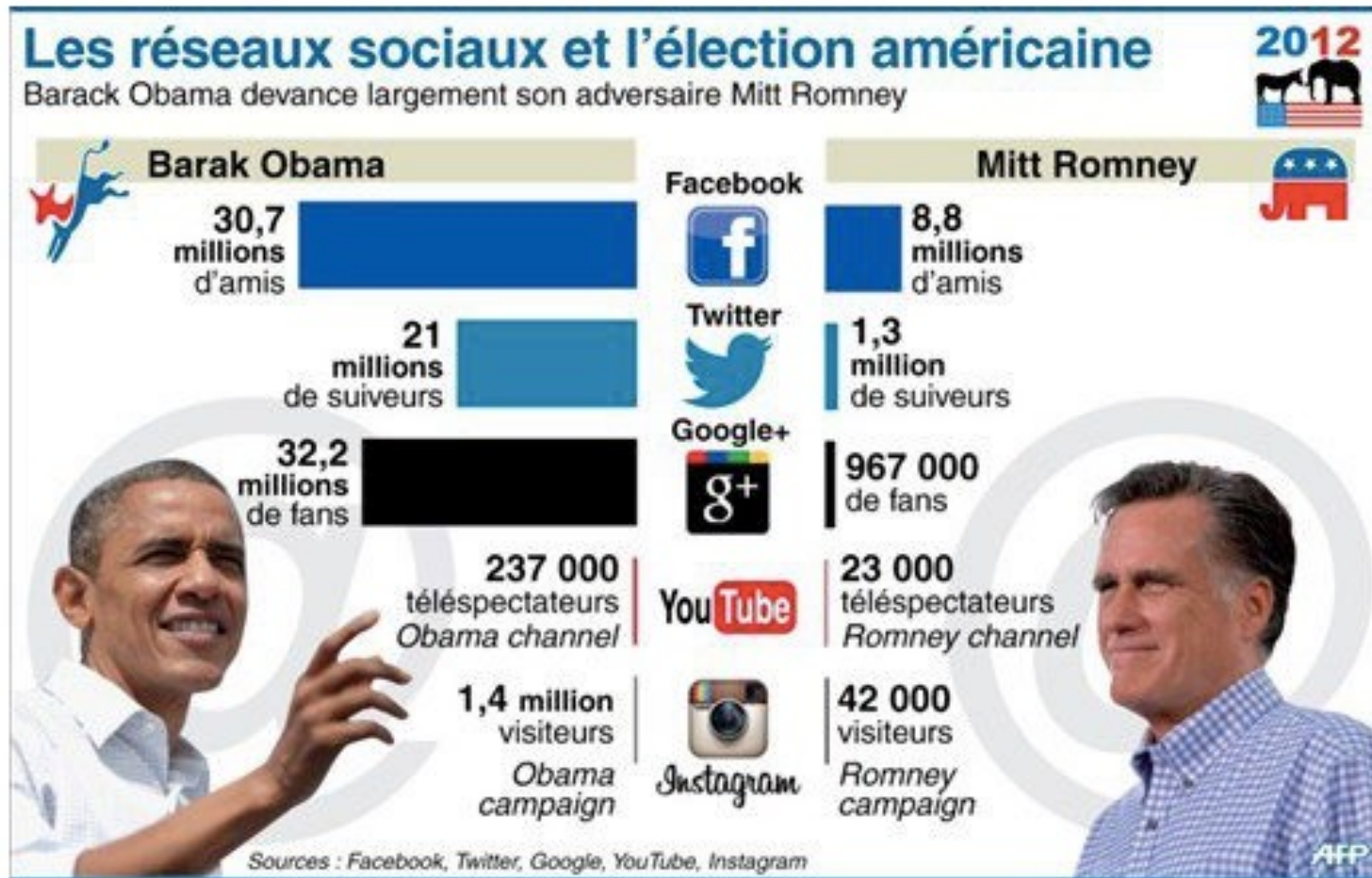
- 81% des internautes ont fait au moins une fois une recherche de produits sur Internet
- 73 à 87% des lecteurs de critiques en ligne (restaurant, hôtel, livres...) disent être significativement influencés dans leur comportement
- Les consommateurs sont prêts à payer plus cher un article qui serait évalué 5 étoiles par rapport à un autre évalué 4 étoiles
- Un tiers des personnes interrogées ont déjà posté un avis sur un produit ou un service

*Pang&Lee, Opinion Mining and Sentiment Analysis, 2008*

*D'après 2 sondages de 2007 et 2008 sur 2000 adultes américains chacun*

# Influence...

6



<https://talleyrandezvous.jimdo.com/lacampagne-presidentielle-aux-etats-unis/>

# Détection d'opinion, pourquoi faire?

7

- Analyser un sentiment vis à vis : d'un film, d'un livre, d'un produit, d'un politicien
  - ▣ Que pensent les clients du dernier produit de ma société?
  - ▣ Comment sont perçus les différents candidats?
  - ▣ Quel est le meilleur resto de Paris?
- Prédiction :
  - ▣ Qui va être élu?
  - ▣ Comment va évoluer la bourse?
  - ▣ Comment améliorer mon offre? Ne pas perdre de client?

# De plus en plus de ressources

8

- Des outils d'analyse disponibles en lignes
- Pour la plupart en langue anglo-saxonne





## Google Product Search



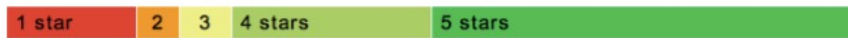
HP Officejet 6500A Plus e-All-in-One Color Ink-jet - Fax / copier / printer / scanner

\$89 online, \$100 nearby ★★★★★ 377 reviews

September 2010 - Printer - HP - Inkjet - Office - Copier - Color - Scanner - Fax - 250 shi

### Reviews

Summary - Based on 377 reviews



What people are saying

|                  |       |  |
|------------------|-------|--|
| ease of use      | ★★★★★ | "This was very easy to setup to four computers." |
| value            | ★★★★★ | "Appre   |
| setup            | ★★★★★ | "Overz   |
| customer service | ★★★★★ | "I DO I  |
| size             | ★★★★★ | "Pretty  |
| mode             | ★★★★★ | "Photc   |
| colors           | ★★★★★ | "Full c  |

## Bing Shopping

### HP Officejet 6500A E710N Multifunction Printer

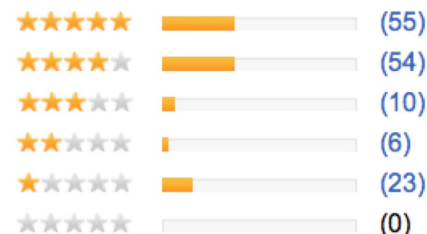
[Product summary](#) [Find best price](#) **Customer reviews** [Specifications](#) [Related items](#)



**\$121.53 - \$242.39** (14 stores)

☐ Compare

Average rating ★★★★★ (144)



Most mentioned

|              |       |      |
|--------------|-------|------|
| Performance  | ★★★★★ | (57) |
| Ease of Use  | ★★★★★ | (43) |
| Print Speed  | ★★★★★ | (39) |
| Connectivity | ★★★★★ | (31) |
| More ▼       |       |      |

Show reviews by source

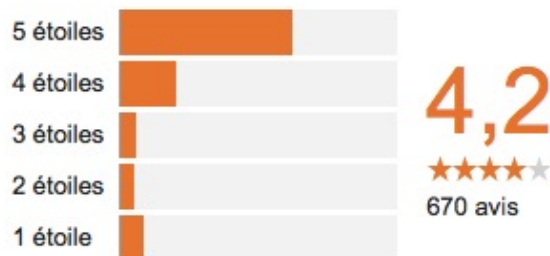
|            |       |
|------------|-------|
| Best Buy   | (140) |
| CNET       | (5)   |
| Amazon.com | (3)   |



**HP Officejet Pro 8715 All-in-One Couleur Jet d'encre - Fax / photocopieur / imprimante / scanner - Français / France**

140 € en ligne ★★★★★ 670 avis sur le produit

## Avis



### ★★★★★ Une HP classique mais de bon goût – 23 avril 2018

GEBA – Avis fourni par Cdiscount  
23 avril 2018

Produit reçu rapidement. Très bien emballé. Installation facile de l'imprimante, du driver et de la connexion wifi à une box. Très rapide. Mode recto verso pour toutes les fonctions. Je n'ai pas encore testé le scanner. Le wifi est efficace car on accède à l'imprimante de n'importe où dans la maison avec un très bon signal.

L'ensemble est un peu imposant en particulier en hauteur probablement en raison du recto-verso intégral. Le bac pour d'alimentation en papier est pratique et protège bien les feuilles. Reste plus qu'à tester la longévité et ça sur les imprimantes ce n'est jamais gagné d'avance... Ceci dit compte tenu du prix d'achat elle sera vite amortie. Un produit recommandable pour peu que l'on ait un peu de place pour la loger et qu'on ne soit pas allergique à HP.

### ★★★★★ Très bonne imprimante HP 8715 – 14 avril 2018

MoiMoi – Avis fourni par  ruecommerce.fr  
14 avril 2018

HP - Office jet Pro 8715 réceptionnée il y a tout juste une semaine.

Je ne peux pas dire si elle est bien, parfaite ou moins bien, cela dépend de chacun face à ses attentes.

Pour ma part elle imprime très vite et très bien. Le scan recto verso est très bon aussi. La prise en main facile. L'écran tactile peut être un peu petit mais très bien et très réactif.

Le seul petit reproche que j'ai à faire est que les feuilles, une fois imprimées ou scannées, ne sont pas super faciles à récupérer. Vraiment rien de bien grave.

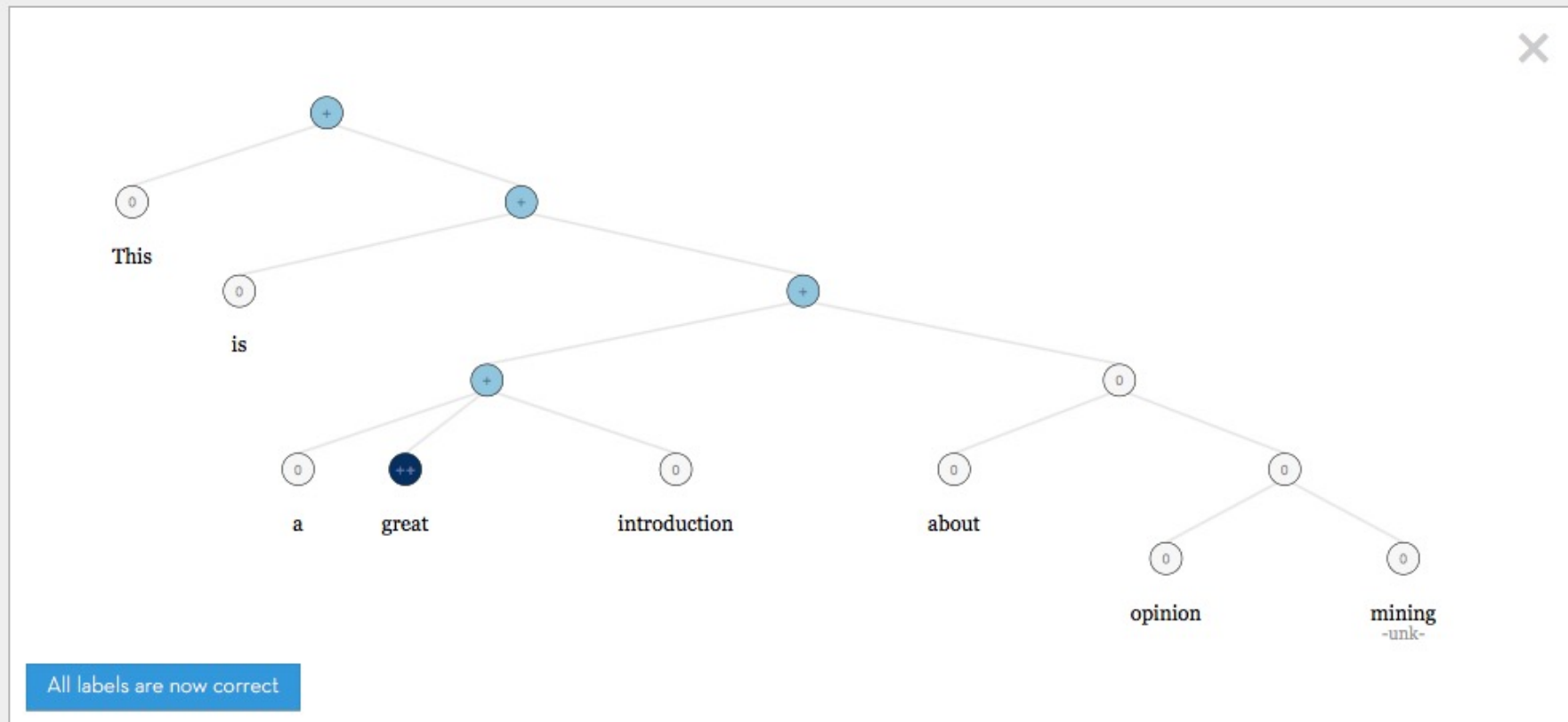


## Sentiment Analysis

[Information](#) | [Live Demo](#) | [Sentiment Treebank](#) | [Help the Model](#) | [Source Code](#)

### Sentiment Trees

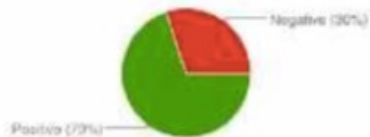
You can double-click on each tree figure to see its expanded version with greater details. There are 5 classes of sentiment classification: **very negative**, **negative**, neutral, **positive**, and **very positive**.



English

## Sentiment analysis for Obama

Sentiment by Percent



Sentiment by Count



## Tweets about: Obama

CharlotteTricia: RT @lizzie\_knott: As if **Obama** is staying at the Grove ?? so close yet so far

Posted: 21 seconds ago

ThePixelated: RT @TruthBooth: Glad to see Britain figuring out **Obama** is nothing but an asshole. Normal Americans have known it for a decade. Get out o?

Posted: 21 seconds ago

cllrbeech: RT @JASEMARKRUTTER: On reflection of **Obama's** threats today; Anyone else feel more determined than ever to VOTE LEAVE EU!Your comments... ht?

Posted: 22 seconds ago

jadeis\_kool: RT @Dory: One of the many reasons I love **Obama** <https://t.co/8UYnStlBGc>

Posted: 24 seconds ago

Maryjocork: RT @ShehabKhan: '**Obama** is a hypocrite for telling the UK what to do,' says Boris in a paper owned by an Australian-born US citizen who tell?

Posted: 26 seconds ago

JVaun: @smerconish @CNN Hard for me to take this bunch seriously, especially when 1 claims to be Dem & say **Obama** has been in WH doing nothing. Huh?

Posted: 26 seconds ago

PamEpley: RT @JohnFromCranber: Donald Trump on Hannity: '#**Obama** is The Worst Thing That Ever Happened to Israel'. My View: AND to USA. #tcot #ccot ht?

Posted: 28 seconds ago

<http://www.sentiment140.com>

<https://www.kaggle.com/kazanov/sentiment140>





# sentiment viz

Tweet Sentiment Visualization



| Search: <input type="text"/> |                 |      |      |   |  |
|------------------------------|-----------------|------|------|---|--|
| Date                         | User            | v    | a    | Tweet   |  |
| 11-05-18 10:42               | AskMrMini       | 5.63 | 4.23 | RT @rtehrani: While iPhone sales <i>remain stagnant</i> , <i>Apple services hit \$10 billion in revenue</i> <a href="https://t.co/Tb85spaV6v">https://t.co/Tb85spaV6v</a> <a href="https://t.co/IF8AuM5ZEg">https://t.co/IF8AuM5ZEg</a>   |  |
| 11-05-18 10:42               | MatthewFoxworth | 4.90 | 4.31 | The iPhone XR <i>camera</i> doesn't let any <i>ugly slip</i> by at the <i>expense of bad photo quality</i> . I'm in a <i>bind</i> .   |  |
| 11-05-18 10:42               | jackD_Reeves    | 5.03 | 4.78 | <i>Think</i> I'd honestly rather <i>be</i> woken up by a <i>lion</i> eating my <i>leg</i> than an iPhone <i>alarm,worst</i> thing ever  |  |
| 11-05-18 10:42               | PUNchayatiii    | 5.67 | 4.24 | Apne Bhakt, pidi, aapiye sab bhole hai. Sirf ss nikalte hai ek dusre ke. A <i>black woman</i> in USA <i>pretended to be trump supporter</i> and <i>raised</i> 150K <i>on</i> twitter for her tuitions and <i>rent</i> (which she is spending <i>on</i> iPhone and <i>stuff</i> ) 🤔🤔🤔 Such Alpha <i>activity</i> needed <i>on</i> Indian twitter too |  |
| 11-05-18 10:42               | halimatvivian   | 3.58 | 4.40 | RT @Abazz_bdx: If you <i>slap</i> me and I <i>vomit</i> iPhone X...How many <i>times</i> would you <i>slap</i> me???🤔🤔  |  |
| 11-05-18 10:42               | ValentineMukaz  | 5.01 | 5.10 | @gameking2018 @t_square25 I hv an iphone 7 ne Ultra XA1. Xa2 ne <i>mate</i> 20 ndo next <i>stop</i> btry re iphone rakadhakwa. Sony handina <i>objection</i> been usin sony sinc 2010.  |  |
| 11-05-18 10:42               | jacksoonmfr     | 6.11 | 4.15 | RT @myself2k4: @ConradiRobin @Rael_The_Ether @T_T_Reviews_YT @oneplus And why would you <i>want</i> that? You just <i>plug</i> and copy... With no need for any other <i>software</i> . My <i>wife</i> had an iPhone and I hated having to <i>use</i> iTunes to <i>get</i> anything <i>out</i> of it...   |  |
| 11-05-18 10:42               | oraekene        | 5.72 | 4.28 | RT @Brayz: <i>Notice</i> how it's <i>big</i> tiddied babes anon is <i>buying</i> iPhone X for <i>on</i> TL. Later, some of you <i>will argue</i> that bress is not <i>life</i> .  |  |
| 11-05-18 10:42               | pacetone        | 6.55 | 4.25 | I <i>got</i> Presentation <i>Kit</i>   <i>Free</i> iPhone Multicolor Mockup <i>on</i> @Gumroad: <a href="https://t.co/XvV6p8Fqkm">https://t.co/XvV6p8Fqkm</a>   |  |
| 11-05-18 10:42               | jkottenstette   | 6.30 | 3.74 | @ATT I need <i>help!</i> <i>Please</i> dm for assistance in getting my <i>cell</i> to <i>work</i> in Japan! I just landed and <i>can't get service</i> . I have an iPhone xs  |  |
| 11-05-18 10:42               | daimmm_boi      | 6.69 | 4.30 | RT @Apple: <i>Welcome</i> to the <i>big</i> screens. iPhone XS and iPhone XS Max.   |  |
| 11-05-18 10:42               | Ann_tunji       | 6.42 | 3.87 | But a <i>woman gifted</i> another <i>woman</i> iPhone X <i>last month</i> . But OK. <a href="https://t.co/LmeLQKTPiO">https://t.co/LmeLQKTPiO</a>   |  |
| 11-05-18 10:42               | ...             | ...  | ...  | ...   |  |

Showing 1 to 242 of 242 entries

# Opinion ou émotion?

14

- Transcription d'un retour client d'une entreprise de téléphonie :

*« oui j'ai été satisfait euh de l'accueil qui m'a été réservé [...] mais vous avez rien d'autre à foutre que de faire des enquêtes de satisfaction et d'emmerder les gens sur leur portable à longueur de temps merci au\_revoir [...] font chier »*

Opinion : positive sur le service client

Emotion : plutôt en colère

*Opinion ou autre?*

# Typologie des états affectifs de Scherer (1984)

15

- **Emotion** : brève réponse de l'organisme en réponse à l'évaluation d'un événement majeur
  - ▣ *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood** : modification diffuse de faible intensité et de longue durée non causée par le sentiment subjectif
  - ▣ *cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances** : position affective envers une autre personne dans une interaction spécifique
  - ▣ *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*
- **Attitudes**: croyances, dispositions envers les objets ou les personnes, relativement durables et « polarisées »
  - ▣ *liking, loving, hating, valuing, desiring*
- **Personality traits**: dispositions de personnalité stables et tendances comportementales typiques
  - ▣ *nervous, anxious, reckless, morose, hostile, jealous*

# Qu'est-ce qu'une opinion?

16

- Expression d'une idée, d'un fait qu'une personne tient pour vrai
- Exemple :
  - ▣ Une opinion – énoncé **subjectif**

« Le dernier iphone d'Apple est vraiment hors de prix! »

- ▣ Un fait – énoncé **objectif**

« Le dernier iphone d'Apple est vendu au prix de 1 629,90€ »



# Qu'est-ce qu'une opinion?

17

- Dictionnaire Larousse :
  - ▣ Jugement, avis, sentiment qu'un individu ou un groupe émet sur un sujet, des faits, ce qu'il en pense
- En NLP, une opinion s'exprime au travers de différents **énoncés** :
  - ▣ un jugement, une croyance, un avis, un sentiment, un point de vue, une attitude, un goût...

# Qu'est-ce qu'une opinion?

18

- Des nuances sur les termes?
  - ▣ Opinion : conclusion réfléchie mais sujette à controverse
  - ▣ Point de vue : opinion subjective très affirmée
  - ▣ Conviction : opinion ferme et sérieuse
  - ▣ Sentiment : opinion reflétant son sentiment intérieur
  - ▣ Croyance : acceptation délibérée et consentement intellectuel
  - ▣ Jugement : appréciation positive ou négative sur quelqu'un ou quelque chose

# Opinion régulières ou comparatives

19

- Deux différents types d'opinion [Jindal and Liu, 2006]

- Opinion régulière

- Expression classique vis à vis d'un produit

« Le Coca Cola a vraiment un super goût. »

- Opinion comparative

- Comparaison entre différentes entités qui partagent un même aspect

« Le Coca Cola a vraiment un meilleur goût que le Pespi. »

→ Préférence vis à vis du Coca

# Opinion explicite ou implicite

20

## □ Expression de l'opinion...

### ▣ Explicite

- Mots exprimant une polarité, opinion bien identifiée...

« Je suis très satisfaite de l'accueil qui m'a été réservé. »

### ▣ Implicite

- Pas de mots exprimant de polarité en particulier... mais
- Une figure de rhétorique utilisée dans une phrase objective

« Après avoir dormi deux jours sur ce matelas, une vallée s'est formée en plein milieu »

→ La vallée est une métaphore : expression d'une opinion négative

# Définition de l'opinion

21

□ Une opinion est un quadruplet  $(g,s,h,t)$  [Liu,2012]

1. Cible g:

- Quel objet ou quelle personne soumis à l'opinion?
- Plus finement : décomposition en Entité/Aspect

2. Sentiment s: \*valeur\* de l'opinion sur g

- Polarité : positive, négative, neutre
- Intensité : de forte à faible (e.g. intensité de 1 à 5)

3. Source h: la personne qui exprime l'opinion

4. Temps t:

- Date à laquelle a été exprimée l'opinion

*Obtenir une représentation complète est très compliquée...*

# Détection d'opinion...

22

- ... Résolution de différents problèmes !
  - ▣ Déterminer la cible
    - Détection d'entités nommées
    - Résolution des co-références
  - ▣ Déterminer l'aspect
    - Extraction d'information
  - ▣ Déterminer la polarité
    - Identification des orientations/sentiments
  - ▣ Déterminer la source
    - Analyse syntaxique ou méta-données

# Différents niveaux d'analyse

23

- Déterminer quel est le niveau de granularité et donc la complexité
  - ▣ *L'opinion est-elle positive?*
  - ▣ *Donner une note entre 1 et 5.*
  - ▣ *Déterminer les éléments de l'opinion : qui pense quoi avec quel intensité de quel aspect de quel objet quand?*
- Généralement, 3 niveaux d'analyse :
  - ▣ Document
  - ▣ Paragraphe/Phrase
  - ▣ Aspect de la cible et polarité

# Document-level

24

- Classer l'opinion générale exprimée dans le document comme *positive* ou *négative*
  - **Implique une seule entité** (non applicable si plusieurs cibles dans le document) **et une opinion**
- Exemple : Classification de « movie review »

*Thumbs up? Sentiment Classification using Machine Learning Techniques* (Pang, Lee and Vaithyanathan, 2002)

  - ▣ Utilisation de techniques classiques : Naives Bayes, MaxEnt, SVM
  - ▣ Des résultats moins probants qu'en classification en thème
    - « More challenging »



# Sentence-level

25

- Classer la phrase (ou l'énoncé) selon 3 (ou 4) classes
  - ▣ Objective : aucune opinion exprimée
  - ▣ Subjective : opinion positive, négative ou mixed (??)
    - ➔ Implique une **seule source** et **éventuellement une opinion** à évaluer
- Exemples :
  - ▣ *Development and use of a gold standard data set for subjectivity classifications* [Wiebe, Bruce and O'Hara, 1999]
    - Définir si un reportage expose un fait ou une opinion
    - Calcul de la mesure Kappa de Cohen (1960)
  - ▣ *Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences* [Yu and Hatzivassiloglou, 2003]
    - Dans un système de Q/A : séparer les opinions des faits
    - En 2 étapes : objective/subjective puis positive/négative

# Fine-grained level

26

- Rechercher la cible et la polarité de l'opinion
  - ▣ Plus de découpage structurelle (documents, paragraphe, phrase, propositions...)
  - ▣ Préciser exactement \*ce\* qui est \*aimé ou pas\*
    - ➔ Implique certainement **plusieurs opinions** sous la forme de couples cible/polarité
- Exemple : Analyse d'avis client sur des produits vendus en ligne
  - ▣ *Feature-based opinion mining and summarization* [Hu and Liu, 2004]
    - Détailler des avis clients sur des produits

# Quelques exemples

27

## □ [Liu,2012]

### ▣ Opinions « fine-grained » multiples exprimées

*« Although the service is not that great, I still love this restaurant »*

■ Target=service, opinion polarity=negative

■ Target=restaurant, opinion polarity=positive

*« The iPhone's call quality is good, but its battery life is short »*

■ Target=call quality, opinion polarity=positive

■ Target=life battery, opinion polarity=negative

## □ [Pereg et. al, 2019] : ABSApp: A Portable Weakly-Supervised Aspect-Based Sentiment Extraction System

[https://drive.google.com/file/d/1BLk0xkjlOqyRhNy4UQEFQpDF\\_KR\\_NMAd/view](https://drive.google.com/file/d/1BLk0xkjlOqyRhNy4UQEFQpDF_KR_NMAd/view)

# Simple ou complexe?

28

- Plusieurs tâches peuvent être définies :
  - ▣ La plus simple : est ce que ce texte exprime une opinion positive ou négative?
  - ▣ Un peu plus complexe : Affecter une note exprimant l'opinion sur une échelle de 1 à 5.
  - ▣ Avancée (très) : Définir quelle est la source de l'opinion, la cible (ou l'entité et l'aspect) de l'opinion et polarité exprimée (+temps)

# Classification supervisée

29

- 3 approches principales :
  - ▣ Méthodes symboliques
    - À base de règles et de lexiques
  - ▣ Méthodes statistiques
    - Utilisation d'algorithme d'apprentissage automatique
  - ▣ Méthodes hybrides
    - Combine connaissance *a priori* et machine learning
    - À la mode : weakly supervised...

# Préparation des données (1)


30


- Étapes de prétraitements
  - ▣ Supprimer des formes non pertinentes :
    - nombres, ponctuation (ironie...), minuscule...
  - ▣ Normaliser :
    - par ex : réduire les mots exagérés
  - ▣ Généraliser les formes de surfaces :
    - Tokenisation, Lemmatisation, Stemming
- Réduire la dimensionnalité

# Méthode à base de lexiques

31

- Exemple très simple :
  - ▣ Soit un lexique de mots polarisés
  - ▣ Compter le nombre de mots positifs et négatifs pour obtenir le score global indiquant la classe

 <sup>+1</sup> It is <sup>+1</sup> fun and easy to do sentiment analysis!

 <sup>-1</sup> I don't like reading all of the <sup>-1</sup> negative Tweets!

- Si plus complexe : nécessite beaucoup d'investissement manuel...

### Algorithm 1: Sentiment Calculation

**Data:** Preprocessed Twitter data

**Result:** Output: Positive, Negative, Neutral

Find the list of sentiment words *SentiList*, its position in the sentence;

Find the list of sentiment negation words *SentiNegat*, its position in the sentence;

Find the list of blind negation words *BlindNegat*, its position in the sentence;

**if** *BlindNegat* **then**

**return** negativity;

**else**

**if** *SentiList* and *SentiNegat* **then**

**foreach** word in the *SentiList* **do**

**if** word is atmost the distance of 2 from *SentiNegat* **then**

                Revert the polarity of the word;

**end**

**end**

**else**

**if** *SentiNegat* **then**

            Add the *SentiNegat* to the negative *SentiList*;

**end**

**end**

**end**

*SentiSum*=0;

**foreach** word in the *SentiList* **do**

*SentiSum*=*SentiSum*+sentiment of word;

**end**

**if** *Hashtag* is present **then**

    Find all the sentiment words in hash tag

## Extrait d'algorithme de classification d'opinion à base de lexiques :

*Serendio: Simple and Practical lexicon based approach to Sentiment Analysis*

[Palanisamy et. al, Sem'Eval 2013]

Table 2: Lexicon Details

| Data type               | Count |
|-------------------------|-------|
| Blind Negation word     | 7     |
| Negation                | 13    |
| Positive sentiment word | 1260  |
| Negative sentiment word | 1703  |
| Split word              | 16    |



# Les lexiques de sentiment anglais

33

- WordNet Affect :
  - ▣ ressource linguistique pour la représentation lexicale de connaissances sur les affects
  - ▣ 1 903 concepts liés à un état mental ou émotionnel
  - ▣ <http://wndomains.fbk.eu/wnaffect.html>
- SentiWordNet :
  - ▣ ressource dédiée aux systèmes de classification de textes d'opinions
  - ▣ Assigne à chaque mot de WordNet 3 scores (+,-,=)
  - ▣ <https://github.com/aesuli/SentiWordNet>
- Et bien d'autres!
  - ▣ General Inquirer, MPQA ...

# Les lexiques de sentiment français

34

- Polarimots
  - ▣ 7 483 noms, verbes, adjectifs et adverbes français associés semi-automatiquement à une polarité positive, négative et neutre
  - ▣ <http://polarimots.lif.univ-mrs.fr/>
- Emotaix
  - ▣ 4921 entrées identifiants émotions, sentiments, humeurs...
  - ▣ <https://centrepsyche-amu.fr/outils-recherche/>
- Blogoscopie
  - ▣ 982 entrées lexicales dont 493 adjectifs
  - ▣ <http://www.lina.univ-nantes.fr/?Lexique-de-l-evaluation.html>
- Casoar
  - ▣ Polarité, force et catégorie sémantique de 2830 mots
  - ▣ <https://projetcasoar.wordpress.com>
- VADER et AFINN
  - ▣ Deux outils d'analyse du sentiment

# Aller plus loin

35

- Créer ou enrichir automatiquement les lexiques
  - ▣ V. Hatzivassiloglou and K. McKeown. 1997.  
*Predicting the semantic orientation of adjectives*. In  
ACL-EACL 1997
    - The first one!! → see next slide
  - ▣ *Enrichissement d'un lexique de termes subjectifs à partir de tests sémantiques*
    - Vernier et Monceau, TAL 2010
    - <https://hal.archives-ouvertes.fr/hal-00472165/document>

1. All conjunctions of adjectives are extracted from the corpus along with relevant morphological relations.
2. A log-linear regression model combines information from different conjunctions to determine if each two conjoined adjectives are of same or different orientation. The result is a graph with hypothesized same- or different-orientation links between adjectives.
3. A clustering algorithm separates the adjectives into two subsets of different orientation. It places as many words of same orientation as possible into the same subset.
4. The average frequencies in each group are compared and the group with the higher frequency is labeled as positive.

# Limites des méthode à base de lexiques

37

- Un même mot polarisé peut avoir des orientations opposées selon le contexte

Les compétences de ce candidat sont élevées  
Le prix de cette imprimante est élevé

- Une phrase avec des mots polarisés n'exprime pas forcément d'opinion... interrogative et conditionnelle

Est-ce que les manteaux de Decathlon sont suffisamment chauds?  
J'aimerais tant trouver un manteau chaud et douillet

- Une phrase sans mot polarisé peut exprimer une opinion (implicite...)

Ce lave vaisselle utilise beaucoup trop d'eau.

# Limites des méthode à base de lexiques

38

## □ Système très naïf :

- ▣ Pas de prise en compte de l'organisation des mots dans la phrase
- ▣ Prise en compte de la négation (sac de mots) ?
- ▣ Difficile prise en compte du contexte (langage figuratif)

Quelle super caméra! Un ensemble de pixels verts s'est formé en haut à droite de l'image au bout de 2 jours

## □ Une maintenance fastidieuse

- ▣ Toujours inclure de nouvelles règles
- ▣ Vérifier qu'il n'y a pas de mauvaises interactions avec les existantes

# Préparation des données (2)

39

- Feature engineering :
    - ▣ Représentation du texte
      - Bag of unigrams/n-grams + TF-IDF
      - Word embeddings
    - ▣ Extraction de descripteurs supplémentaires
      - Classiques : POS, mot gouverneur, nombre de mots, ...
      - Spécifiques : détection d'émoticônes, utilisation de lexique de mots polarisés...
- Obtenir une classification efficace

# Méthodes à base d'IA

40

- Application d'un algorithme d'apprentissage supervisé
  - ▣ Naïves Bayes, Max Ent, Régression linéaire, SVM...
  - ▣ **Réseaux de neurones...**
    - Reconnus comme performants en SA : CNN, LSTM et Bi-LSTM
      - *Convolutional neural networks for sentence classification* [Kim, EMNLP 2014]
      - *Neural Sentiment Classification with User and Product Attention* [Chen et al. , EMNLP 2016]
    - Recherches plus récentes : Embeddings de sentiments
      - *Learning sentiment-specific word embedding for twitter sentiment classification* [Tang et al. , ACL 2014]
      - *Refining word embeddings for sentiment analysis* [Yu et al., EMNLP 2017]

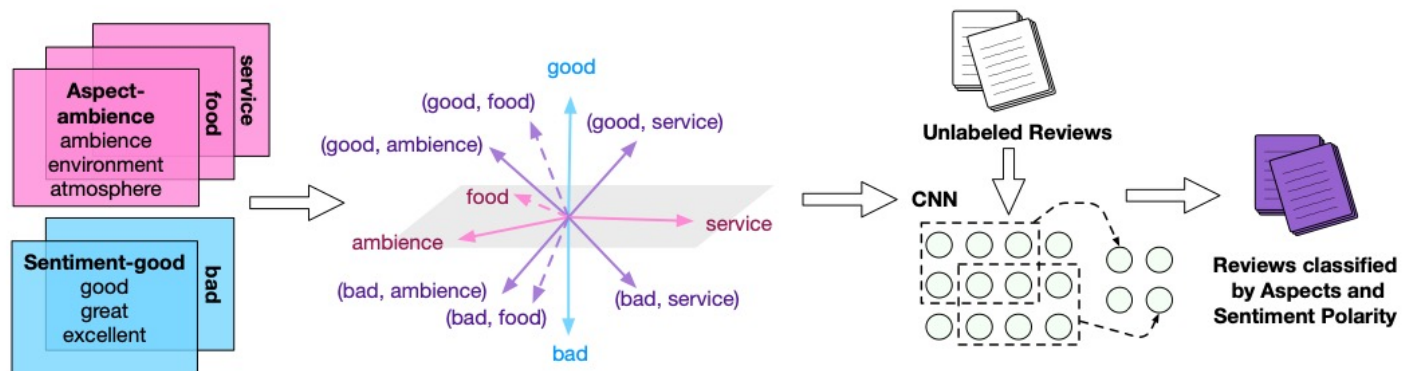


# Méthodes à base d'IA

41

## □ Aspect Based + Weakly Supervised

- Purpura et. al., *WS4ABSA: An NMF-Based Weakly-Supervised Approach for Aspect-Based Sentiment Analysis with Application to Online Reviews*, 21st Int. Conf. on Discovery Science, 2018.
  - Utiliser word2vec + TF.IDF pour élargir un ensemble de seeds
  - Utiliser Non-neg. Matrix Factorization pour extraire les aspects
- Huang et. al., *Weakly-Supervised Aspect-Based Sentiment Analysis via Joint Aspect-Sentiment Topic Embedding*
  - <https://aclanthology.org/2020.emnlp-main.568.pdf>



# Complexité car langage naturel

42

- Nécessité de traiter du langage naturel, et toutes les difficultés qui l'accompagnent!
  - ▣ Résolution de co-référence

« *Le dernier i-phone vient de sortir. Encore un qui n'est pas donné!!* »
  - ▣ Traitement de la négation
  - ▣ Désambiguïsation du sens des mots
  - ▣ Disfluences si langage oral

# Complexité car subjectivité

43

- Subjectivité= Avis possiblement différent entre 2 pers.
  - Incohérence de l'annotation de référence?
- Notion d'accord inter-annotateurs
  - ▣ Idée : Si deux personnes annotent le même document, la probabilité qu'ils soient en accord sur l'annotation doit être supérieure à un accord obtenu au hasard
- Mesure Kappa :
  - ▣ J. Cohen. 1960. *A coefficient of agreement for nominal scales.*

# Accord inter-annotateur

- *Kappa statistic* : Mesure du degré de concordance de deux annotations
- Ratio qui permet d'évaluer la différence entre :
  - ▣ Accord réel observé entre les annotations
  - ▣ Accord aléatoire
- Mesure comprise entre -1 et 1
  - ▣  $Kappa > 0,8$  est considéré excellent
  - ▣  $Kappa < 0,4$  médiocre

# Accord inter-annotateur

- Se calcule à partir de la matrice de confusion
- Accord inter-annotateur proposé entre hypothèse et référence

$$\kappa = \frac{P_{AccordJuges} - P_{AccordHasard}}{1 - P_{AccordHasard}}$$

■ Avec :

$$P_{AccordJuges} = \frac{\sum_i \#Case_{ii}}{n}$$

$$P_{AccordHasard} = \frac{\sum_i \#Col_i * \#Ligne_i}{n^2}$$

et n le nombre d'instances

# Quels autres défis?

46

- Opinion (très) implicite
  - ▣ Objectivité ne veut pas dire sans opinion...

*« Nous avons acheté cette voiture la semaine dernière et l'essuie glace vient de tomber! »*
- Opinion comparative, avec contexte nécessaire!

*« Mon accès internet est aussi rapide qu'au temps du minitel »*  
*« Ce n'est clairement pas le fils d'Einstein »*
- Langage figuratif

*« J'adore ce cours passionnant de détection d'opinion »*

# Langage figuratif

47

- En opposition avec langage littéral – strictement conforme au sens des mots –
  - Déformation du sens
- Le langage figuratif, tout un style!
  - ▣ Ironie : contradiction entre la pensée de l'auteur (du locuteur) et le sens littéral exprimé
  - ▣ Sarcasme : moquerie ironique souvent blessante
  - ▣ Et aussi humour, satire, parodie...

# Spécificité du langage figuratif

48

- Utilisation du contexte obligatoire
  - ▣ Contexte proche (même phrase)
  - ▣ Contexte plus large
    - Exemple d'un dialogue
      - Alizée : « On a interro de fouille de texte aujourd'hui »
      - Pierre : « Super!! On va trop bien s'amuser! »
  - ▣ Connaissance *a priori*
    - Exemple d'un tweet
      - « *Pour ceux qui ne le savaient pas : Le prix des #carburants augmentera encore en janvier. Quel joie! #BonneAnnee* »



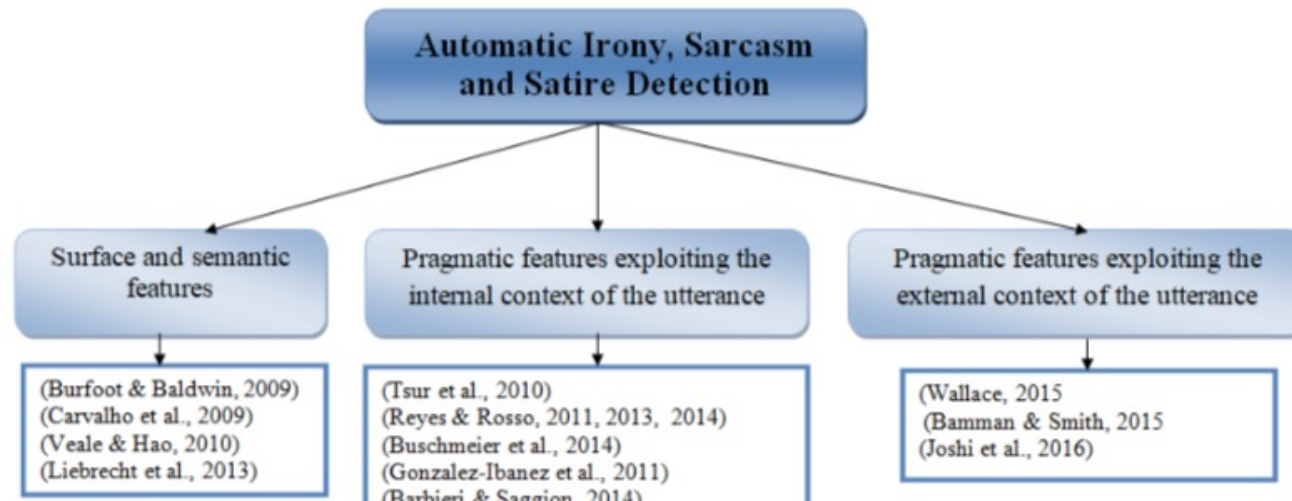
# Détection de l'ironie dans les tweets

49

- Domaine de + en + populaire en NLP
- Quels défis?
  - ▣ Quelles différences entre texte littéraire et tweet?
  - ▣ Quels indices linguistiques marquent l'ironie?
  - ▣ Comment modéliser ces indices?
- Annotation automatique du corpus de référence
  - ▣ [Gonzalez-Ibanez et al., 2011] : « the best judge of whether a tweet is intended to be sarcastic is the author of the tweet »
  - ▣ Utilisation des hashtags spécifiques qui se trouvent en fin de tweet : #ironie, #sarcasme...

# Principales approches

50



External  
knowledge needed



## Internal context

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

## Surface features

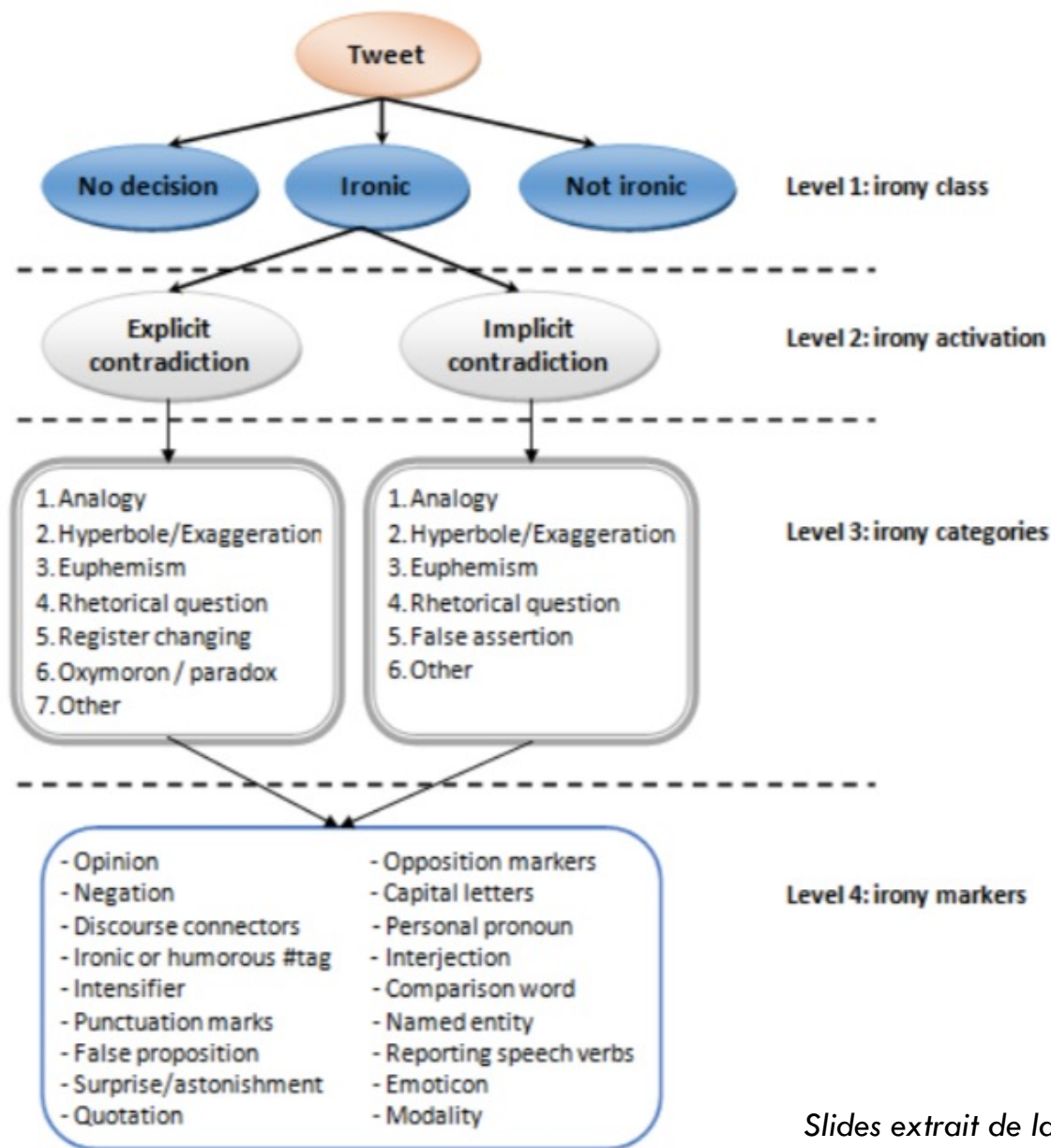
- #Valls a appris la mise sur écoute de #Sarkozy en lisant le journal !!!  
Heureusement qu'il n'est pas Ministre de l'Intérieur :-)

## External context

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

# Schéma d'annotation multi-niveaux

[Karoui,2017]



# Marqueurs d'intensité

52

| Types d'opération |              | Intensité/Quantité  |
|-------------------|--------------|---|
| Indéterminée      | Faible       | *peu, un peu, *ne... guère,<br>à peine, faiblement, légèrement      |
|                   | Modéré       | moyennement, modérément, plutôt,<br>passablement, presque/quasiment |
|                   | Forte        | très, fort, bien, si, tellement,<br>beaucoup, fortement, drôlement  |
|                   | Extrême      | extrêmement, énormément<br>extraordinairement                       |
| Relativité        | Excès        | trop, excessivement   |
|                   | Adéquation   | assez, suffisamment   |
|                   | Insuffisance | pas assez, trop peu, insuffisamment                                 |
| Totalisante       |              | tout, totalement, entièrement, absolument                           |

Extrait de la thèse de Lei Zhang, « Analyse automatique d'opinion : problématique de l'intensité et de la négation pour l'application à un corpus journalistique », 2013

# Les campagnes d'évaluation - anglais

53

- TREC (Text REtrieval Conference) - NIST
  - ▣ Blog Track de 2006 à 2010
    - 2010 : tâche 1 objectif/subjectif
  - ▣ Microblog de 2011 à 2015 (=tweet)
- SemEval (Semantic Evaluation) - ACL
  - ▣ Sentiment analysis : 2007, 2010, 2013-14, 2016-2019
  - ▣ En 2019 :
    - Task 5: HatEval: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter
    - Task 6: OffensEval: Identifying and Categorizing Offensive Language in Social Media

# Les campagnes d'évaluation - français

54

- DEFT (Défi Fouille de Textes) [2005-2017]
  - ▣ Atelier conjoint à TALN, organisé par le LIMSI
  - ▣ Sur l'opinion : (2007, 2009, 2015, 2017, 2018)
    - ▣ 2007 : détection de l'opinion exprimée dans un texte de retranscription de débats parlementaires
      - Positif ou négatif
    - ▣ 2015 : fouille d'opinion, de sentiment et d'émotion dans des messages postés sur Twitter
      - Classe spécifique
    - ▣ 2018 : Recherche d'information et analyse de sentiments dans des tweets sur les transports en IDF
      - Marqueurs de sentiments et cible

# Bibliographie

56

- Opinion Mining and Sentiment Analysis", Foundations and Trends® in Information Retrieval: Vol. 2 [Pang and Lee,2008] <http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf>
- **Sentiment Analysis and Opinion Mining** [Liu, 2012] <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>
- Analyse automatique d'opinions - États des lieux et perspectives [Benamara, 2016] [sur demande...](#)
- Détection automatique de l'ironie dans les contenus générés par les utilisateurs [Jihen Karoui,2017] <https://hal.archives-ouvertes.fr/tel-01615868>

# Webographie

57

En plus des différentes références dans les slides...

- <https://www.kdnuggets.com/2018/03/5-things-sentiment-analysis-classification.html>
- <https://www.displayr.com/sentiment-analysis-simple/>
- <https://books.openedition.org/oep/214>
- <https://monkeylearn.com/sentiment-analysis/>