Text Mining CBS 2019

Lecture 3: Subjectivity mining
Piek Vossen

Overview

- Part I: What is subjectivity
- Part II: Subjectivity mining
- Part III: Tools and resources

Part I
What is subjectivity in language?

- bella is the picture of health with boundless energy until a few days before she dies . this is absolutely and completely ridiculous and an insult to every family whose mother has suffered through the horrible pains of a death by cancer .

nltk_data/corpora/sentence_polarity/rt-polarity.neg

Part I
What is subjectivity in language?

- bella is the picture of health with **boundless energy** until a few days before **she dies** . this is absolutely and **completely ridiculous** and an **insult** to every **family** whose **mother** has **suffered** through the **horrible pains** of a **death** by **cancer** .

nltk_data/corpora/sentence_polarity/rt-polarity.neg

- **Explicit sentiment**: boundless energy, completely ridiculous, insult, horrible
- **Implicit sentiment**: dies, suffer, pains, death, cancer
- **Holders**:
  - Author about the picture and the **participants** mentioned in the text
  - **Participants** of the text: she, family, mother about the cancer, the pain and the death
- But also what is mentioned and what is not mentioned about the picture is subjective (ag
  setting)

Part I
What is subjectivity in language?

- A Colombia government trade official has urged the business community to aggressively diversify its activities and stop relying so heavily on coffee. Samuel Alberto Yohai, director of the Foreign Trade Institute, INCOMEX, said private businessmen should not become what he called "mental hostages" to coffee, traditionally Colombia's major export.

nltk_data/corpora/reuters/test/15198

Part I
What is subjectivity in language?

- A Colombia government trade **official** has **urged** the business community to **aggressively** **diversify** its **activities** and **stop** **relying** so heavily on **coffee**. **Samuel Alberto Yohai**, director of the Foreign Trade Institute, INCOMEX, said private businessmen should not become what he called "mental **hostages**" to coffee, traditionally Colombia's **major** export.

nltk_data/corpora/reuters/test/15198

Opinion

Holder | **Sentiment** | **Target**

official — **urged** — **aggressively** — **diversify activities**

urged — **stop** — **rely on coffee**

Product reviews

- [t]**excellent** **picture quality** / color canon powershot g3[+3]##i recently purchased the canon powershot g3 and am **extremely satisfied** with the **purchase** . use[+2]##the camera is **very easy** to **use** , in fact on a recent trip this past week i was asked to take a picture of a vacationing elderly group .

nltk_data/corpora/product_reviews_1/Canon_G3.txt

Opinion

Sentiment	**Aspects**
excellent	**picture quality**
very easy	**use**
extremely satisfied	**purchase**

Tweets & Irony?

- {"contributors": null, "coordinates": null, "text": "Everything in the **kids section** of IKEA is **so cute**. **Shame** I'm nearly 19 in 2 months :(", "user": {"screen_name": "EveHollyHousley", "time_zone": "London", "profile_background_image_url": "http://pbs.twimg.com/profile_background_images/776873880/f89d8aa869414e41eefd804284a3d95c.jpeg", "profile_background_image_url_https": "https://pbs.twimg.com/profile_background_images/776873880/f89d8aa869414e41eefd804284a3d95c.jpeg", "default_profile_image": false, "url": null, "profile_text_color": "333333", "following": false, "listed_count": 0, "entities": {"description": {"urls": []}}, "utc_offset": 3600, "profile_sidebar_border_color": "FFFFFF", "name": "Eve", "favourites_count": 4759, "followers_count": 450, "location": "Manchester", "protected": false, "notifications": false, "profile_image_url_https": "https://pbs.twimg.com/profile_images/620302844093181952/M7IP4ZHa_normal.jpg", "profile_use_background_image": false, "profile_image_url": "http://pbs.twimg.com/profile_images/620302844093181952/M7IP4ZHa_normal.jpg", "lang": "en", "statuses_count": 4384, "friends_count": 541, "profile_banner_url": "https://pbs.twimg.com/profile_banners/383849833/1435267039", "geo_enabled": true, "is_translator": false, "contributors_enabled": false, "profile_sidebar_fill_color": "DDEEF6", "created_at": "Sun Oct 02 16:42:30 +0000 2011", "verified": false, "profile_link_color": "0084B4", "is_translation_enabled": false, "has_extended_profile": false, "id_str": "383849833", "follow_request_sent": false, "profile_background_color": "C0DEED", "default_profile": false, "profile_background_tile": true, "id": 383849833, "description": "Lauv it \u2022 18 \u2022 Music Student"}, "retweet_count": 0, "favorited": false, "entities": {"hashtags": [], "user_mentions": [], "urls": [], "symbols": []}, "source": "Twitter for iPhone", "truncated": false, "geo": null, "in_reply_to_status_id_str": null, "is_quote_status": false, "in_reply_to_user_id_str": null, "place": null, "in_reply_to_status_id": null, "in_reply_to_screen_name": null, "lang": "en", "retweeted": false, "in_reply_to_user_id": null, "created_at": "Fri Jul 24 10:42:48 +0000 2015", "metadata": {"iso_language_code": "en", "result_type": "recent"}, "favorite_count": 0, "id_str": "624530162890219521", "id": 624530162890219521}

nltk_data/corpora/twitter_samples/negative_tweets.json

Ekman (1972, ..)

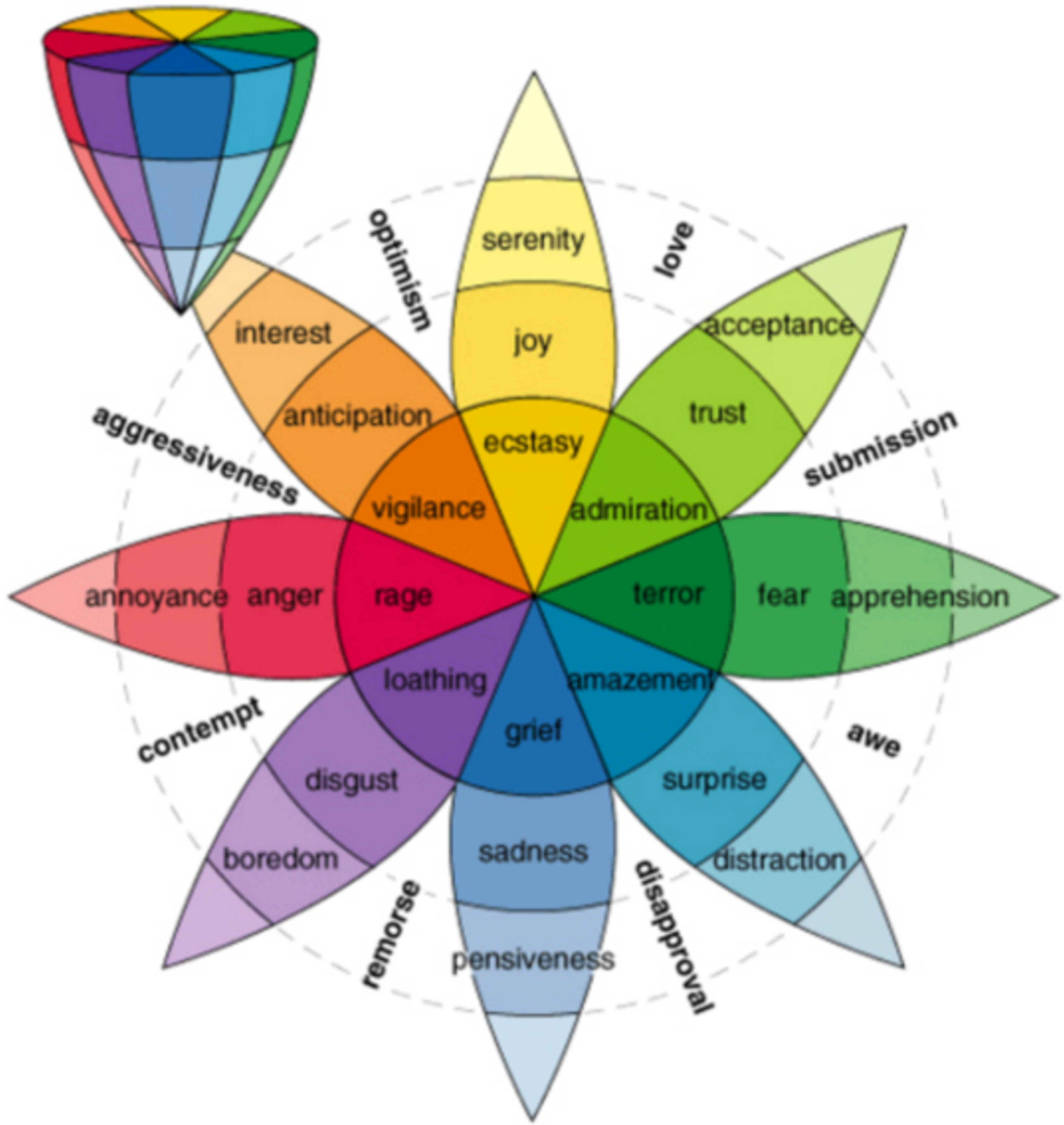
- 6 Basic emotions in many different cultures from facial expressions related to described situations

- Anger
- Disgust
- Fear
- Happiness
- Sadness
- (Surprise)

- Extension but not all facially expressed:

- Amusement
- Contempt
- Contentment
- Embarrassment
- ExcitementGuilt
- Pride in achievement
- Relief
- Satisfaction
- Sensory pleasure
- Shame

Plutchik’s Wheel of Emotions (1980)



http://wndomains.fbk.eu/wnaffect.html

A-Labels and corresponding example synsets

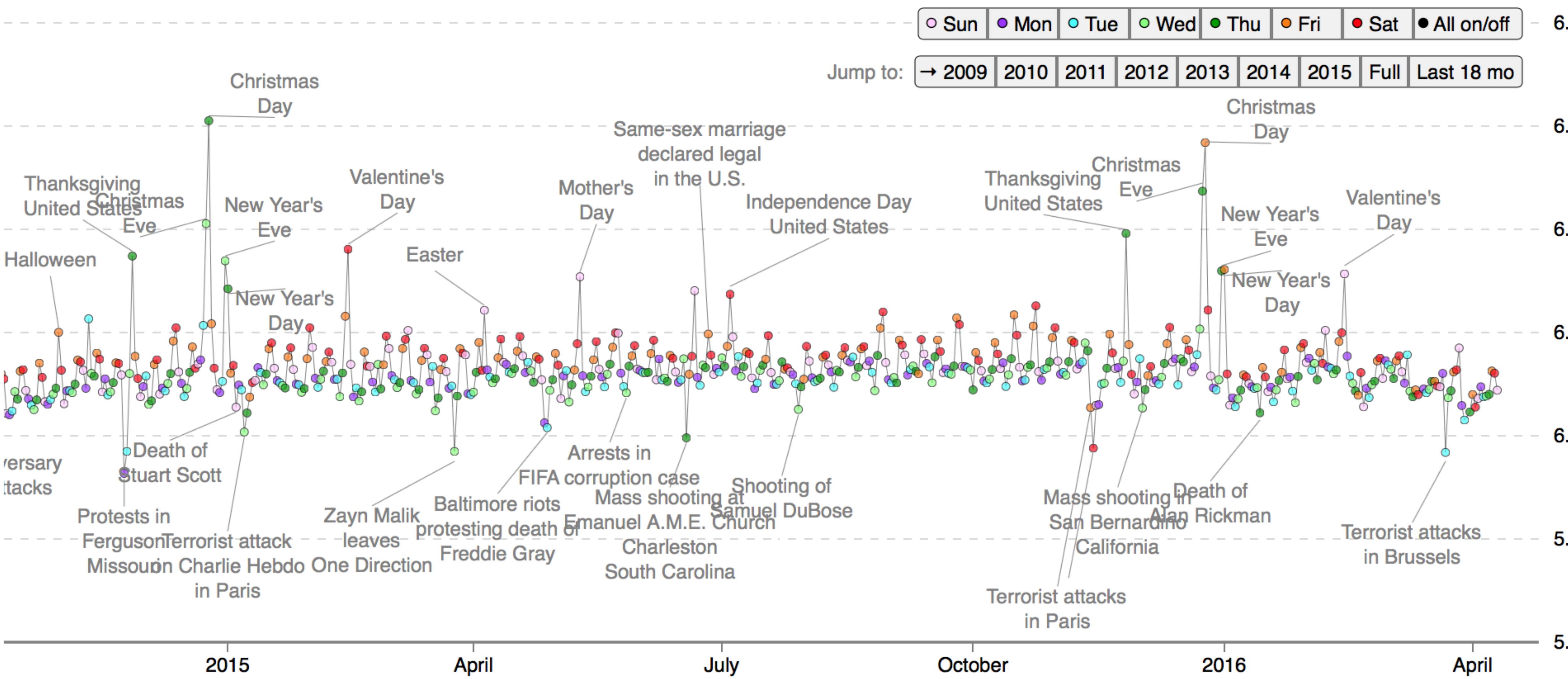
A-Labels	Examples
EMOTION	*noun anger#1, verb fear#1*
MOOD	*noun animosisy#1, adjective amiable#1*
TRAIT	*noun aggressiveness#1, adjective competitive#1*
COGNITIVE STATE	*noun confusion#2, adjective dazed#2*
PHYSICAL STATE	*noun illness#1, adjective all in#1*
HEDONIC SIGNAL	*noun hurt#3, noun suffering#4*
EMOTION-ELICITING SITUATION	*noun awkwardness#3, adjective out of danger#1*
EMOTIONAL RESPONSE	*noun cold sweat#1, verb tremble#2*
BEHAVIOUR	*noun offense#1, adjective inhibited#1*
ATTITUDE	*noun intolerance#1, noun defensive#1*
SENSATION	*noun coldness#1, verb feel#3*

Terminology

- **Opinion**: I like him
- **Subjectivity** (broad term that covers all forms of opinion mining, sentiment analysis, stance)
- **Sentiment**: blaah
- **Attitude** (see subjectivity)
- **Stance** : I support the anti-vaccination movement
- **Polarity**: (like sentiment)
- **Aspects-facets-features-properties**: This iPhone has an incredible good battery
- **Argumentation**: I like this iPhone because it has a very powerful battery
- **Emotion**: response to a situation that is deemed important: anger, joy, sadness

hedonometer.org

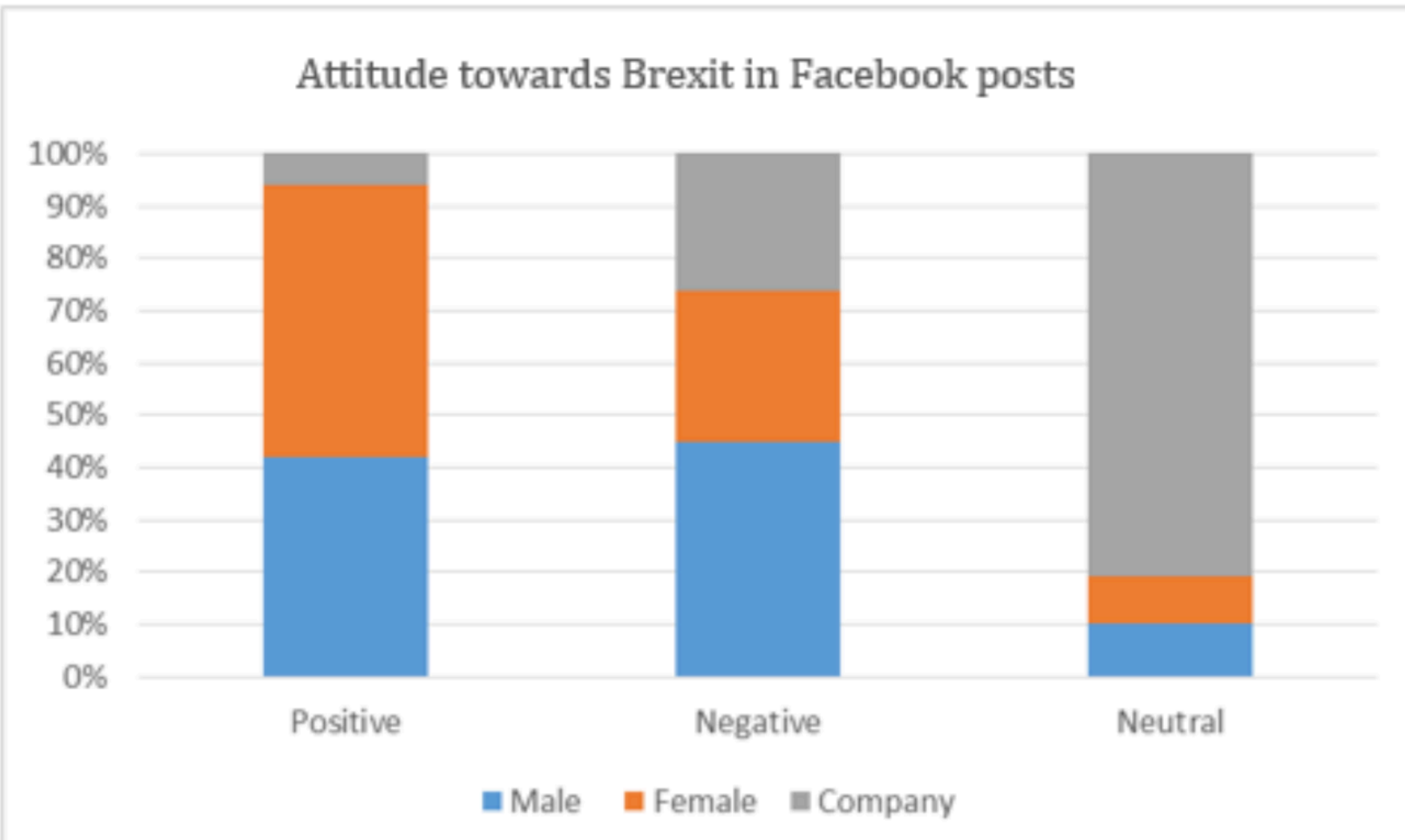
Average Happiness for Twitter



https://www.venndigital.co.uk/blog/the-eu-referendum-brexit-or-bremain-78807/

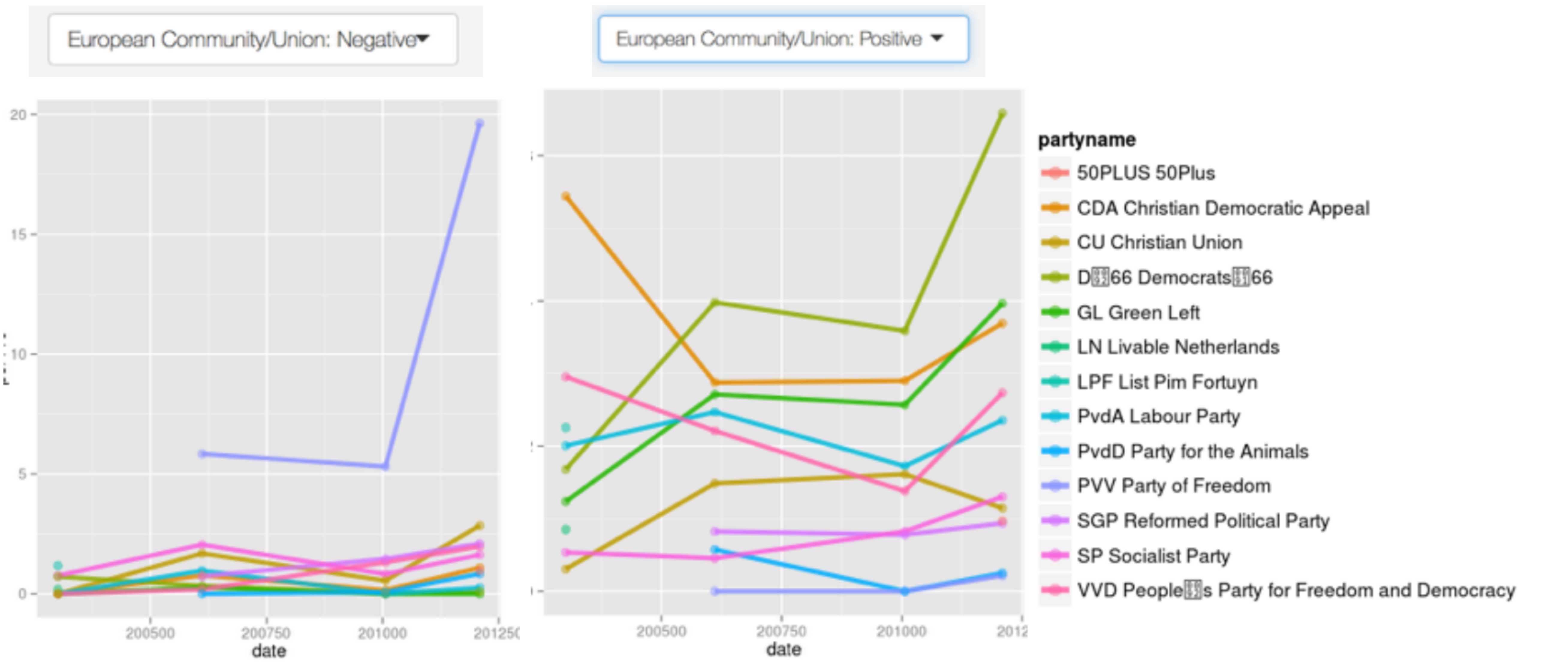
pro/against Brexit (2016)

Interestingly, females were most likely to post **for** Brexit (68%) followed by men (55%), while companies posted mainly neutral content (71%). 36% of the posts made by men were actually **against** Brexit while only 23% of women voiced a negative opinion.



The manifesto project:

The Manifesto Project provides the scientific community with parties' policy positions derived from a content analysis of parties' electoral manifestos. It covers over 1000 parties from **1945 until today** in over **50 countries** on five continents



Netherlands 2000-2015

https://manifesto-project.wzb.eu

Love & Hate Letters

348 posts from: lovingyou.com

279 fragments from the millennium project: https://ratbags.com/rsoles/categories.htm

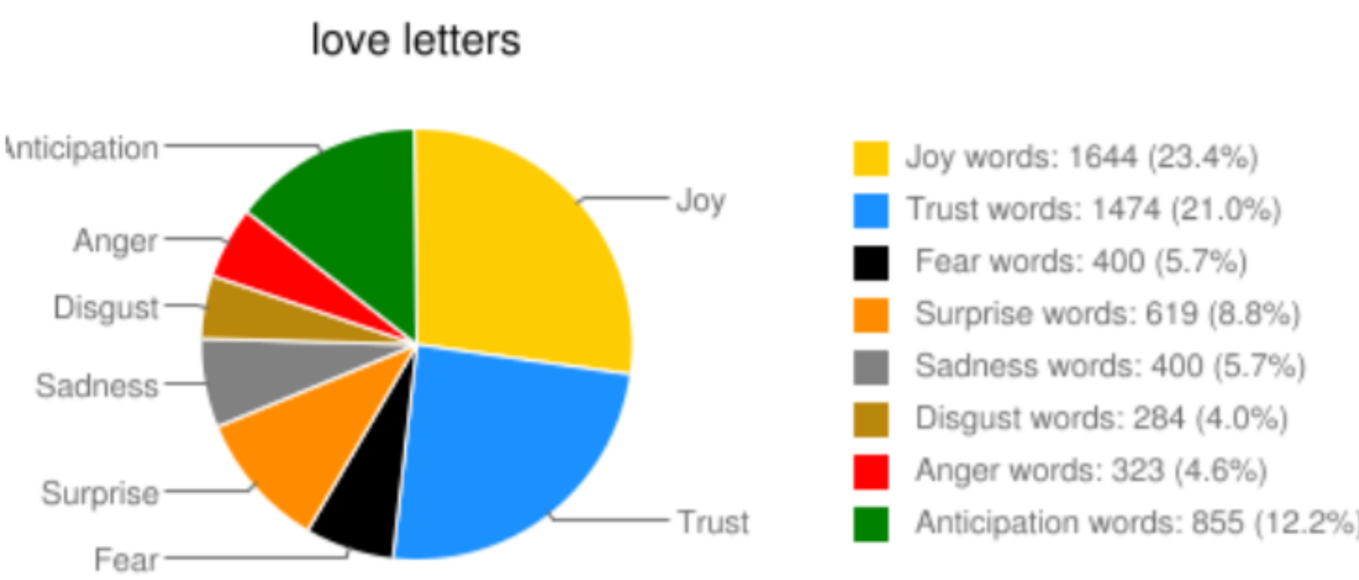
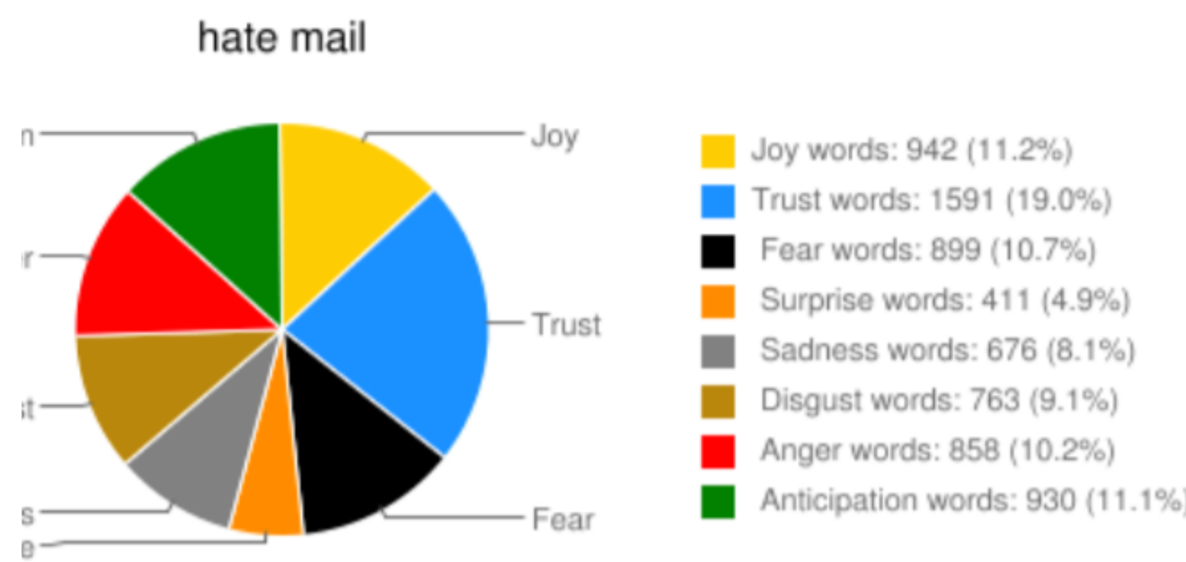


Figure 5: Percentage of emotion words in the love letters corpus.



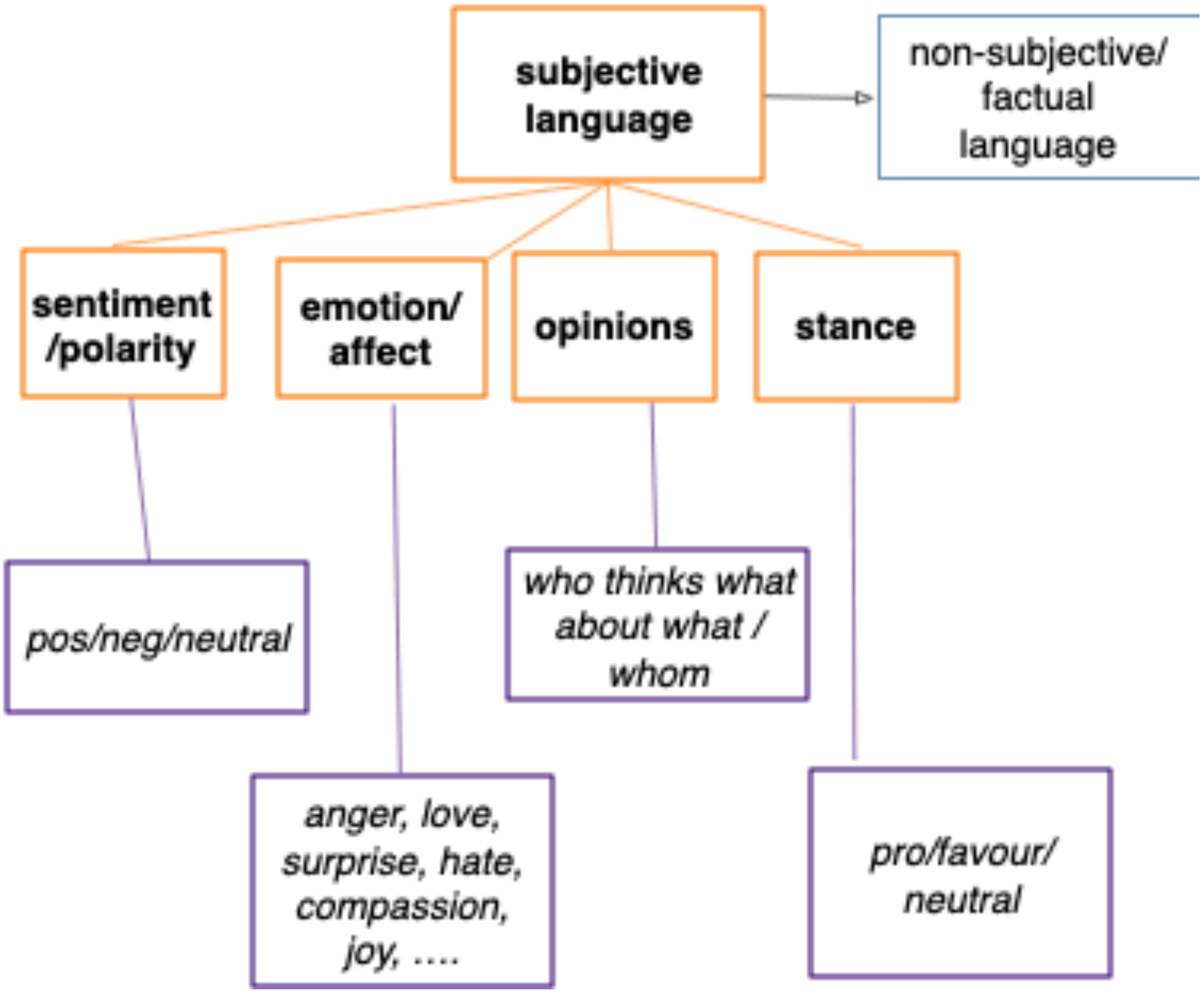
e 6: Percentage of emotion words in the hate mail s.

Saif M. Mohammad, 2012

What is Subjectivity Mining?

- Software for automatically extracting **opinions, emotions and sentiment** in text.
- It allows us to track attitudes and feelings on the web. People write **blogs, comments, reviews and tweets** about all sorts of different topics
- We can track products, brands and people for example and determine whether they are **positively or negatively** on the web
- It allows individuals to get **an opinion an a global scale**

Subjectivity mining



Levels of Analysis

- Corpus or set of documents
- Document: one sentiment per document
- Sentence/ statement
  - Several opinions, on several topics by several opinion holders per document
- Opinions on entities, topics or aspects

Methods

- Rule-based approach , knowledge based
  - Identify subjective sentences, exclude objective sentences
  - Count all positive and negative words (from a lexicon) per unit
  - Process negations
  - Process intensifiers: very, extremely, terribly
  - Aggregate sentiment at document level, corpus level, etc. (in order to generate opinions on a generic scale)
- Machine learning
  - need a lot of training data
  - classification problem
  - Decide on the features to use (feature engineering)

Opinion mining

- Preprocessing: tokenisation, sentence splitting, syntactic parser
- Detect holder
    - *My wife did not like the staff* (e.g. *like, think, say, claim, believe, feel* are source-introducing predicates)
    - *These idiots do not know how to run a hotel*
- Detect the opinion expression
    - *like, idiots*
- Detect the target
    - *staff, hotel*
- Determine the sentiment or emotion
    - *negative*

Aspect based sentiment analysis

	Sentence3	Sentence4
Entity (= opinion	Canon G12	Canon G12
Aspects (= related to	Picture quality	Weight
Sentiment (=opinion	Positive	Negative
Opinion holder	John Smith	Wife of John Smith
Time	September 2010	September 2010

- (1) I bought a Canon G12 camera six months ago. (2) I simply love it. (3) The picture quality is amazing. (4)However, my wife thinks it is too heavy for her. (John Smith, september 2010)

Difficulties

- hard to define what the task is both for humans and machines
  - fine grained opinion mining: holders, agents
  - emotion mining (6 basic emotions)
  - explicit and implicit opinions
- complex process that depends on other modules in the pipeline (error propagation)
- genres vary a lot: news, microblogs, reviews,

Do we agree on opinions?

- **Hotel reviews**
  - Document level - 3 classes (positive, negative, neutral) - 0.87 Kappa
- **Black Pete debate**
  - Tweet level - 4 classes / 3 annotators : 0.65 (a1-a2); 0.57 (a1-a3); 0.55 (a2-a3) Kappa
- **News annotations**
  - 0.70 Kappa on opinion expressions

Review text	User rating	Ann1	Ann2
The hotel is **fantastic**, but the area around the hotel is very very noisy.	4	-2	3
Good Hotel, in Down Town. Really for business. For holidays I had choose something else.	4	-2	-2

Opinion Annotation

Washington (AFP) - Obama said Thursday that he had included openly gay athletes in the US Olympic delegation to show the

32 United States would not abide discrimination in sport or anywhere else.

His comments, in an interview with NBC on the eve of the Sochi Winter Olympics, came after a senior Russian official activists

20 should <u>not promote</u> gay rights during the Olympics following the passage of controversial anti-gay legislation in Russia.

4 Obama <u>picked</u> several openly gay former athletes in the US delegation to the opening of the games and pointedly did not dispatch a cabinet-level official or a member of his family to Sochi

Black Pete debate

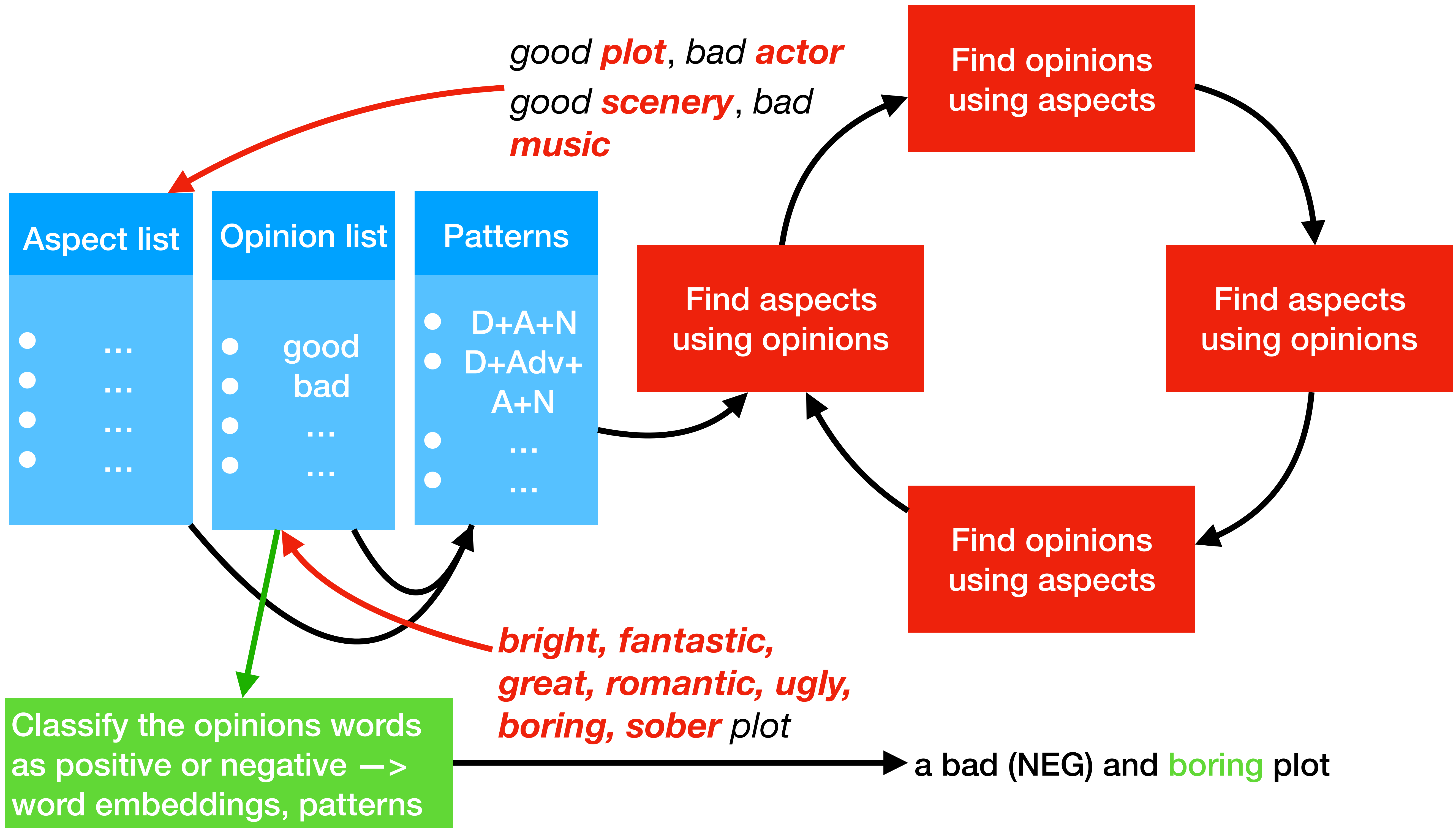
Why should zwarte piet go away? 1. He is not a nigger but he is black from the chimney. 2. Why screw up my favourite feast	Black Pete stays
If you don't like Black Pete then go back to your own country	Black Pete stays
Black Pete is okay ? Imagine that our neighbors to the east(Germany) had a child party with dancing clowns in striped	Black Pete goes
Black Pete has to leave the Netherlands because it is racism	Black Pete goes
@GJK1979 Typically Dutch, we complain about the weather and about black Pete ;-)	Stop discussion
@ElmoSwek I was at DeWereld Draait Door(Dutch tv show) about the discussion of black Pete. Disgusting guy	Off topic
Lois throws a pepernoot (dutch candy) across the class room and shouts I always wanted to be black Pete"	Off topic

Master Thesis Lars Meijers (VUA, 2013)

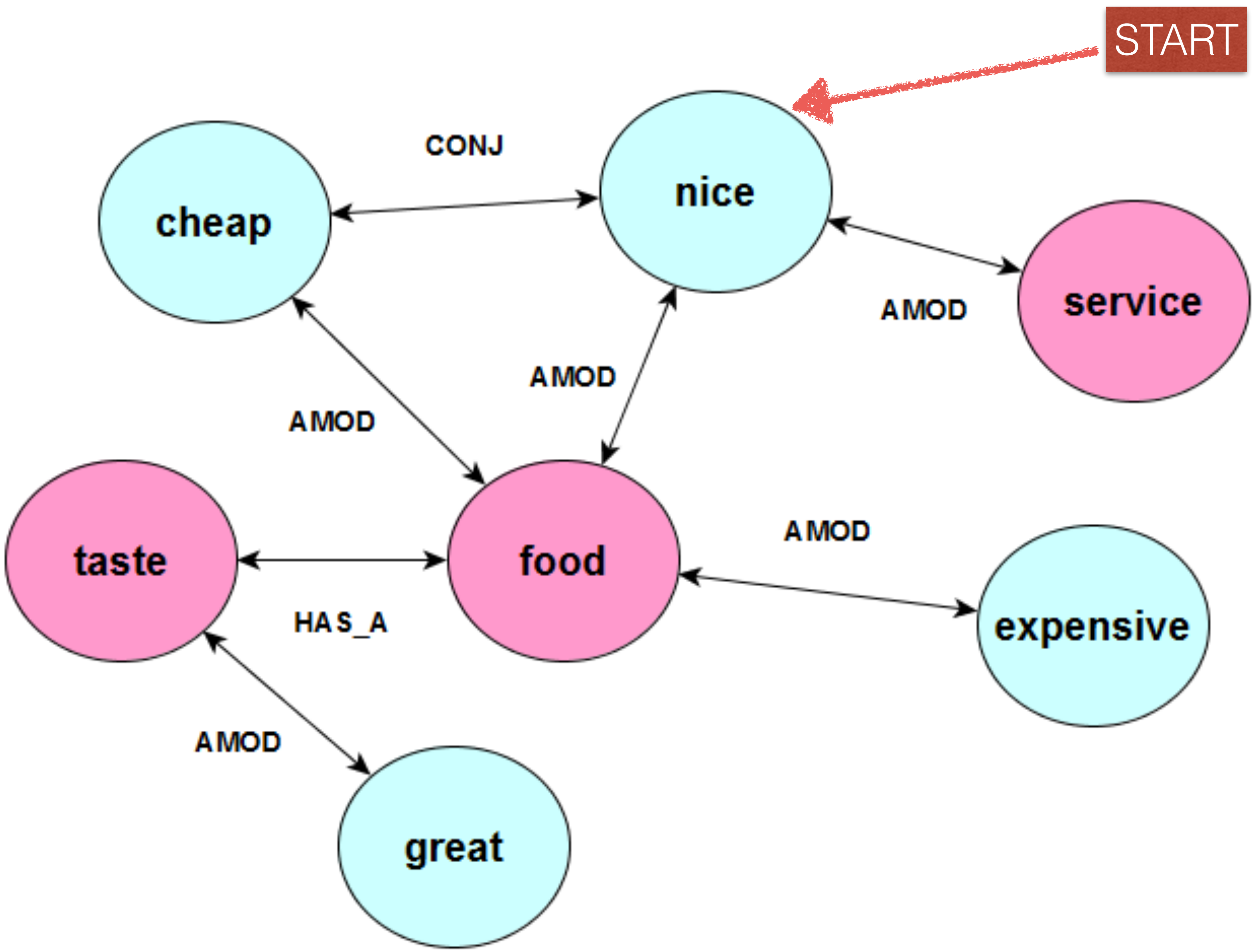
Context dependence

- A **cold** person
- A **cold** soda
- A **cold** shower
  - to cool off
  - because the boiler broke (again!)
- Low prices, low ceilings, cheap fast food, cheap fast car

Double Propagation



Double propagation



Top 15 ranked aspect terms for restaurants	Top 15 ranked aspect terms for laptops	Top 15 ranked aspect terms for hotels
1- food	1- battery life	1- hotel
2- service	2- keyboard	2- room
3- staff	3- screen	3- staff
4- bar	4- feature	4- service
5- drink	5- price	5- food
6- table	6- machine	6- view
7- menu	7- toshiba laptop	7- stay
8- dish	8- windows	8- breakfast
9- atmosphere	9- performance	9- pool
10- pizza	10- use	10- floor
11- meal	11- battery	11- area
12- bartender	12- program	12- location
13- price	13- speaker	13- bed
14- server	14- key	14- beach
15- dinner	15- hard drive	15- bar

Document level sentiment classification with and without preprocessing

- Ex. (2) De kamers zijn niet schoon en hebben geen eigen badkamer. (*The rooms are not clean and do not have an own bathroom*)

78.3
- Rule and lexicon-based: negation flips vote
- Machine Learning Naive Bayes, 1171 hotel reviews

82.3
- Bag-of-WORDS: de kamers zijn niet schoon en hebben geen eigen badkamer

83.6
- Bag-of-words + LEMMA: de kamer zijn niet schoon en hebben geen eigen badkamer (the room are not clean and have no own bedroom)

84.7
- Bag-of-Words + sentiment lexicon and rules: de kamer zijn niet schoon **neg_tag** en hebben geen eigen badkamer **negator_tag** (the room are not clean **neg_tag** and have no own bathroom **negator_tag**

Features for sentiment mining

- **Features**:
  - Words (bag-of-words)
  - N-grams
  - Parts-of-speech (e.g. Adjectives and adjective-adverb combinations)
  - Opinion words (lexicon-based: dictionary or corpus)
  - Valence intensifiers and shifters (for negation); modal verbs; ...
  - Syntactic dependency
- **Feature selection based on**
  - frequency
  - information gain
  - Odds ratio (for binary-class models)
  - mutual information
- **Feature weighting**
  - Term presence or term frequency
  - Inverse document frequency (TF.IDF)
  - Term position : e.g. title, first and last sentence(s)

Machine Learning methods

Table 6
Distribution of articles based on intelligent techniques applied.

Applied techniques	#Articles	Articles' references
SVM	55	[21,26,29,33,44–46,50,51,53,54,57,58,66,67,73,76,77,86,88,90,91,94,95,97,101,108,109,111,114,116,118,125,131,148,157,160,163,165,167,169,172,176,177,183,195,197,200,209,210,212,214,225,228,240]
Dictionary based approaches (DBA)	41	[13,18,23–25,35,36,47,55,64,67–69,85,96,110,112,117,126,127,158,170,171,175,183,193–196,202,203,206,207,209,210,213,216,218,220,229,241]
NB	28	[33,46,50,54,56,73,80,86,90,94,98,101,111,114,116,118,121,124,125,131,148,156,163,167,197,209,217,228]
NN	11	[48,50–53,57,76,101,116,213,226]
DT	9	[53,66,73,76,86,94,116,118,209]
Maximum entropy	8	[33,46,54,60,63,66,148,156,174]
Logistic regression	9	[53,77,88,99,116,118,163,197,220]
Linear regression	8	[8,18,70,75,222–224,231]
Ontology	8	[30,41,62,98,182,194–196]
LDA	8	[61,92,107,185,189,191,182,240]
Random forest	4	[77,81,228,210,228]
SVR	5	[118,123,130,155,227]
CRF and rCRP	5	[37,88,93,186,190]
Boosting	4	[12,118,179,197]
SVM-SMO	4	[76,97,118,169]
Fuzzy logic	3	[23,41,62,213]
Rule miner	4	[37,100,112,217]
EM	3	[56,59,155]
K-medoids	1	[52]
RBF NN	1	[130]

Table 5
Sentiment classification accuracy reported on common datasets.

S#	Dataset	Articles	Obtained result
1	Pang and Lee [167]	[156]	92.70% accuracy
2		[112]	90.45% F_1
3		[169]	90.2% accuracy
4		[35]	89.6% accuracy
5		[54]	87.70% accuracy
6		[46]	87.4% accuracy
7		[50]	86.5% accuracy
8		[26]	85.35% accuracy
9		[162]	81% F_1
10		[124]	79% accuracy & 86% F_1
11		[61]	76.6% accuracy
12		[69]	76.37% accuracy
13		[48]	75% precision
14		[98]	79% precision
15	Pang et al. [33]	[109]	Approx. 90% accuracy
16		[165]	88.5% accuracy
17		[172]	87% accuracy
18		[33]	82.9% accuracy
19		[156]	78.08% accuracy
20		[180]	75% accuracy
21		[48]	60% precision
22		[195]	86.04%
23	Blitzer et al. [149]	[45]	84.15% accuracy
24		[99]	80.9% (Avg.) accuracy
25		[54]	85.15% (Avg.) Max. 88.65% accuracy on Kitchen reviews
28		[165]	88.7% accuracy
29		[61]	71.92% accuracy

Ravi & Ravi 2015

(Some) Tools

- OpinionFinder: subjective sentences , source (holder) of the subjectivity and words that are included in phrases expressing positive or negative sentiments: http://mpqa.cs.pitt.edu

- Basic sentiment tokenizer plus some tools, by Christopher Potts: http://sentiment.christopherpotts.net

- Sentiment for Tweets: VADER (for tweets); https://github.com/cjhutto/vaderSentiment

- Linguistic Inquiry and Word Count, Psychological features: http://liwc.wpengine.com

(Some) Tools

- SentiStrength (sentistrength.wlv.ac.uk)
- TheySay (apidemo.theysay.io)
- Sentic (sentic.net/demo)
- Sentdex (sentdex.com)
- Lexalytics (lexalytics.com)
- Sentilo (wit.istc.cnr.it/stlab-tools/sentilo)
- nlp.stanford.edu/sentiment

Lexicons

- Bing Liu's opinion lexicon: http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html
- MPQA subjectivity lexicon: http://www.cs.pitt.edu/mpqa/
- SentiWordNet
  - Project homepage: http://sentiwordnet.isti.cnr.it
  - Python/NLTK interface: http://compprag.christopherpotts.net/wordnet.html
- WordNet Affect: http://wndomains.fbk.eu/wnaffect.html
- Harvard General Inquirer: http://www.wjh.harvard.edu/~inquirer/
- SenticNet: http://sentic.net
- NRC (emotions expressed by words): https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

(Some) datasets

- Data from Lillian Lee's group: http://www.cs.cornell.edu/home/llee/data/
- Data from Bing Liu: http://www.cs.uic.edu/~liub/
- Large movie review dataset: http://ai.stanford.edu/~amaas/data/sentiment/
- Pranav Anand & co. (http://people.ucsc.edu/~panand/data.php)
    - Internet Argument Corpus
    - Annotated political TV ads
    - Focus of negation corpus
    - Persuasion corpus (blogs)
- Data on AFS:
    - /afs/ir/data/linguistic-data/mnt/mnt4/PottsCorpora
      README.txt, Twitter.tgz, imdb-english-combined.tgz, opentable-english-processed.zip
    - /afs/ir/data/linguistic-data/mnt/mnt9/PottsCorpora
      opposingviews, product-reviews, weblogs
- Twitter data collected and organized by Moritz!
  /afs/ir.stanford.edu/data/linguistic-data/mnt/mnt3/TwitterTopics/

From Potts (2013), p.5

More datasets

- SNAP review datasets: http://snap.stanford.edu/data/
- Yelp dataset: http://www.yelp.com/dataset_challenge/
- More on Twitter datasets, including critical appraisal: Saif et al. (2013)
- Kaggle: https://www.kaggle.com/

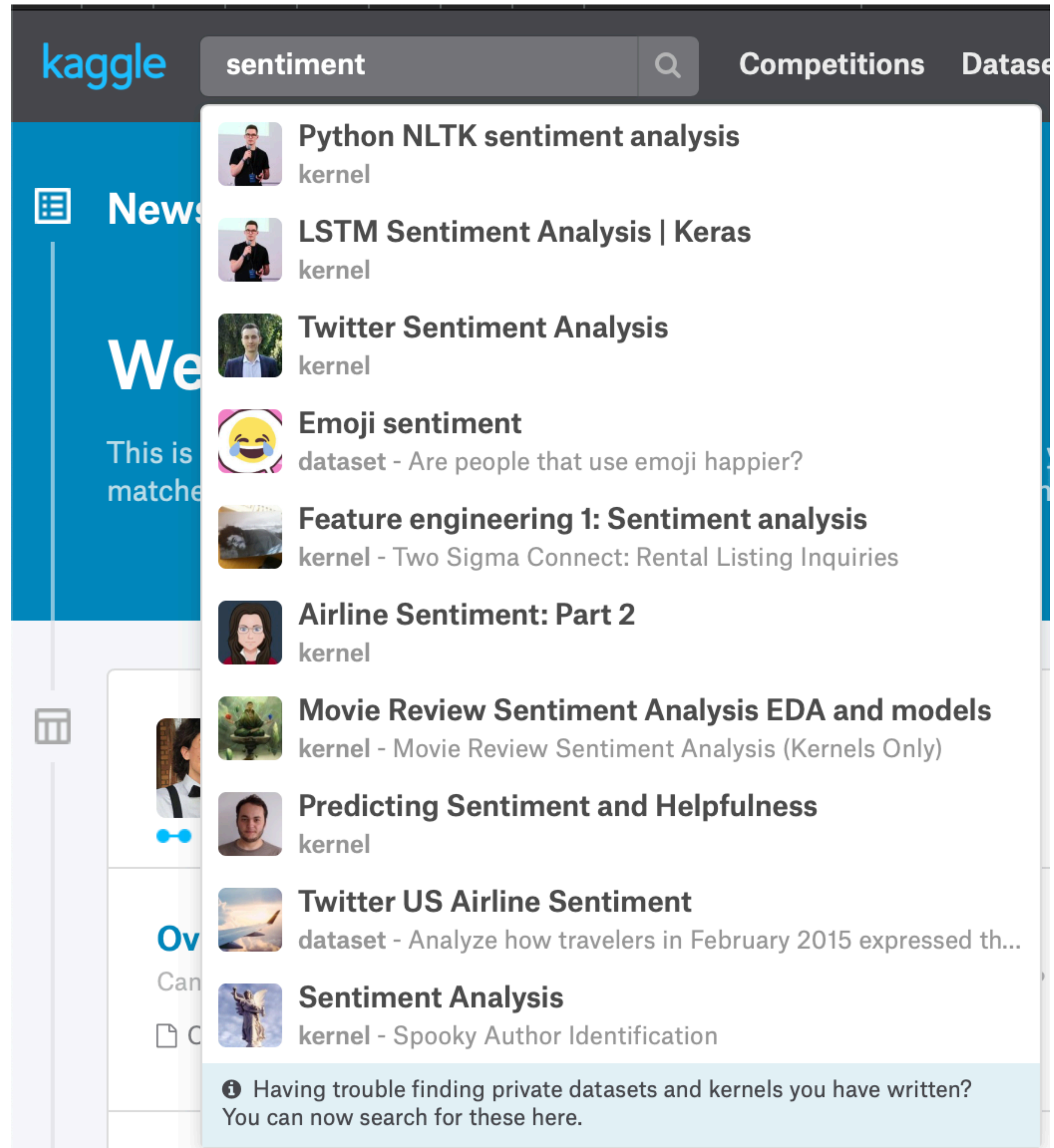


Table 9
List of publicly available datasets.

S#	Data set	Type	Lang.	Web resource	Details
1	Stanford large movie data set	Movie Reviews	English	http://ai.stanford.edu/~amaas/data/sentiment/	Movie Reviews
2	COAE2008	Product Reviews	Chinese	http://ir-china.org.cn/coae2008.html	2739 documents for movie, education, finance, economics, house, computer, mobile phones, etc. 1525 +ve, 1214 −ve
3	Boacar	Car Reviews	Chinese	http://www.riche.com.cn/boacar/	11 type of car TradeMarks and total review 1000 words, having 578 POS, 428 reviews
4	[187]	Reviews, forums	English	http://sifaka.cs.uiuc.edu/~wang296/Data/	Accessed: 27 August, 2014
5	[188]	Reviews	English	http://uilab.kaist.ac.kr/research/WSDM11	Aspect oriented dataset. Accessed: 18 December, 2014
6	Movie-v2.0	Movie Reviews	English	http://www.cs.cornell.edu/people/pabo/movie-review-data/	Data size: 2000 Positive: 1000 Negative: 1000
7	Multi-domain	Multi-domain	English	http://www.cs.jhu.edu/~mdreze/datasets/sentiment	
8	SkyDrive de Hermit Dave	Spanish Word Lists	Spanish	https://skydrive.live.com/?cid=3732e80b128d016f&id=3732E80B128D016F%213584	
9	TripAdvisor	Reviews	Spanish	http://clic.ub.edu/corpus/es/node/106	18,000 customer reviews on hotels and restaurants from Hopinion
10	[38]	Multi-Domain	English	www2.cs.uic.edu/~liub/FBS/sentiment-analysis.html	6800 opinion words on 10 different products
11	TBOD [144]	Reviews	English		Product Review on Cars, Headphones, Hotels
12	[68]	Product Reviews	English	http://www.lsi.us.es/_fermin/index.php/Datasets	Product Reviews from Epinion.com on headphones 587 reviews, hotels 988 revi… and cars 972 reviews
13	[148]	Movie Reviews	Turkish	http://www.win.tue.nl/~mpechen/projects/smm/#Datasets	5331 positive and 5331 negative reviews on movie
14	[148]	Product Reviews	Turkish	http://www.win.tue.nl/~mpechen/projects/smm/#Datasets	700 +ve &700 −ve reviews on books, DVD, electronics, kitchen appliances
15	ISEAR	English sentences	English	www.affective-sciences.org/system/files/page/2636/ISEAR.zip	The dataset contains 7666 such statements, which include 18,146 sentences, 449,060 running words.
16	[149]	Product Reviews	English	http://www.cs.jhu.edu/~mdredze/datasets/sentiment/	Amazon reviews on 4 domain (books, DVDs, electronics, kitchen appliances)
17	DUC data, NIST	Texts	English	http://www-nlpir.nist.gov/projects/duc/data.html, http://www.nist.gov/tac/data/index.html	Text summarization data
18	[70]	Restaurant and Hotel Reviews	English	http://uilab.kaist.ac.kr/research/WSDM11	Restaurant and Hotel Reviews from Amazon and Yelp
19	[114]	Restaurant Reviews	Cantonese	http://www.openrice.com	Reviews on restaurant
20	[125]	Biographical Articles	Dutch	http://www.iisg.nl/bwsa	574 Biographical articles
21	Spinn3r dataset	Multi-Domain	English	http://www.icwsm.org/2011/data.php	
22	[86]	Ironic Dataset	English	http://users.dsic.upv.es/grupos/nle/	3163 ironic reviews on five products
23	HASH [179]	Tweets	English	http://demeter.inf.ed.ac.uk	31,861 Pos tweets, 64,850 Neg tweets, 125,859 Neu tweets
24	EMOT [179]	Tweets and Emoticons	English	http://twittersentiment.appspot.com	230,811 Pos & 150,570 Neg tweets
25	ISIEVE [179]	Tweets	English	www.i-sieve.com	1520 Pos tweets, 200 Neg tweets, 2295 Neu tweets
26	[177]	Tweets	English	e-mail: apoorv@cs.columbia.edu	11,875 tweets
27	[52]	Opinions	English	http://patientopinion.org.uk	2000 patient opinions
28	[96]	Tweets	English	http://goo.gl/UQvdx	667 tweets
29	[39]	Movie Reviews	English	http://ai.stanford.edu/~amaas/data/sentiment/	50,000 movie reviews
30	[164]	Tweets	English	http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip	
31	[210]	Spam Reviews	English	http://myleott.com/op_spam	400 deceptive and 400 truthful reviews in positive and negative category. Las… Accessed by: 12 April, 2015
32	[230]	Sarcasm and nasty reviews	English	https://nlds.soe.ucsc.edu/iac	1000 discussions, ~390,000 posts, and some ~73,000,000 words

Ravi & Ravi 2015

Survey Literature

- Ronen Feldman: Techniques and applications for sentiment analysis. Commun. ACM 56(4): 82-89 (2013).
- Bing Liu, Lei Zhang: A Survey of Opinion Mining and Sentiment Analysis. Mining Text Data 2012: 415-463.
- Bo Pang, Lillian Lee: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval 2(1-2): 1-135 (2007).
- Potts (2013). Introduction to Sentiment Analysis. http://www.stanford.edu/class/cs224u/slides/2013/cs224u-slides-02-26.pdf
- Mikalai Tsytsarau, Themis Palpanas: Survey on mining subjective data on the web. Data Min. Knowl. Discov. 24(3): 478-514 (2012)
- Saif M. Mohammad, (2012) From once upon a time to happily ever after: Tracking emotions in mail and books, Decision Support Systems 53 (2012) 730–741
- Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-Based Systems, 89, 14-46.

The OpeNER project

- https://www.opener-project.eu
- Opinion detection and Named Entity recognition for 6 European languages
- Sentiment lexicons for 6 European languages
- Demonstrator: http://tour-pedia.org/about/

A hotel review

Nothing special really. Comfortable and clean but very boring decor in comparison to other NH hotels. I stayed in NH in Brussels and Zurich and I really liked them because of their modern and stylish design and big rooms. This one was just like any other hotel. Basic rooms with basic and dull decor - bit disappointing. The customer service was average. The rate was very expensive and I still had to pay for Internet and 20 euros for breakfast!!! It was good but way overpriced! The best thing about the hotel was the location - city centre, 2min from a metro stop.

A hotel review

Nothing special really. Comfortable and clean but very boring decor in comparison to **other NH hotels**. I stayed in **NH** in **Brussels** and **Zurich** and I really liked **them** because of **their** modern and stylish design and big rooms. **This one** was just like any **other** hotel. Basic rooms with basic and dull decor - bit disappointing. The customer service was average. The rate was very expensive and I still had to pay for Internet and **20 euros** for breakfast!!! It was good but way overpriced! The best thing about **the hotel** was **the location** - city centre, 2min from a metro stop.

Named entities

- this NH hotel
- NH Brussels
- NH Zurich

http://en.wikipedia.org/wiki/NH_Hoteles
http://dbpedia.org/page/NH_Hoteles
http://en.wikipedia.org/wiki/Brussels
http://en.wikipedia.org/wiki/Zurich

A hotel review

- Nothing special really. Comfortable and clean but very boring decor in comparison to **other NH hotels**. I stayed in **NH** in **Brussels** and **Zurich** and I really liked **them** because of **their** modern and stylish design and big rooms. **This one** was just like any **other** hotel. Basic rooms with basic and dull decor - bit disappointing. The customer service was average. The rate was very expensive and I still had to pay for Internet and **20 euros** for breakfast!!! It was good but way overpriced! The best thing about **the hotel** was **the location** - city centre, 2min from a metro stop.

Named entities:
- this NH hotel
- NH Brussels
- NH Zurich

Co-references:
- other
- them
- this one

Properties:
- décor
- design
- room
- clean (hygiene)
- service
- rate
- Internet
- breakfast

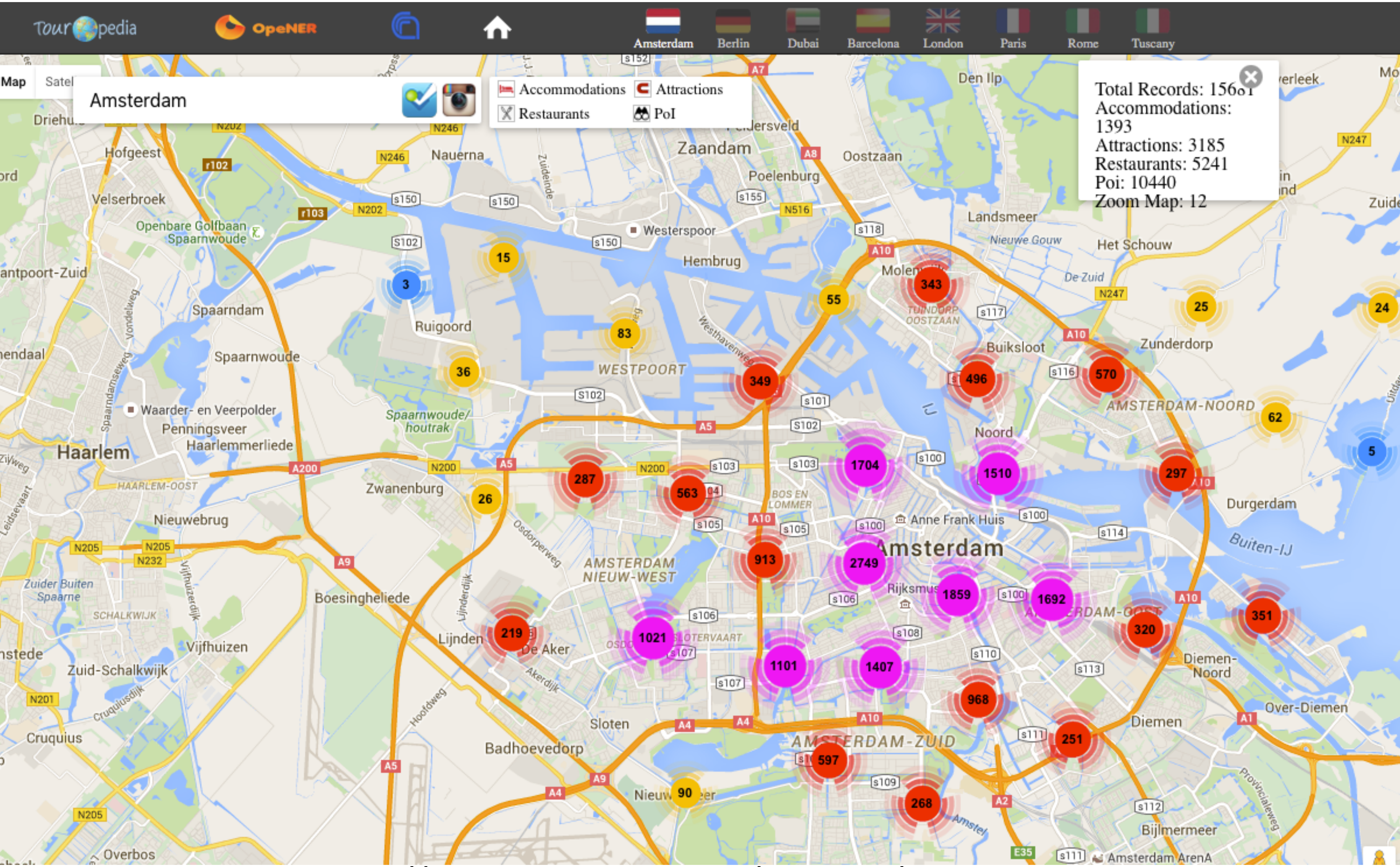
Sentiments:
- nothing special really
- comfortable and clean
- boring
- really liked
- modern and stylish
- big
- just
- basic and dull

A hotel review

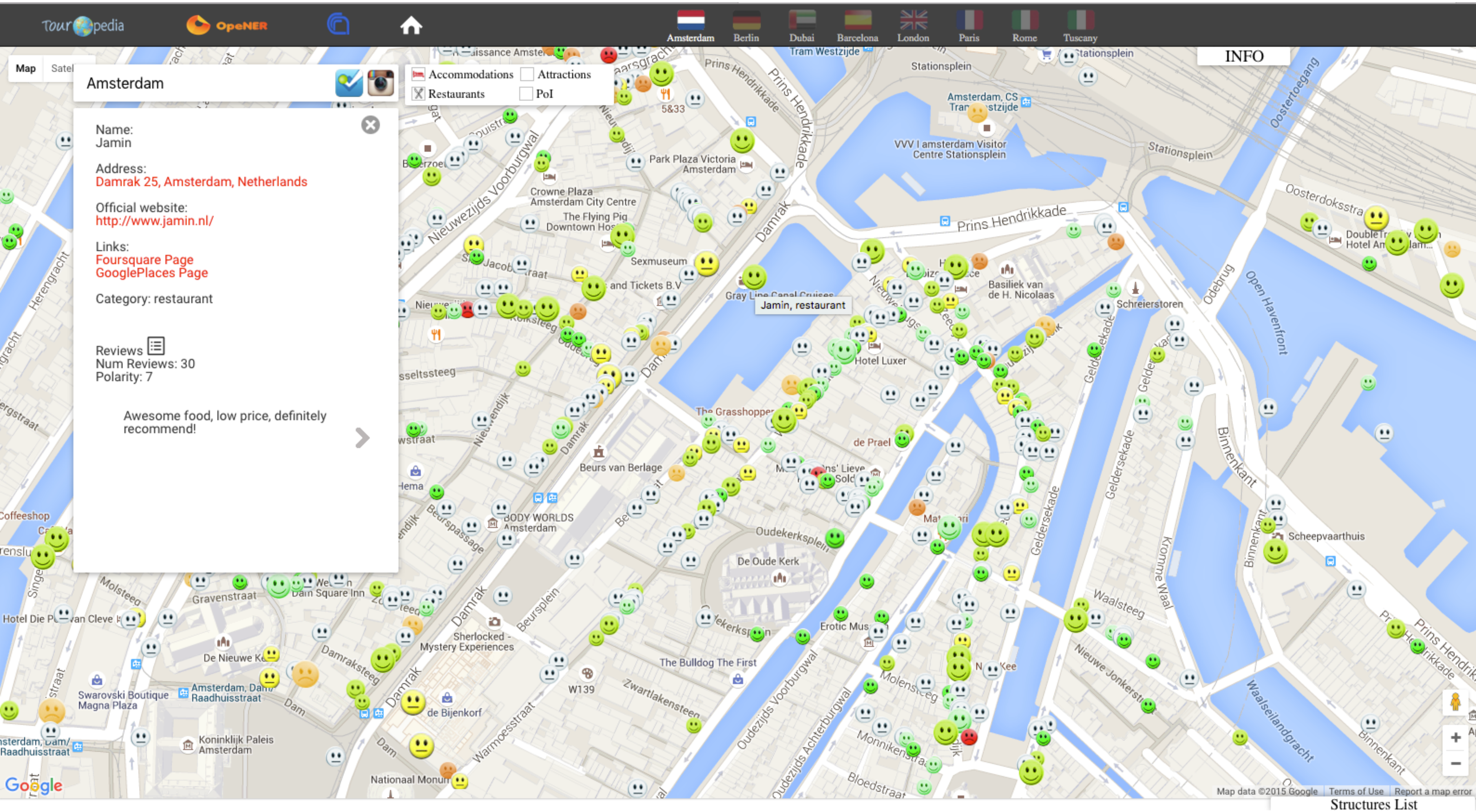
- Nothing special really. Comfortable and clean but very boring decor in comparison to **other NH hotels**. I stayed in **NH** in **Brussels** and **Zurich** and I really liked **them** because of **their** modern and stylish design and big rooms. **This one** was just like any **other** hotel. Basic rooms with basic and dull decor - bit disappointing. The customer service was average. The rate was very expensive and I still had to pay for Internet and **20 euros** for breakfast!!! It was good but way overpriced! The best thing about **the hotel** was **the location** - city centre, 2min from a metro stop.

Scale 1 (negative) to 5 (positive)

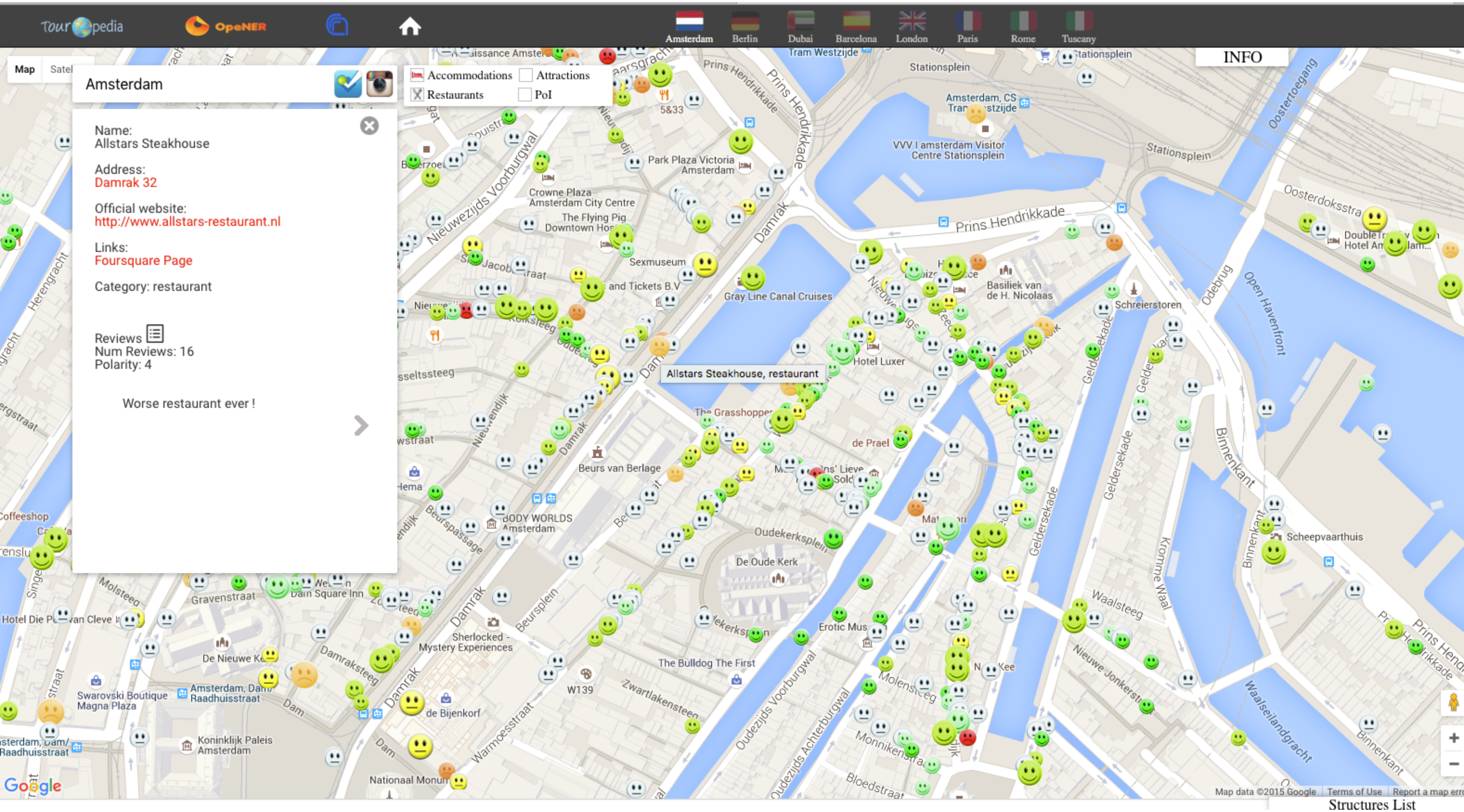
	Overall	Design	Room	Service	Price	Location	Transport
This NH	2	1,5	2	3	1	4	4
NH Brussels	4	4	4				
NH Zurich	4	4	4				



http://tour-pedia.org/about/index.html



Tourpedia
OpeNER
Amsterdam
Berlin
Dubai
Barcelona
London
Paris
Rome
Tuscany
Map
Amsterdam
Accommodations
Attractions
Restaurants
PoI
INFO
Name:
Jamin
Address:
Damrak 25, Amsterdam, Netherlands
Official website:
http://www.jamin.nl/
Links:
Foursquare Page
GooglePlaces Page
Category: restaurant
Reviews
Num Reviews: 30
Polarity: 7
Awesome food, low price, definitely recommend!
Jamin, restaurant
Structures List



Tour pedia
OpeNER
Amsterdam
Berlin
Dubai
Barcelona
London
Paris
Rome
Tuscany
Map
Amsterdam
Accommodations
Attractions
Restaurants
PoI
INFO
Name:
Allstars Steakhouse
Address:
Damrak 32
Official website:
http://www.allstars-restaurant.nl
Links:
Foursquare Page
Category: restaurant
Reviews
Num Reviews: 16
Polarity: 4
Worse restaurant ever !
Allstars Steakhouse, restaurant
Structures List

Cross-lingual/cultural sentiment

- 2,486 expressions in hotel reviews
- annotated in 6 languages
- 8 aspect groups (food, clean, price, behaviour, general evaluation, size, location, noise, size),

GENERAL-EVALUATION 669		
en	definitely recommend;	strong positive
es	recomendariamos sin duda alguna	positive
fr	recommander chaleureusement	strong positive
it	consiglio vivamente	positive

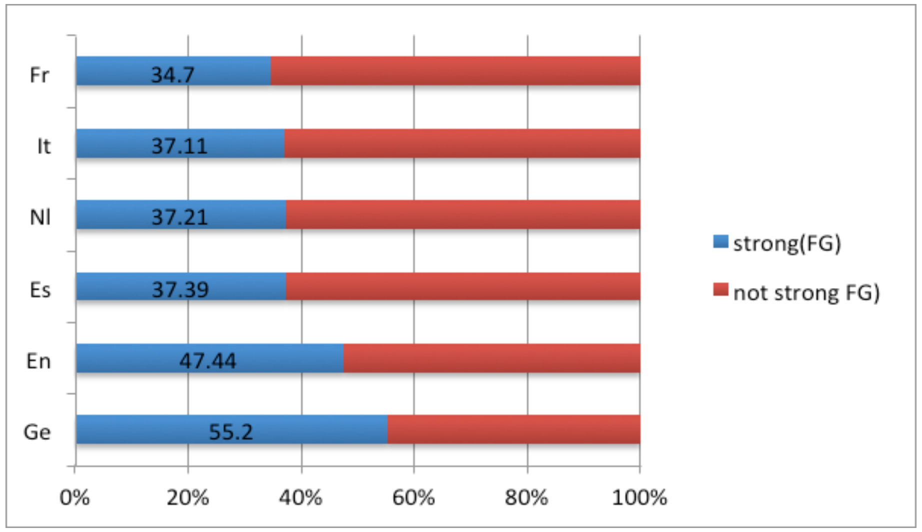
NOISE 1117		
en	a lot of noise	strong negative
fr	énormement de bruit	strong negative
it	molto rumore	negative
nl	veel lawaai	negative
nl	ontzettend veel lawaai	strong negative

http://opener-fp7project.rhcloud.com

Cultural Normalization

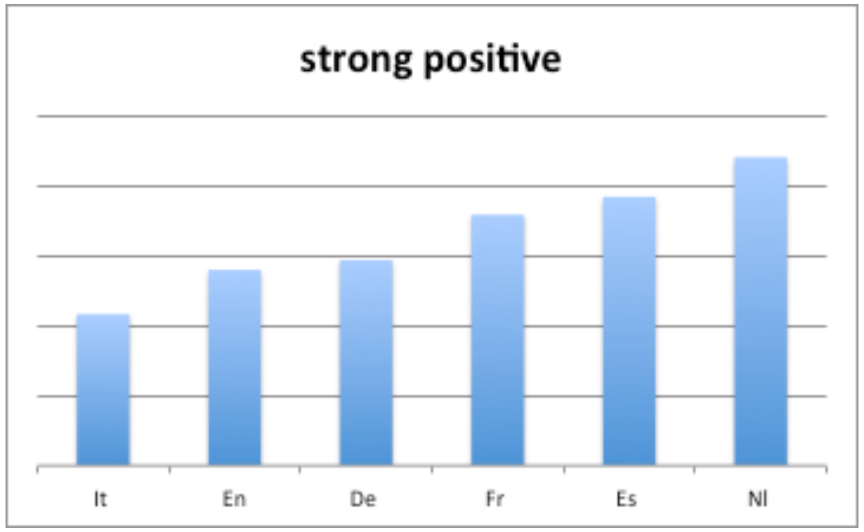
Differences in ratings in hotel reviews:

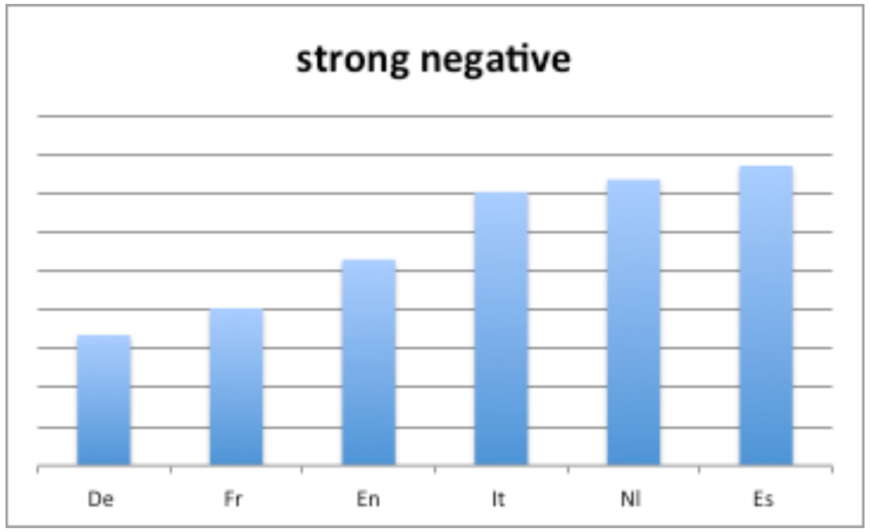
- English and Germans give more extreme (strong positive and strong negative) ratings than the other cultures



Differences in language use in hotel reviews:

- Dutch and Spanish are more expressive as they show a relative over-use of strong expressions
- German and English are less expressive as they show a relative under-use of strong expressions





=> People from German and English culture are quite strong when expressing their emotions in ratings, but they seem to downgrade their emotions when expressing them with language

Reviewer & reader ratings

- The hotel seems rather outdated. The breakfast room is just not big enough to cope with the Sunday-morning crowds.


Reviewer & reader ratings

- The hotel seems rather outdated. The breakfast room is just not big enough to cope with the Sunday-morning crowds.
- Maks and Vossen (RANLP-2013)

Review rating
7

Reader rating
negative

Target= hotel, holiday

Target= aspects in the text

9% - 37% sentiment mismatch at document level