Sentiment Analysis: modern methods, resources and applications

Davide Buscaldi,
LIPN - Université Paris 13

davide.buscaldi@lipn.univ-paris13.fr

Cagliari
16/05/2017 - 23/05/2017

Plan - day 2

- Fine-grained sentiment analysis
- Recent advances:
  - Deep learning and Sentiment Analysis
  - Machine reading and Sentiment Analysis
    - Lived Experiences
- Detecting fake reviews
- Afternoon: discussion on day 1 experiences and final test

Fine-grained Sentiment Analysis

Fine-grained Sentiment Analysis

- Levels of detail:
  - Assigning a polarity to a full text vs. assigning a polarity to specific items (**aspects**)
  - Assigning a binary polarity vs. assigning an **intensity** or strength of the opinion
  - Assigning a positive/negative score vs. detailing different types of emotions in the positive/negative spectrum

Aspects

DONE IT AGAIN

By brentalexis on 14 May 2017

Size: 128 GB | Style: iPhone 7 Plus | Colour: Jet Black | **Verified Purchase**

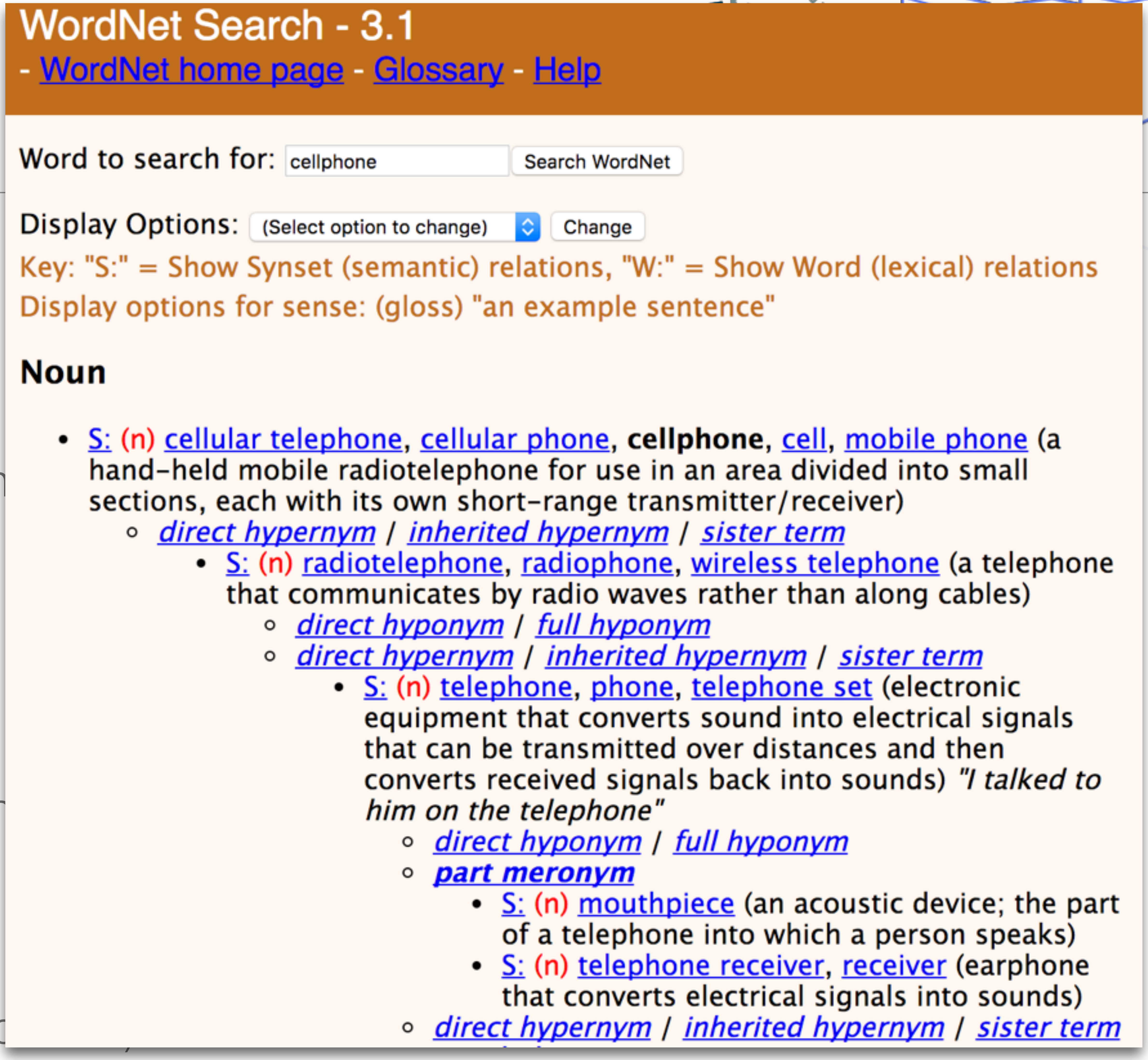
Firstly let me give credit to Amazon for their excellent service. I went into a well known store which is on my driving route, thinking that I could just walk away with a newly purchased iPhone 7, but to be told that they did not have one and it would take two days to have one delivered to my home at a time that was inconvenient to me! On the same evening I went on to Amazon made my order and had it delivered to a local collection point the next day by 1pm. Brilliant !

What about the iPhone 7? Well I have had a iPhone 4, a couple of iPhone 5's and now a 7. Well for me it is progress and in particular I think that this is just the right screen size, especially as I am of a mature age and sometimes find the text size a bit challenging. Most things are in the right places and there are new features which I will consider setting up as i get more familiar. The finger print reader is a good idea, as I have had this on one of my laptops for many years. It is very easy to set up. The main negative for me is the price is perhaps a bit too high however Amazon was cheaper than the shop I first tried so do shop around. However I hope to use this for a few years and so will get good value that way. So if your ready for an upgrade or a new kid on the block, I recommend it

https://presidentielle2017.politoscope.org/dashboard

Open Issues

- Detecting aspects:
  - aspects vary from product to product
    - Using knowledge bases with *part-of* relations (n
  - aspects in free text: it can be almost everything
- Weighting aspects:
  - adjectives may mean different things depending on
    - *high price* vs. *high performance*
  - Polarity may not be clearly stated (e.g. political disc
  - Evaluating comparative opinions



WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: cellphone Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

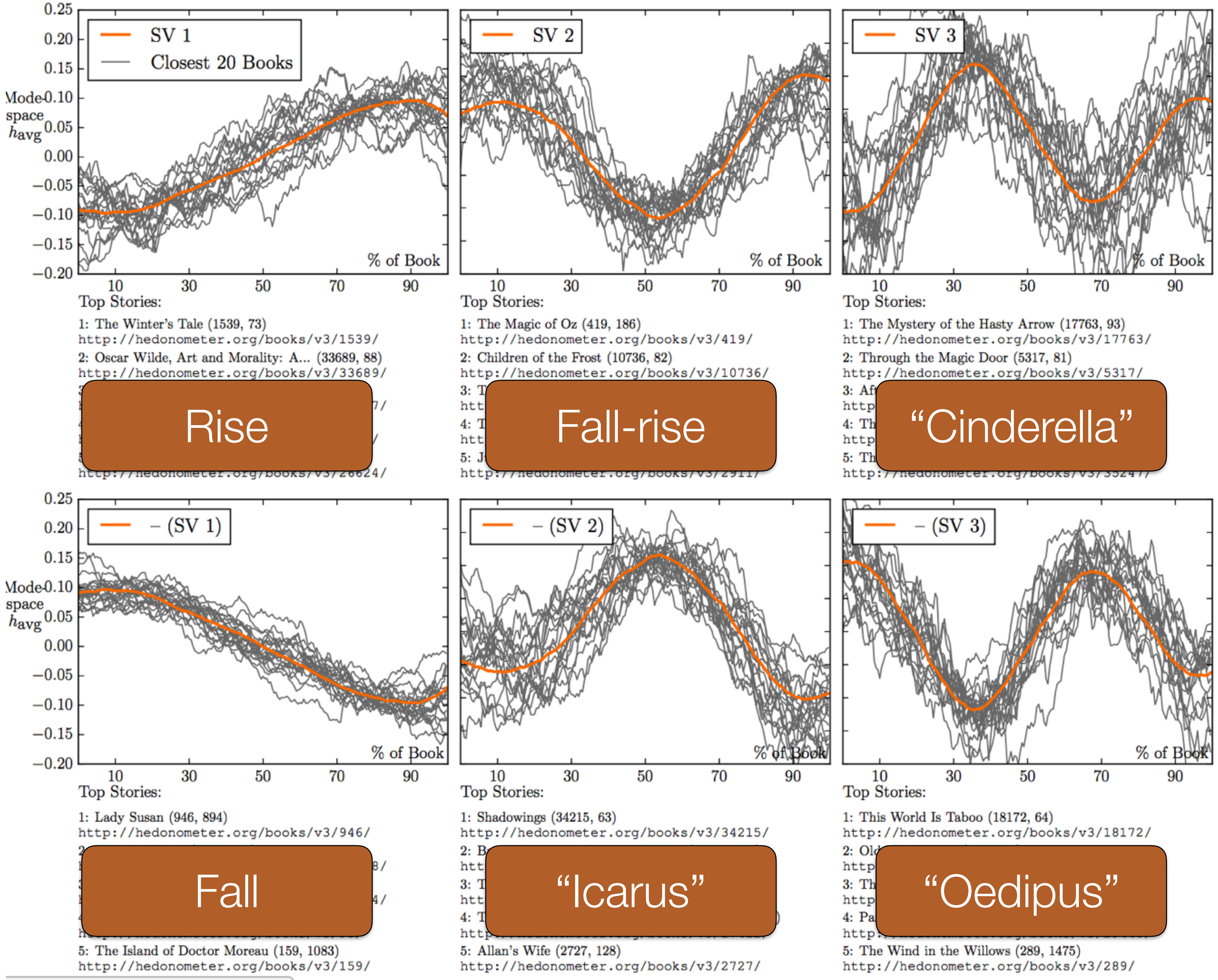
Noun

- S: (n) cellular telephone, cellular phone, **cellphone**, cell, mobile phone (a hand-held mobile radiotelephone for use in an area divided into small sections, each with its own short-range transmitter/receiver)
  - *direct hypernym* | *inherited hypernym* | *sister term*
    - S: (n) radiotelephone, radiophone, wireless telephone (a telephone that communicates by radio waves rather than along cables)
      - *direct hyponym* | *full hyponym*
      - *direct hypernym* | *inherited hypernym* | *sister term*
        - S: (n) telephone, phone, telephone set (electronic equipment that converts sound into electrical signals that can be transmitted over distances and then converts received signals back into sounds) *"I talked to him on the telephone"*
          - *direct hyponym* | *full hyponym*
          - ***part meronym***
            - S: (n) mouthpiece (an acoustic device; the part of a telephone into which a person speaks)
            - S: (n) telephone receiver, receiver (earphone that converts electrical signals into sounds)
          - *direct hypernym* | *inherited hypernym* | *sister term*

Opinion/emotion strength

- How to quantify the strength of an opinion?
    - A possibility is to use dictionaries that provide a score and not just polarity
    - Still a problem with negation and complex expressions

	labMT
superb	7.68
amazing	7.66
excellent	8.00
good	7.20
not bad	2.64 (bad)
I’ve seen better	7.00 (better)



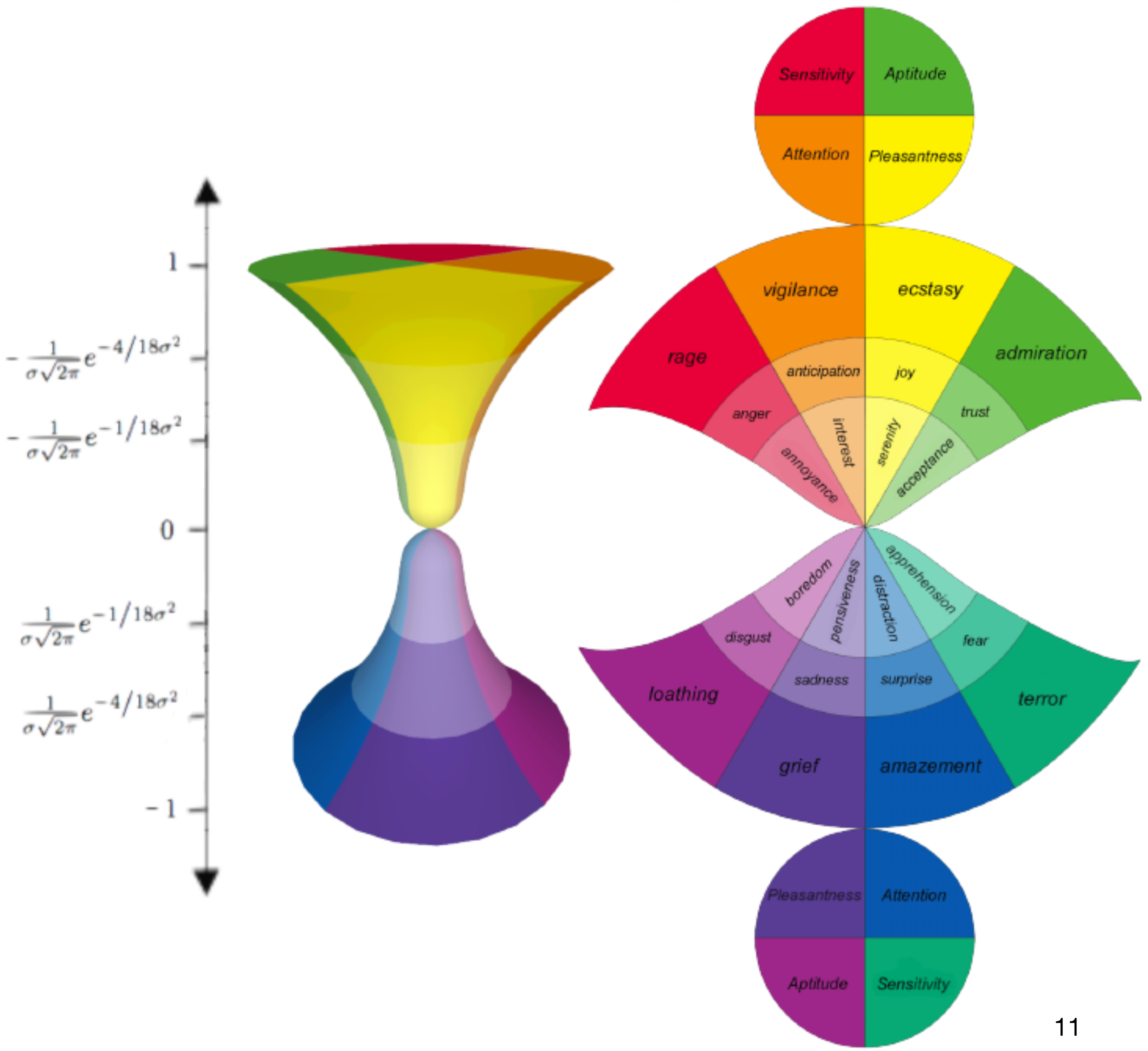
- Interesting findings are still possible with the limited analysis tools we have
- Example: *plotting avg. happiness in proj. Gutenberg books*
- (from Andy Reagan, hedonometer.org project)
- More emotions, more dimensions to analyze text

SentiWordNet

- Home page: http://sentiwordnet.isti.cnr.it/
- All WordNet *synsets* automatically annotated for degrees of positivity, negativity, and neutrality/objectiveness
  - [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

sentic.net

- http://sentic.net/
- 50,000 concepts with an intensity assigned
- -1 is extreme negativity and +1 is extreme positivity
- Different “axes” of emotions:
    - rage <-> terror
    - admiration <-> loathing
    - ecstasy <-> grief
    - vigilance <-> amazement



NRC Affect Intensity Lexicon

- Home page: http://saifmohammad.com/WebPages/AffectIntensity.htm
- Similarly to sentic.net , different emotions:
  - anger, fear, joy, sadness
  - real values from 0 (not associated) to 1 (strongly associated)
- Around 6000 entries
- Emotion Intensity task at WASSA2017
  - http://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html

NRC Affect Intensity Lexicon

Word	Anger	Word	Fear	Word	Joy	Word	Sadness
outraged	0.964	*horror*	0.923	*sohappy*	0.868	*sad*	0.844
brutality	0.959	*horrified*	0.922	*superb*	0.864	*suffering*	0.844
satanic	0.828	*hellish*	0.828	*cheered*	0.773	*guilt*	0.750
hate	0.828	*grenade*	0.828	*positivity*	0.773	*incest*	0.750
violence	0.742	*strangle*	0.750	*merrychristmas*	0.712	*accursed*	0.697
molestation	0.742	*tragedies*	0.750	*bestfeeling*	0.712	*widow*	0.697
volatility	0.687	*anguish*	0.703	*complement*	0.647	*infertility*	0.641
eradication	0.685	*grisly*	0.703	*affection*	0.647	*drown*	0.641
cheat	0.630	*cutthroat*	0.664	*exalted*	0.591	*crumbling*	0.594
agitated	0.630	*pandemic*	0.664	*woot*	0.588	*deportation*	0.594
defiant	0.578	*smuggler*	0.625	*money*	0.531	*isolated*	0.547
coup	0.578	*pestilence*	0.625	*rainbow*	0.531	*unkind*	0.547
overbearing	0.547	*convict*	0.594	*health*	0.493	*chronic*	0.500
deceive	0.547	*rot*	0.594	*liberty*	0.486	*injurious*	0.500
unleash	0.515	*turbulence*	0.562	*present*	0.441	*memorials*	0.453
bile	0.515	*grave*	0.562	*tender*	0.441	*surrender*	0.453
suspicious	0.484	*failing*	0.531	*warms*	0.391	*beggar*	0.422
oust	0.484	*stressed*	0.531	*gesture*	0.387	*difficulties*	0.421
ultimatum	0.439	*disgusting*	0.484	*healing*	0.328	*perpetrator*	0.359
deleterious	0.438	*hallucination*	0.484	*tribulation*	0.328	*hindering*	0.359

WASSA2017 Emotion Intensity Task

```
30001	Oh dear an evening of absolute hilarity I don't think I have laughed so much in a long time! 😂 joy 0.958
30002	Been waiting all week for this game ❤❤❤ #cheer #friday ❤	joy 0.940
30003	@gardiner_love : Thank you so much, Gloria! You're so sweet, and thoughtful! You just made my day more joyful! I love
30004	I feel so blessed to work with the family that I nanny for ❤ nothing but love &amp; appreciation, makes me smile.	jo
30005	Today I reached 1000 subscribers on YT!! , #goodday, #thankful	joy 0.926
30006	@Singaholic121 Good morning, love! Happy first day of fall. Let's make some awesome #autumnmemories #annabailey #laugh
30007	#BridgetJonesBaby is the best thing I've seen in ages! So funny, I've missed Bridget! #love	#TeamMark	joy 0.922
30008	Just got back from seeing @GaryDelaney in Burslem. AMAZING!! Face still hurts from laughing so much	joy 0.920
30009	@IndyMN I thought the holidays could not get any more cheerful, and then I met you. #TheNiceBot joy 0.917
30010	I'm just still . So happy .\nA blast	joy 0.917
30011	It's meant to be!! #happy #happy	joy 0.917
30012	💥Yeah!! PAUL!!💥	#glorious #BB18	joy 0.917
30013	My morning started off amazing!! Hopefully the whole day is going as i want it to go!\n #GreatDay	joy 0.917
30014	😱 @cailamarsai you've had me 😂 😂 the whole time watching @black_ishABC after you've lost your #glasses! It was #hila
30015	@iamTinaDatta love you so much #smile 😊😊	joy 0.896
30016	@WyoWiseGuy @LivingVertical however, REI did offer me the job today as well. Can't believe how exponentially freaking
30017	2 days until #GoPackGo and 23 days until #GoGipeGo..... I'm so excited! #smiling	joy 0.880
30018	@TheMandyMoore You are beyond wonderful.  Your singing prowess is phenomenal but damn... I'm just elated to watch you
30019	@luckiiCHARM_ Luckii, I'm changing in so many ways bc of Him!! It's a scary but joyful feeling, making me so strong.
30020	@JoshNoneYaBiz I love parody accounts! Well done. Vote for #Trump. #lol #hilarious	joy 0.875
30021	When you wake up from a dream laughing at something stupid, and that makes you laugh more #hilarious	joy 0.875
30022	now that I have my future planned out, I feel so much happier #goals #life #happy #igotthis #yay	joy 0.875
30023	Online now !!!:) all day come play with me !! I'm happy happy horny playful sweet sour;)	joy 0.875
30024	@grahnort wonderful experience watching you yesterday at. @BBCLetItShine thankyou for the #laughter joy 0.872
30025	@itsyourgirl_Z happy birthday :) have a blessed day love from Toronto :) #bday	joy 0.868
30026	@WSJNordics You make the world a more joyful place. #TheNiceBot joy 0.864
```

DepecheMood

- Staiano, J., & Guerini, M. (2014). "DepecheMood: a Lexicon for Emotion Analysis from Crowd-Annotated News". Proceedings of ACL-2014.
  - https://github.com/marcoguerini/DepecheMood/releases
  - Demo: http://www.depechemood.eu/DepecheMood.html
- 8 emotions:
  - Afraid, amused, angry, annoyed, don't care, happy, inspired, sad
- 37770 word#POS pairs
- Less structured than sentic.net

DepecheMood

Lemma#PoS	AFRAID	AMUSED	ANGRY	ANNOYED	DONT_CARE	HAPPY	INSPIRED	SAD
abandon#n	0.209483077	0.125982808	0.130890532	0.072843713	0.130934074	0.118185311	0.110374289	0.101306195
abandon#v	0.113040387	0.139867299	0.121110886	0.144178754	0.119478846	0.089469002	0.127728865	0.145125962
abandoned#a	0.265976822	0.145815915	0.137806835	0.075302352	0.045844968	0.066166939	0.117848934	0.145237235
abandonment#n	0.138557646	0.107824234	0.126134359	0.182632744	0.046223373	0.133494138	0.182890371	0.082243135
abase#v	0.01785788	0.329667643	0.014215326	0.390906879	0.205607479	0.009589575	0.011808594	0.020346624
abate#v	0.35330477	0.125446177	0.064422149	0.049593322	0.026719391	0.179533436	0.043882925	0.157097831
abatement#n	0	0	0	0	0	0	0	1
abattoir#n	0	0.107486122	0.427845191	0.05079894	0	0.006421894	0.312361802	0.095086051
abaya#n	0.084147353	0.094263175	0.159432449	0.208539693	0.089018221	0.173648182	0.08908686	0.101864068
abb#n	0	0	0	0	0	0.408762703	0.591237297	0
abbess#n	0.014618277	0.090075734	0.139638481	0.084997971	0.012946788	0.090413546	0.567309204	0
abbey#n	0.023838954	0.267072902	0.014921008	0.055135702	0.294950326	0.162656496	0.146357558	0.035067054
abbot#n	0	0.400405963	0	0	0	0	0.599594037	0
abbreviate#v	0.064353583	0.079626721	0.520633943	0.120342781	0.077396996	0.10619352	0.006282595	0.02516986
abbreviated#a	0	0	0.15018768	0	0.586851272	0.083020778	0.124882525	0.055057745
abbreviation#n	0	0.198100971	0.021371991	0.305357094	0.302985158	0.009743161	0.162441623	0
abc#n	0.140420663	0.183108079	0.104657442	0.111712951	0.162196158	0.099893444	0.057754743	0.140256521
abcs#n	0	0	0	0	0	0.378520052	0.621479948	0
abdicate#v	0.021830494	0.090266704	0.097424626	0.118051999	0.155117201	0.051472619	0.097738718	0.36809764
abdication#n	0.076755149	0.113853787	0.000908661	0.166095535	0.366677227	0.051648735	0.102992579	0.121068326
abdomen#n	0	0.110856757	0.083037907	0	0.372935259	0.19250516	0.031696789	0.208968129
abdominal#a	0.209880087	0.017795553	0.048140442	0.014781564	0	0.059184056	0.426159769	0.22405853
abduct#v	0.165277173	0.044713348	0.240804314	0.08759287	0.063032584	0.101039934	0.070716991	0.226822786
abducting#a	0.093759496	0	0.493451984	0	0	0.261918248	0.116340506	0.034529766
abduction#n	0.209340365	0.024344219	0.230863548	0.080585028	0.103416394	0.084056756	0.083265993	0.184127698
abductor#n	0.214912518	0.033844829	0.224400153	0.080764046	0.043783342	0.046363579	0.059758504	0.296173029

Deep Learning and Sentiment Analysis

Deep Learning

- Standard machine learning: using human-designed features and representations

- Deep Learning algorithms: learn multiple levels of representation of increasing complexity/abstraction

Output layer

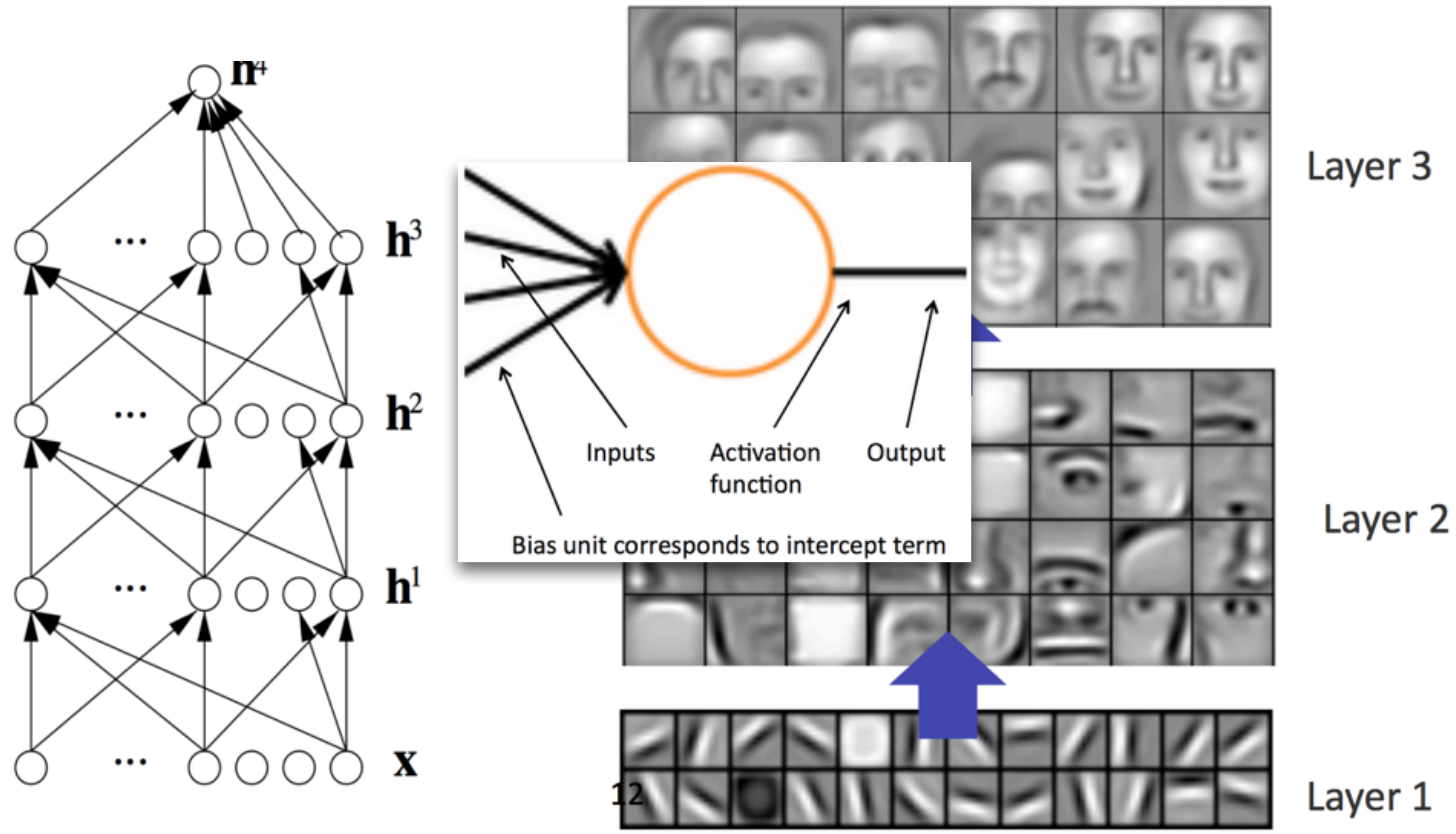
Here predicting a supervised target

Hidden layers

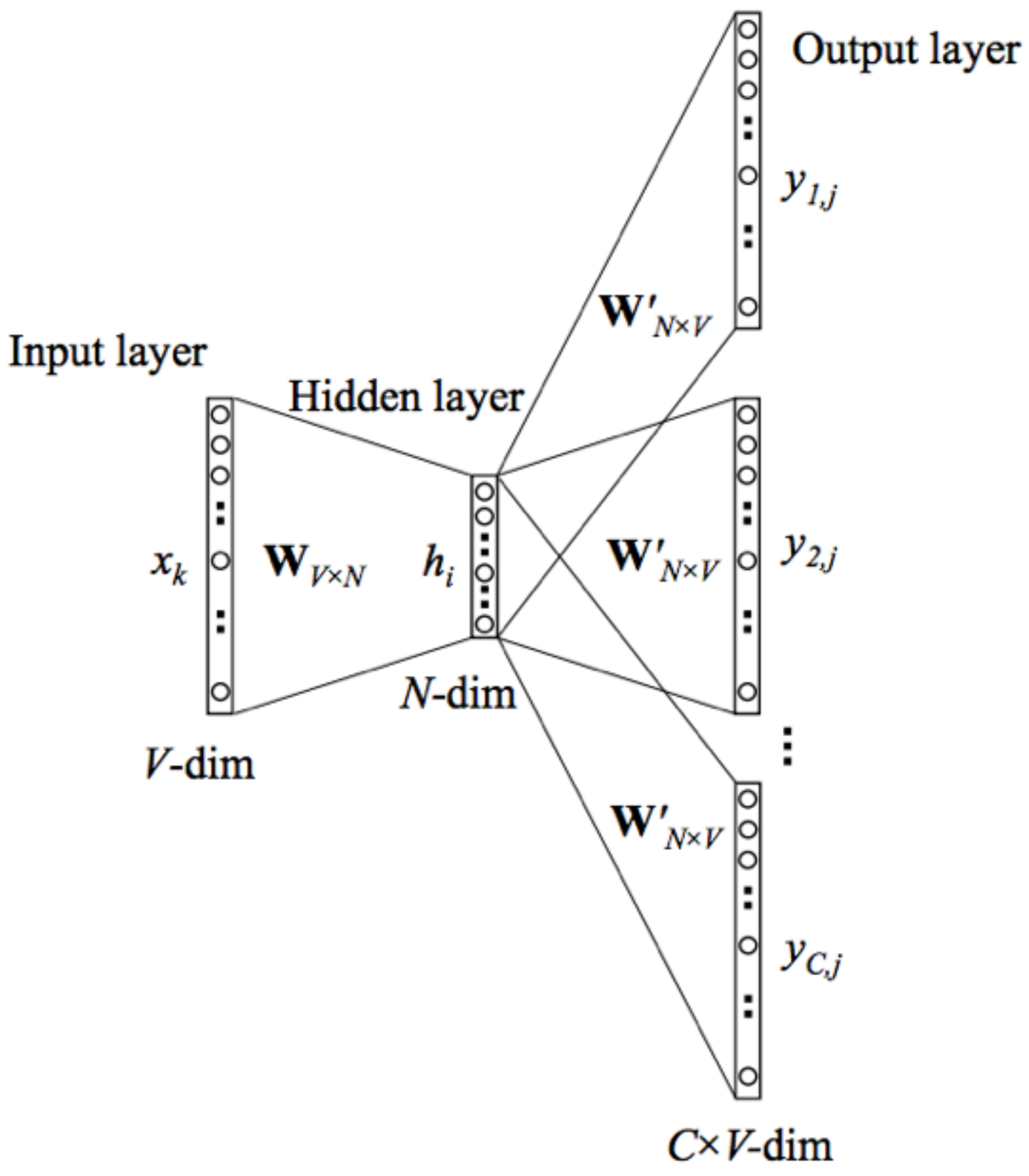
These learn more abstract representations as you head up

Input layer

3 Raw sensory inputs (roughly)



Deep Learning and text: word2vec



- Input vectors:
  - *One-hot representations*
  - V = (Monday, Tuesday, is, a, today)
    - “Monday Monday”= [2 0 0 0 0]
    - “today is a Monday” = [1 0 1 1 1]
    - “today is a Tuesday” = [0 1 1 1 1]
    - “Is a Monday today” = [1 0 1 1 1]
- The vectors in the matrix between the input and hidden layer are the *word vectors* (also known as *word embeddings*)

- Word2Vec: https://code.google.com/p/word2vec/
- Mikolov’s COLING tutorial Using Neural Networks for Modelling and Representing Natural Languages :
  - http://www.coling-2014.org/tutorials.php

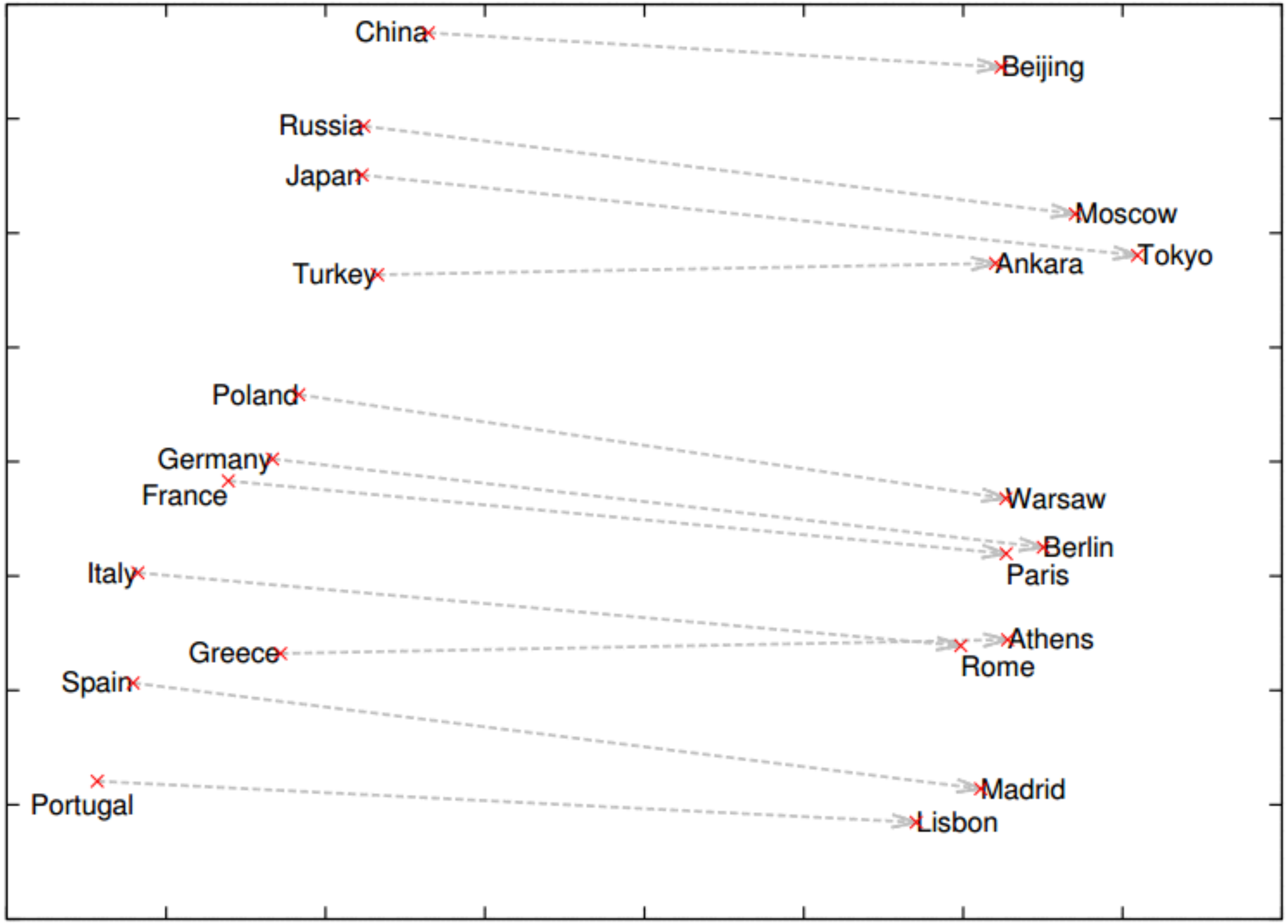
How to train a model

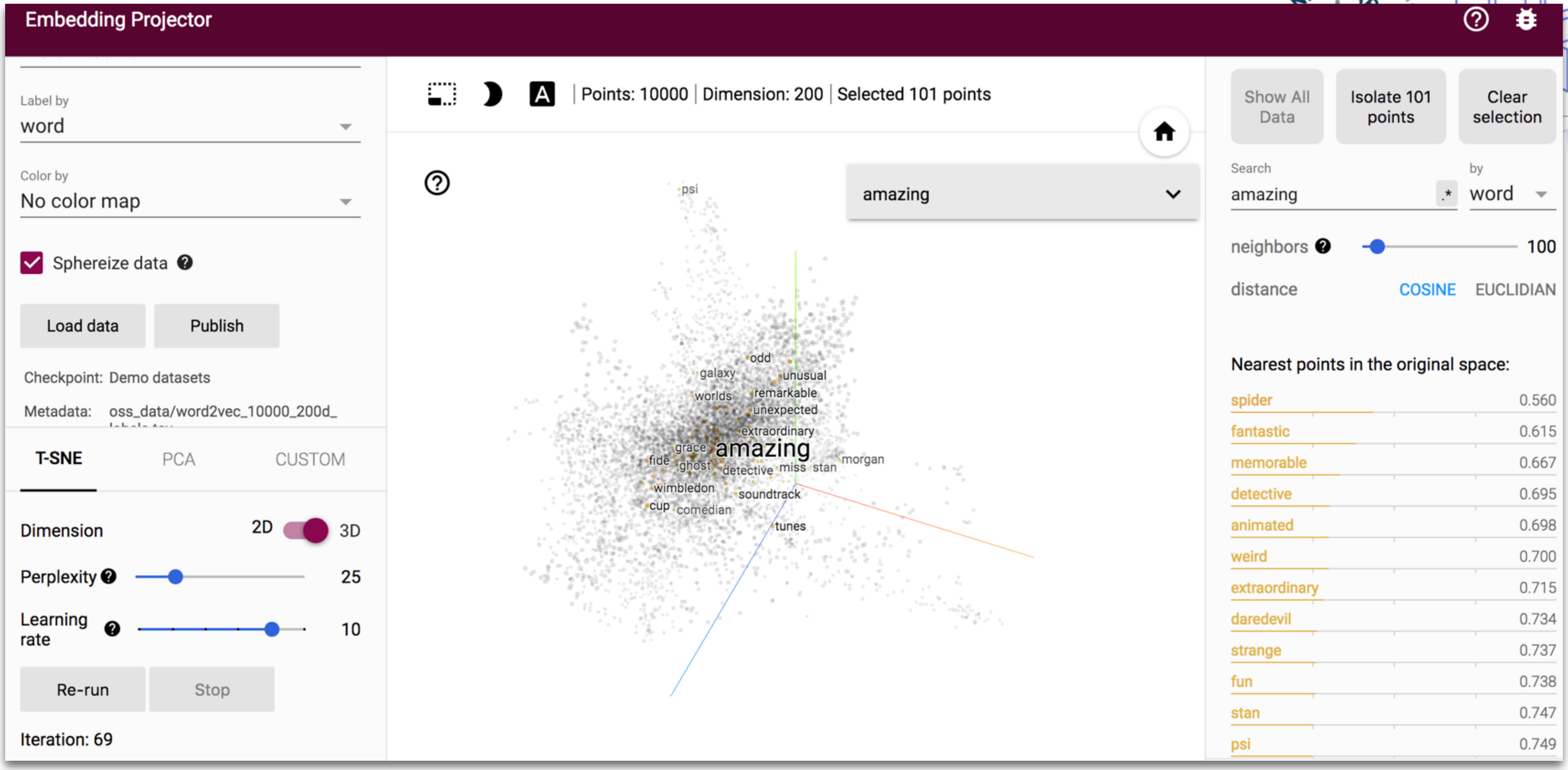
- Prepare a large enough corpus of text
    - Typical choices: full wikipedia, WaCky (http://wacky.sslmit.unibo.it/doku.php)
- Choose the size of vectors and the model (skip-grams, CBOW, etc.)
- train the model with **word2vec**, **GloVe**, or **fasttext**
    - **GloVe** (Global Vectors) is a similar implementation from Stanford University
        - https://nlp.stanford.edu/projects/glove/
    - **Fasttext** is an improved version of word2vec with character-level features
        - https://github.com/facebookresearch/fastText
- Use the resulting vectors as word representations

Some interesting properties

Expression	*Nearest token*
Paris - France + Italy	Rome
bigger - big + cold	colder
sushi - Japan + Germany	bratwurst
Cu - copper + gold	Au
Windows - Microsoft + Google	Android
Montreal Canadiens - Montreal + Toronto	Toronto Maple Leafs

Some interesting properties





http://projector.tensorflow.org/

Some pre-trained vectors

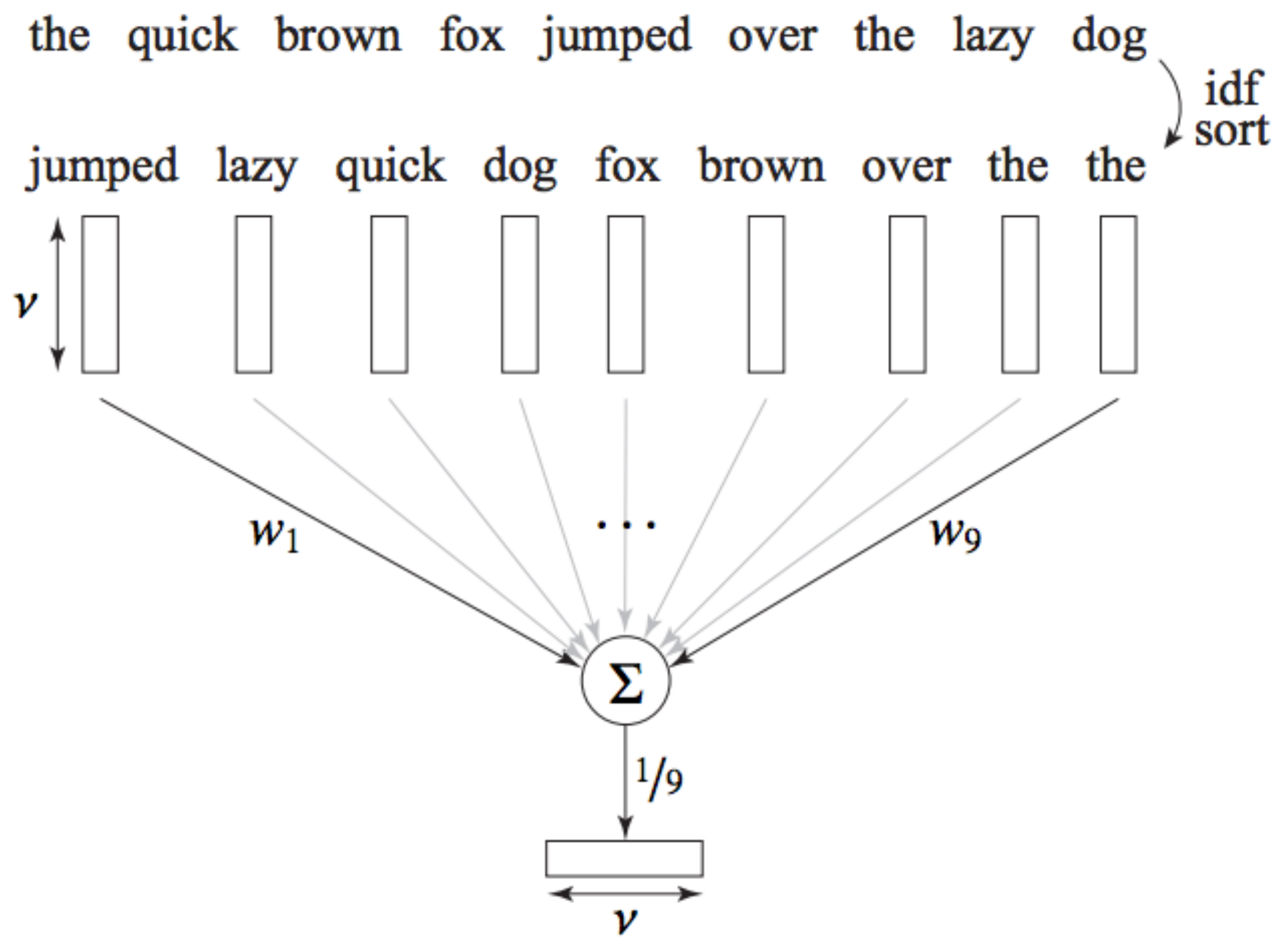
- The original pre-trained vectors trained on part of Google News dataset (about 100 billion words). The model contains 300-dimensional vectors for 3 million words and phrases
  - https://github.com/mmihaltz/word2vec-GoogleNews-vectors
- Fasttext vectors for 294 languages (Facebook research):
  - https://github.com/facebookresearch/fastText/blob/master/pretrained-vectors.md
- GloVe vectors including vectors from 2B twitter dump:
  - https://nlp.stanford.edu/projects/glove/

Consequences for Sentiment Analysis

- **Dimensionality reduction of the feature space**
  - Bag-of-words: thousands of features (number of words in the training sets)
  - Vectors: n features, usually $100 < n < 400$ (the dimensions of vector space)
- **Less OOV (out-of-vocabulary) words**
  - The size of the corpus to build the vector space is greater (by a wide margin) than most training sets for SA
- Bringing in external knowledge
  - The knowledge (word meaning) is acquired from the corpora in an unsupervised way

Semantic Vectors: from words to sentences

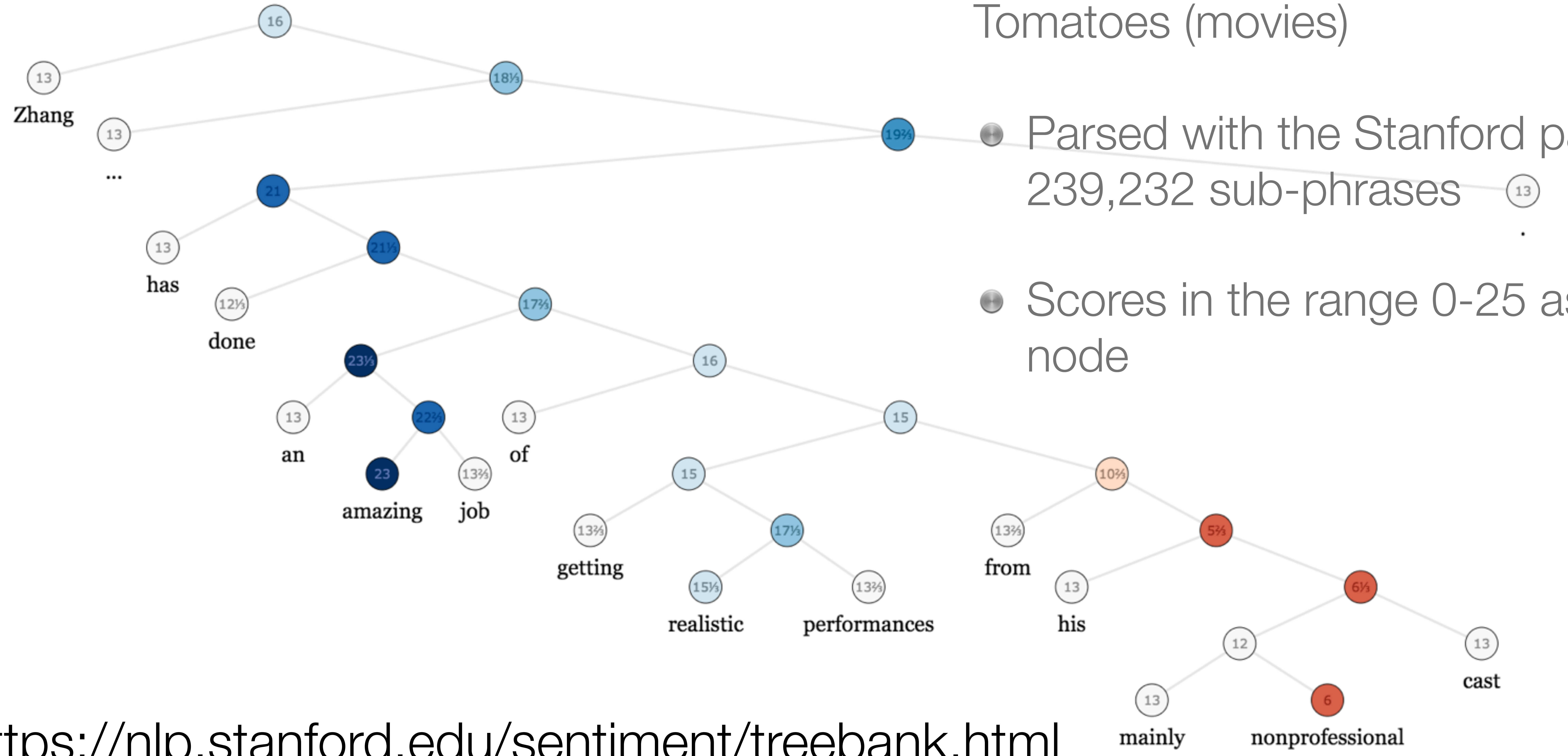
- One of the problems is to transform a sequence of word vectors into a sentence representation
  - Typical solutions:
    - Average of word vectors
    - max (per-dimension) of word vectors
    - min, max.min concatenation



- De Boom et al., “Representation learning for very short texts using weighted word embedding aggregation” (2016)

Stanford Sentiment Treebank

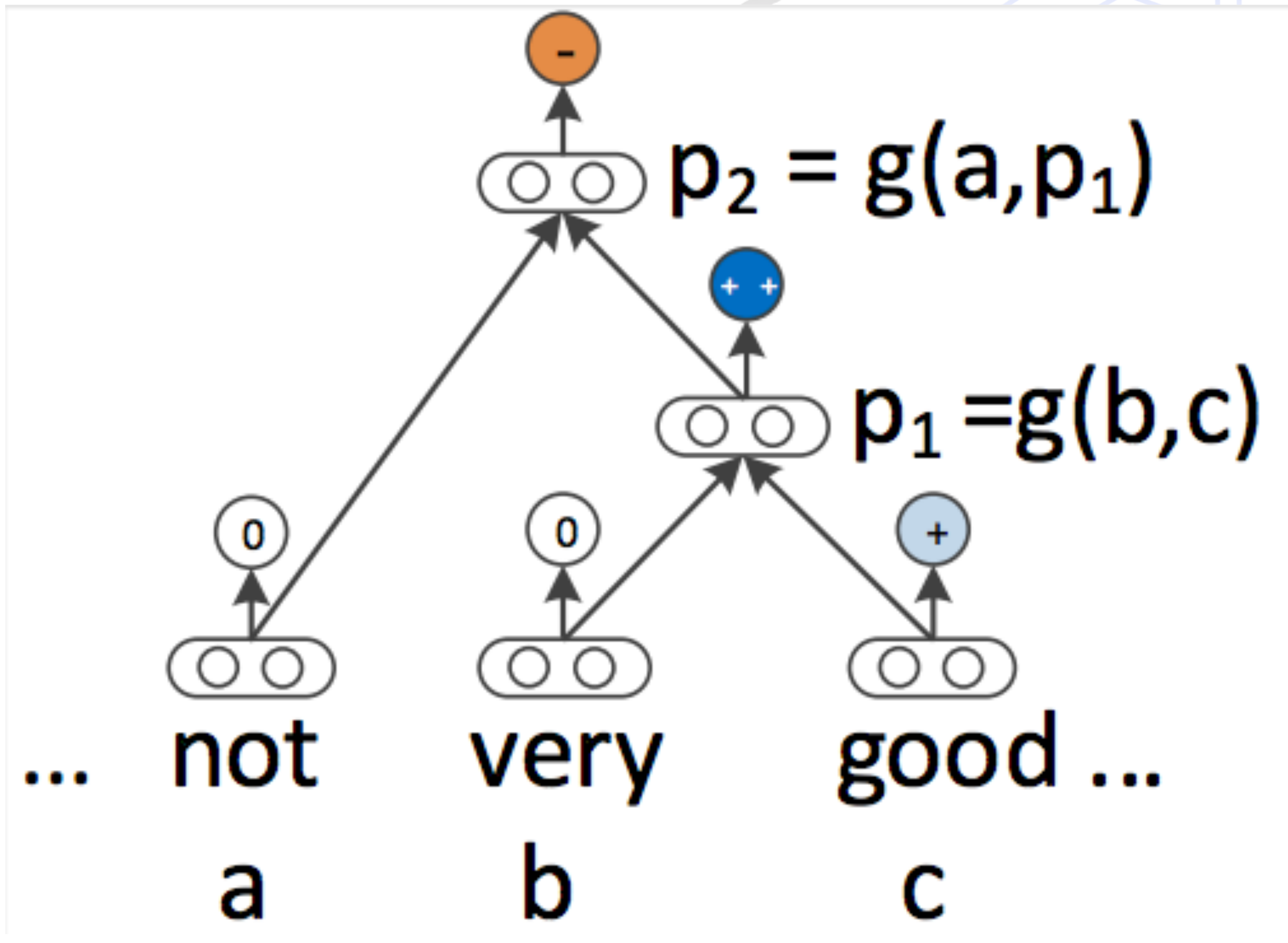
- 11,855 sentences taken from Rotten Tomatoes (movies)
  - Parsed with the Stanford parser into 239,232 sub-phrases
  - Scores in the range 0-25 assigned to node



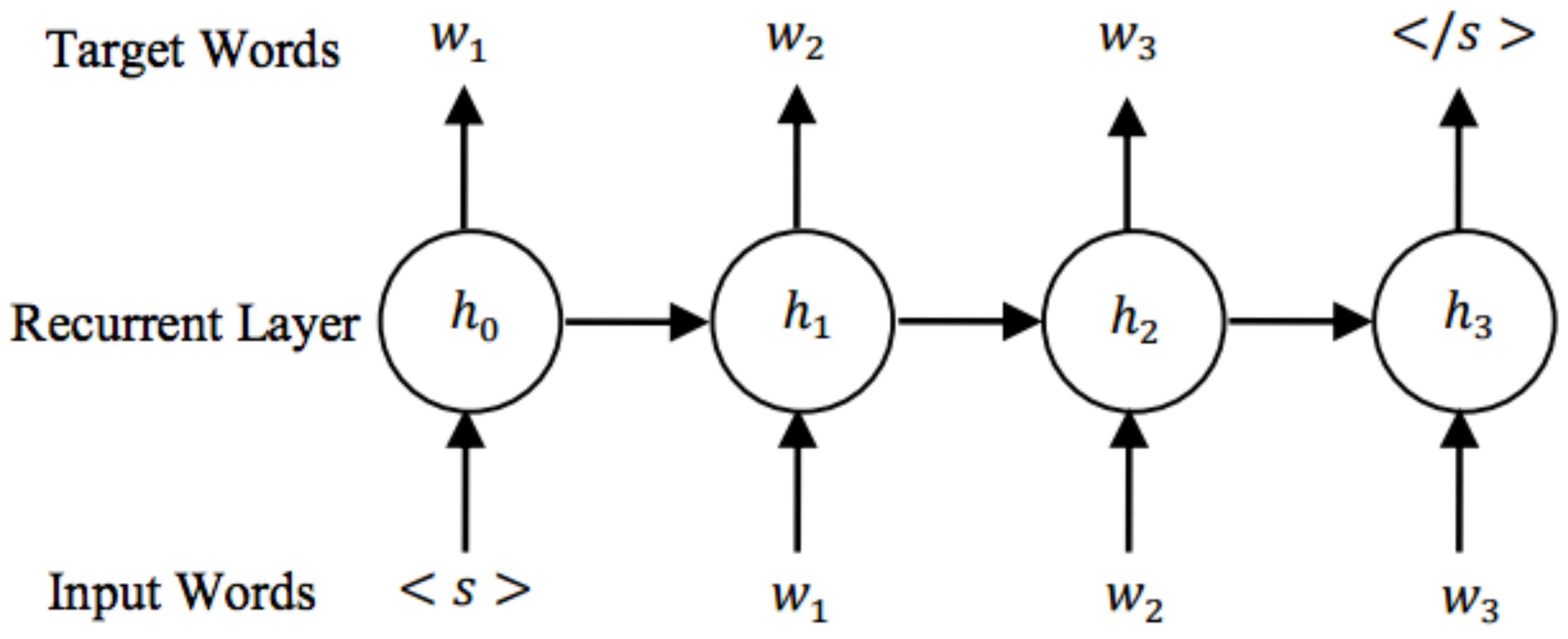
https://nlp.stanford.edu/sentiment/treebank.html

Some DL-based methods

- Socher et al., 2013: Recursive Neural Tensor Networks
  - Using node vectors to predict the polarity of parent nodes
  - Very effective on negations
  - ~90% accuracy on SST



- Mousa, Schuller, 2017: Long Short-Term Memories
  - DL sequential method to predict the class of a word given previous observations
  - More flexible than HMM and CRFs
  - ~90% accuracy on IMDB corpus



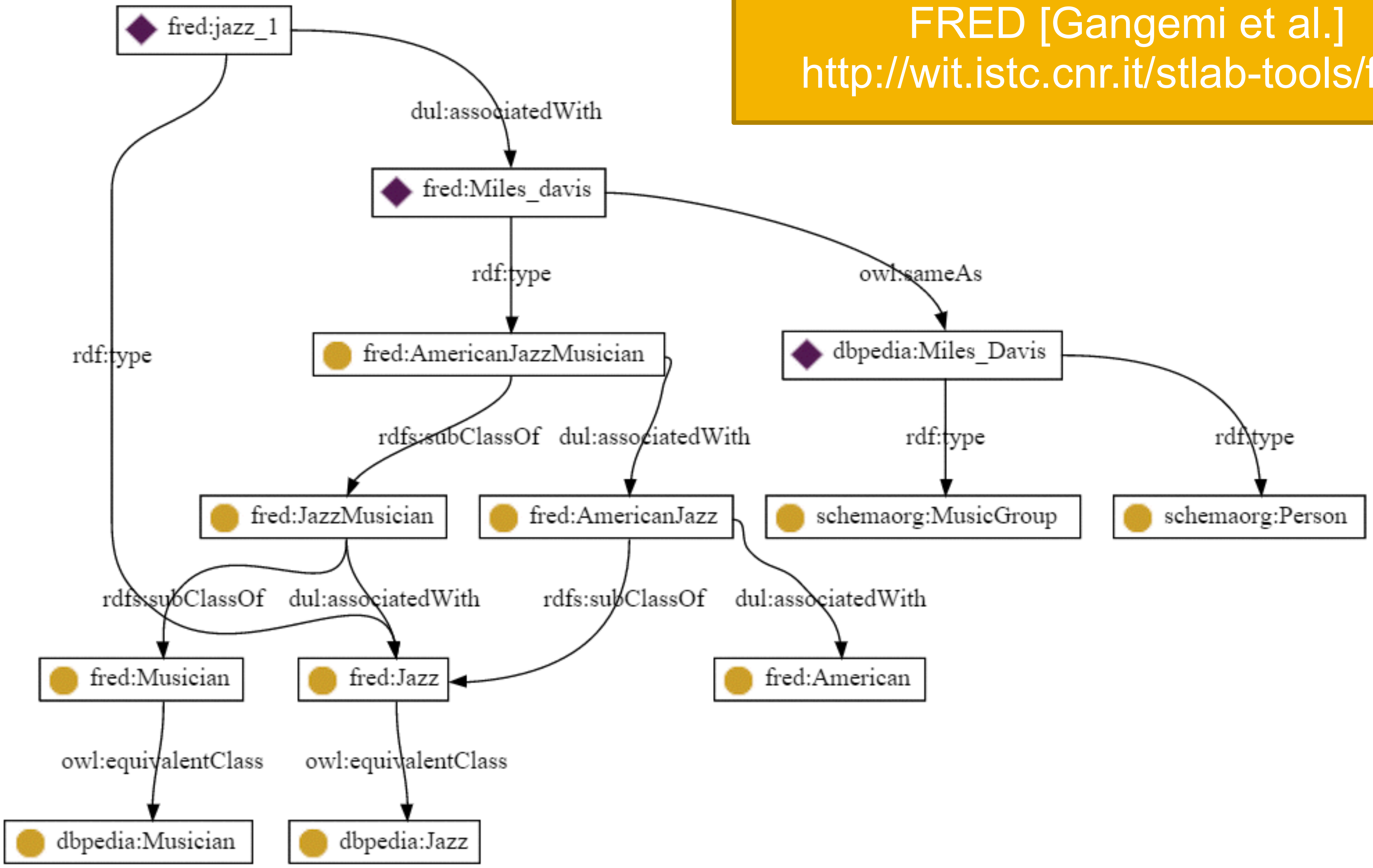
Machine Reading and Sentiment Analysis

Explicit Knowledge

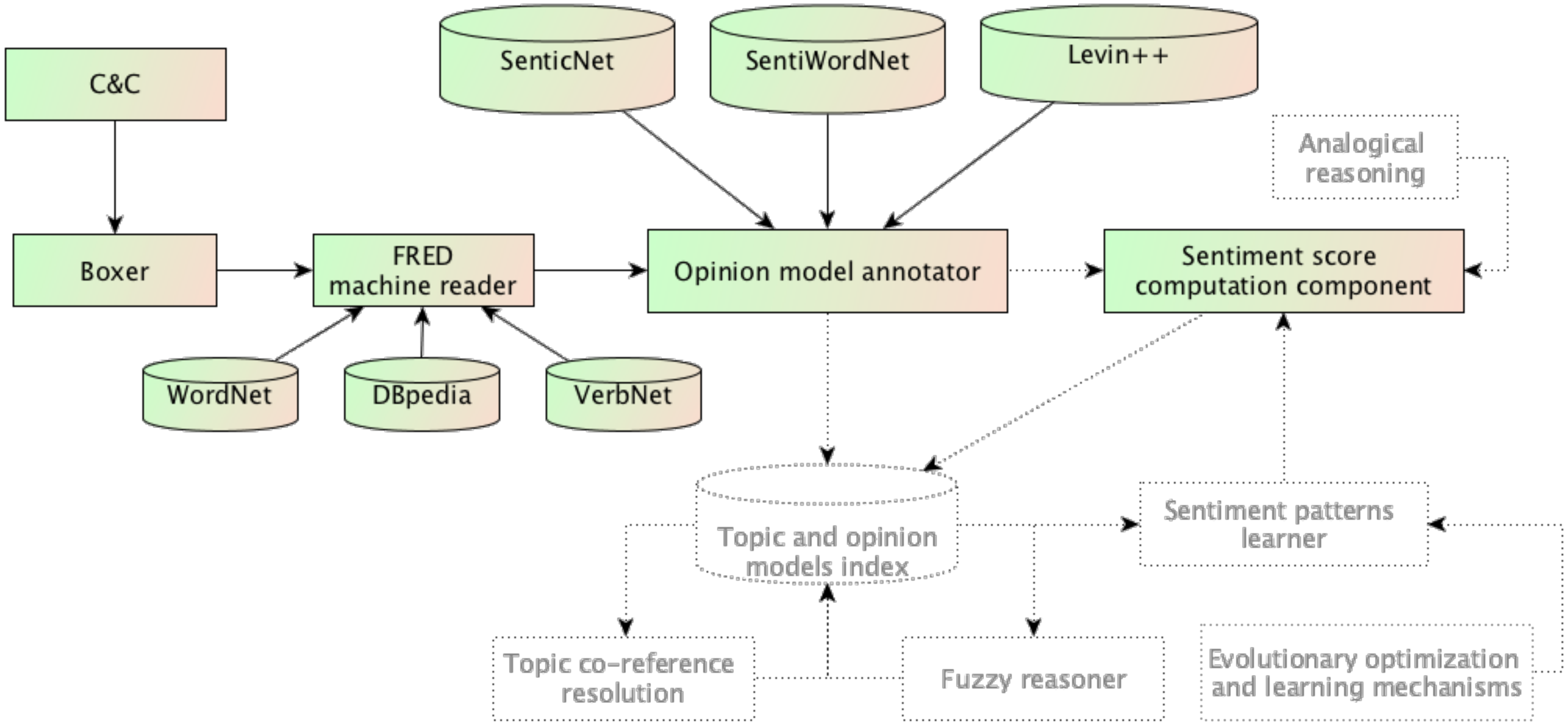
- DL and Semantic Vectors allow to introduce external knowledge into the Sentiment Analysis process
- There’s a more direct way to introduce such kind of knowledge: by accessing knowledge repositories
    - Need to link knowledge to words in the proper way
- *Machine Readers* perform this task

“Miles Davis was an American jazz musician”.

FRED [Gangemi et al.]
http://wit.istc.cnr.it/stlab-tools/fred/



Putting Sentiment into FRED: Sentilo



http://wit.istc.cnr.it/stlab-tools/sentilo/

Reforgiato Recupero et al., 2015 “Sentilo: Frame-Based Sentiment Analysis”

Sentilo tasks

- Holder identification: if the holder is explicitly mentioned in a sentence, it is identified and represented in a RDF graph;
- Topic detection: it identifies all topics that are in the scope of the expressed opinion and represents them as a RDF graph;
- Holders and topics resolution: it resolves holders and topics identities on the Semantic Web;
- Sentiment score computation: it calculates the sentiment score for each topic, considering also their inter-relationships and dependencies - if any

Example

- “Zhang did an amazing job of getting realistic performance from a mainly unprofessional cast.”

Example

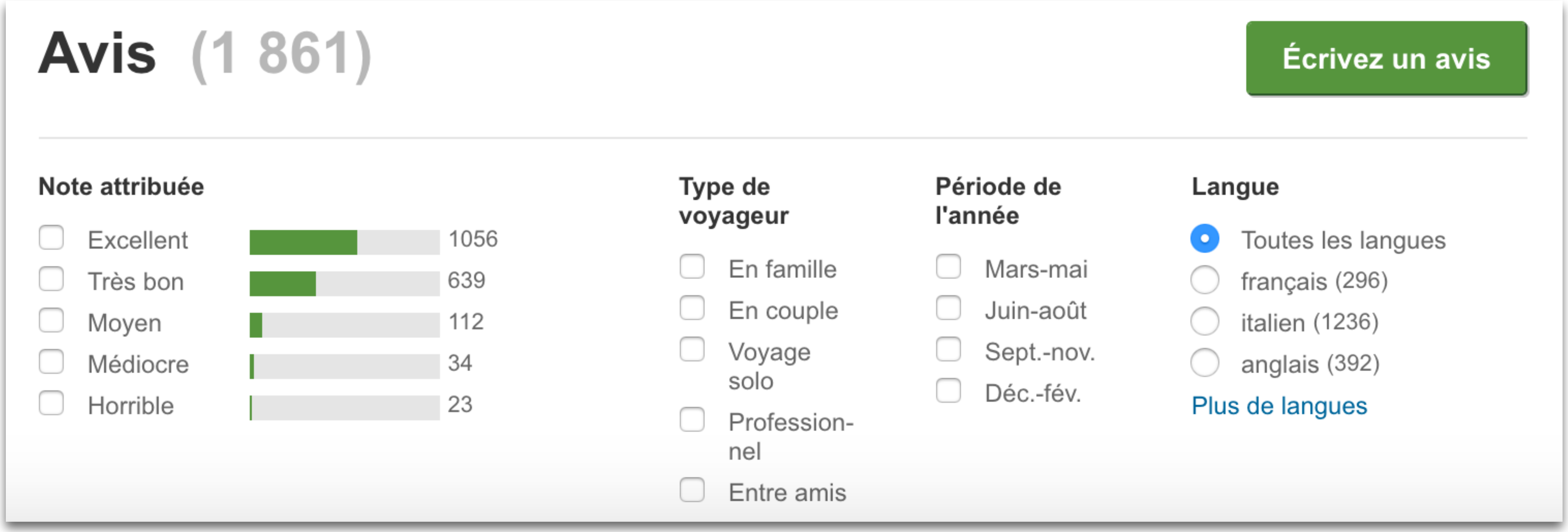

Sentilo
Powered by STLAB
Home
HowTo
Measure sentilometers of a sentence
Under the hood
Enter an opinion sentence to analyze and then click on the SENTILO button
Zhang did an amazing job of getting realistic performance from a mainly unprofessional cast.
SENTILO
Pos
0
0.2
0.4
0.6
0.8
1
Neg
0
0.2
0.4
0.6
0.8
1

A case study: Lived Experiences

Ehab Hassan Ph.D. Thesis

Advancing through the jungle of opinions

- Taking a decision about booking a service, or buying an item, is becoming more and more time-consuming



Advancing through the jungle of opinions

- Two needs:
  - Summarize judgments (from thousands of opinions to an overall view of the product/services)
  - Find opinions of like-minded customers, or in similar conditions
    - Tendance to trust more people who has similar experiences

Lived Experiences - a definition

- All experiences are related to “something that happened”
  - Consequence: Lived Experiences are a special case of **events**
- Tentative definition:
  - An event mentioned in a review, where the author is among the participants
- Examples:
  - I ate in the hotel restaurant / We slept very well
  - I arrived early and they had a room waiting
- Not LE:
  - The rooms were clean and large / The hotel staff was helpful

Example - Lived Experiences

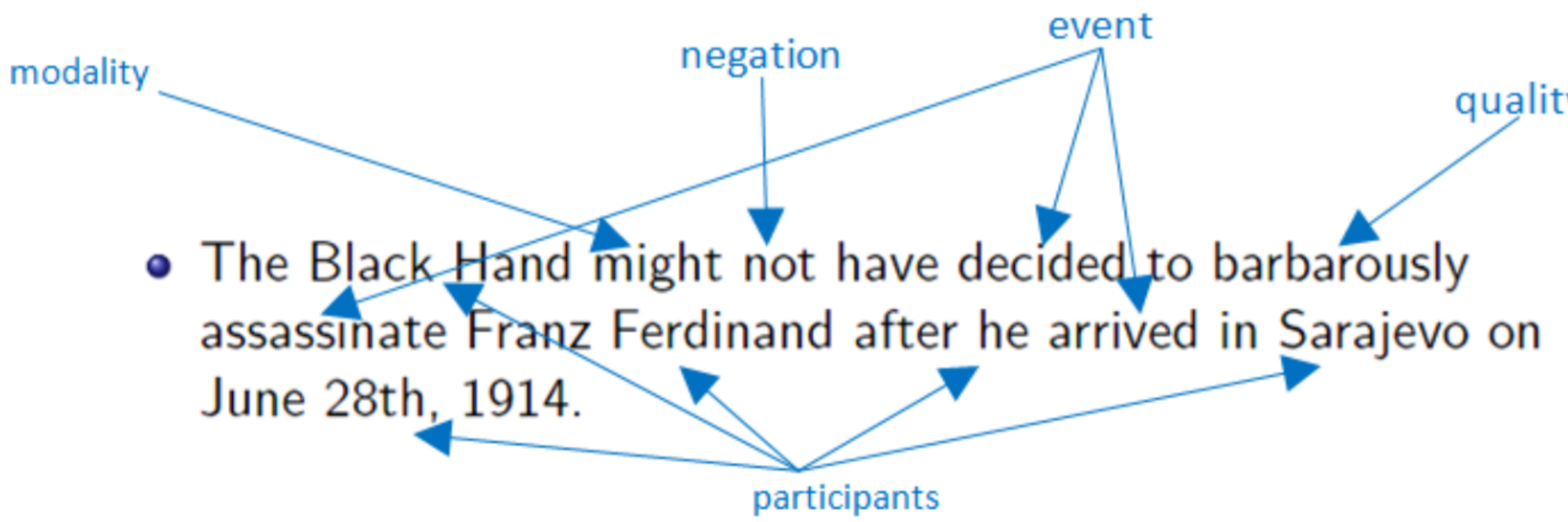
The view from this hotel's rooms is quite stunning. And that's what make it very special, possibly better than the next door 4 star hotel and than many other hotels in Paris. The bedrooms interior decor is extremely nice. (1) **I asked for a room overlooking the pantheon and I got it**. (2) **My deluxe room was number 32, and was tastefully decorated with a classic and beautiful Pierre Frey wallpaper, and an extra day bed**. The bath had bathtub-shower combination and was separated from the toilet. (3) **If you book directly through the hotel, you'll get a voucher for a free drink upon arrival**. It was a bit cold at night at some point, maybe because it's March and the heating is not constantly on anymore. Each room has its own heating control, though. Strongly recommended.

Example - not Lived Experiences

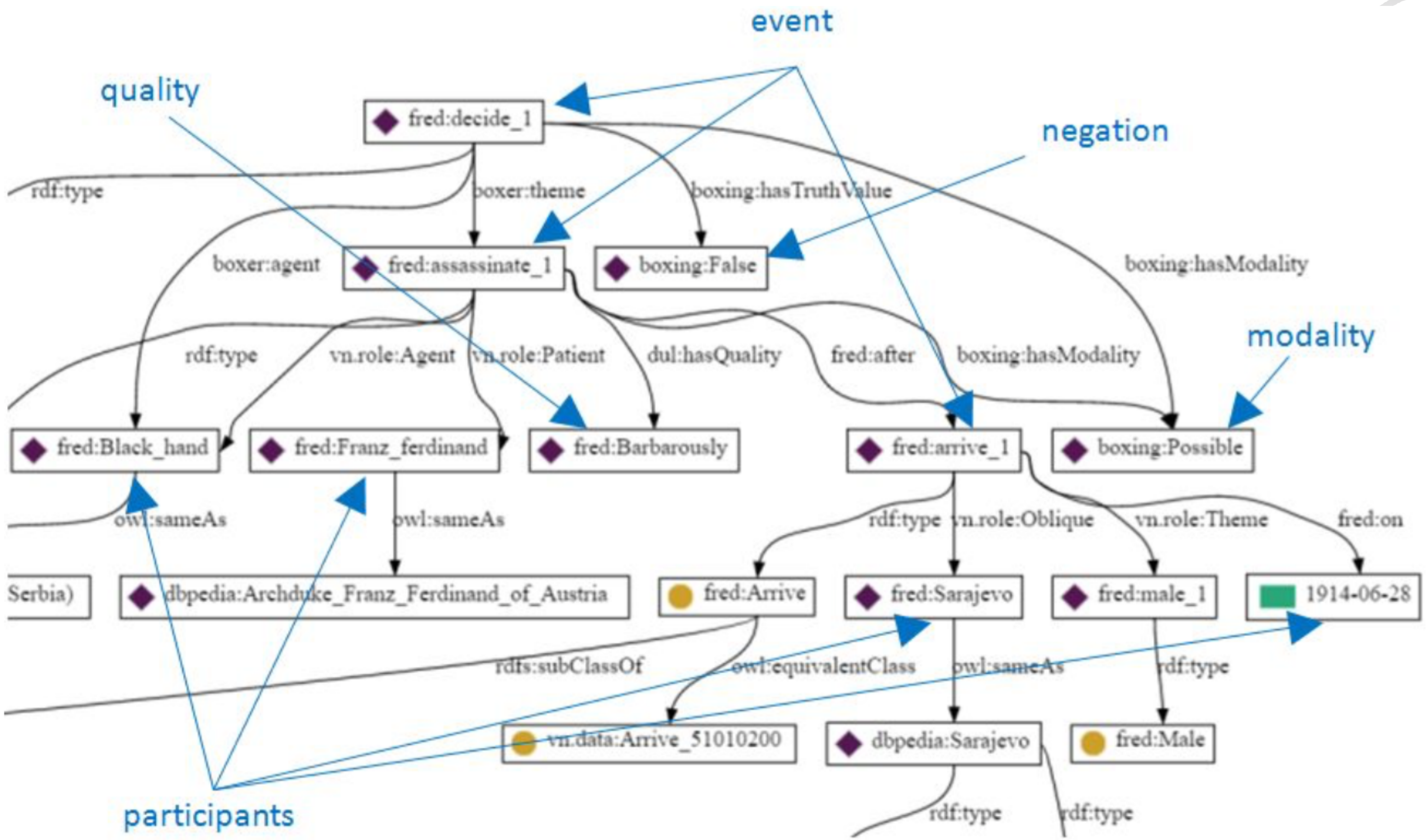
The view from this hotel's rooms is quite stunning. And that's what make it very special, possibly better than the next door 4 star hotel and than many other hotels in Paris. The bedrooms interior decor is extremely nice. I asked for a room overlooking the pantheon and I got it. My deluxe room was number 32, and was tastefully decorated with a classic and beautiful Pierre Frey wallpaper, and an extra day bed. The bath had bathtub-shower combination and was separated from the toilet. If you book directly through the hotel, you'll get a voucher for a free drink upon arrival. It was a bit cold at night at some point, maybe because it's March and the heating is not constantly on anymore. Each room has its own heating control, though. Strongly recommended.

How to extract Lived Experiences?

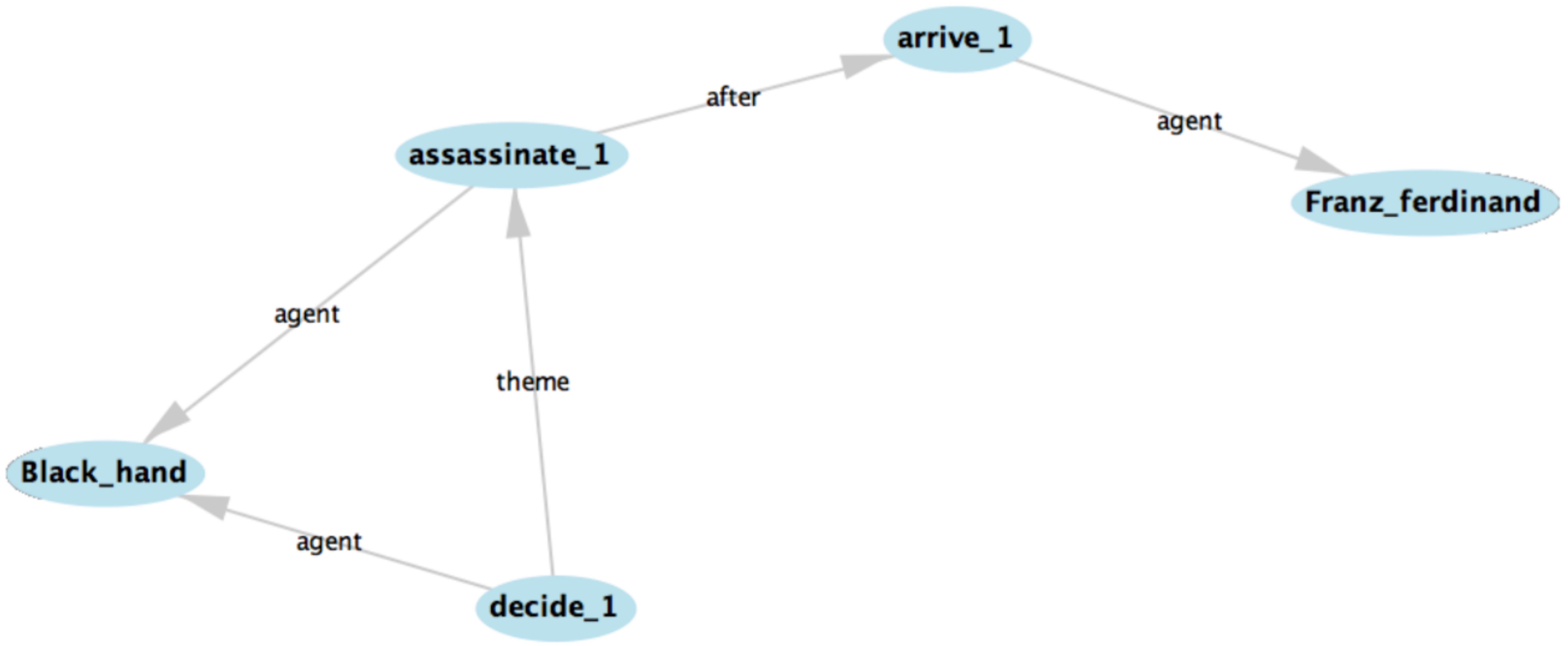
- LEs are linked to events -> event extraction



FRED as an Event Extraction tool

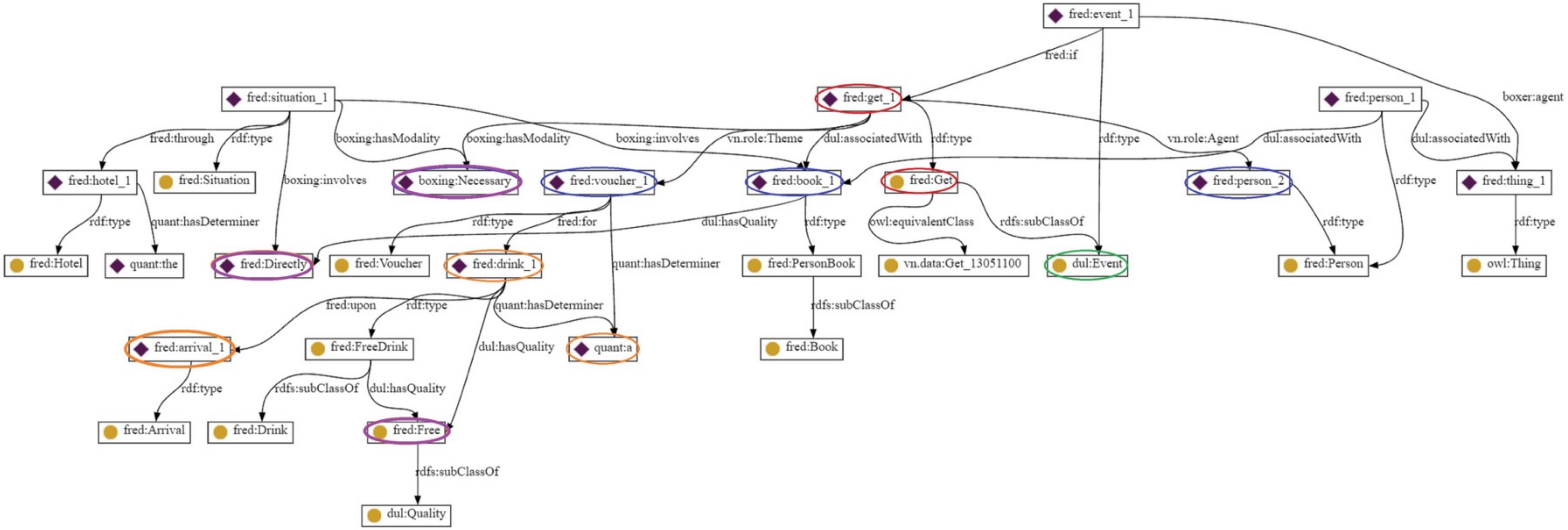


Event sub-graph



LE Extraction

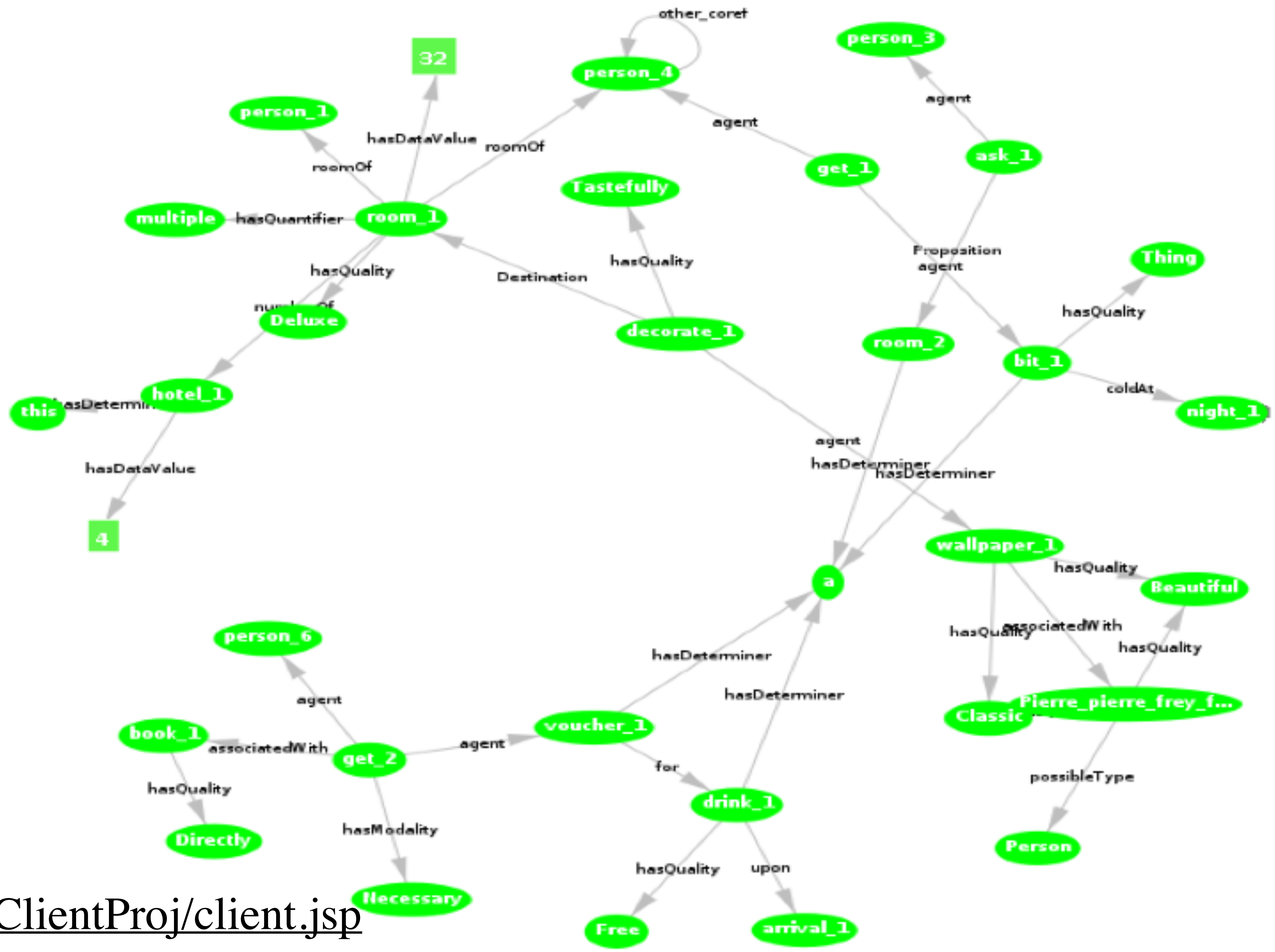
- Events with "personal" characteristics (writer or 1,2 p.pronouns as participants)



"If you book directly through the hotel, you'll get a voucher for a free drink upon arrival."

LE Graph

- Events with the writer among the participants

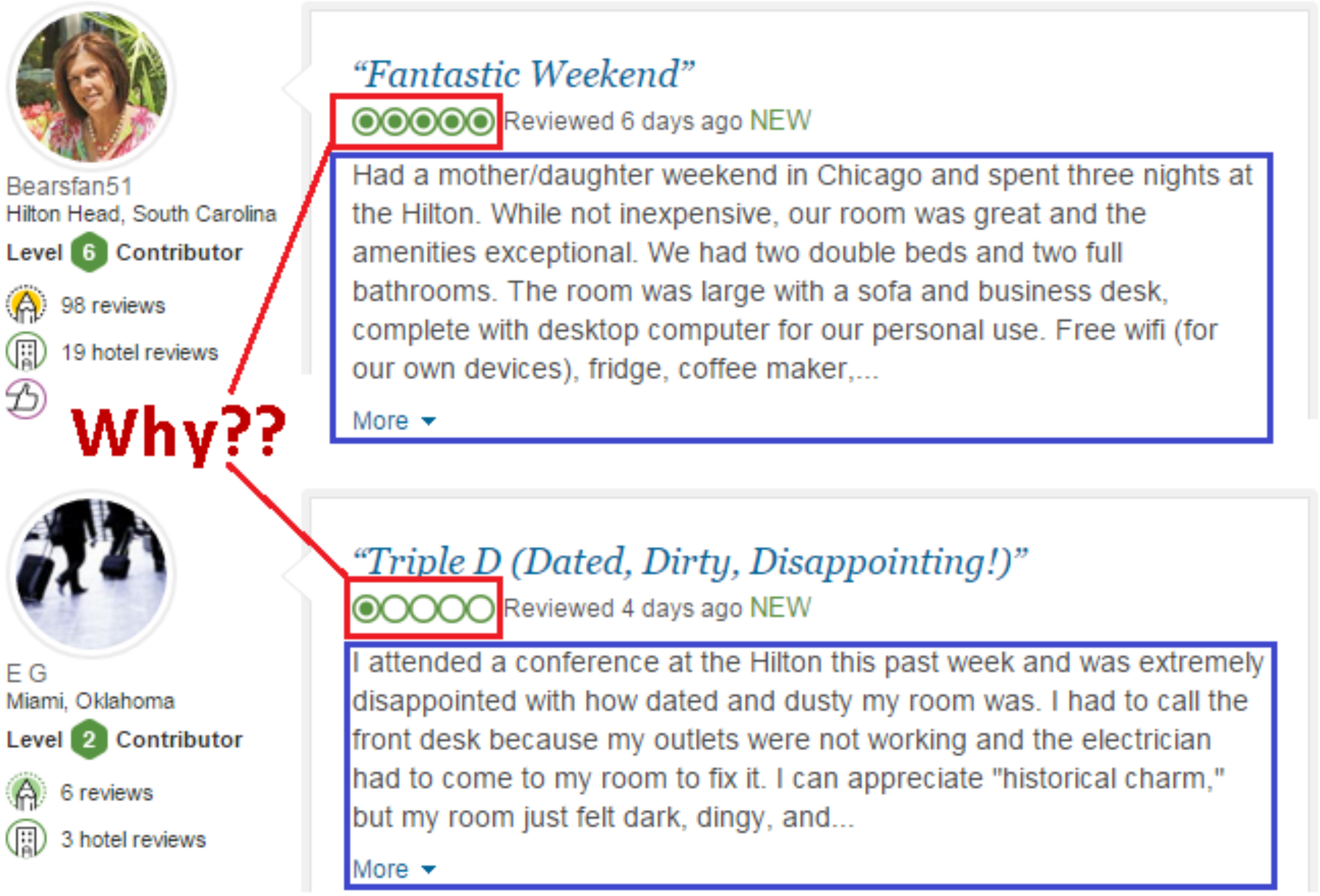


https://lipn.univ-paris13.fr/ClientProj/client.jsp

Correlating events with rating

- One doubt about Lived Experiences is that if we don't consider the parts of the sentence that have attributes we risk to lose the informative part of the review

- However, experiments show that events and ratings are correlated, so events can provide a justification of the score assigned to the review



Experiments

- Ott dataset of positive and negative reviews (800, 400 positive and 400 negative)
- All LE Eve: all extracted lived experience events
- (LE Eve)σ = 0.7: remove shared LE events
- All LE Eve Part: all extracted lived experience events and their participants
- (LE Eve Part)σ = 0.7: removed some common lived experience events and their participants

Naïve Bayes:

SVM:

The case of Deceptive Opinions

Opinion spam

- Mislead readers by creating false opinions
    - i.e., opinion not based on a real experience with the reviewed product/service
- Sometimes opinions are automatically generated (click-farms)
- Sometimes they are written to emulate real ones

Test: truthful or deceptive?

- I want to make this review in order to comment on the excellent service that my mother and I received on the Serenade of the Seas, a cruise line for Royal Caribbean. There was a lot of things to do in the morning and afternoon portion for the 7 days that we were on the ship. We went to 6 different islands and saw some amazing sites! It was definitely worth the effort of planning beforehand. The dinner service was 5 star for sure. One of our main waiters, Muhammad was one of the nicest people I have ever met. However, I am not one for clubbing, drinking, or gambling, so the nights were pretty slow for me because there was not much else to do. Either than that, I recommend the Serenade to anyone who is looking for excellent service, excellent food, and a week full of amazing day-activities!

Review A

Test: truthful or deceptive?

- High Points: Guacamole burger was quite tall; clam chowder was tasty. The decor was pretty good, but not worth the downsides. Low Points: Noisy, noisy, noisy. The appetizers weren't very good at all. And the service kind of lagged. A cross between Las Vegas and Disney world, but on the cheesy side. This Cafe is a place where you eat inside a plastic rain forest. The walls are lined with fake trees, plants, and wildlife, including animatronic animals. A flowing waterfall makes sure that you won't hear the conversations of your neighbors without yelling. I could see it being fun for a child's birthday party (there were several that occurred during our meal), but not a place to go if you're looking for a good meal.

Review B

The answer

- The truthful review was

Review A

- The best human result on the Ott dataset is 60% accuracy
- Ott dataset: http://myleott.com/op_spam/
- Ott et al., 2011: “Finding Deceptive Opinion Spam by Any Stretch of the Imagination”

How to detect spam opinions?

- Review content:
  - Lexical features such as word n-grams, part-of-speech n-grams, and other lexical attributes.
  - Content and style similarity of reviews from different reviewers.
- Semantic inconsistencies: for example, a reviewer wrote "My wife and I bought this car ..." in one review and then in another review he/she wrote "My husband really love …"
- Reviewer abnormal behaviors:
  - Public data available from Web sites, e.g., reviewer id, time of posting, frequency of posting, first reviewers of products, and many more.
- Product related features: E.g., product description

Lexical analysis

- “I” more prominent than other pronouns
- Bias towards positive emotions
- Polarization of emotion strength
  - “Love” rather than “like”, etc.
- Less diversity among different contexts
- However: spammers may improve their techniques…

LIWC (hotel)		LIWC (doctor)	
deceptive	truthful	deceptive	truthful
i	AllPct	Sixletters	present
family	number	past	AllPct
pronoun	hear	work	social
Sixletters	we	health	shehe
see	space	i	number
posemo	dash	friend	time
certain	human	posemo	we
leisure	exclusive	feel	you
future	past	perceptual	negemo
perceptual	home	leisure	Period
feel	otherpunct	insight	relativ
comma	negemo	comma	ingest
cause	dash	future	money

Conclusions

Sentiment Analysis: the future

- Clearly two directions of research:
    - Deep Learning + acquisition of knowledge from large corpora (distributional hypothesis)
        - Still fails at recognizing polarity switches and irony
    - Dictionaries + knowledge repositories (Frege model of meanings)
        - Needs good linking capabilities and wide coverage resources

Sentiment Analysis: open problems

- Figurative language (Irony, Sarcasm, etc.)
- Rich contextual information
  - When we name an entity, we evoke all of its context too
  - Ex: Macron -> president of France, liberal, former socialist, etc…
- Comparative opinions
- Sentiment is a continuously evolving and malleable matter

“The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without any emotions.”

–Marvin Minsky