Linnæus University Gavagai

# Sentiment, stance and applications of distributional semantics

Maria Skeppstedt

# Sentiment, stance and applications of distributional semantics

# Sentiment analysis (opinion mining)

- Aims at determining the attitude of the speaker/ writer
  * In a chunk of text (e.g., a document or sentence)
  * Towards a certain topic
- Categories:
  * Positive, negative, (neutral)
  * More detailed

# Movie reviews

- Excruciatingly unfunny and pitifully unromantic.
- The picture emerges as a surprisingly anemic disappointment.
- A sensitive, insightful and beautifully rendered film.
- The spark of special anime magic here is unmistakable and hard to resist.
- Maybe you'll be lucky , and there'll be a power outage during your screening so you can get your money back.

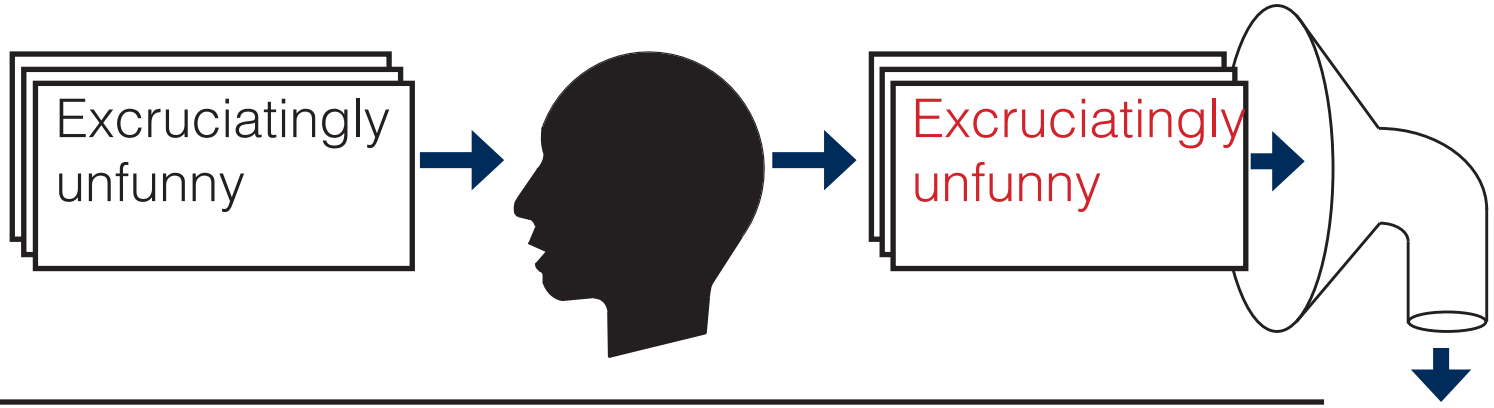
# Methods for sentiment analysis

- Train a machine learning classifier
- Use lexicon and rules
- (or combine the methods)

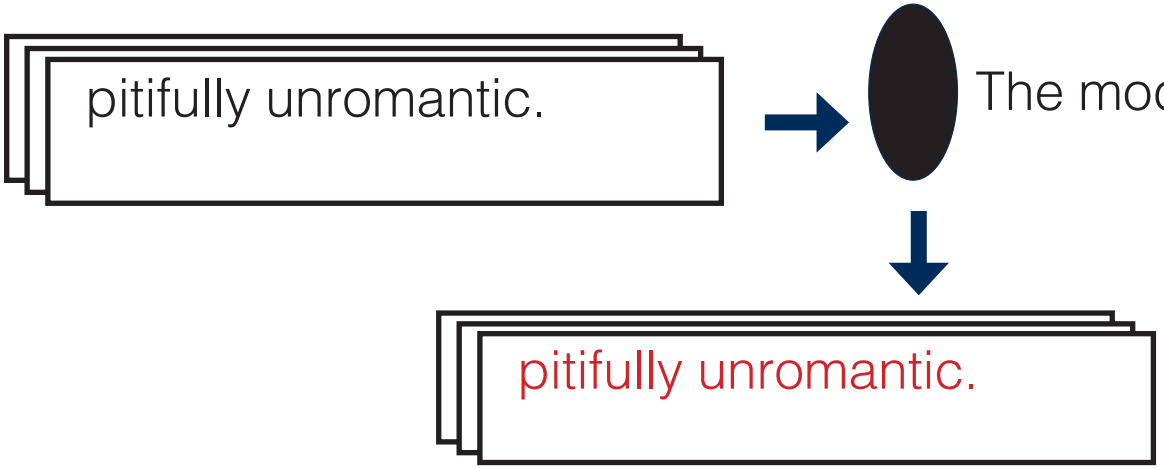
# Train a machine learning classifier

Linnæus University Gavagai

## Annotating text and training the model




## Applying the model



# Examples of machine learning Frameworks

- Scikit-learn
- NLTK (natural language toolkit)
- CRF++

# Methods for sentiment analysis

- Train a machine learning classifier
  * See for instance Socher et al. (2013) that classified English movie review sentences into positive and negative sentiment.
  * Trained their classifier on 11,855 manually labelled sentences.

(R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank)

# Methods based on lexicons and rules

- Lexicons for different polarities
- Sometimes with different strength on the words

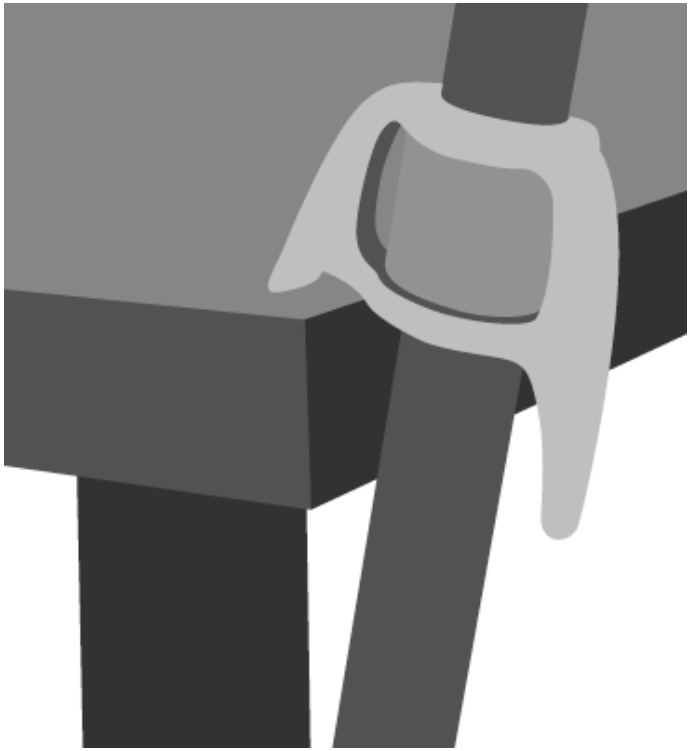
  * Does not need the large amount of training data that is typically required for machine learning methods

  * More flexible
  (e.g, to add new sentiment poles and new languages)

# Commercial use of sentiment analysis

I want to retrieve everything positive and negative that is said about the products I sell.

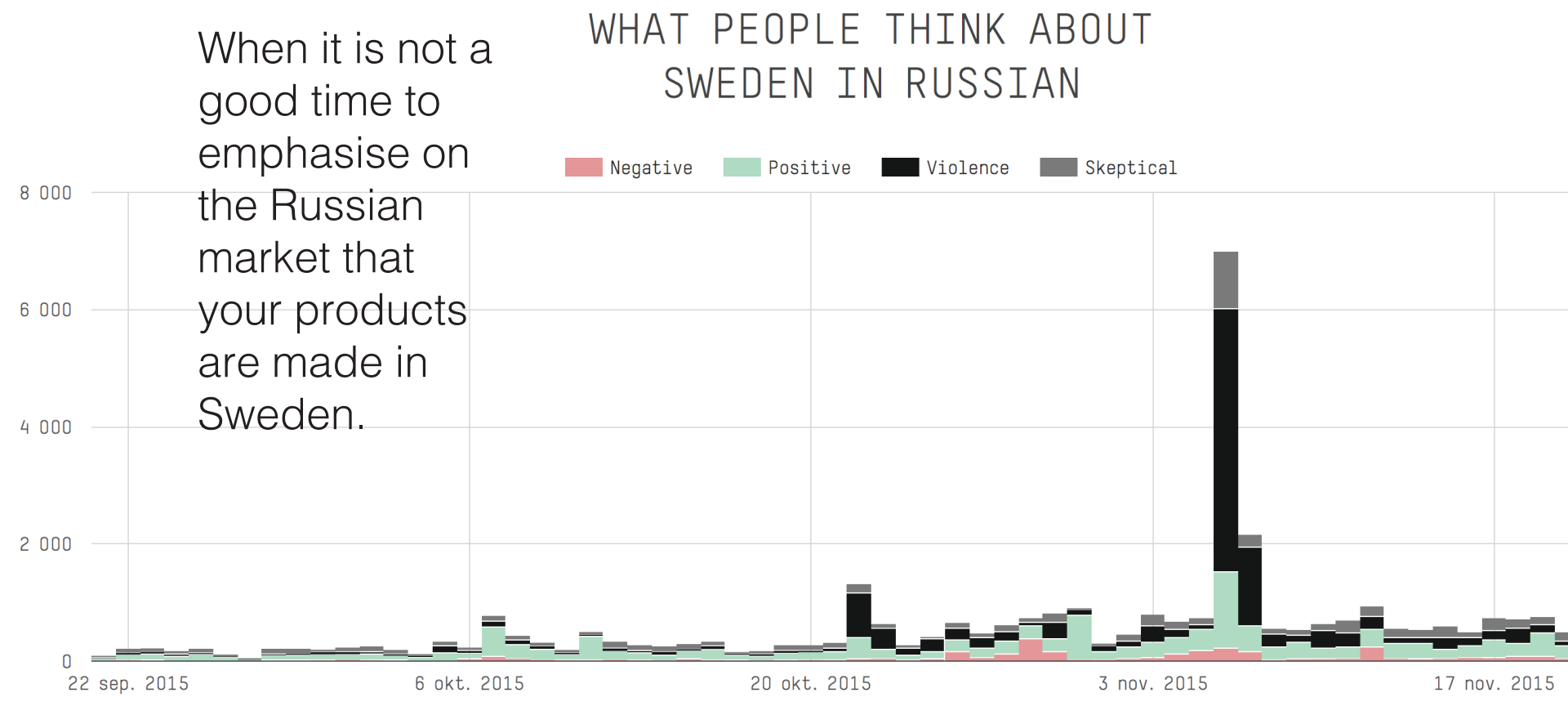
- My Dad absolutely loved these. Solved his problem of canes being propped against furniture only to fall over.
- Holder works fine. The product description is lacking a  photo of the other side. The holder is open on the back side.This is how you attach it to the cane.



Linnæus University Gavagai

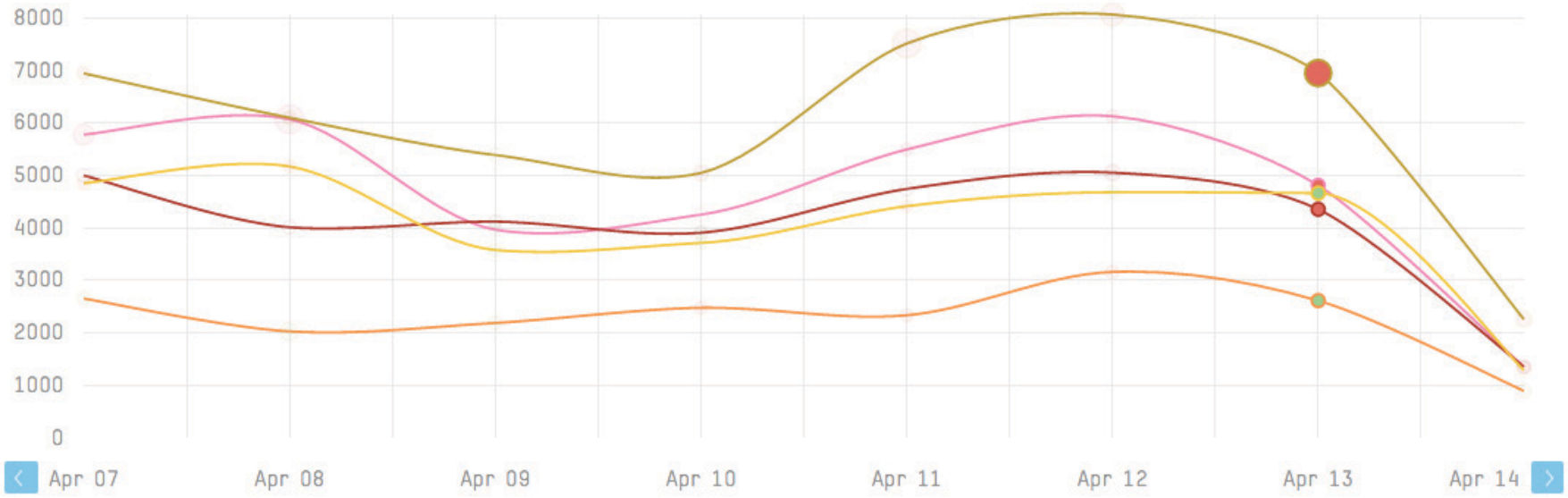
# Also looking at the big picture (data aggregated from many authors)

When it is not a good time to emphasise on the Russian market that your products are made in Sweden.




SOURCES: NEWS MEDIA AND SOCIAL MEDIA IN RUSSIAN

# Also, aggregated sentiment on e.g., political issues




**HILLARY CLINTON**
Occurrence 4808 mentions
Time April 13
18% more negative than positive
NEW YORK (AP) " Presidential candidate Hillary Clinton ...

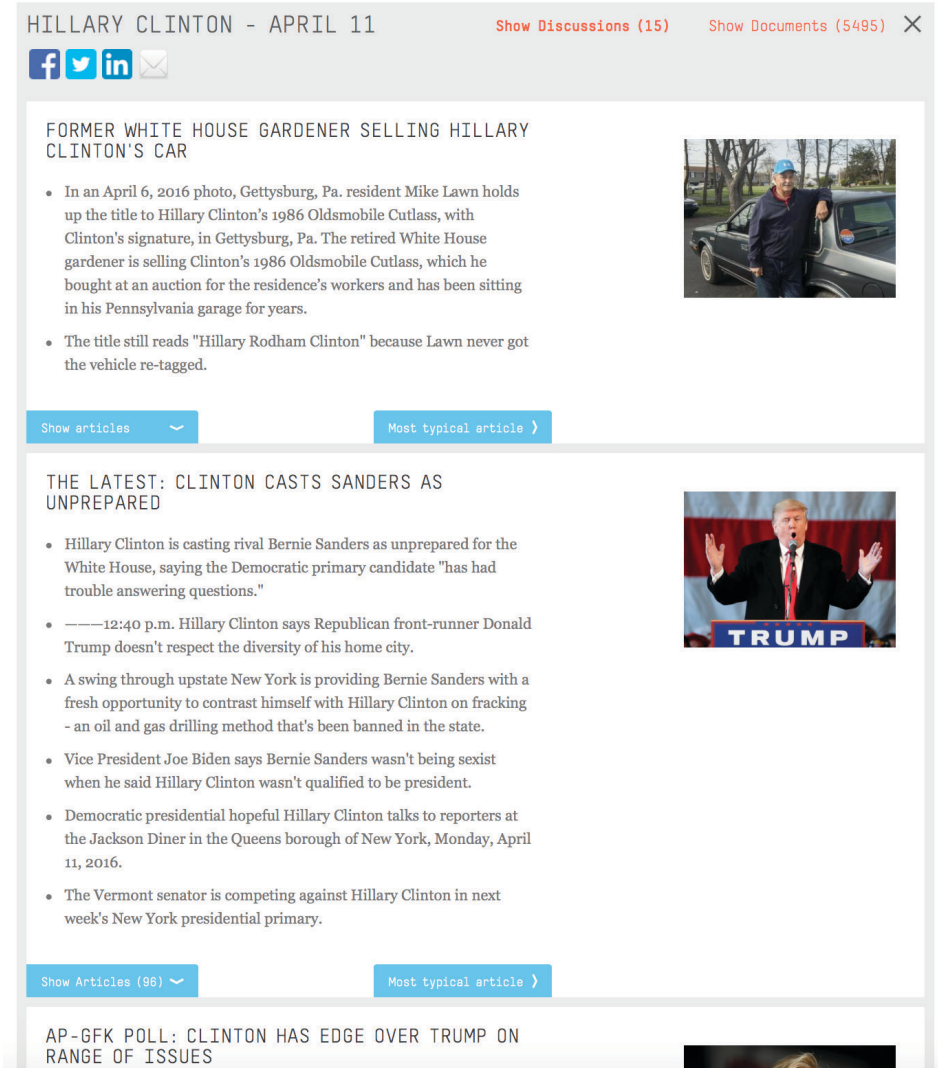
**TED CRUZ**
Occurrence 4353 mentions
Time April 13
8% more negative than positive
The analysis shows, at least at this point, that Clinto...

**BERNIE SANDERS**
Occurrence 4661 mentions
Time April 13
5% more positive than negative
Democratic presidential candidate, Sen. Bernie Sanders,...

**DONALD TRUMP**
Occurrence 6949 mentions
Time April 13
38% more negative than positive
Fox says its anchor, Megyn Kelly, met Wednesday with Re...

# Gavagai monitor

- The two previous examples showed Gavagai monitor
- Different targets
  * Sweden
  * Hillary Clinton, Ted Cruz ..
- Several sentiment poles (called tonality):
  * Positive, negative, violence, skeptical

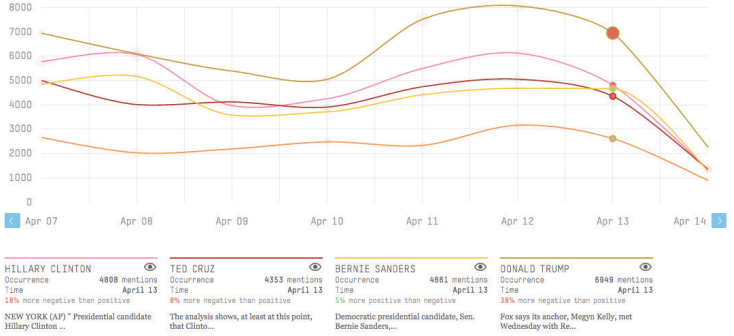



Linnæus University

Gavagai

# Gavagai monitor

Try example at:
https://monitor.gavagai.se/sectors/408




Explainer video:

https://www.youtube.com/watch?v=edTxOKDoCtA&feature=youtu.be

# Sentiment, stance and applications of distributional semantics

# Stance detection

- Position on an issue.
- E.g. SemEval-2106 shared task: ‘In favour’, ‘against’ or ‘none of these’ towards a target.
  * Target e.g., ‘Climate Change is a Real Concern’
  * ‘In favour’ and ‘against’ not the same as ‘positive’ and ‘negative’ sentiment. E.g. ‘I believe global warming has mostly positive effects’. (‘positive’ and ‘against’)

http://alt.qcri.org/semeval2016/task6/

# Language components used when expressing stance

The StaViCTA project at Linnaeus University. A collaboration between Gavagai and researchers at Lund and Linnaeus University.

- Sentiment (**A** is good) (**B** is bad)
- Uncertainty (**A** might be true)
- Contrast (I believe **A**, but I do not believe **B**)
- Hypothetical (If **A** then **B** would happen)
- Agreement/disagreement (I can't agree on **A** )

# Used methods for detecting stance components

- Manually annotate a lot of text to train a machine learning classifier
  - Takes a lot of time
  - Boring
- Possible solution 1: Use lexicons and rules instead of machine learning
- Possible solution 2: Use active learning

# Active learning

- The standard method is to randomly select training data for the classifier
- When active learning is used, there are different methods for selecting good training samples
  * I.e., training samples that are useful for the classifier.

# Sentiment, stance and applications of distributional semantics

# Distributional semantics

Recap from a previous lecture
(about Random Indexing)

- Expressions that occur in similar context tend to have a similar meaning.
(The distributional hypothesis.)
  * ‘I usually have a cup of coffee in the morning’
  * ‘I usually have a cup of tea in the evening’
  * ‘I drink too much coffee’
  * ‘It’s better to drink tea in the evening’

# Represent distribtuional semantics with a vector space

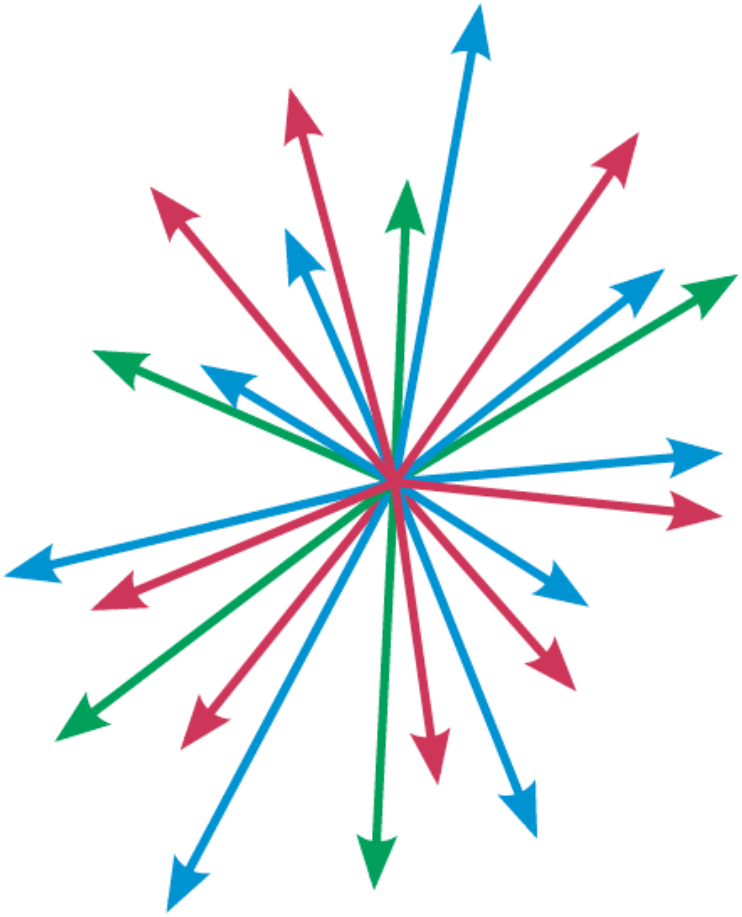
* I usually have a cup of **coffee** in the morning
* I usually have a cup of **tea** in the evening

Context vectors

| | cup | morning | evening | ... | |
|---|---|---|---|---|---|
| **coffee** | [1 | 1 | 0 | | ] |
| **tea** | [1 | 0 | 1 | | ...] |
| **term_1** | [0 | 0 | 0 | 20 ... | |
| **...** | | | | | |
| **term_n** | [0 | 0 | 0 | 2 ... | ] |

* The context vectors represent the context in which a term has occurred.
* Here, a context window of 1+1 (removed stop words).

# The vector space model



A geometrical meta-phor for meaning. Similarity of two terms are measured by the distance between their vectors.

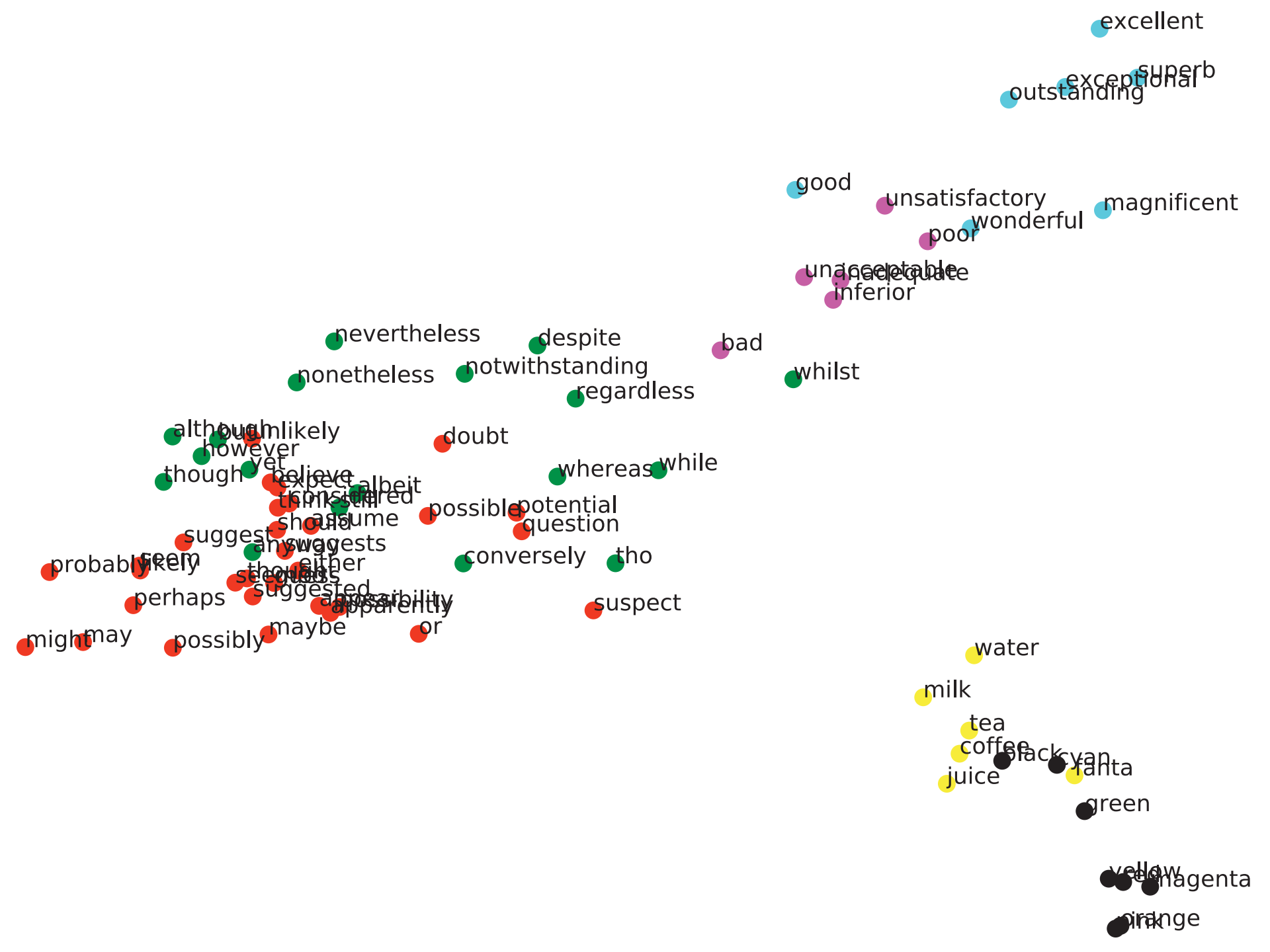
# Methods for building this vector space

A number of different methods for creating the models. Typically represent the the vector space with smaller matrixes. Examples are:

- Random indexing
- Word2vec

A new word for vector spaces are word embeddings.

See the following blog posthttp://www.gavagai.se/blog/2015/09/30/a-brief-history-of-word-embeddings/


excellent
superb
exceptional
outstanding
good
unsatisfactory
magnificent
wonderful
poor
unacceptable
inadequate
inferior
nevertheless
despite
bad
notwithstanding
whilst
nonetheless
regardless
although
however
though
yet
believe
doubt
whereas
while
albeit
possible
potential
question
assume
should
suggest
anyway
suggests
probably
likely
seem
either
conversely
tho
perhaps
suggested
apparently
suspect
might
may
possibly
maybe
or
water
milk
tea
coffee
black
cyan
fanta
juice
green
yellow
magenta
orange
pink

# Gavagai living lexicon

Try it out at: http://lexicon.gavagai.se

# Gavagai living lexicon, answers

(pan african, organ systems)
- hot chocolate
- hot cocoa
- baked goods
- steaming hot
- bubble tea
- doughnuts
- donuts
- donut shop

(drinking, kabul afghanistan)
- chamomile tea
- ginger tea
- herbal teas

(connoisseur)
- black tea
- green tea
- herbal tea
- matcha tea

(pan african, beans)
- black coffee
- mint tea
- earl grey
- coffe
- brewed coffee
- hot coffee
- coffee
- orange juice
- cold coffee
- warm milk

The lexicon is constantly learning, so it might have changed.

# Gavagai API

**KEYWORDS** : EXTRACT SALIENT CONCEPTS FROM A COLLECTION OF TEXTS. ORDER BY NUMBER OF OCCURRENCES.

Show/Hide | List Operations | Expand Operations | Raw

**LEXICON** : LOOKUP A WORD IN GAVAGAI LIVING LEXICON. Show/Hide | List Operations | Expand Operations | Raw

**STORIES** : MULTI-LANGUAGE STORY EXTRACTION FROM LARGE AMOUNTS OF TEXT INTO CONVENIENTLY SHORT AND EXACT EXECUTIVE SUMMARIES. AVAILABLE FOR MULTIPLE LANGUAGES.

Show/Hide | List Operations | Expand Operations | Raw

**TONALITY** : MULTIDIMENSIONAL TONALITY ANALYSIS - WE COMPUTE SCORES FOR POSITIVITY, NEGATIVITY, FEAR, HATE, LOVE, SKEPTICISM, VIOLENCE, AND DESIRE IN ALL SUPPORTED LANGUAGES.

Show/Hide | List Operations | Expand Operations | Raw

**TOPICS** : MULTI-LANGUAGE TOPIC EXTRACTION FROM LARGE AMOUNTS OF TEXT INTO CONVENIENTLY SHORT AND EXACT EXECUTIVE SUMMARIES. AVAILABLE FOR MULTIPLE LANGUAGES. ALLOWS FOR IGNORING AND MERGING OF TERMS.

Show/Hide | List Operations | Expand Operations | Raw

# Tonality, json input, over HTTP

{“language”: “en”,

“texts”: [{“body”: “West Africa is suffering the worst Ebola out-break in history, with more than 17,100 illnesses and at least 6,000 deaths this year.”,

“id”: “http://www.kmph-kfre.com/story/27541218/ameri-can-coming-to-us-to-be-evaluated-for-ebola”

},{“body”: “An American health-care worker in West Africa who may have been infected with #Ebola is being flown to Emory. West Africa is suffering the worst Ebola outbreak in history, with more than 17,100 illnesses and at least 6,000 deaths this year.”,

“id”: “https://twitter.com/DoctorYasmin/sta-tus/540277718294605824”}]}

# Tonality, json output

TonalityResponse {

texts (array[ScoredText]): An array of texts scored by tonality}

ScoredText {

id (string): The unique identifier of the text,

tonality (array[Score]): A list of tones and their scores.}

Score {tone (string): The name of the tone.,

score (number): The score given for this tone, based on lexical analysis. A positive number.,

normalizedScore (number): The normalized score given for this tone. A number within the range 0-1.}

---

Tones: POSITIVITY, NEGATIVITY, FEAR, HATE, LOVE, SKEPTICISM, VIOLENCE, AND DESIRE

# Gavagai explorer

https://www.youtube.com/watch?v=4qSNbWDpTiM