Analysing Language in Conversation

Matthew Purver

Centre for Intelligent Sensing
Queen Mary University of London

CIS centre for intelligent sensing

Queen Mary University of London

Analysing Language in Conversation

- Basic Techniques
  - Language Modelling
  - Grammars
- Basic NLP Tasks
  - Document classification
  - Summarisation
- Sentiment and Emotion
- Topic Modelling
- Dialogue and Coordination
  - Dialogue acts
  - Dialogue management: ISU vs POMDPs
  - You
- Incrementality

Some Tasks We Might Attempt

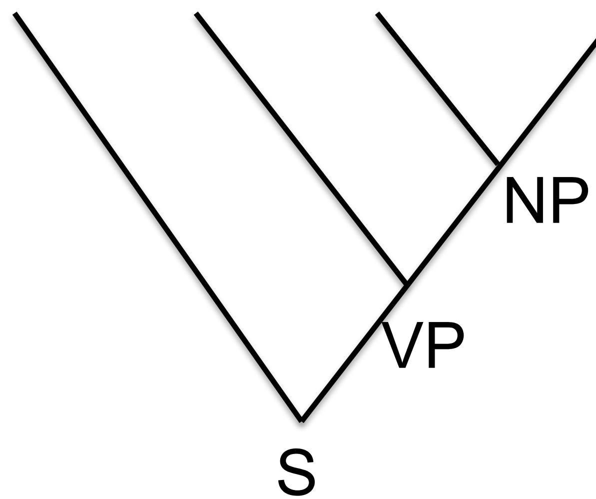
- What might we want to sense from language?

- What people like & dislike
  - Sentiment analysis
- What people are talking about / interested in
  - Topic detection & tracking
- What people think
  - Opinion mining
- What people are going to do
  - Decision detection, behaviour prediction
- How to hold up our end of a conversation
  - Human-computer dialogue systems

Some Basic Aspects of Language

words:	mary	hires	a	detective
parts of speech:	PN	VBZ	DET	CN

syntax:

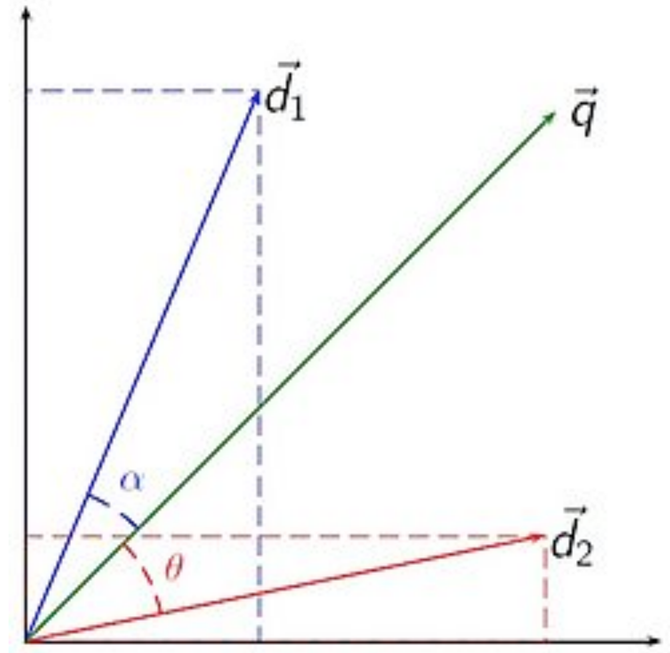


semantics: $\exists$ x.detective(x) & hire(mary,x)

e,x | subj(e,mary) & hire(e) & obj(e,x) & det(x)

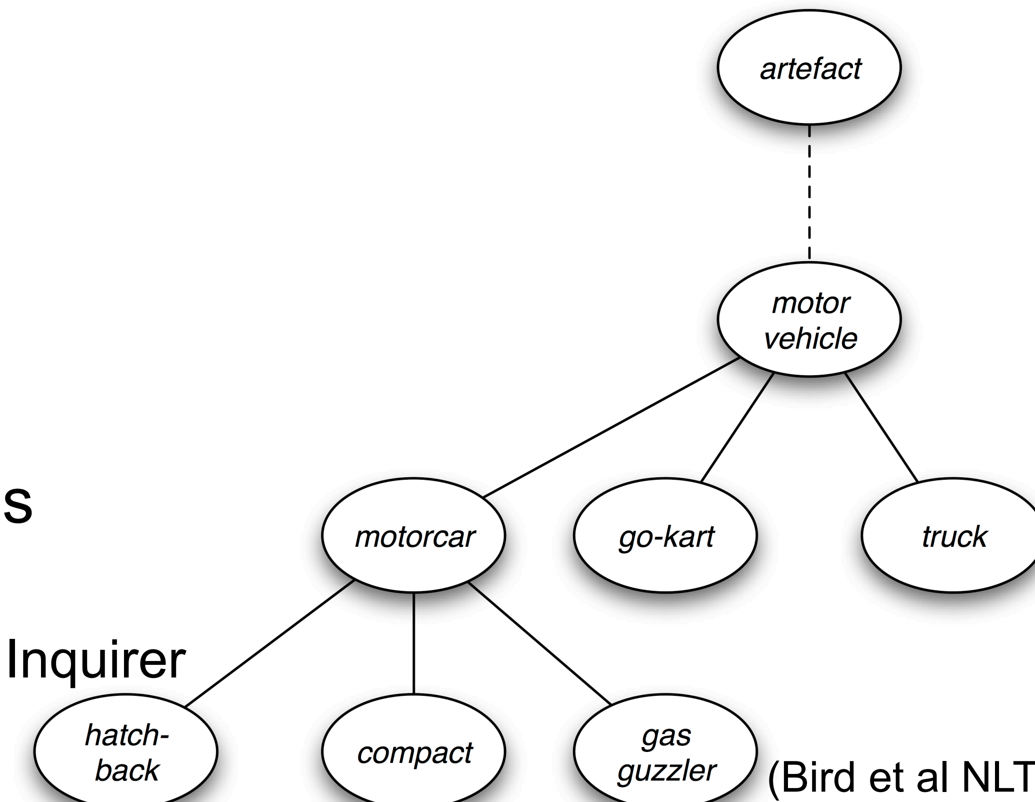
Words

- We can characterise a text in terms of its words
- Vector space models
  - words = dimensions
- Good for:
  - Information retrieval
  - Document similarity
  - Document classification



Words

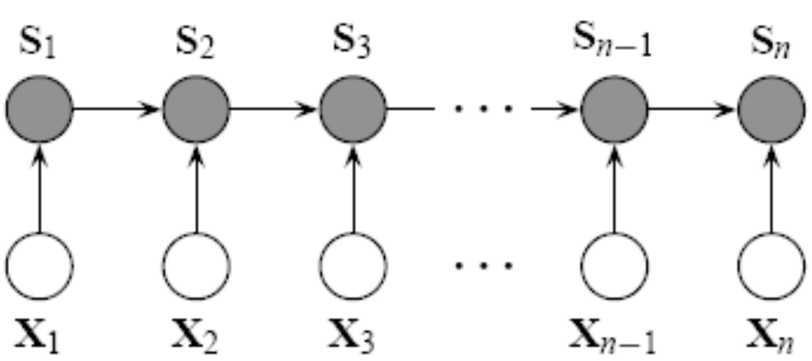
- Good for:
  - dog bites man
  - dog chases man
  - dog bites cat
  - cat eats fish
- Bad for:
  - puppy bites man
  - cat bites man
  - fish bites man
- Dictionaries / ontologies
  - WordNet
  - SentiWordNet, General Inquirer



(Bird et al NLTK)

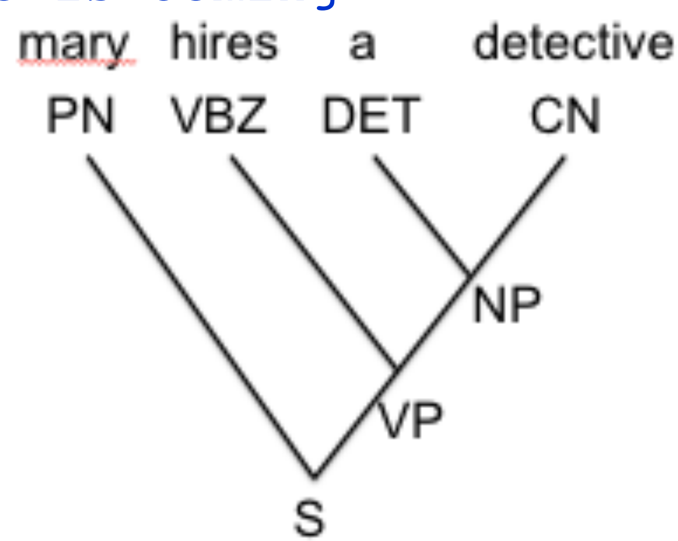
Sequences

- Bad for: `dog bites man` vs. `man bites dog`
- Language Modelling
  - probabilistic models of word sequence $w_1, w_2, w_3, \ldots, w_i$
    - $P(w_i \mid w_{i-1} \ldots w_1)$
    - approximate as bigrams: $P(w_i \mid w_{i-1}) \times P(w_{i-1} \mid w_{i-2}) \times \ldots$
    - Or trigrams: $P(w_i \mid w_{i-1}, w_{i-2}) \times P(w_{i-1} \mid w_{i-2}, w_{i-3}) \times \ldots$
- Part-of-Speech Tagging
  - probabilistic models of word-tag sequence associations
  - HMMs
  - CRFs
  - …



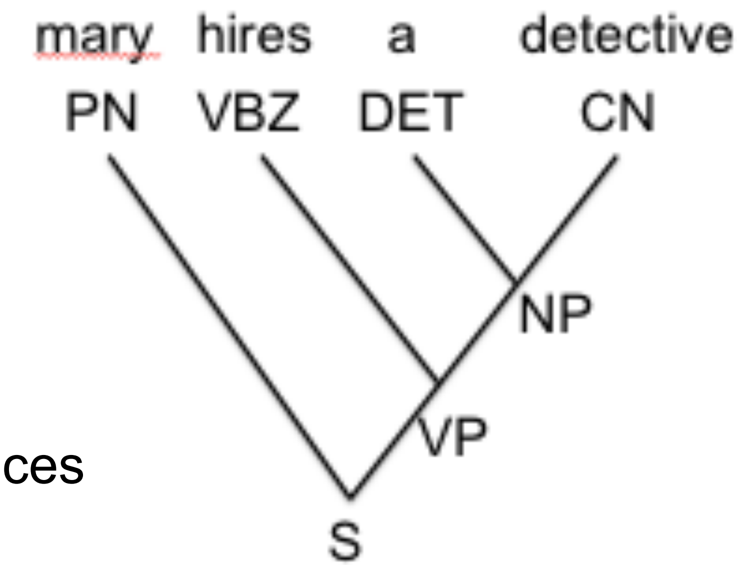
Sentence Structure

- Sequence doesn't capture everything!
  - `man bites dog`
  - `no man bites dog`
  - `almost no man bites dog`
  - `i don't think man bites dog ...`
- "Long-distance dependencies"
  - `who do you think we need to see?`
  - `the man I sold the car to is coming`
  - negation
  - wh-movement
  - clause structure



Sentence Structure

- Structural dependencies
- Meaning (semantics)
- Syntactic & semantic parsing
  - (Probabilistic) grammar rules:
    - S → NP VP
    - VP → VBZ NP
    - NP → DET CN
    - …
  - (Probabilistic) parsing algorithm
  - Syntax-semantics correspondences
  - Good accuracy c. 80-90%
    - See e.g. (Clark & Curran, 2007)
  - But someone has to write (much of) it by hand …



Knowledge-rich vs knowledge-poor

- Decades of research in machine learning:
  - language modelling
  - document classification
  - grammar induction
  - …
  - robust, but mostly quite shallow

- Decades of research in building resources:
  - dictionaries (eg. sentiment)
  - word & concept ontologies (similarity)
  - grammars (structure, meaning)
  - …
  - deep, but mostly language- and/or domain-specific

We Need To Talk About Dialogue

- But what happens when we look at **dialogue**?
  - Human-computer dialogue
  - Human-human dialogue
  - Social media interaction

- Dialogue is informal
  - do we know how people talk?
- Dialogue has structure
  - high-level topical structure
  - low-level dialogue structure
- Dialogue is incremental
  - people process language word by word

We Need To Talk About Dialogue

- But what happens when we look at **dialogue**?
  - Human-computer dialogue
  - Human-human dialogue
  - Social media

- Dialogue is informal
  - do we know how people talk?
- Dialogue has structure
  - high-level topical structure
  - low-level dialogue structure
- Dialogue is incremental
  - people process language word by word

Informality

- Language can be unpredictable
  - vocabulary, grammar, spelling …
  - the interview was uuhhhh it was alright
  - we reckon we reckon yeah she looks like erm oi oi oi you know oi oi oi you know the Fraggles yeah?
  - Nyt alexx tweetdreamsh RT @JDBAustralia: Goodnight everyone, i will tweet you all tomorrow <3
  - I just said look you know silly really cos I mean he knew I had a couple of people erm you know Monday and Tuesday before Monday and Tuesday and er you know you got erm you need a couple of people as well so if you don't mind coming over

Language Change

- And it refuses to stay still:
  - `I was goin o2 sleep buht, im UP lol.`
  - `Im Not Goin o2 Be Sad o2day Imah $MILE , Jus o4 Big Bruhh !`
  - `LOL IM BOR3D @ENYCHARM YU GOIN O2 DA M33TING?`

Speech Recognition

- And we're still not that good at ASR:

```
do you have the comments cetera and uh the
the other is
you don't have
i do you want
oh we of the time align said is that
i you
well fifty comfortable with the computer
mmm
oh yeah that's the yeah that
sorry like we're set
make sure we captive that so this deviates
```

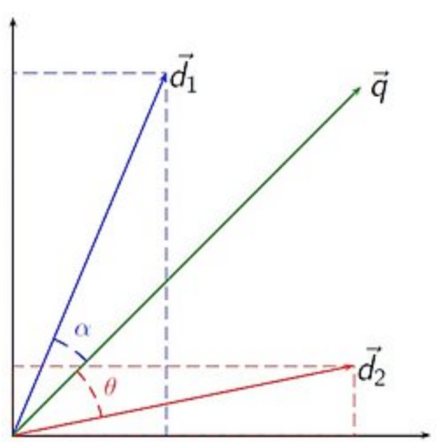
- Can we use standard methods?

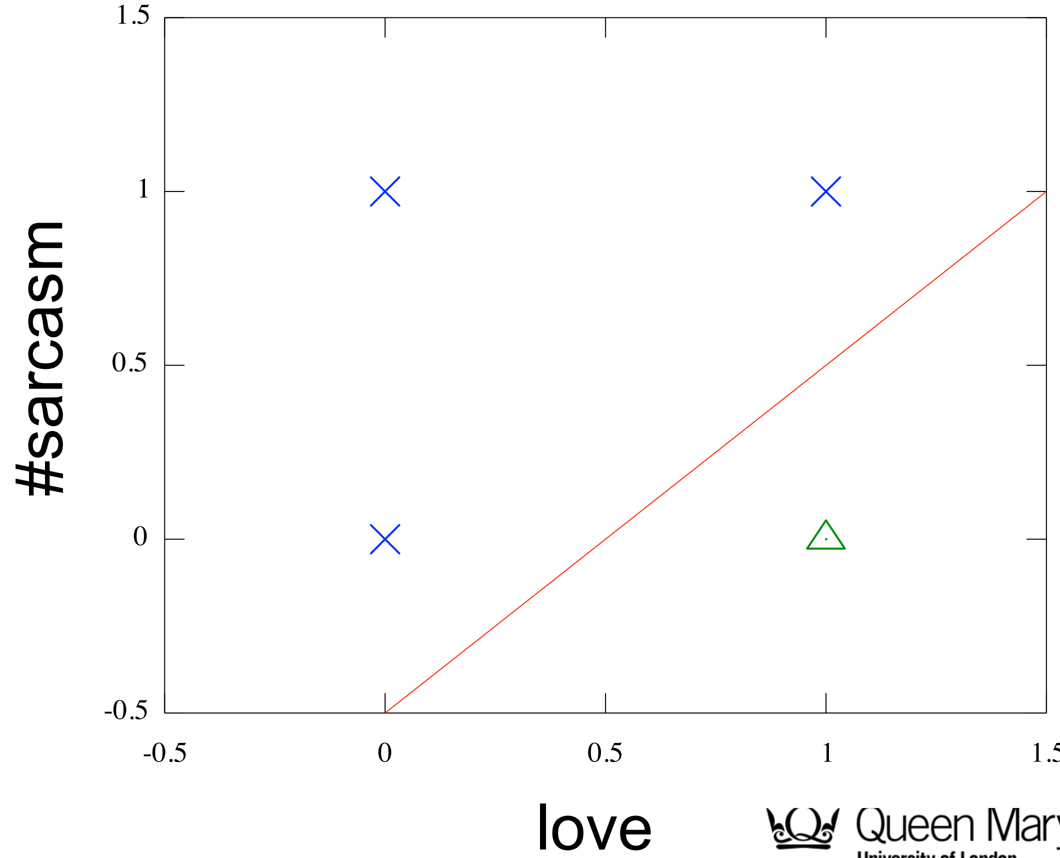
Supervised Machine Learning

- We need robustness
- We need adaptivity
  - machine learning
- E.g. supervised discriminative classification:
  - We know the categories
  - We have lots of labelled examples
  - Some idea of discriminative features
    - e.g. words
  - Machine learns to distinguish the categories itself

Sentiment Analysis

- Binary classification e.g. with SVMs
  - **i love @justinbieber #sarcasm**





Problem 1: Structural dependencies

- Structural dependencies
  - (remember questions, relative clauses etc)
  - but syntactic parsing just isn't realistic here
  - part-of-speech tagging perhaps (Gimpel et al, 2011)
- E.g. negation
  - `justin's new hair is nice`
  - `justin's new hair is not nice`
  - n-gram features?
  - `justin's new hair is not really that nice`
  - `justin's new hair is not really all that nice`
  - negated word features?
  - `justin's new hair is not really_n that_n nice_n`
  - get a LOT of data
    - See e.g. http://sentiment.christopherpotts.net

Problem 2: Labelling

- What if we don't know what the categories are?
  - (we'll look at this in a minute)
- Even if we do, how do we label our data?
  - how long does it take us to label 1,000,000 texts?
    - well, there's crowdsourcing …
  - how reliably can we label a short text?
    - `I hope XXXX is okay, he hasn't texted me all day`
    - `I'm worried people wouldn't turn up hahhaha`

Distant Supervision

- Find some **author**-generated conventions
- Treat them as “noisy” labels:
  - Nyt alexx tweetdreamsh RT @JDBAustralia: Goodnight everyone, i will tweet you all tomorrow <3
  - Gets so #angry when tutors don’t email back... Do you job idiots! :@
  - 考完它我就能回家啦~[鼓掌][鼓掌][鼓掌][鼓掌]开心O(∩_∩)O~~

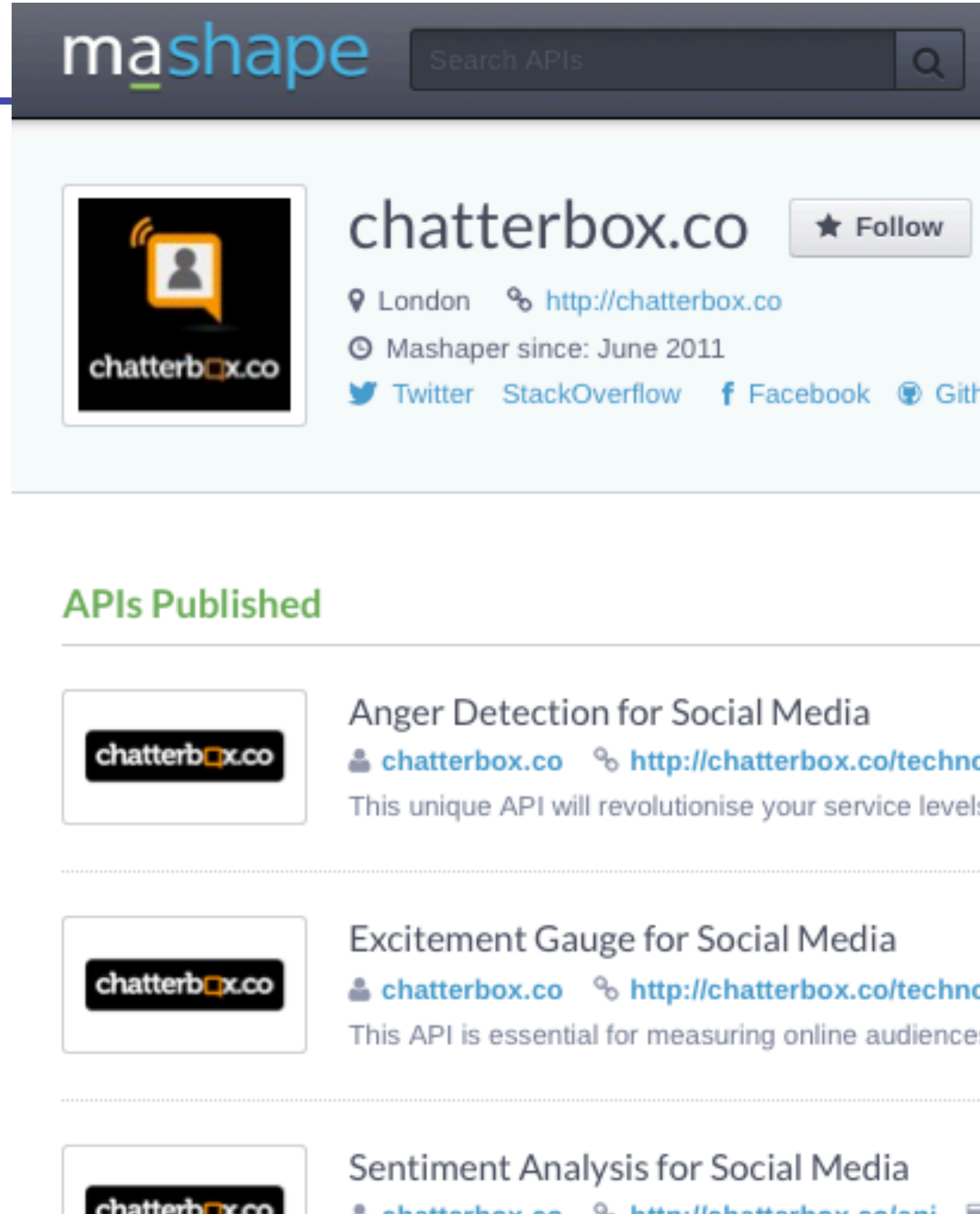
Distant Supervision

- Find some **author**-generated conventions
- Treat them as "noisy" labels:
  - Nyt alexx tweetdreamsh RT @JDBAustralia: Goodnight everyone, i will tweet you all tomorrow
  - Gets so angry when tutors don't email back... Do you job idiots!
  - 考完它我就能回家啦~ 开心 ~~
- Learn a model as normal:
  - Sentiment (positive/negative)
  - Emotions (angry/happy/sad/worried)
  - Opinions (agreement/disagreement)
- Bootstrapping, active learning, co-training, …

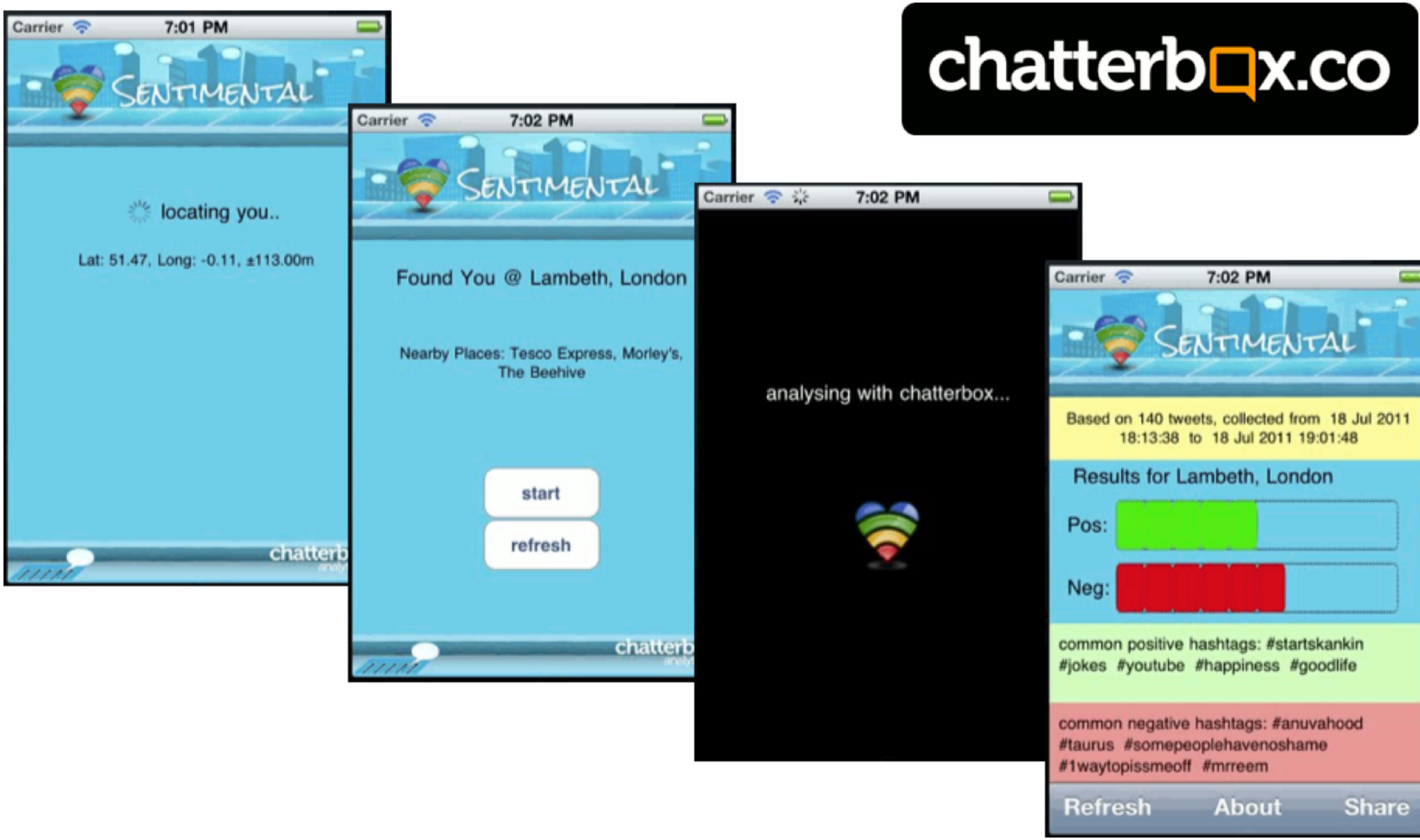
Quick Plug

APIs available free on mashape.com/chatterbox-co

See, or even cite, (Purver & Battersby, EACL 2012)

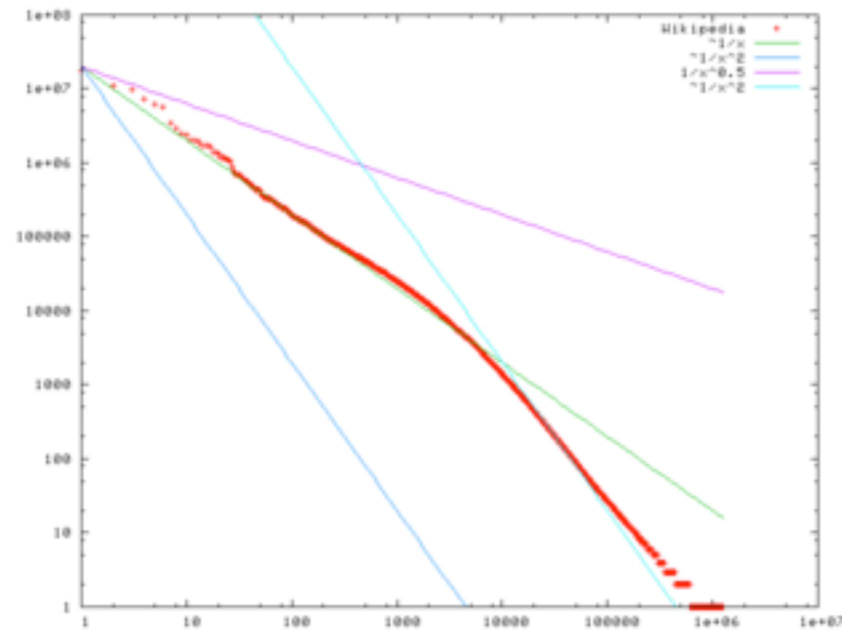


iPhone app: *Sentimental*



chatterbox.co

Problem 3: Dimensionality



- Language follows Zipf's Law:
- Corollaries:
  - More data = more features
  - … = more rare features
  - … = more chance correlations
- Can be hard to constrain feature space

Schizophrenia & Prediction

- Study with SMD: 128 out-patients with schizophrenia
  - transcripts of therapy dialogues with clinicians
  - measured outcomes:
    - symptom scales
    - patient satisfaction, patient & doctor evaluations
    - adherence to treatment 6 months later
  - learn models of language and dialogue structure
  - classify patients in terms of outcomes (Howes et al, 2012, 2013)

Dr:	Rather than **the diazepam** which I don't think is going to do you any good
P:	**the valium**

Dr:	Yeh, it doesn't happen in real life does it?
P:	**What do you mean by real life?**
Dr:	**You can't - there are no messages coming from the television to people are there?**

Feature Dimensionality & Selection

- Small dataset (128 points), very large feature space
  - unigrams, bigrams, trigrams
- Raw feature set: poor performance
- Select features: information gain over whole set:

	Baseline	Words	High-level
PANSS *positive*	51.1	87.0*	56.5*
PANSS *negative*	49.6	87.8*	56.5*
PANSS *general*	48.4	91.1*	54.0
PEQ *emotions*	51.9	89.1*	53.5
PEQ *communication*	50.8	79.8*	52.4
PEQ *comm. barriers*	51.6	90.6*	51.6
PEQ *overall*	50.8	90.6*	53.9
Adherence	73.2	91.1*	63.4
Adherence (balanced)	53.5	93.0*	52.1

Table 2: Percentage accuracies vs feature set

Feature Dimensionality & Selection

- Select features: information gain over **training sets**:
  - Poor performance again!
  - Training set contains rare but apparently indicative words
- Select features: as above but **exclude rare features**:

Features	P (%)	R (%)	F (%)
Class of interest	28.9	100.0	**44.8**
Best features	70.3	70.3	**70.3**

- Human psychiatrist given same task:

Data	P (%)	R (%)	F (%)
Text transcripts	60.3	79.6	**68.6**
Transcripts + video	69.6	88.6	**78.0**

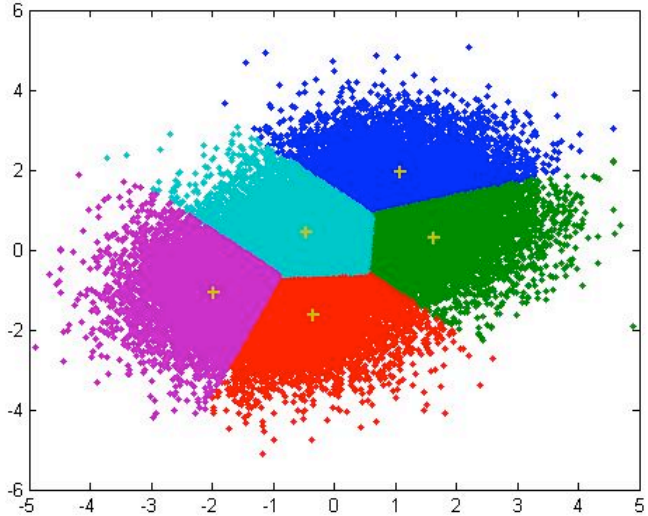
Classification Approaches

- We can get quite a long way
  - mostly using words and sequences of words
  - learning associations from data

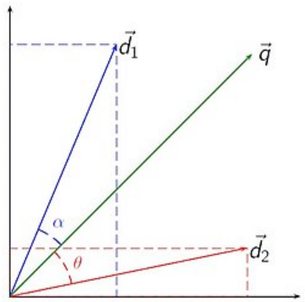
- But:
- We're limited in linguistic nuance
- We have a general labelling problem
- We have a general feature selection problem
- We have to know what we're looking for
- We don't really find out anything about meaning

Topic Modelling

- We can characterise language by **topic**
  - Tells us something about meaning
  - Dimensionality reduction
- Knowledge-rich methods:
  - Predefined ontologies e.g. WordNet-, Wikipedia-based
- Unsupervised methods:
  - Helps with the labelling problem
  - Helps with the conversational language problem

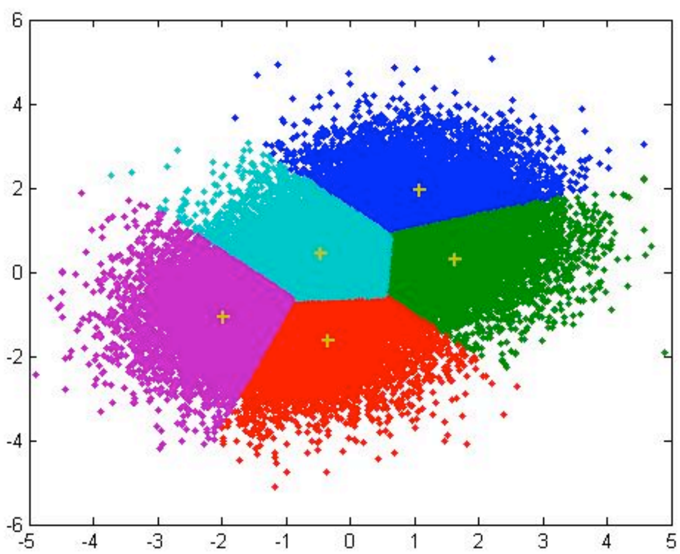


Topic Modelling



- Unsupervised methods:
  - Helps with the labelling problem
  - Helps with the conversational language problem
- Based around geometric model of language
- E.g. term-document matrix as a space:

$$
\begin{array}{cc}
 & \begin{array}{ccc} \phantom{x_{1,1}} & \mathbf{d}_j \\ & \downarrow & \end{array} \\
\mathbf{t}^T \rightarrow & \begin{bmatrix} x_{1,1} & \dots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \dots & x_{m,n} \end{bmatrix}
\end{array}
$$



Latent Semantic Analysis (Landauer et al 1998)

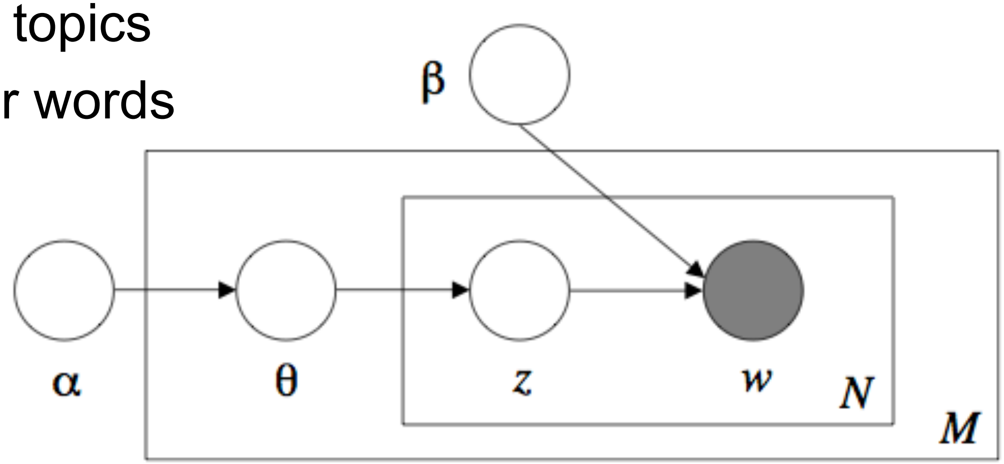
- Term-document matrix:

$$\begin{array}{cc} & \begin{array}{c} \mathbf{d}_j \\ \downarrow \end{array} \\ \mathbf{t}^T \rightarrow & \begin{bmatrix} x_{1,1} & \dots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \dots & x_{m,n} \end{bmatrix} \end{array}$$

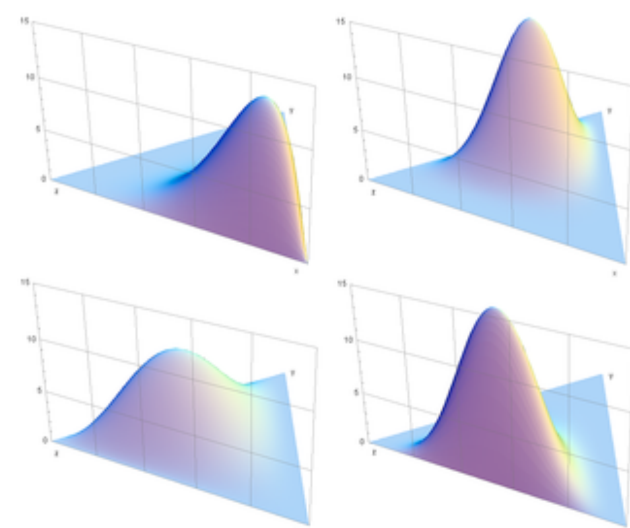
- Singular value decomposition
  - singular vectors as "topics" (cf. PCA)
- Successful in document classification
- Successful for dimension reduction
- Not very interpretable
- Not very good for multi-topic documents, ambiguous words

Latent Dirichlet Allocation (Blei et al, 2003)

- Probabilistic generative model:
  - documents as mixtures of topics
  - topics as distributions over words
  - basic assumptions about types of distributions



- Successful in topic discovery, document classification



We Need To Talk About Dialogue

- But what happens when we look at **dialogue**?
  - Human-computer dialogue
  - Human-human dialogue
  - Social media

- Dialogue is informal
  - do we know how people talk?
- Dialogue has structure
  - high-level topical structure
  - low-level dialogue structure
- Dialogue is incremental
  - people process language word by word

Dialogue structure

- Documents might contain mixtures of topics
  - But at least we know where they start & stop
- Conversation moves from one topic to another!

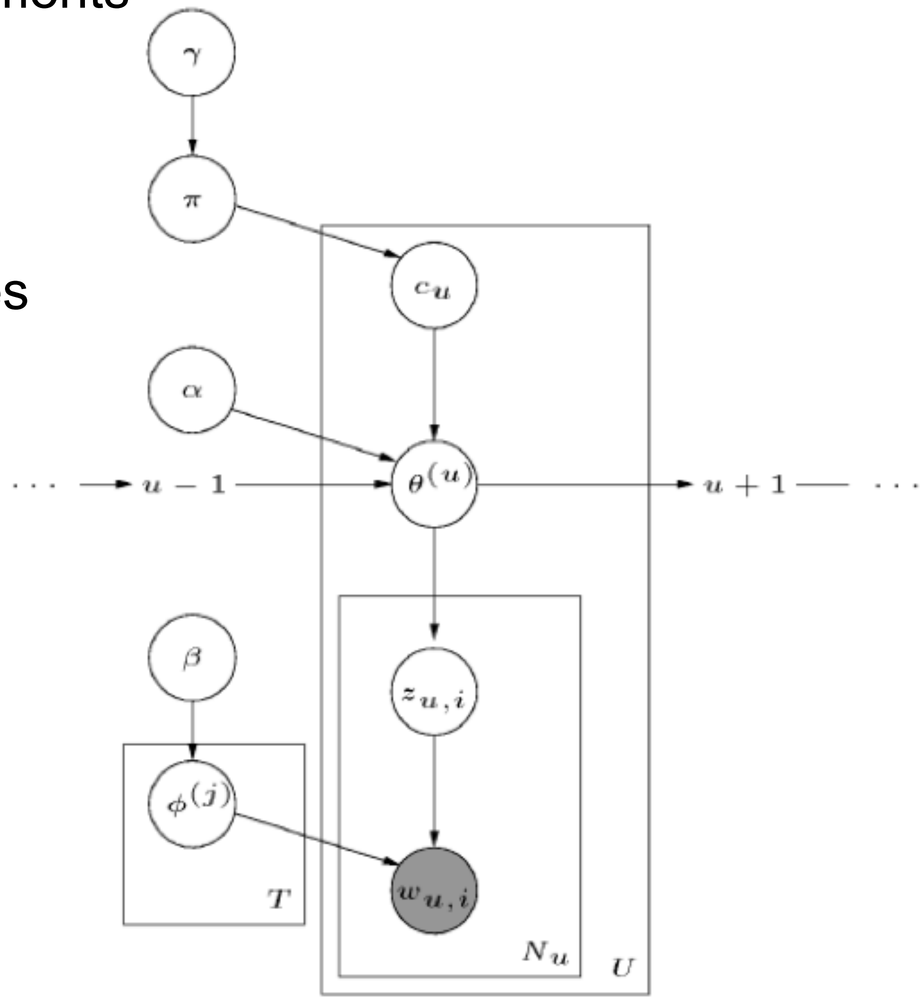
  - …
  - `sounds good`
  - `ok`
  - `so`
  - `what about next week's deadline?`
  - …

Topic segmentation

- Look for characteristics of **boundaries**
  - Key phrases
  - Pauses, overlaps, disfluencies
  - Speaker changes
  - Gesture/posture/gaze changes
  - Subtitles? Screen shifts?
  - e.g. (Beeferman et al, 1999)
- Look for **changes in vocabulary**
  - compare sliding windows e.g. (Hearst, 1993)
  - explicit sequence models e.g. HMMs (Tur et al, 2001)
- Combine the two
  - e.g. LCSeg (Galley et al, 2003)

Joint topic structure modelling

- Learn a model of topics & segmentation together
  - Dialogues as sequences of segments
  - Segments as mixtures of topics
  - Switching states
    - (Purver et al, 2006)
  - Associate with boundary features
    - (Dowman et al, 2008)



CIS centre for intelligent sensing

Schizophrenia dialogues

- Correlations with manually derived topics
- Correlations with outcomes
- Outcome prediction: adherence 66%, HAS Dr 76%

Topic 0	feel low alright mood long drug feeling tired time confidence con
Topic 4	voices pills mood cannabis telly voice shaking chris control inside
Topic 5	letter health advice letters council copy send dla cpn problems ho
Topic 7	church voice voices hear medication sister bad hearing taking felt
Topic 9	school children kids back september oclock gonna phone social s
Topic 10	weight months medication stone risk lose eat write gp hasnt exer
Topic 11	place support work centre gotta job stress feel psychologist theyl
Topic 12	door house police thought ring knew worse wall hadnt sat comin
Topic 13	doctor alright years nice ill anxious write long sit eye heart ring l
Topic 14	drug taking milligrams hundred doctor night time medication voi
Topic 15	sort medication work drugs kind team issues drink alcohol things
Topic 16	mum place brother tablets died dad depot house meet money liv
Topic 17	people life drug make care lot friends dry camera live cope thing
Topic 18	alright house drink drinking money alcohol god drugs living basic

Social media dialogue

- Topics around the Barbican on Twitter:
  - What people say & what people say to them

- exhibition, duchamp, bride, bachelors, new, dancing, #duchamp2013, enjoyed, 27, pop, glad, listen, artists, learn, fascinating, street, glass, preview, review, guy
- rain, room, #rainroom, queue, worth, hour, hours, random, #barbican, experience, international, centre, wait, quite, long, today, pic, actually, possible, recommend
- live, orchestra, april, bbc, moving, doing, symphony, london, concert, beethoven, performing, 70s, tonight, performed, 60s, highly, alexander, nevsky, review, musical
- music, hall, concert, play, london, sure, review, classical, remember, symphony, arts, orchestra, final, reviews, royal, tonight, summer, 14, trust, route
- richard, ii, playing, david, tennant, cage, rauschenberg, cunningham, hand, 10, johns, london, january, idea, company, #saharasoul, dates, price, design, center
- art, season, check, gallery, fun, head, duchamp, marcel, 25, tea, modern, 4pm, tate, email, em, lights, contemporary, cover, social, doran
- saturday, 20, soul, february, bar, sahara, trip, drinking, stories, spring, rd, #mali, blues, mali, village, today, 2013, road, jan, london
- theatre, want, dance, wow, world, opening, read, set, train, graduation, instead, 15, form, selling, today, page, magical, visit, kids, fri

Dialogue Structure

- What if we want to get more fine-grained?
  - opinions, questions & answers, agreements, decisions, …
- Searle, Austin (1960s): people ***do things*** *with words*
  - **speech acts / dialogue acts**

  - What's the capital of Burkina Faso?
  - Ouagadougou
  - Ouagadougou?
  - Right.

Dialogue Structure

- What if we want to get more fine-grained?
  - opinions, questions & answers, agreements, decisions, …
- Searle, Austin (1960s): people ***do things*** *with words*
  - **speech acts / dialogue acts**

  - I think we should go to the lecture.
  - No.

  - I don't think we should go to the lecture.
  - No.

We Need To Talk About Dialogue

- But what happens when we look at **dialogue**?
  - Human-computer dialogue
  - Human-human dialogue
  - Social media

- Dialogue is informal
  - do we know how people talk?
- Dialogue has structure
  - high-level topical structure
  - low-level dialogue structure
- Dialogue is incremental
  - people process language word by word

Dialogue Act Tagging

- Assign each utterance a functional tag
  - (i.e. the dialogue-level equivalent of part-of-speech tagging?)
- Multi-class classification
  - given some taxonomy e.g. ASK, ASSERT, CLARIFY, GREET …
  - features could be words, ngrams, etc, but also:
    - Syntax
    - Prosodic, acoustic
    - Context
- Sequence models (e.g. HMMs, CRFs)
  - Sequence is important e.g. ASK > ANSWER
- A well-known task: accuracies OK e.g. 60-80%, but:
  - depends on dataset
  - depends on DA taxonomy
  - rare classes do much worse, often 3-4% accuracy!

Dialogue

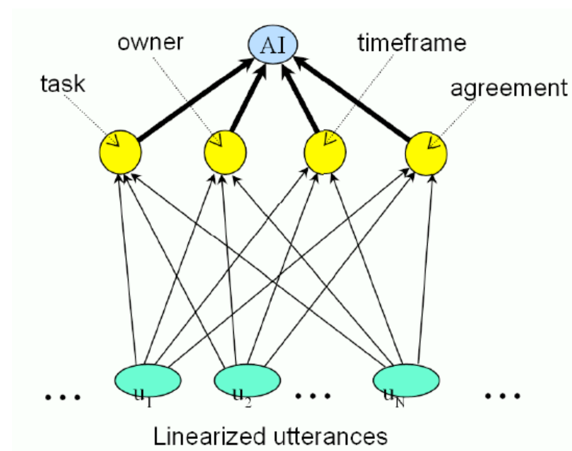
- We can have long-distance dependencies here too:
  - What's the capital of Burkina Faso?
    - Hmm. Bamako?
      - Isn't that in Mali?
      - Oh yes
    - No, Ouagadougou
      - Ouaga-what?
      - Ouagadougou
    - Oh OK
  - Yes that sounds right
- More complex models required e.g. (Ginzburg, 2011)'s KoS, (Asher & Lascarides, 2003)'s SDRT, (Poesio & Rieser)'s PTT
- But no robust computational application yet …

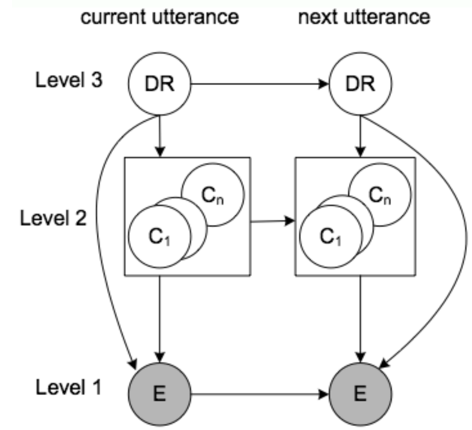
Decision Detection

A:	not really. So there was the notion of the preliminary patent that uh
B:	yeah it is a cheap patent
…	…
A:	yeah and it is really broad you er don’t have to
B:	yeah
C:	I actually think we should apply right away
D:	yeah I think that is a good idea
C:	I think you should I mean like this week start moving in that direction
…	…
A:	mhmm
D:	right

Decision Detection

- Classify utterances as decision-related?
  - Accuracy c. 30% (Hsueh & Moore, 2007; Morgan et al, 2006)
- Look for decision-related topic segments?
  - Accuracy c. 60%, but coarse-grained
- Use dialogue structure
  - Hierarchical classifiers c. 60%
    - (Fernandez et al, 2008)
  - Dynamic Bayesian Networks c.80%
    - (Bui & Peters, 2010)





Dialogue Systems

- If we can assign dialogue acts, we can have dialogues!
  - Sensible sequences of dialogue acts
  - With sensible content

```
– which film is showing at 8pm?
– at which cinema?
– the odeon
– avatar is showing at 8pm at the odeon
– how much is a ticket?
– 10 pounds
```

State-based systems & beyond

- Rule/state-based approaches:
  - VoiceXML
  - Information-state update
  - POMDPs
  - Robust, learnable (e.g. by reinforcement learning)

- Beyond state-based approaches:
  - Semantic representations
  - Information-state update rules
  - This means going back to deep, knowledge-rich methods …

We Need To Talk About Dialogue

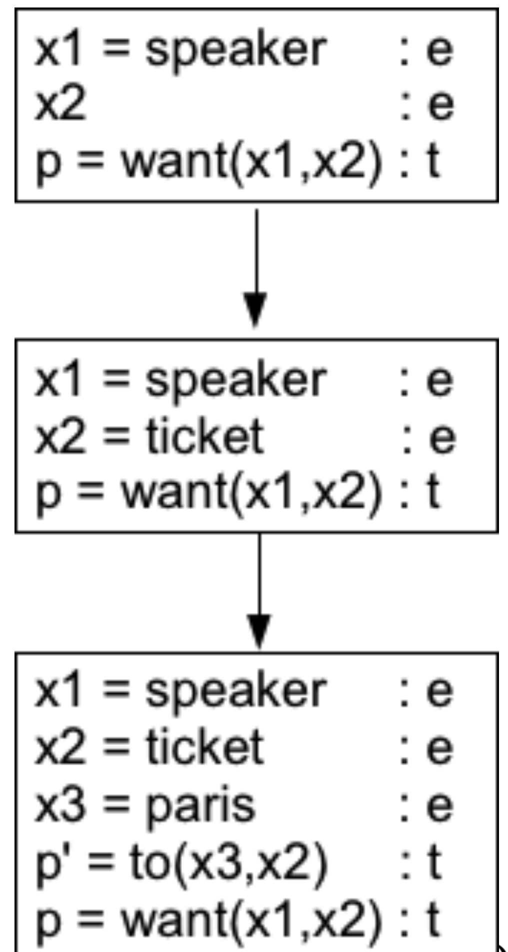
- But what happens when we look at **dialogue**?
  - Human-computer dialogue
  - Human-human dialogue
  - Social media

- Dialogue is informal
  - do we know how people talk?
- Dialogue has structure
  - high-level topical structure
  - low-level dialogue structure
- Dialogue is incremental
  - people process language word by word

Incrementality

- People don’t speak in complete sentences:

A: I want to go to er
B: yes
A: to London
B: London?
A: sorry no Paris, in March



Incremental Grammar Induction

- Induction from semantics
  - for an incremental grammar
  - with incremental learning
  - 80% coverage & accuracy
  - DYLAN system
  - (Eshghi et al, 2013)

x = john
y = mary
p = upset(x,y)