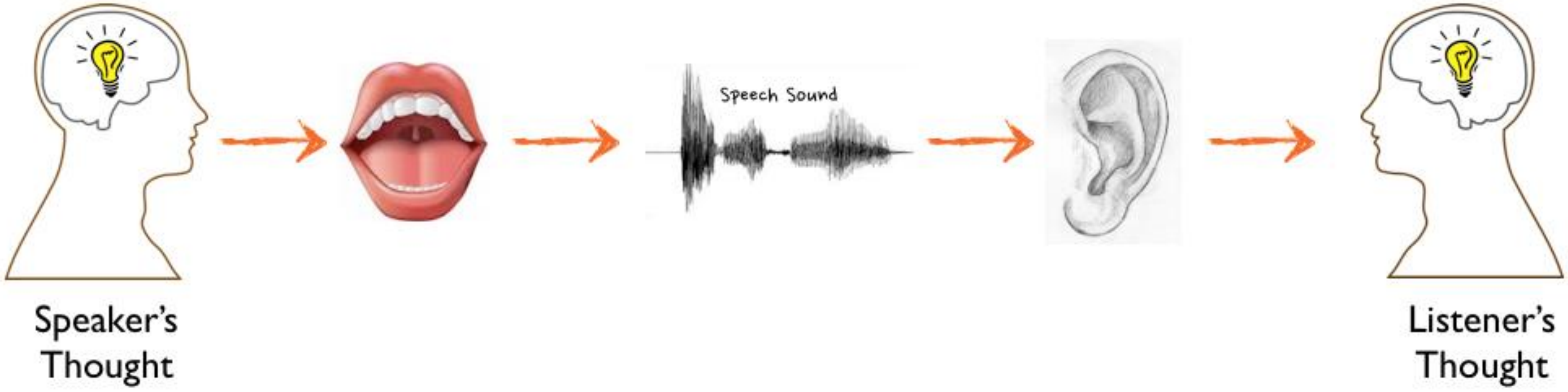
Semantics

What does it mean to have "meaning"?



source: http://www.mimicmethod.com/flow-101-day-1.html

What is meaning?

- **Compositional**

  Sentence meaning =

  function (meaning (word1), meaning(word2)…)

- **Holistic**

  Sentence meaning = function (context, word constraints)

- Key issue : do words have "meaning" ? [role of context]

Word meanings

- Ram fell through the window
- The window broke

APERTURE

PHYSICAL OBJECT

- Wordnet: window, N : 8 senses

- 1. (72) window -- (a framework of wood or metal that contains a glass windowpane and is built into a wall or roof to admit light or air)
- 2. (6) window -- (a transparent opening in a vehicle that allow vision out of the sides or back; usually is capable of being opened)
- 5. windowpane, window -- (a pane of glass in a window; "the ball shattered the window")
- 7. window -- (an opening in the wall of a building (usually to admit light and air); "he stuck his head in the window")

Sentences and Context

- a. John was going to commit suicide. GOAL
  b. He got the rope on Tuesday. PLAN

Sentences and Context

- a. The window broke — CAUSE

  b. Ram fell through it — CONSEQUENCE

- a. Sita saw Ravan.
- b. She greeted him.
- c. He asked for a glass. She gave it to him.

ANAPHORA = DISCOURSE REFERENTS

Lexical Semantics (Compositional)

- Words have a basic meaning, which is composed in sentences

- Sense variations : e.g.
- Bank = river's edge vs. Bank = financial institution

- Senses often run into one another
- E.g. window – as aperture or physical object
  newspaper – organization / object / information

Levels of semantics

- Language Processing Stack

Complexity Increases

Pragmatics

|

Discourse Semantics

|

Lexical Semantics

|

Syntax

|

Morphology

More World Knowledge

More Linguistic knowledge

Specification of Meaning

Many Meanings — Words — **Under specified**

Words Inside a Sentence

Words Inside a Discourse

Single Meaning — **Fully specified**

- other words in sentence context reduces meaning variation. (Composition)
- other sentences in discourse constrains sentence meaning. (Discourse)

Formal Models

Formal Semantics

- Declarative Sentences: Assign Truth Values
- Non-Declarative: inferential connections
- Interpretation function: Semantics of Words -> composition → semantics for complex expressions
  - Model-Theoretic: Map phrases / words → model
    - [Montague PTQ]
  - Truth-Theoretic: Conditions under which sentence is true. [Tarski, Davidson]

Model Theory

- Montague grammar :
  - Handles FRAGMENT of language
  - Syntax – define expression structure
  - Translation – into logical structure
  - Model-Theory : meanings as sets / individuals (PN) → Denotata
- Modern versions of Montague grammar – avoid “translation”

Montagovian Translation [1973]

A student sleeps

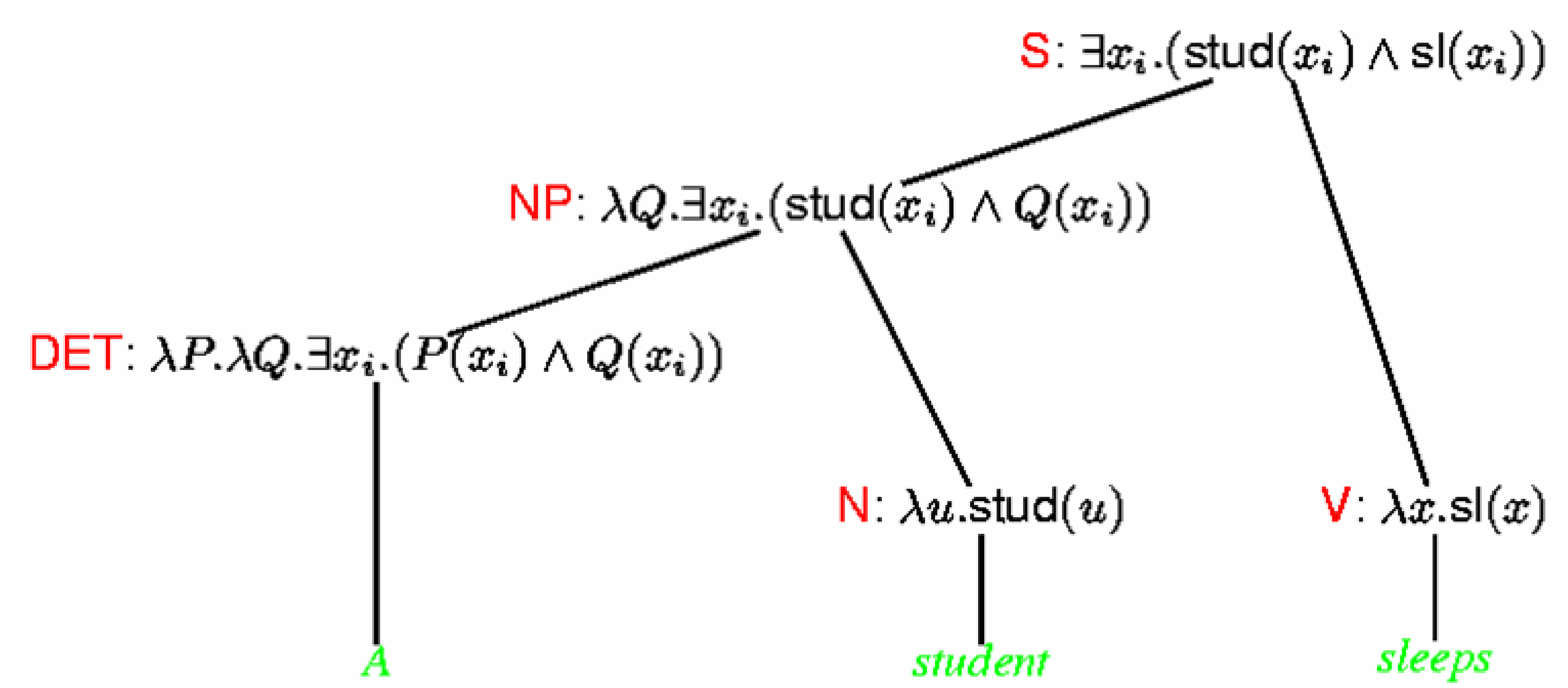
Lexicon:

student, N: $\lambda u.\mathsf{stud}(u)$

sleep, V : $\lambda x.\mathsf{sl}(x)$

a, DET : $\lambda P.\lambda Q.\exists x_i.(P(x_i) \wedge Q(x_i))$

Montagovian Translation [1973]



[Kohlhase]

The role of Context

- Charles Morris and Rudolf Carnap: 3-fold division of the theory of language:
  - syntax : relations between expressions
  - semantics: relations between expressions and what they stand for
  - pragmatics: relations between expressions and those who use it
- [Peregrin 98]
  - Internal Challenge (deictic - demonstrative/ anaphora)
  - External Challenge (function rather than designation)

Commitment of Grammar

Cognitive Grammar:

- Try to make sense of
  - polysemy (systematically related linguistic forms),
  - inference,
  - historical change,
  - gesture,
  - language acquisition
  - iconicity in signed languages.

[Lakoff/Johnson p.80]

Semantic Lexicons

Frame Elements for frame Ingestion

Frame Elements	Type
Degree	Peripheral
Ingestibles	Core
Ingestor	Core
Instrument	Peripheral
Manner	Peripheral
Means	Peripheral
Place	Peripheral
Source	Peripheral
Time	Peripheral

Lexical Units in : *Ingestion*

Lexical Units for Ingestion

English	Hindi	Bangla
breakfast.v	नाश्ता	prAtarAsh v
Consume.v	भोग करना	bhog k.v
drink.v	पी	khA.v
eat.v	खा	khA.v
feast.v	भोज करना	bhoj k .v
feed.v	खिला	khAoyA.v
gulp.v	निगल	gelA.v
have.v	ले	Neo.v
munch.v	चबा	chebA.v
nibble.v	कुतर	ThokrA.v
sip.n	घूंट	chumuk.n
sip.v	घूंट लेना	Chumuk de.v

Generative Lexicon

Traditional view: Adjective modifies noun

GL: Adj semantics is underspecified – is modified by noun semantics

e.g. fast car

fast lane

fast typist

Sentiment Analysis

Positive or negative movie review?

- unbelievably disappointing

- Full of zany characters and richly applied satire, and some great plot twists

- this is the greatest screwball comedy ever filmed

- It was pathetic. The worst part about it was the boxing scenes.

Google Product Search

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

$89 online, **$100** nearby ★★★★☆ 377 reviews

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

Reviews

Summary - Based on 377 reviews

1 star	2	3	4 stars	5 stars

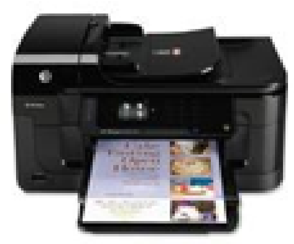
What people are saying

ease of use	"This was very easy to setup to four computers."
value	"Appreciate good quality at a fair price."
setup	"Overall pretty easy setup."
customer service	"I DO like honest tech support people."
size	"Pretty Paper weight."
mode	"Photos were fair on the high quality mode."
colors	"Full color prints came out with great quality."

Bing Shopping

HP Officejet 6500A E710N Multifunction Printer

Product summary | Find best price | **Customer reviews** | Specifications | Related items



$121.53 - $242.39 (14 stores)

☐ Compare

Average rating ★★★½☆ (144)

★★★★★	(55)
★★★★☆	(54)
★★★☆☆	(10)
★★☆☆☆	(6)
★☆☆☆☆	(23)
☆☆☆☆☆	(0)

Most mentioned

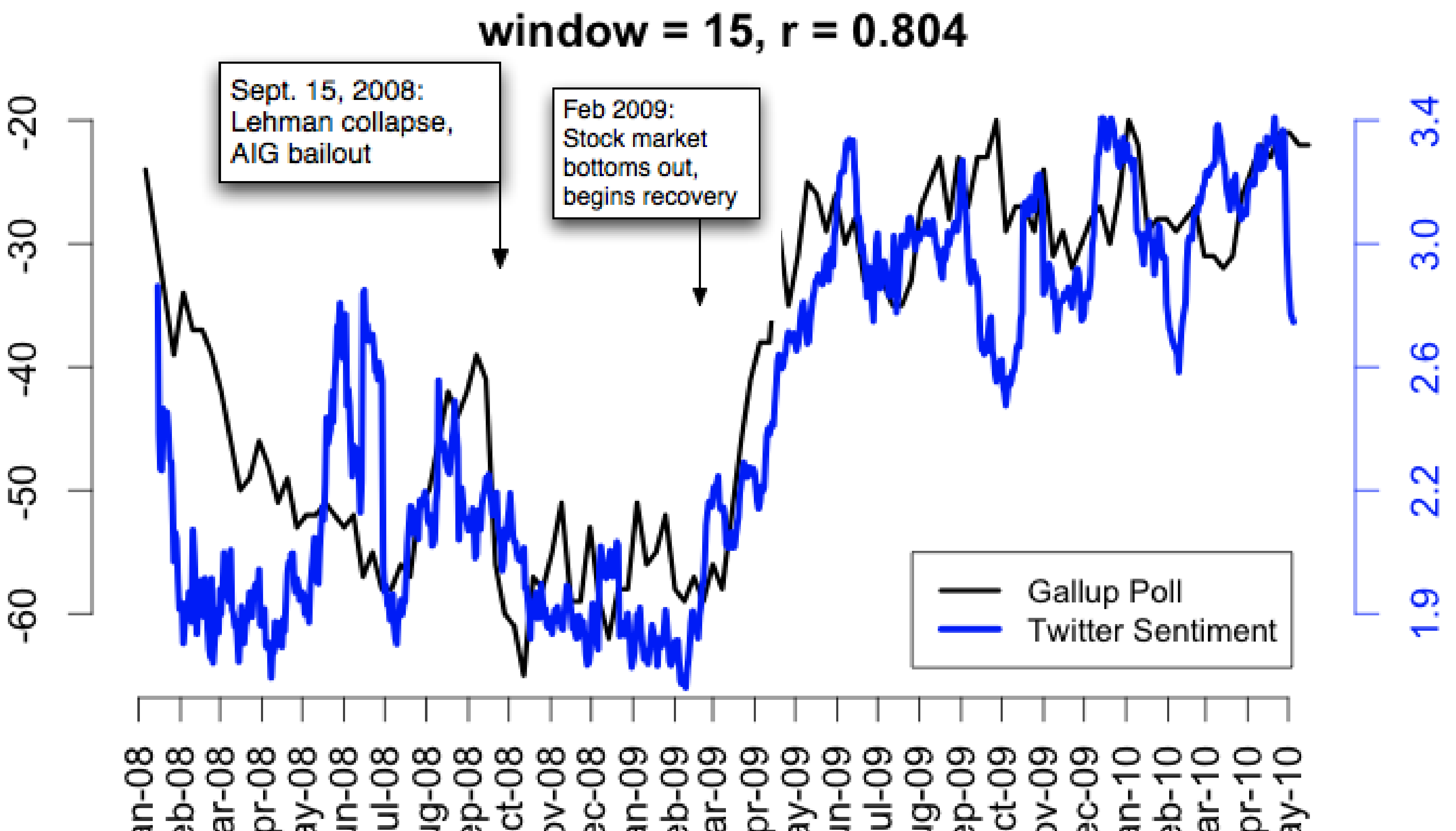
Performance	(57)
Ease of Use	(43)
Print Speed	(39)
Connectivity	(31)
More ▼	

Show reviews by source

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

Twitter sentiment versus Gallup Poll of Consumer Confidence

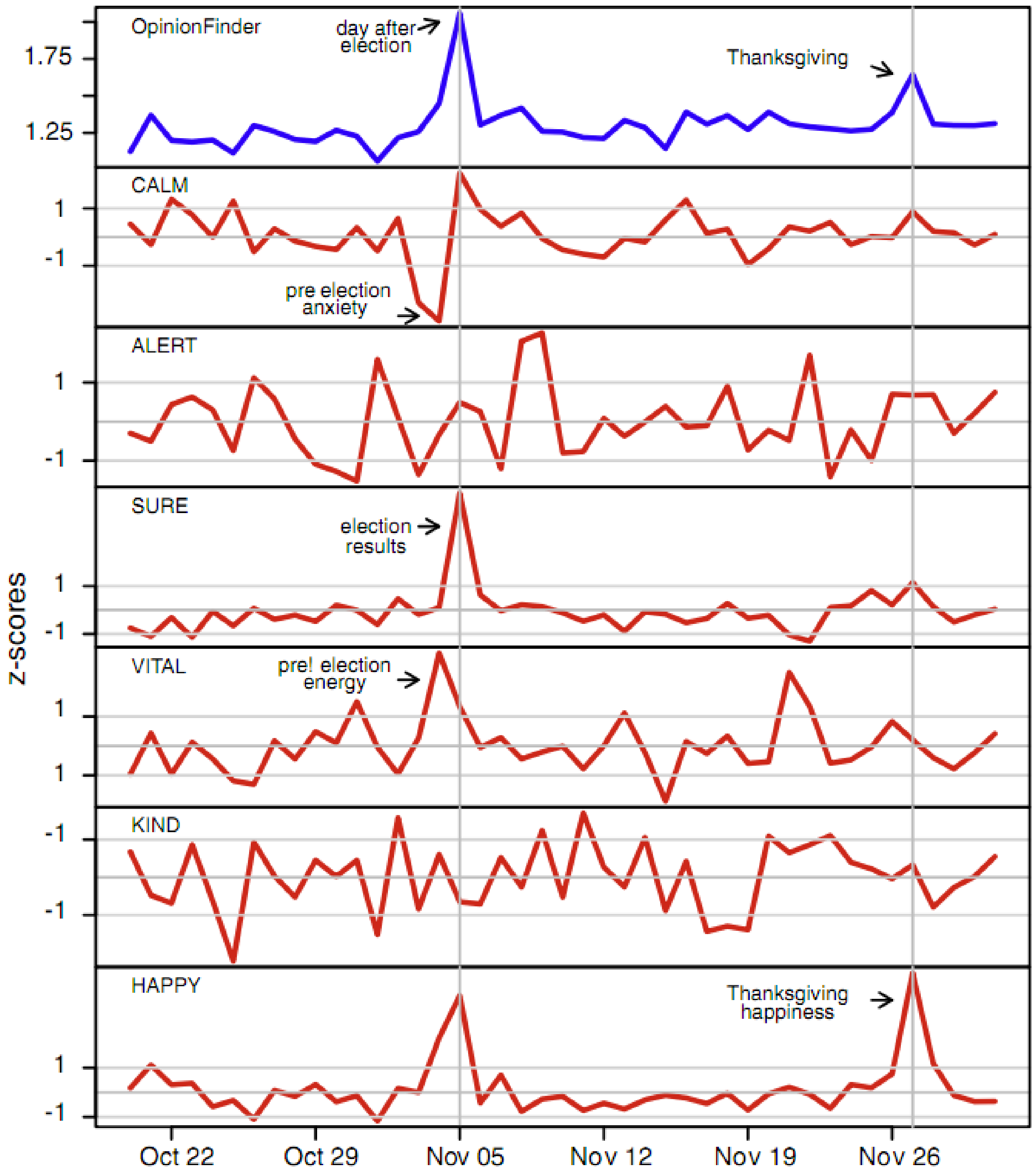
Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In ICWSM-2010



Twitter sentiment:

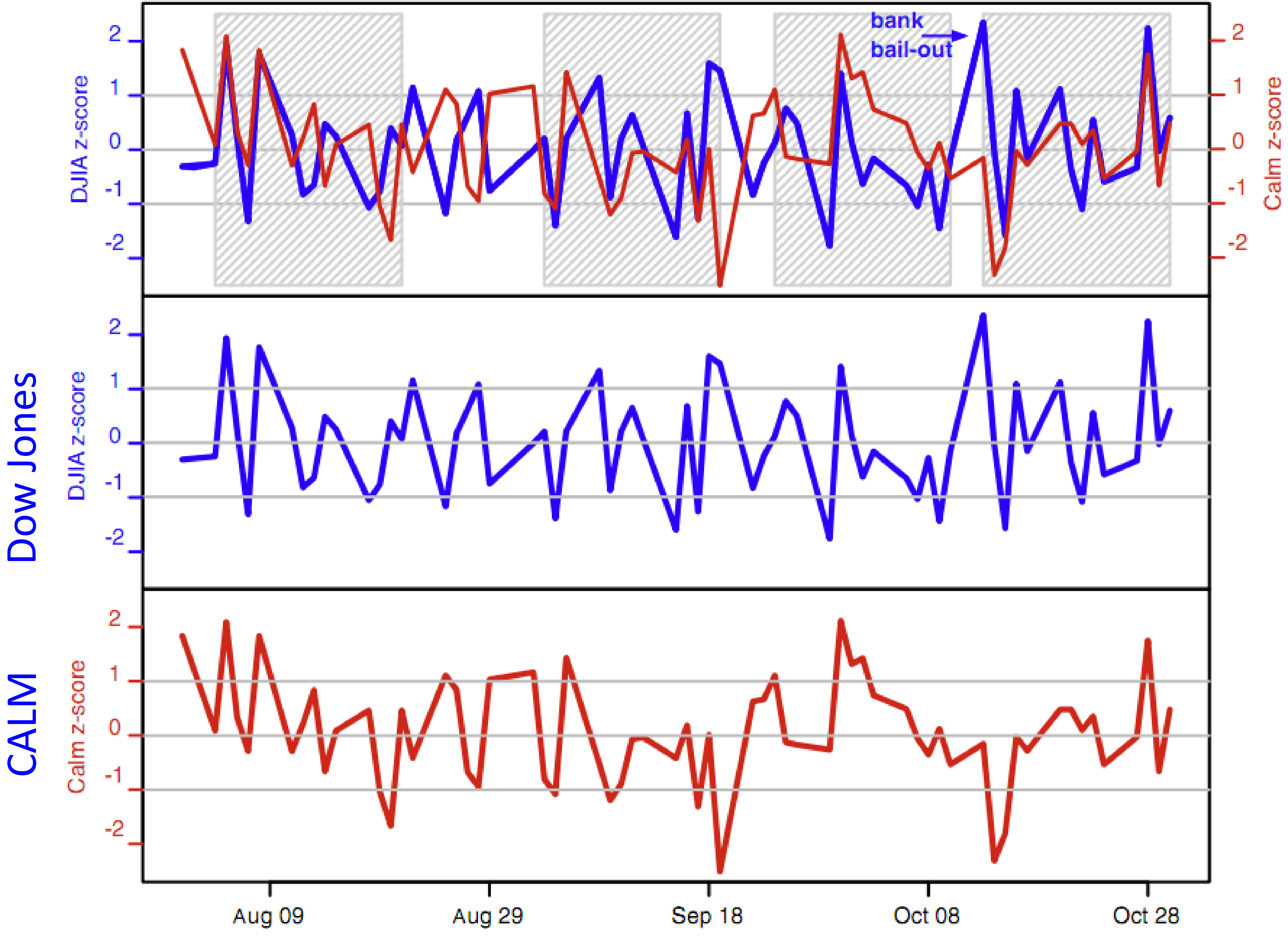
Johan Bollen, Huina Mao, Xiaojun Zeng. 2011. [Twitter mood predicts the stock market,](#)

Journal of Computational Science 2:1, 1-8. 10.1016/j.jocs.2010.12.007.



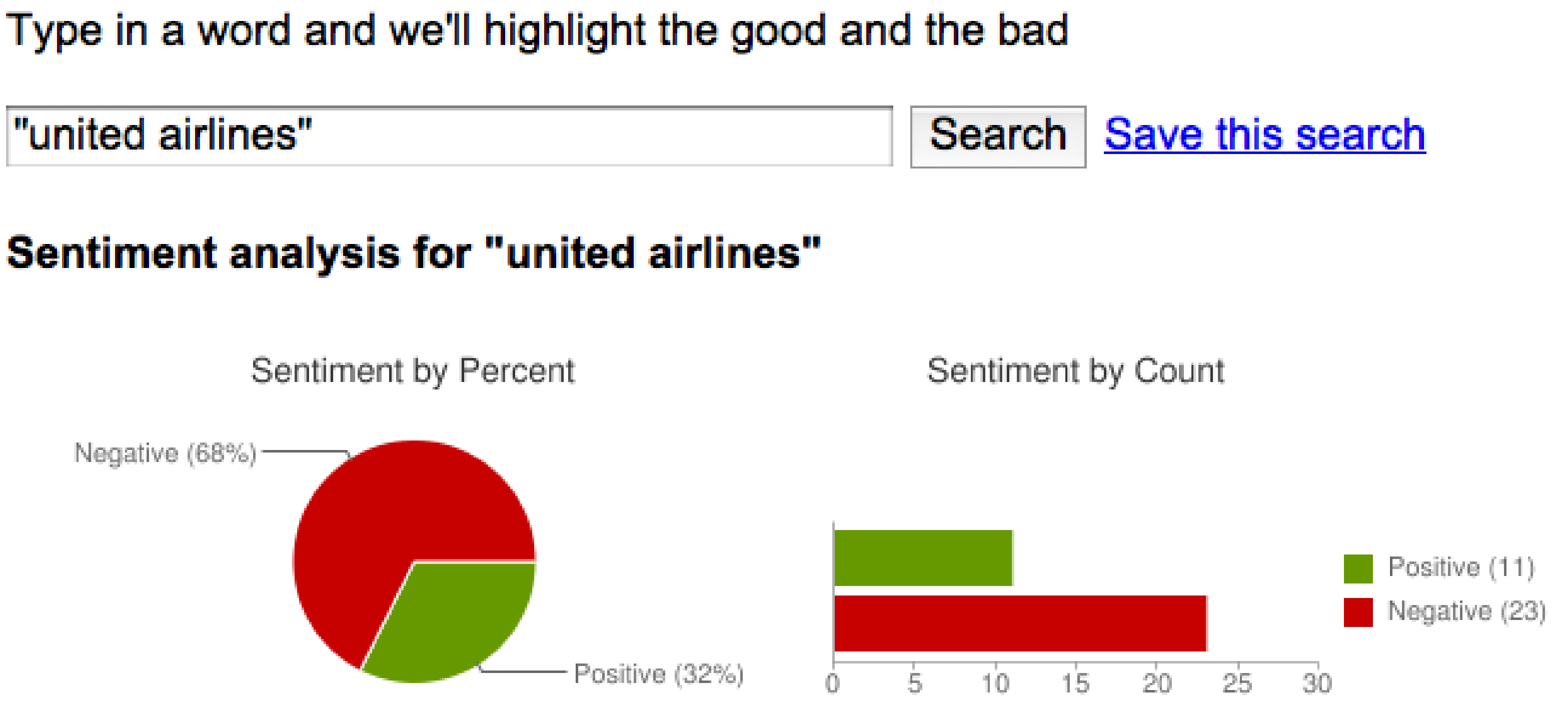
Bollen et al. (2011)

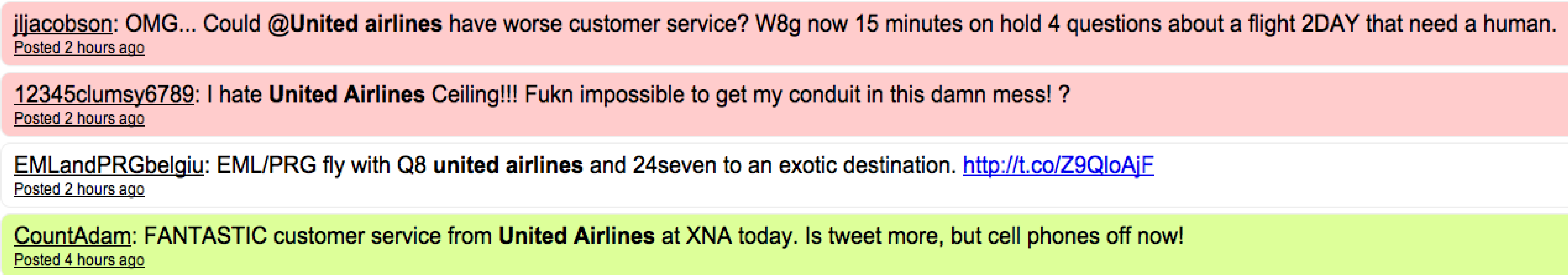
- CALM predicts DJIA 3 days later
- At least one current hedge fund uses this algorithm



Target Sentiment on Twitter

- [Twitter Sentiment App]
- Alec Go, Richa Bhayani, Lei Huang. 2009. Twitter Sentiment Classification using Distant Supervision





Sentiment analysis has many other names

- Opinion extraction
- Opinion mining
- Sentiment mining
- Subjectivity analysis

Scherer Typology of Affective States

- **Emotion**: brief organically synchronized … evaluation of a major event
  - *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood**: diffuse non-caused low-intensity long-duration change in subjective feeling
  - *cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances**: affective stance toward another person in a specific interaction
  - *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*
- **Attitudes**: enduring, affectively colored beliefs, dispositions towards objects or persons
  - *liking, loving, hating, valuing, desiring*
- **Personality traits**: stable personality dispositions and typical behavior tendencies
  - *nervous, anxious, reckless, morose, hostile, jealous*

Scherer Typology of Affective States

- **Emotion**: brief organically synchronized … evaluation of a major event
  - *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood**: diffuse non-caused low-intensity long-duration change in subjective feeling
  - *cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances**: affective stance toward another person in a specific interaction
  - *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*
- **Attitudes: enduring, affectively colored beliefs, dispositions towards objects or persons**
  - *liking, loving, hating, valuing, desiring*
- **Personality traits**: stable personality dispositions and typical behavior tendencies
  - *nervous, anxious, reckless, morose, hostile, jealous*

Sentiment Analysis

- Sentiment analysis is the detection of **attitudes**

  "enduring, affectively colored beliefs, dispositions towards objects or persons"

  1. **Holder (source)** of attitude
  2. **Target (aspect)** of attitude
  3. **Type** of attitude
     - From a set of types
       - *Like, love, hate, value, desire,* etc.
     - Or (more commonly) simple weighted **polarity**:
       - *positive, negative, neutral,* together with *strength*
  4. **Text** containing the attitude
     - Sentence or entire document

Sentiment Analysis

- Simplest task:
  - Is the attitude of this text positive or negative?
- More complex:
  - Rank the attitude of this text from 1 to 5
- Advanced:
  - Detect the target, source, or complex attitude types

Sentiment Analysis

A Baseline Algorithm

Sentiment Classification in Movie Reviews

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP-2002, 79—86.
Bo Pang and Lillian Lee. 2004. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. ACL, 271-278

- Polarity detection:
  - Is an IMDB movie review positive or negative?
- Data: *Polarity Data 2.0:*
  - http://www.cs.cornell.edu/people/pabo/movie-review-data

Baseline Algorithm (adapted from Pang and Lee)

- Tokenization
- Feature Extraction
- Classification using different classifiers
  - Naïve Bayes
  - MaxEnt
  - SVM

Sentiment Tokenization Issues

- Deal with HTML and XML markup
- Twitter mark-up (names, hash tags)
- Capitalization (preserve for words in all caps)
- Phone numbers, dates
- Emoticons
- Useful code:
  - Christopher Potts sentiment tokenizer
  - Brendan O'Connor twitter tokenizer

Potts emoticons

```
[<>]?                       # optional hat/brow
[:;=8]                      # eyes
[\-o\*\']?                  # optional nose
[\)\]\(\[dDpP/\:\}\{@\|\\]  # mouth
|                           #### reverse orientation
[\)\]\(\[dDpP/\:\}\{@\|\\]  # mouth
[\-o\*\']?                  # optional nose
[:;=8]                      # eyes
[<>]?                       # optional hat/brow
```

Extracting Features for Sentiment Classification

- How to handle negation
  - `I` **`didn't`** `like this movie`

    vs
  - `I really like this movie`
- Which words to use?
  - Only adjectives
  - All words
    - All words turns out to work better, at least on this data

Negation

Das, Sanjiv and Mike Chen. 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In Proceedings of the Asia Pacific Finance Association Annual Conference (APFA).

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP-2002, 79—86.

Add NOT_ to every word between negation and following punctuation:

```
didn't like this movie , but I
```

```
didn't NOT_like NOT_this NOT_movie but I
```

Reminder: Naïve Bayes

$$c_{NB} = \underset{c_j \in C}{\operatorname{argmax}} P(c_j) \prod_{i \in positions} P(w_i \mid c_j)$$

$$\hat{P}(w \mid c) = \frac{count(w,c)+1}{count(c)+|V|}$$

Binarized (Boolean feature) Multinomial Naïve Bayes

- Intuition:
  - For sentiment (and probably for other text classification domains)
  - Word occurrence may matter more than word frequency
    - The occurrence of the word *fantastic* tells us a lot
    - The fact that it occurs 5 times may not tell us much more.
  - Boolean Multinomial Naïve Bayes
    - Clips all the word counts in each document at 1

Boolean Multinomial Naïve Bayes: Learning

- From training corpus, extract *Vocabulary*

- Calculate $P(c_j)$ terms
  - For each c_j in C do

    $docs_j \leftarrow$ all docs with class $= c_j$

$$P(c_j) \neg \frac{|docs_j|}{|\text{total \# documents}|}$$

- Calculate $P(w_k \mid c_j)$ terms
  - Remove duplicates in each doc:
    - For each word type w in doc_j
      - Retain only a single instance of w
  - $Text_j \leftarrow$ single doc containing all $docs_j$
  - For each word w_k in *Vocabulary*

    $n_k \leftarrow$ # of occurrences of w_k in $Text_j$

$$P(w_k \mid c_j) \neg \frac{n_k + a}{n + a|Vocabulary|}$$

Boolean Multinomial Naïve Bayes on a test document *d*

- First remove all duplicate words from *d*
- Then compute NB using the same equation:

$$c_{NB} = \underset{c_j \in C}{\operatorname{argmax}} P(c_j) \prod_{i \in positions} P(w_i \mid c_j)$$

Normal vs. Boolean Multinomial NB

Normal	Doc	Words	Class
Training	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Test	5	Chinese Chinese Chinese Tokyo Japan	?

Boolean	Doc	Words	Class
Training	1	Chinese Beijing	c
	2	Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Test	5	Chinese Tokyo Japan	?

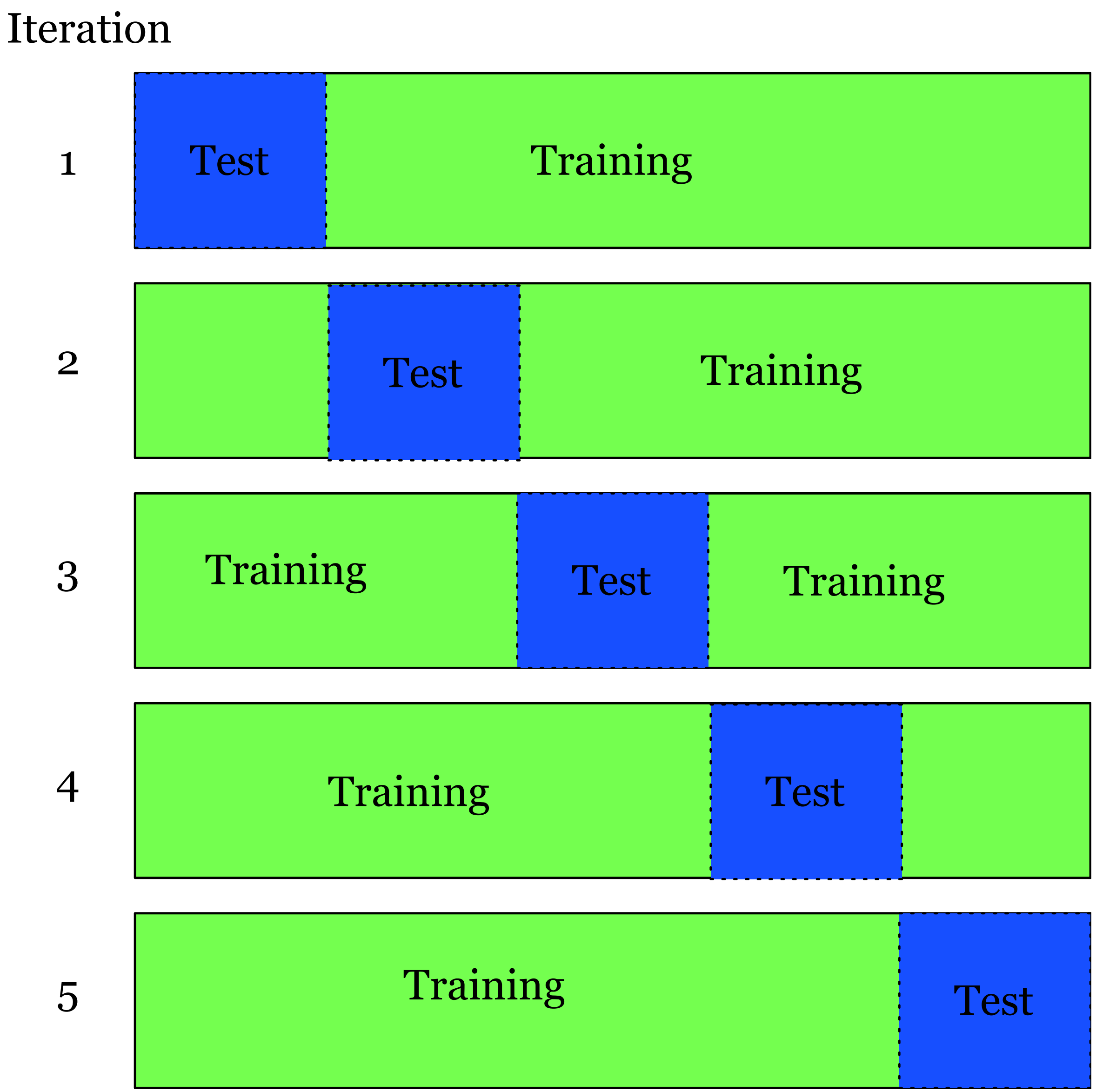
Binarized (Boolean feature) Multinomial Naïve Bayes

B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP-2002, 79—86.
V. Metsis, I. Androutsopoulos, G. Paliouras. 2006. Spam Filtering with Naive Bayes – Which Naive Bayes? CEAS 2006 - Third Conference on Email and Anti-Spam.
K.-M. Schneider. 2004. On word frequency information and negative evidence in Naive Bayes text classification. ICANLP, 474-485.
JD Rennie, L Shih, J Teevan. 2003. Tackling the poor assumptions of naive bayes text classifiers. ICML 2003

- Binary seems to work better than full word counts
  - This is **not** the same as Multivariate Bernoulli Naïve Bayes
    - MBNB doesn't work well for sentiment or other text tasks
- Other possibility: log(freq(*w*))

Cross-Validation

- Break up data into 10 folds
  - (Equal positive and negative inside each fold?)
- For each fold
  - Choose the fold as a temporary test set
  - Train on 9 folds, compute performance on the test fold
- Report average performance of the 10 runs



Other issues in Classification

- MaxEnt and SVM tend to do better than Naïve Bayes

Problems:
What makes reviews hard to classify?

- Subtlety:
  - Perfume review in *Perfumes: the Guide*:
    - "If you are reading this because it is your darling fragrance, please wear it at home exclusively, and tape the windows shut."
  - Dorothy Parker on Katherine Hepburn
    - "She runs the gamut of emotions from A to B"

Thwarted Expectations and Ordering Effects

- "This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it **can't hold up**."
- Well as usual Keanu Reeves is nothing special, but surprisingly, the very talented Laurence Fishbourne is **not so good** either, I was surprised.

Sentiment Lexicons : Disagreements

Christopher Potts, Sentiment Tutorial, 2011

	Opinion Lexicon	General Inquirer	SentiWordNet	LIWC
MPQA	33/5402 **(0.6%)**	49/2867 **(2%)**	1127/4214 **(27%)**	12/363 **(3%)**
Opinion Lexicon		32/2411 **(1%)**	1004/3994 **(25%)**	9/403 **(2%)**
General Inquirer			520/2306 **(23%)**	1/204 **(0.5%)**
SentiWordNet				174/694 **(25%)**
LIWC				

Sentiment via bag of Words

Analyzing the polarity of each word in IMDB

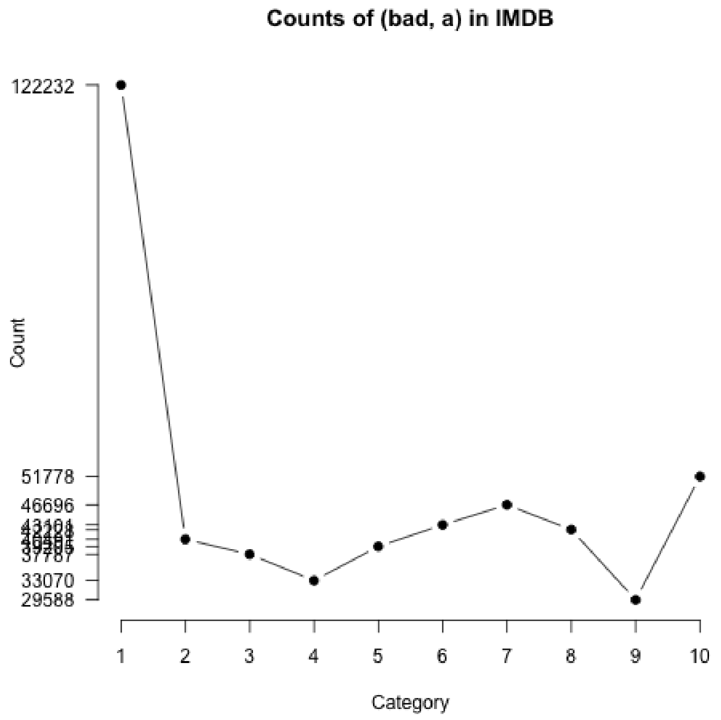
Potts, Christopher. 2011. On the negativity of negation. SALT 20, 636-659.

- How likely is each word to appear in each sentiment class?
- Count("bad") in 1-star, 2-star, 3-star, etc.
- But can't use raw counts:
- Instead, **likelihood:**

$$P(w|c) = \frac{f(w,c)}{\sum_{w \in c} f(w,c)}$$

- Make them comparable between words
  - **Scaled likelihood:**

$$\frac{P(w|c)}{P(w)}$$



Polarity analysis

बजट की कमी से फिल्म मनोरंजक नहीं हो पाई है।

अगर मीडिया में आरक्षण फिल्म के बहाने ठोस बहस आरंभ होती तो सोच-विचार को नए आयाम मिलते, लेकिन हम फिजूल विवादों में उलझ कर रह गए।

जन अभिनय का उनका यह अभियान प्रशंसनीय है।

Polarity analysis : datasets

- IMDB Movie Reviews

  Pos: 25,000

  Neg : 25,000

  Unlabeled : 50,000

- Amazon Product Reviews

  Watches : 30.8mb [68.4K reviews]

  Electronics : 728mb [1242K]

  MP3 : 27.7MB [31K]

- Hindi film reviews: 700 reviews

80-20 ratio for training and testing

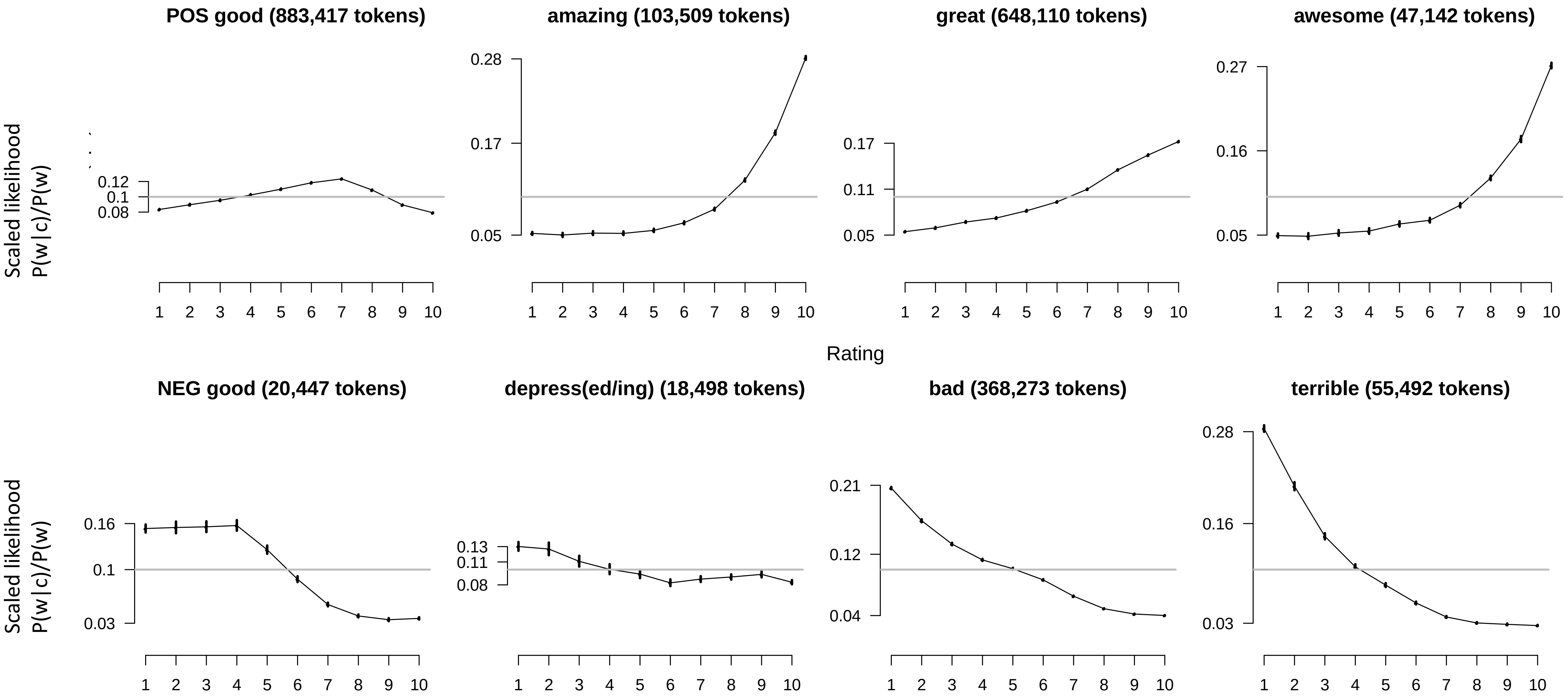
Document Modeling : tf-idf

Term Frequency-Inverse Document Frequency(tf-idf) Model

- Document d_i represented by $v_{d_i} \in \mathbb{R}^{|V|}$
- Each element in v_{d_i} is the product of term frequency and inverse document frequency: $tfidf(t, d) = tf(t, d) \times \log(\frac{\|D\|}{df(t)})$
- Gives weights to terms which are less frequent and hence important
- **Drawbacks:**
    - High-dimensionality
    - Ignores word ordering
    - Ignores word context
    - Very sparse

Analyzing the polarity of each word in IMDB

Potts, Christopher. 2011. On the negativity of negation. SALT 20, 636-659.

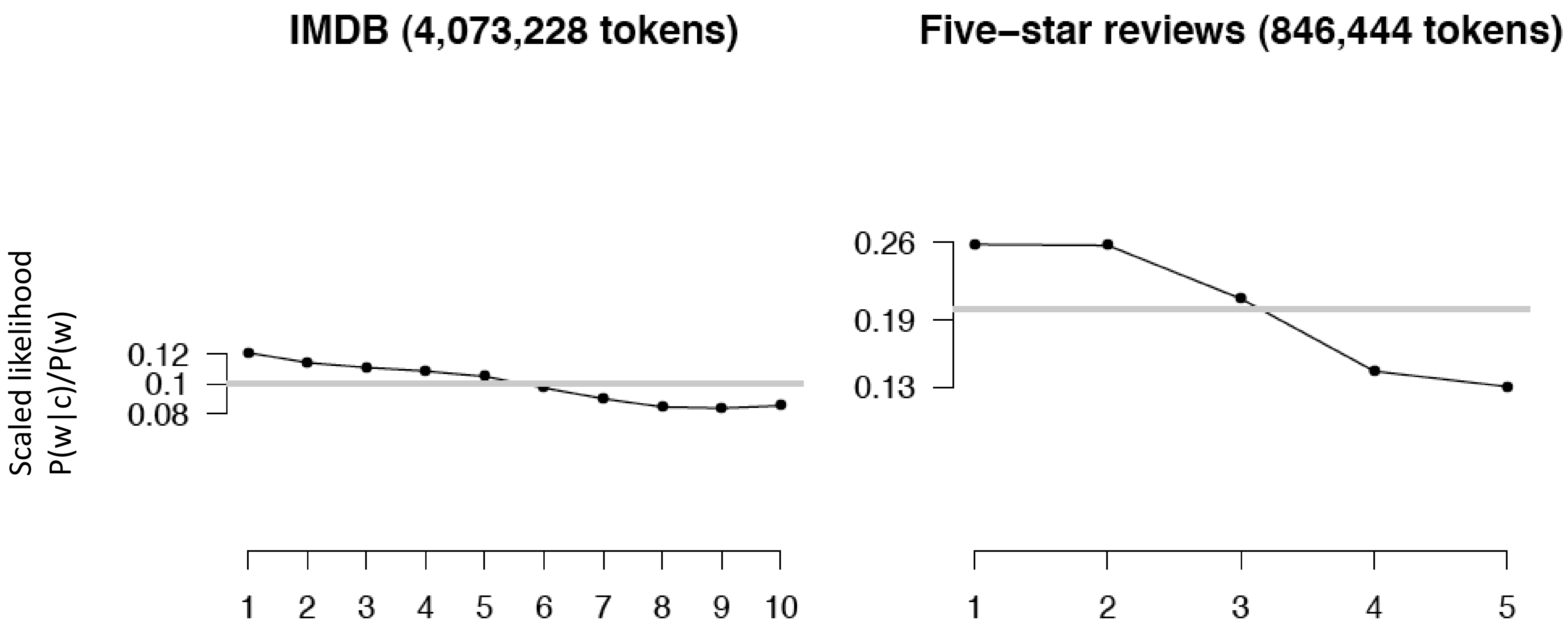


Other sentiment feature: Logical negation

Potts, Christopher. 2011. On the negativity of negation. SALT 20, 636-659.

- Is logical negation (*no, not*) associated with negative sentiment?
- Potts experiment:
  - Count negation (*not, n’t, no, never*) in online reviews
  - Regress against the review rating

Potts 2011 Results:
More negation in negative sentiment

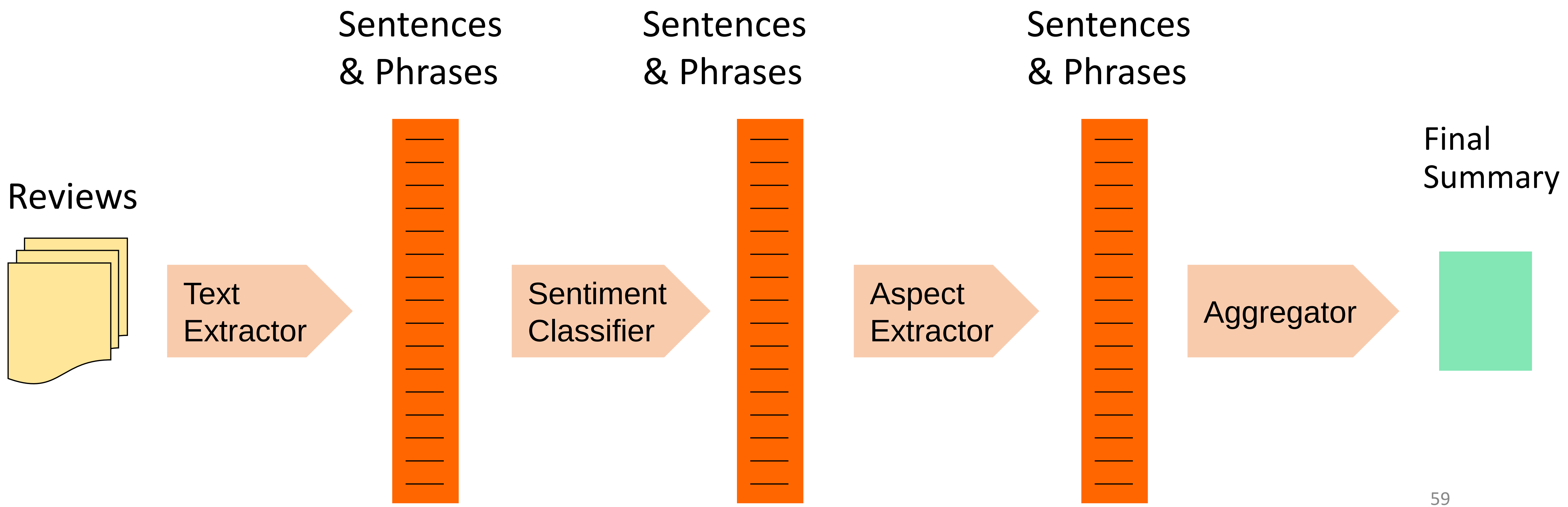


Finding aspect/attribute/target of sentiment

- The aspect name may not be in the sentence
- For restaurants/hotels, aspects are well-understood
- Supervised classification
  - Hand-label a small corpus of restaurant review sentences with aspect
    - food, décor, service, value, NONE
  - Train a classifier to assign an aspect to asentence
    - “Given this sentence, is the aspect *food, décor, service, value,* or *NONE*”

Putting it all together: Finding sentiment for aspects

S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. Reis, and J. Reynar. 2008. Building a Sentiment Summarizer for Local Service Reviews. WWW Workshop



Baseline methods assume classes have equal frequencies!

- If not balanced (common in the real world)
  - can’t use accuracies as an evaluation
  - need to use F-scores
- Severe imbalancing also can degrade classifier performance
- Two common solutions:
  1. Resampling in training
     - Random undersampling
  2. Cost-sensitive learning
     - Penalize SVM more for misclassification of the rare thing

Summary on Sentiment

- Generally modeled as classification or regression task
  - predict a binary or ordinal label
- Features:
  - Negation is important
  - Using all words (in naïve bayes) works well for some tasks
  - Finding subsets of words may help in other tasks
    - Hand-built polarity lexicons
    - Use seeds and semi-supervised learning to induce lexicons

Scherer Typology of Affective States

- **Emotion**: brief organically synchronized … evaluation of a major event
  - *angry, sad, joyful, fearful, ashamed, proud, elated*
- **Mood**: diffuse non-caused low-intensity long-duration change in subjective feeling
  - *cheerful, gloomy, irritable, listless, depressed, buoyant*
- **Interpersonal stances**: affective stance toward another person in a specific interaction
  - *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*
- **Attitudes**: enduring, affectively colored beliefs, dispositions towards objects or persons
  - *liking, loving, hating, valuing, desiring*
- **Personality traits**: stable personality dispositions and typical behavior tendencies
  - *nervous, anxious, reckless, morose, hostile, jealous*

Computational work on other affective states

- **Emotion**:
  - Detecting annoyed callers to dialogue system
  - Detecting confused/frustrated versus confident students
- **Mood**:
  - Finding traumatized or depressed writers
- **Interpersonal stances**:
  - Detection of flirtation or friendliness in conversations
- **Personality traits**:
  - Detection of extroverts

Detection of Friendliness

Ranganath, Jurafsky, McFarland

- Friendly speakers use collaborative conversational style
  - Laughter
  - Less use of negative emotional words
  - More sympathy
    - `That's too bad` `I'm sorry to hear that`
  - More agreement
    - `I think so too`
  - Less hedges
    - `kind of` `sort of` `a little …`

Sentiment via Word Vectors

Word Vector Models

Distributed Representation of Words(Mikolov et al., 2013b)

- Each word $w_i \in V$ is represented using a vector $v_{w_i} \in \mathbb{R}^k$
- The vocabulary V can be represented by a matrix $V \in \mathbb{R}^{k \times |V|}$
- Vectors (v_{w_i}) should encode the semantics of the words in vocabulary
- **Drawbacks**:
    - Ignores exact word ordering
    - Cannot represent documents as vectors without composition

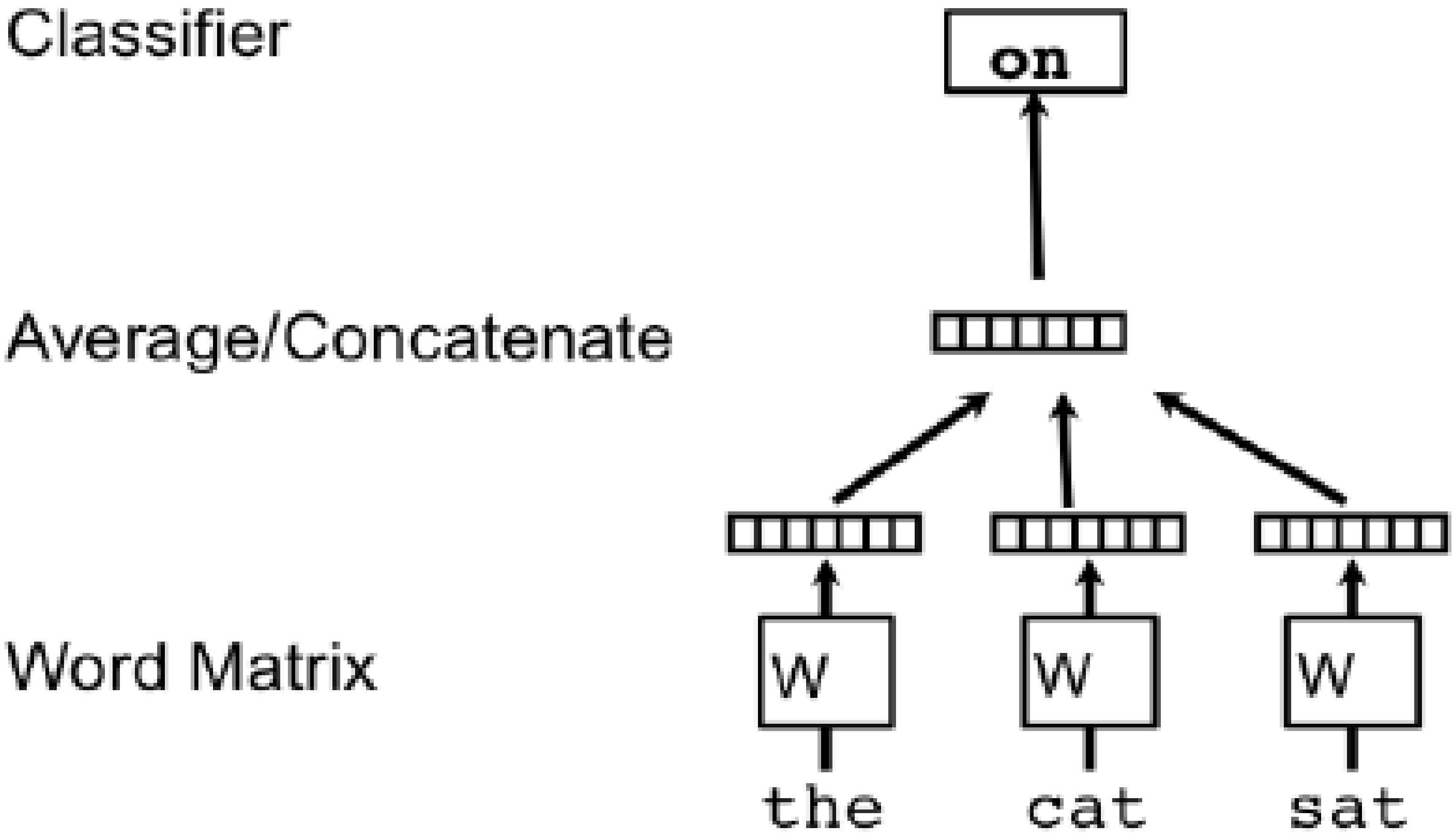
Vector Composition

जन अभिनय का उनका यह अभियान प्रशंसनीय है।

$$S(x) = c_1 w_1(x) \Theta c_2 w_2(x) \Theta c_3 w_3(x) \Theta c_4 w_4(x) \ldots \Theta c_k w_k(x)$$

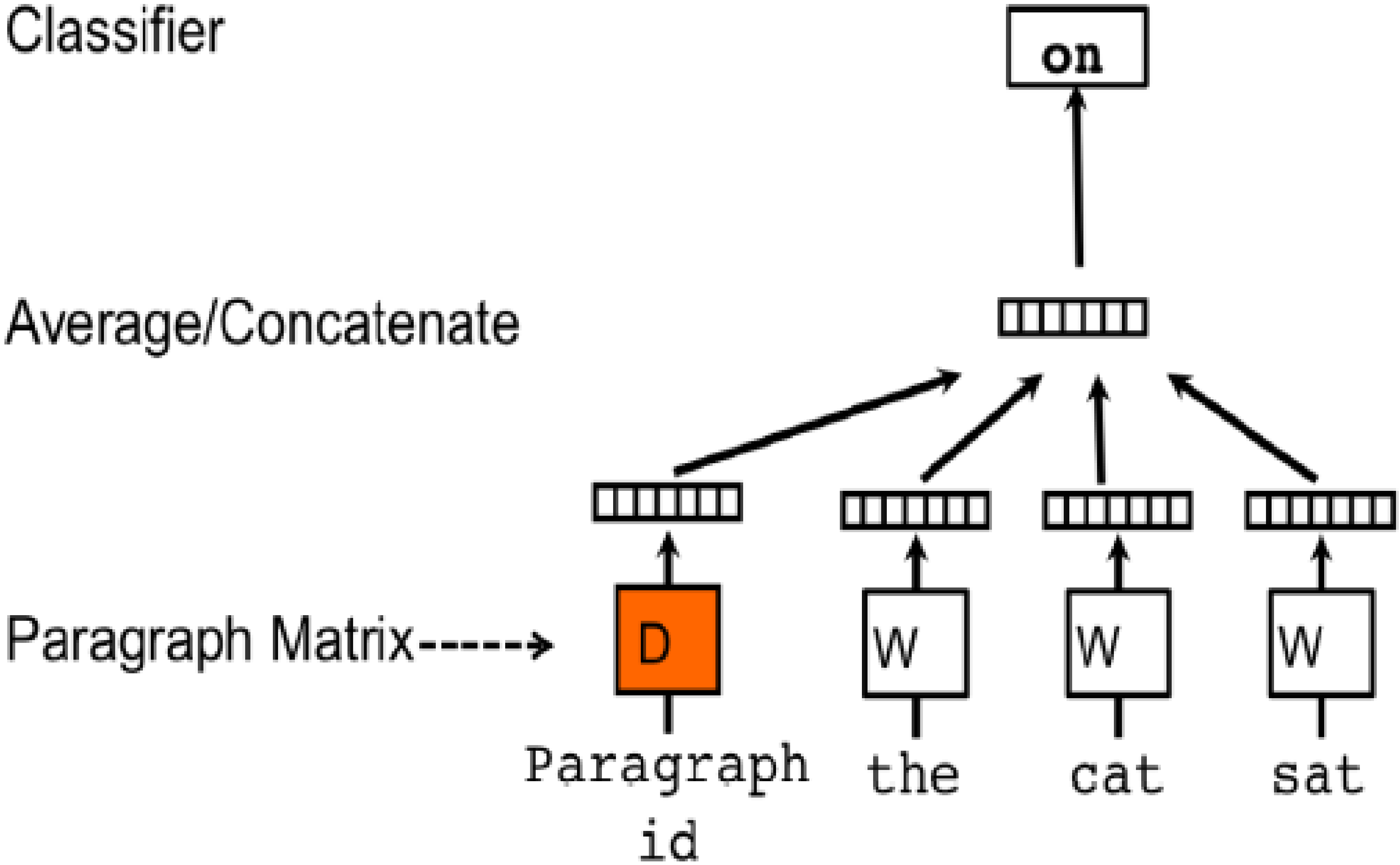
Composition	Accuracy
Average	88.42
Weighted Average	88.41
Multiplication	50.30

Paragraph Vector Models



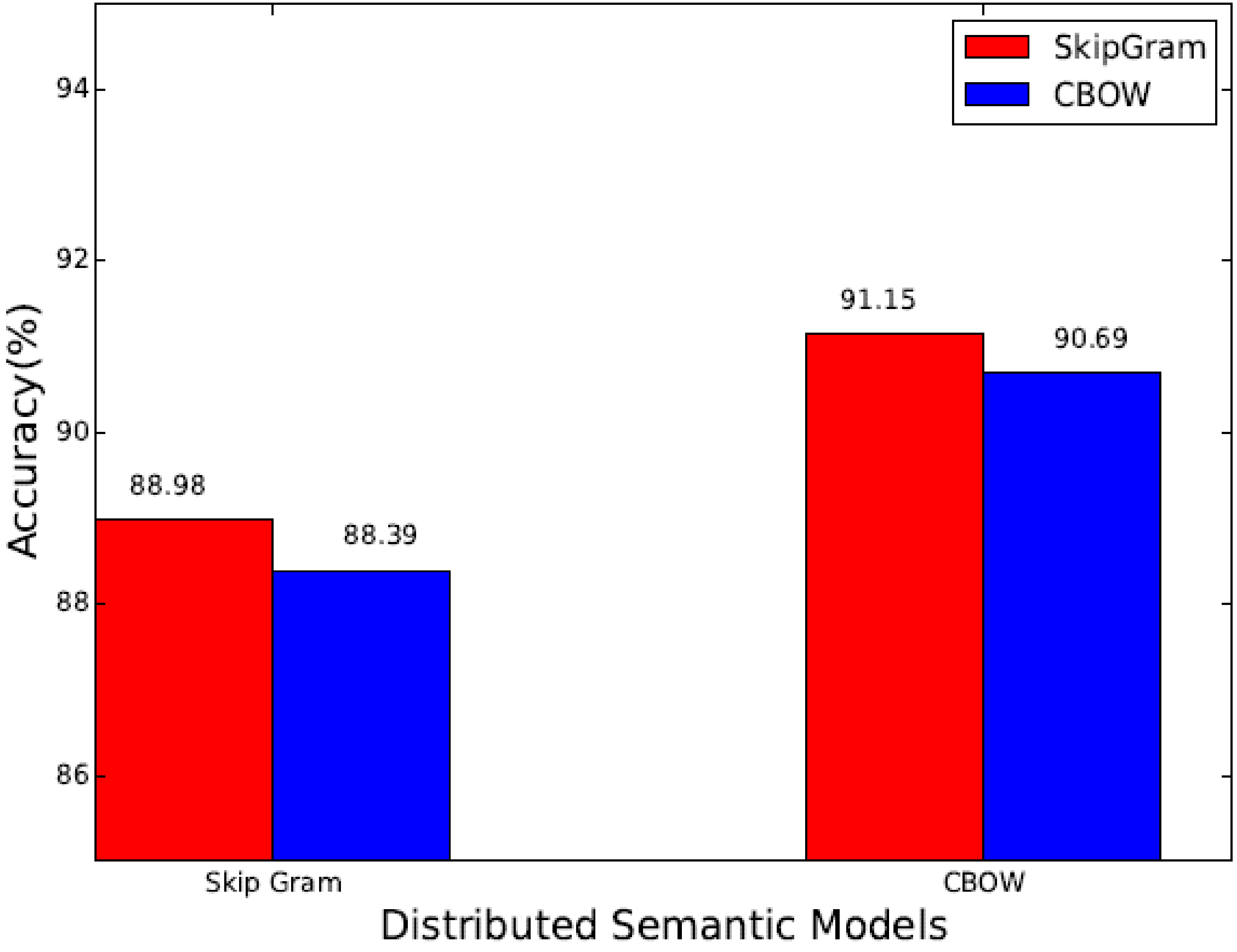
[Le and Mikolov 14]

Paragraph Vector Models



[Le and Mikolov 14]

Word2vec variants



Weighted average vs other models

Method	Accuracy
Maas et al.(2011)	88.89
NBSVM-bi (Wang & Manning, 2012)	91.22
NBSVM-uni (Wang & Manning, 2012)	88.29
SVM-uni (Wang & Manning, 2012)	89.16
Paragraph Vector (Le and Mikolov(2014))	92.58
WordVector+Wiki(Our Method)	88.60
WordVector+TfIdf(Our Method)	89.03
WordVector Averaging+TfIdf+Document Vector	**93.91**

Table 6.1: Results on IMDB Movie Review Dataset

[singh & mukerjee 15]

Semantic Role Labelling

Semantic Role Labeling

Who did *what* to *whom* at *where*?

The police officer detained the suspect at the scene of the crime

Agent | Predicate | Theme | Location

Paraphrasing

XYZ corporation **bought** the stock.

They **sold** the stock to XYZ corporation.

The stock was **bought** by XYZ corporation.

The **purchase** of the stock by XYZ corporation...

The stock **purchase** by XYZ corporation...

A Shallow Semantic Representation: Semantic Roles

Predicates (bought, sold, purchase) represent an **event**

semantic roles express the abstract role that arguments of a predicate can take in the event

More specific ⟷ More general

buyer **agent** **proto-agent**

Semantic Roles

Getting to semantic roles

Neo-Davidsonian event representation:

Sasha broke the window

Pat opened the door

$$\exists e,x,y\; Breaking(e) \wedge Breaker(e,Sasha) \wedge BrokenThing(e,y) \wedge Window(y)$$

$$\exists e,x,y\; Opening(e) \wedge Opener(e,Pat) \wedge OpenedThing(e,y) \wedge Door(y)$$

Subjects of break and open: **Breaker** and **Opener**

Deep roles specific to each event (breaking, opening)

Hard to reason about them for NLU applications like QA

Thematic roles

- **Breaker** and **Opener** have something in common!
  - Volitional actors
  - Often animate
  - Direct causal responsibility for their events
- Thematic roles are a way to capture this semantic commonality between *Breakers* and *Eaters*.
- They are both AGENTS.
- The *BrokenThing* and *OpenedThing*, are THEMES.
  - prototypically inanimate objects affected in some way by the action

Thematic roles

- One of the oldest linguistic models
  - Indian grammarian Panini between the 7th and 4th centuries BCE
- Modern formulation from Fillmore (1966,1968), Gruber (1965)
  - Fillmore influenced by Lucien Tesnière's (1959) *Éléments de Syntaxe Structurale,* the book that introduced dependency grammar
  - Fillmore first referred to roles as *actants* (Fillmore, 1966) but switched to the term *case*

Thematic roles

- A typical set:

Thematic Role	**Definition**	**Example**
AGENT	The volitional causer of an event	*The waiter* spilled the soup.
EXPERIENCER	The experiencer of an event	*John* has a headache.
FORCE	The non-volitional causer of the event	*The wind* blows debris from the mall into our yards.
THEME	The participant most directly affected by an event	Only after Benjamin Franklin broke *the ice*...
RESULT	The end product of an event	The city built a *regulation-size baseball diamond*...
CONTENT	The proposition or content of a propositional event	Mona asked *"You met Mary Ann at a supermarket?"*
INSTRUMENT	An instrument used in an event	He poached catfish, stunning them *with a shocking device*...
BENEFICIARY	The beneficiary of an event	Whenever Ann Callahan makes hotel reservations *for her boss*...
SOURCE	The origin of the object of a transfer event	I flew in *from Boston*.
GOAL	The destination of an object of a transfer event	I drove *to Portland*.

Thematic grid, case frame, θ-grid

Example usages of "break"

John *broke the window.*
AGENT THEME

John *broke the window with a rock.*
AGENT THEME INSTRUMENT

The rock *broke the window.*
INSTRUMENT THEME

The window broke.
THEME

The window was broken by John.
THEME AGENT

thematic grid, case frame, θ-grid

Break:

AGENT, THEME, INSTRUMENT.

Some realizations:

AGENT/Subject, THEME/Object
AGENT/Subject, THEME/Object, INSTRUMENT/PP_{with}
INSTRUMENT/Subject, THEME/Object
THEME/Subject

Diathesis alternations (or verb alternation)

Doris gave the book to Cary.

AGENT THEME GOAL

Doris gave Cary the book.

AGENT GOAL THEME

Break: AGENT, INSTRUMENT, or THEME as subject

Give: THEME and GOAL in either order

Dative alternation: particular semantic classes of verbs, “verbs of future having” (*advance*, *allocate*, *offer*, *owe*), “send verbs” (*forward*, *hand*, *mail*), “verbs of throwing” (*kick*, *pass*, *throw*), etc.

Levin (1993): 47 semantic classes (“**Levin classes**”) for 3100 English verbs and alternations. In online resource VerbNet.

Problems with Thematic Roles

Hard to create standard set of roles or formally define them

Often roles need to be fragmented to be defined.

- Levin and Rappaport Hovav (2015): two kinds of INSTRUMENTS
  - **intermediary instruments** that can appear as subjects
    - The cook opened the jar with the new gadget.
    - The new gadget opened the jar.
  - **enabling instruments** that cannot
    - Shelly ate the sliced banana with a fork.
    - *The fork ate the sliced banana.

Alternatives to thematic roles

1. **Fewer roles**: generalized semantic roles, defined as prototypes (Dowty 1991)
   PROTO-AGENT
   PROTO-PATIENT

2. **More roles**: Define roles specific to a group of predicates

   PropBank

   FrameNet

Semantic Role Labeling

The Proposition Bank (PropBank)

PropBank

- Palmer, Martha, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106

PropBank Roles

Following Dowty 1991

Proto-Agent

- Volitional involvement in event or state
- Sentience (and/or perception)
- Causes an event or change of state in another participant
- Movement (relative to position of another participant)

Proto-Patient

- Undergoes change of state
- Causally affected by another participant
- Stationary relative to movement of another participant

PropBank Roles

- Following Dowty 1991
  - Role definitions determined verb by verb, with respect to the other roles
  - Semantic roles in PropBank are thus verb-sense specific.
- Each verb sense has numbered argument: Arg0, Arg1, Arg2,…

  Arg0: PROTO-AGENT
  Arg1: PROTO-PATIENT
  Arg2: usually: benefactive, instrument, attribute, or end state
  Arg3: usually: start point, benefactive, instrument, or attribute
  Arg4 the end point
  (Arg2-Arg5 are not really that consistent, causes a problem for labeling)

PropBank Frame Files

agree.01

Arg0: Agreer
Arg1: Proposition
Arg2: Other entity agreeing

Ex1: [$_{\text{Arg0}}$ The group] *agreed* [$_{\text{Arg1}}$ it wouldn't make an offer].
Ex2: [$_{\text{ArgM-TMP}}$ Usually] [$_{\text{Arg0}}$ John] *agrees* [$_{\text{Arg2}}$ with Mary] [$_{\text{Arg1}}$ on everything].

fall.01

Arg1: Logical subject, patient, thing falling
Arg2: Extent, amount fallen
Arg3: start point
Arg4: end point, end state of arg1
Ex1: [$_{\text{Arg1}}$ Sales] *fell* [$_{\text{Arg4}}$ to $25 million] [$_{\text{Arg3}}$ from $27 million].
Ex2: [$_{\text{Arg1}}$ The average junk bond] *fell* [$_{\text{Arg2}}$ by 4.2%].

Advantage of a ProbBank Labeling

` **increase.01** "go up incrementally"

Arg0: causer of increase

Arg1: thing increasing

Arg2: amount increased by, EXT, or MNR

Arg3: start point

Arg4: end point

This would allow us to see the commonalities in these 3 sentences:

[$_{\text{Arg0}}$ Big Fruit Co.] increased [$_{\text{Arg1}}$ the price of bananas].

[$_{\text{Arg1}}$ The price of bananas] was increased again [$_{\text{Arg0}}$ by Big Fruit Co.]

[$_{\text{Arg1}}$ The price of bananas] increased [$_{\text{Arg2}}$ 5%].

Modifiers or adjuncts of the predicate: Arg-M

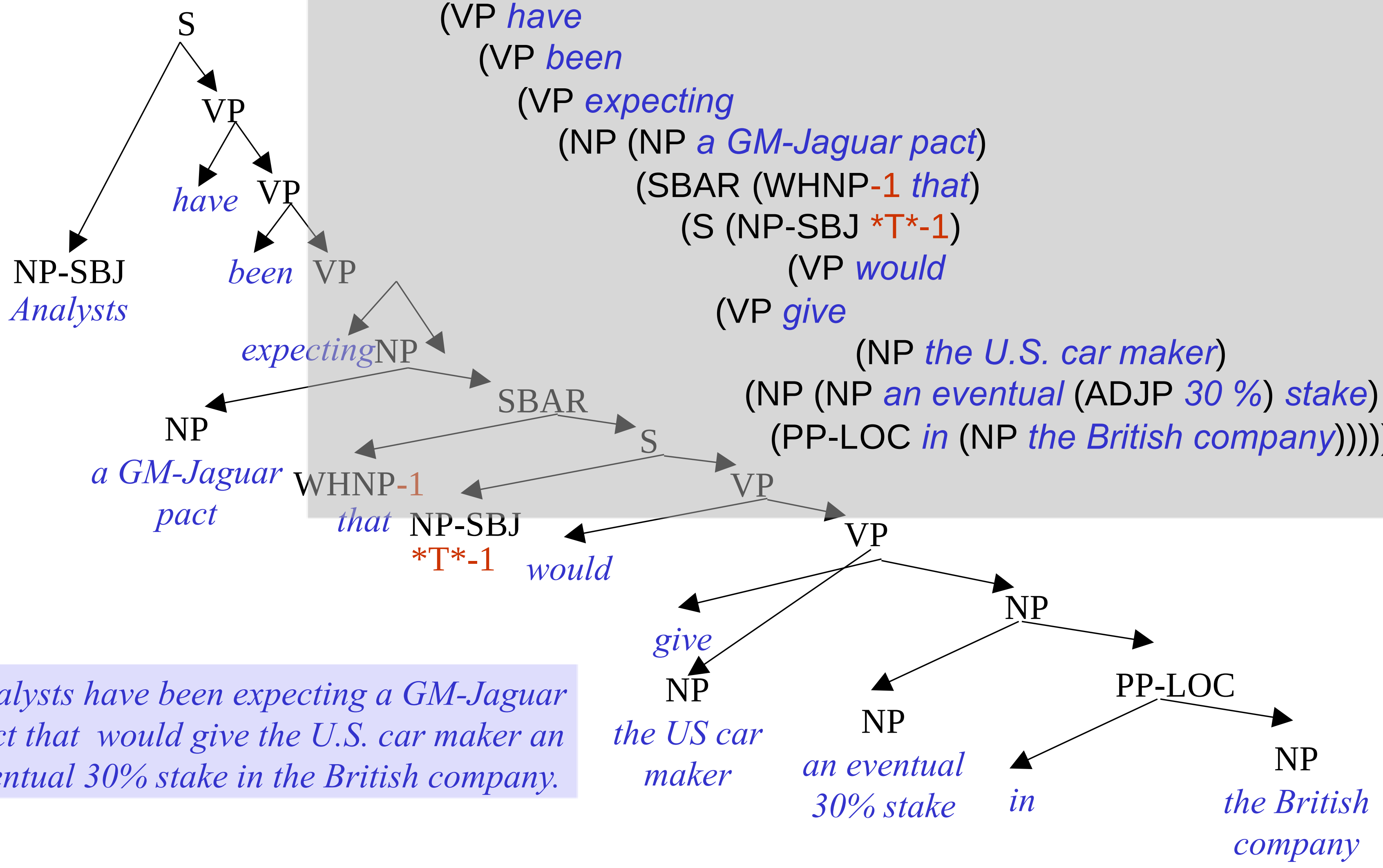
ArgM-TMP	when?	yesterday evening, now
LOC	where?	at the museum, in San Francisco
DIR	where to/from?	down, to Bangkok
MNR	how?	clearly, with much enthusiasm
PRP/CAU	why?	because ... , in response to the ruling
REC		themselves, each other
ADV	miscellaneous	
PRD	secondary predication	...ate the meat raw

PropBanking a Sentence

Martha Palmer 2013

A sample parse tree

```
(S (NP-SBJ Analysts)
   (VP have
      (VP been
         (VP expecting
            (NP (NP a GM-Jaguar pact)
                (SBAR (WHNP-1 that)
                   (S (NP-SBJ *T*-1)
                      (VP would
                   (VP give
                         (NP the U.S. car maker)
                (NP (NP an eventual (ADJP 30 %) stake)
                   (PP-LOC in (NP the British company))))))))))))
```

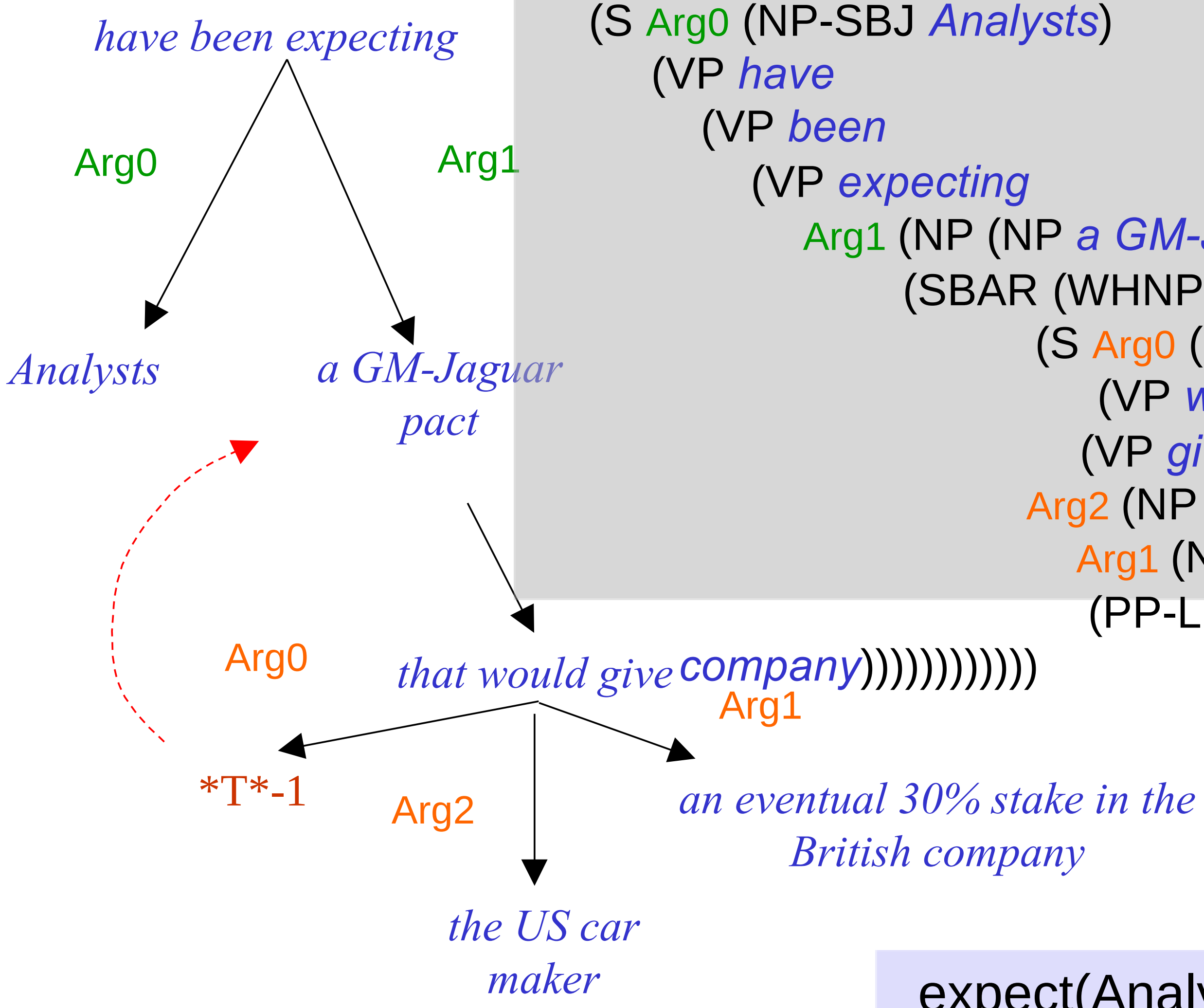


Analysts have been expecting a GM-Jaguar pact that would give the U.S. car maker an eventual 30% stake in the British company.

The same parse tree PropBanked

Martha Palmer 2013

(S Arg0 (NP-SBJ *Analysts*)
(VP *have*
(VP *been*
(VP *expecting*
Arg1 (NP (NP *a GM-Jaguar pact*)
(SBAR (WHNP-1 *that*)
(S Arg0 (NP-SBJ *T*-1)
(VP *would*
(VP *give*
Arg2 (NP *the U.S. car maker*)
Arg1 (NP (NP *an eventual* (ADJP *30 %*) *stake*)
(PP-LOC *in* (NP *the British*
company))))))))))))



expect(Analysts, GM-J pact)
give(GM-J pact, US car maker, 30% stake)

Annotated PropBank Data

- Penn English TreeBank,
  OntoNotes 5.0.
  - Total ~2 million words
- Penn Chinese TreeBank
- Hindi/Urdu PropBank
- Arabic PropBank

2013 Verb Frames Coverage
Count of word sense (lexical units)

Language	***Final Count***
English	**10,615***
Chinese	**24, 642**
Arabic	**7,015**

From Martha Palmer 2013 Tutorial

Plus nouns and light verbs

Example Noun: *Decision*

← Roleset: Arg0: decider, Arg1: decision…

← "…[your$_{ARG0}$] [decision$_{REL}$]
[to say look I don't want to go through this anymore$_{ARG1}$]"

Example within an LVC: *Make a decision*

← "…[the President$_{ARG0}$] [made$_{REL-LVB}$]
the [fundamentally correct$_{ARGM-ADJ}$]
[decision$_{REL}$] [to get on offense$_{ARG1}$]"

Slide from Palmer 2013

Composing Word Vectors

Corpus

- Cleaned-up Wikipedia corpus – oct 13 : 1.7 billion tokens
- Lemmatize → stem forms
- Context words: Top 10K words, after stopwords.
- Sentence boundary = context window.

- Co-occurrence matrix : $M = |w| \times |C|$

Word-Word matrix (raw counts)

sugar, a sliced lemon, a tablespoonful of	**apricot**	preserve or jam, a pinch each of,
their enjoyment. Cautiously she sampled her first	**pineapple**	and another fruit whose taste she likened
well suited to programming on the digital	**computer**.	In finding the optimal R-stage policy from
for the purpose of gathering data and	**information**	necessary for the study authorized in the

	aardvark	computer	data	pinch	result	sugar	…
apricot	0	0	0	1	0	1	
pineapple	0	0	0	1	0	1	
digital	0	2	1	0	1	0	
information	0	1	6	0	4	0	
…	…						

Co-occurrence vectors : Weighting

$$\mathrm{tTest}(\vec{w}_i, c_j) = \frac{p(w_i, c_j) - p(w_i)p(c_j)}{\sqrt{p(w_i)p(c_j)}}$$

Values: [-1,1]

(often ~= 0)

$$\mathrm{PPMI}(\vec{w}_i, c_j) = p(w_i, c_j) \log \left(\frac{p(w_i, c_j)}{p(w_i)p(c_j)} \right)$$

: [0, ∞]

normalize $\vec{w} := \lambda \frac{\vec{w}}{||\vec{w}||_2}$

[polajnar & clark 14]

Co-occurrence vectors : Weighting

EACL-14

Improving Distributional Semantic Vectors through Context Selection and Normalisation

Tamara Polajnar
University of Cambridge
Computer Laboratory
tp366@cam.ac.uk

Stephen Clark
University of Cambridge
Computer Laboratory
sc609@cam.ac.uk

Abstract

Distributional semantic models (DSMs) have been effective at representing semantics at the word level, and research has re-

al., 2012). Evaluation is conducted by comparing the word similarity predicted by the model with the gold standard using a correlation test such as Spearman's ρ.

While words and perhaps some frequent

[polajnar & clark 14]

Context Selection (CS)

- Keep only the N highest-weighted context words (sparsify)
- Select these cj to maximize correlation across all words in the evaluation dataset

Word Vectors via SVD

$$\mathbf{M} \quad = \quad \mathbf{U} \;\; \boldsymbol{\Sigma} \;\; \mathbf{V'}$$

$$l \times c \qquad\qquad l \times l \;\; l\text{x}c \;\; c \times c$$

Keep top k eigenvectors : $\mathbf{U_k} \;\; \boldsymbol{\Sigma_k} \;\; \mathbf{V'_k} \;=\; \mathbf{[lxk]} \;\; \mathbf{[kxk]} \;\; \mathbf{[kxc]}$

k-Word vectors : eigenvectors of $U_k \, \boldsymbol{\Sigma_k}$

Evaluating Word Vector models

Word-pair similarity – gold standards

MEN [Bruni etal 2012] : 3000 word pairs

WS-353 [Finkelstein + 2002] : 353 pairs

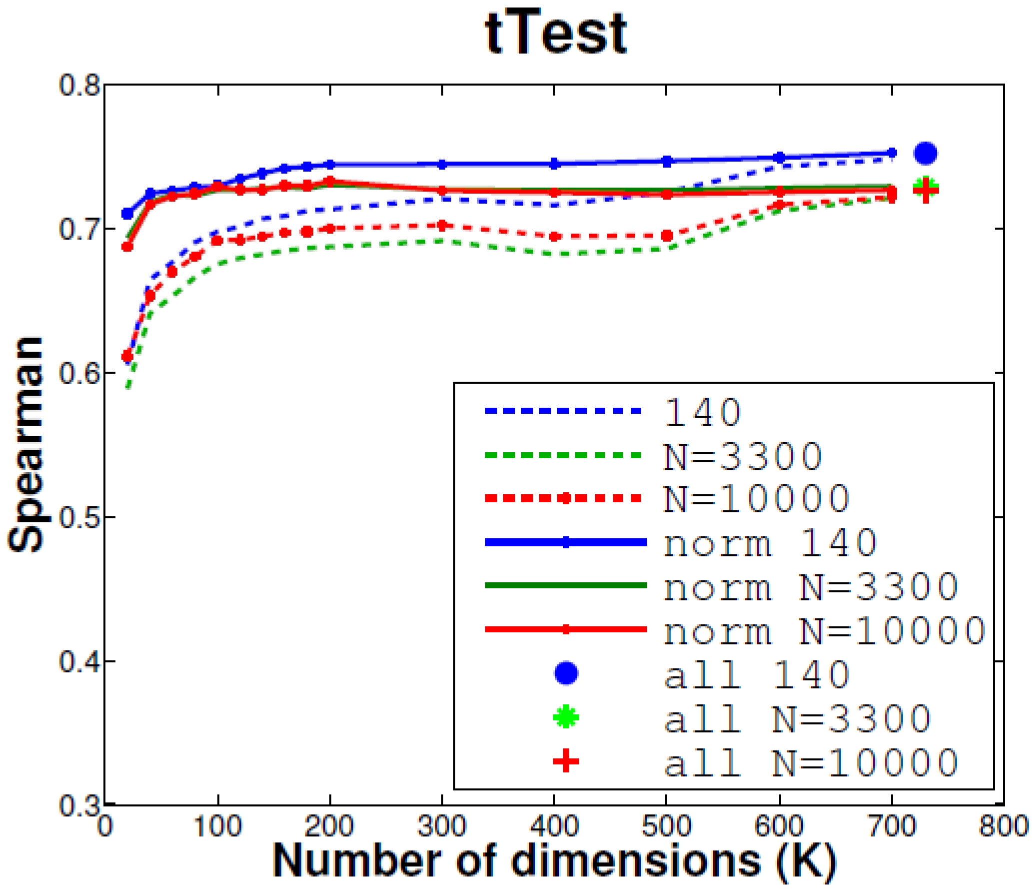
WS-Sim [Agirre etal 09] : small

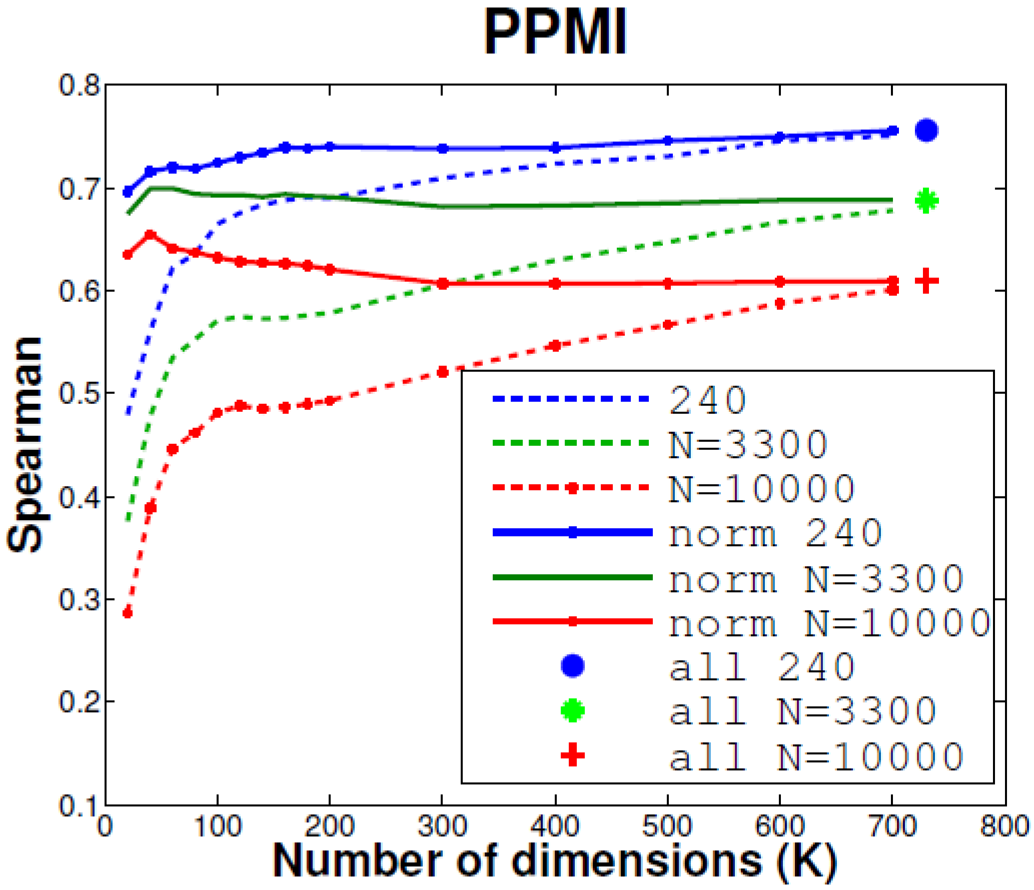
SimLex-999 [Hill etal 2014] : distinguish semantic similarity from association

Turney 12 : two different WVs for similarity vs association

hill-reichart-14_simlex-999-semantic-similarity-evaluation

Evaluating Word Vector models





Blue = tuned for sparseness

[polajnar & clark 14]

Word Vector Composition Operators

Sum	$\vec{x} + \vec{y} = \{\vec{x}_i + \vec{y}_i\}_i$
Prod	$\vec{x} \odot \vec{y} = \{\vec{x}_i \cdot \vec{y}_i\}_i$
Kron	$\vec{x} \otimes \vec{y} = \{\vec{x}_i \cdot \vec{y}_j\}_{ij}$
Conv	$\vec{x} \circledast \vec{y} = \left\{\sum_{j=0}^{n} (\vec{x})_{j\%n} \cdot (\vec{y})_{(i-j)\%n}\right\}_i$

mitchell-lapata-10_composition-in-distributional-models-of-semantics

Evaluating Composition : (t-test)

- Phrasal similarity dataset : mitchell / lapata 2010

Oper		N=140	N=3300	N=10000
sum	ttest	0.40 **(0.41)**	0.40 (0.40)	0.40 (0.40)
	SVD_{100}	0.37 **(0.42)**	0.35 **(0.41)**	0.37 (0.40)
prod	ttest	0.32 (0.32)	**0.40 (0.40)**	0.32 (0.32)
	SVD_{100}	0.25 (0.23)	0.23 (0.23)	0.21 (0.23)
kron	SVD_{100}	0.31 (0.34)	0.34 **(0.38)**	0.29 (0.32)
	SVD_{700}	**0.39 (0.39)**	**0.37 (0.37)**	0.30 (0.30)
conv	RI_{512}	0.10 (0.12)	**0.26** (0.21)	0.25 (0.25)
	RI_{1024}	0.22 (0.15)	0.29 (0.27)	0.25 (0.26)
	RI_{4096}	0.16 (0.19)	**0.33 (0.34)**	0.28 (0.30)

RI = random indexing to a lower-D space

polajnar-clark-14_improving-distributional-vectors-via-normalisation

Evaluating Composition : (PPMI)

- Phrasal similarity dataset : mitchell / lapata 2010

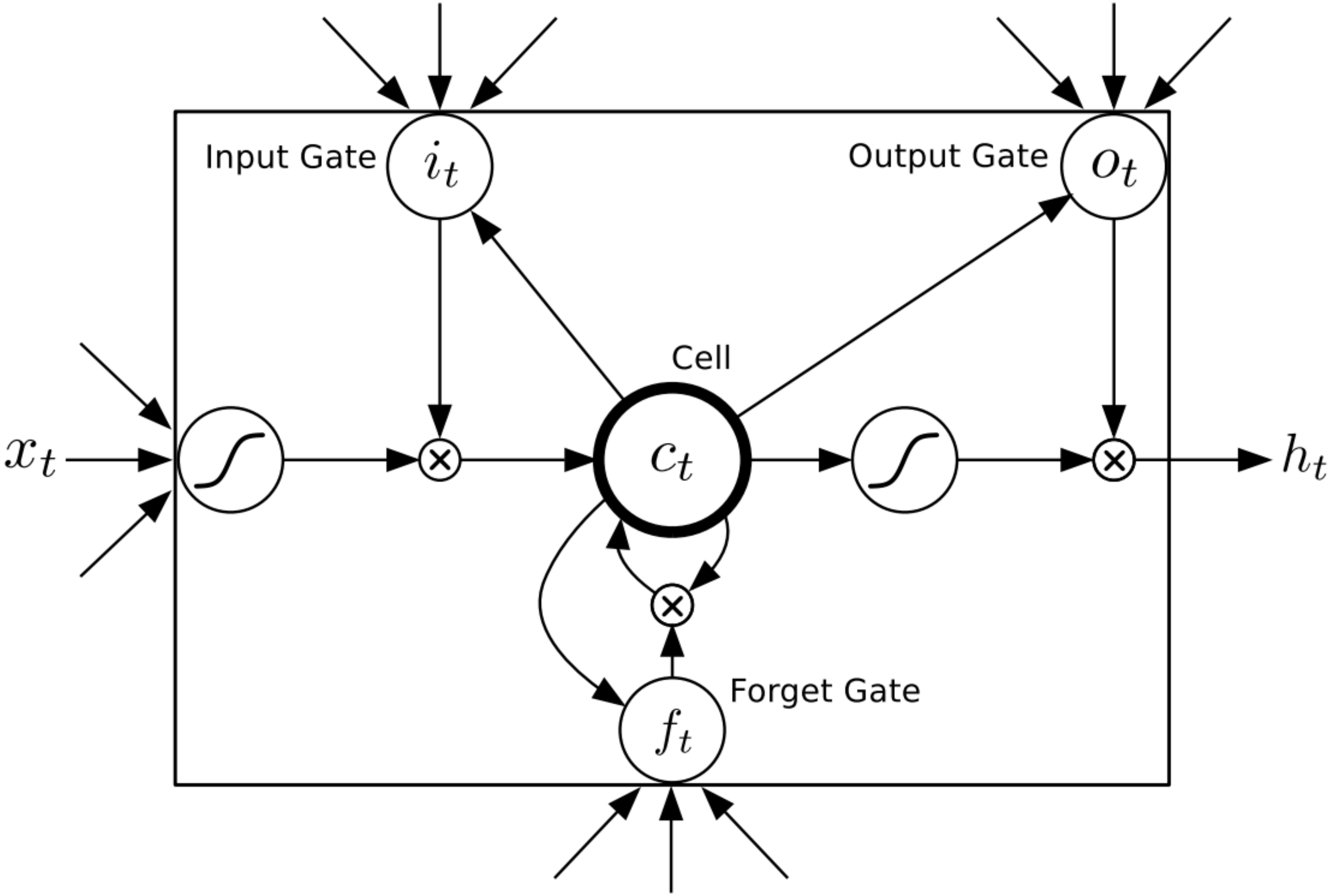
Oper		N=240	N=3300	N=10000
sum	ppmi	**0.40** (0.39)	**0.40 (0.39)**	0.29 (0.29)
	SVD_{100}	**0.40** (0.40)	0.38 **(0.40)**	0.29 (0.30)
prod	ppmi	0.28 (0.28)	**0.40 (0.40)**	0.30 (0.30)
	SVD_{100}	0.23 (0.17)	0.18 (0.22)	0.14 (0.12)
kron	SVD_{100}	0.37 (0.30)	0.36 **(0.38)**	0.27 (0.27)
	SVD_{700}	**0.38 (0.37)**	**0.37 (0.37)**	0.26 (0.26)
conv	RI_{512}	0.09 (0.09)	0.27 **(0.30)**	0.25 (0.24)
	RI_{1024}	0.08 (0.14)	0.33 **(0.37)**	0.25 (0.27)
	RI_{4096}	0.18 (0.19)	**0.37 (0.38)**	0.27 (0.27)

polajnar-clark-14_improving-distributional-vectors-via-normalisation

Sequence Models (syntax)

Long Short-Term Memory

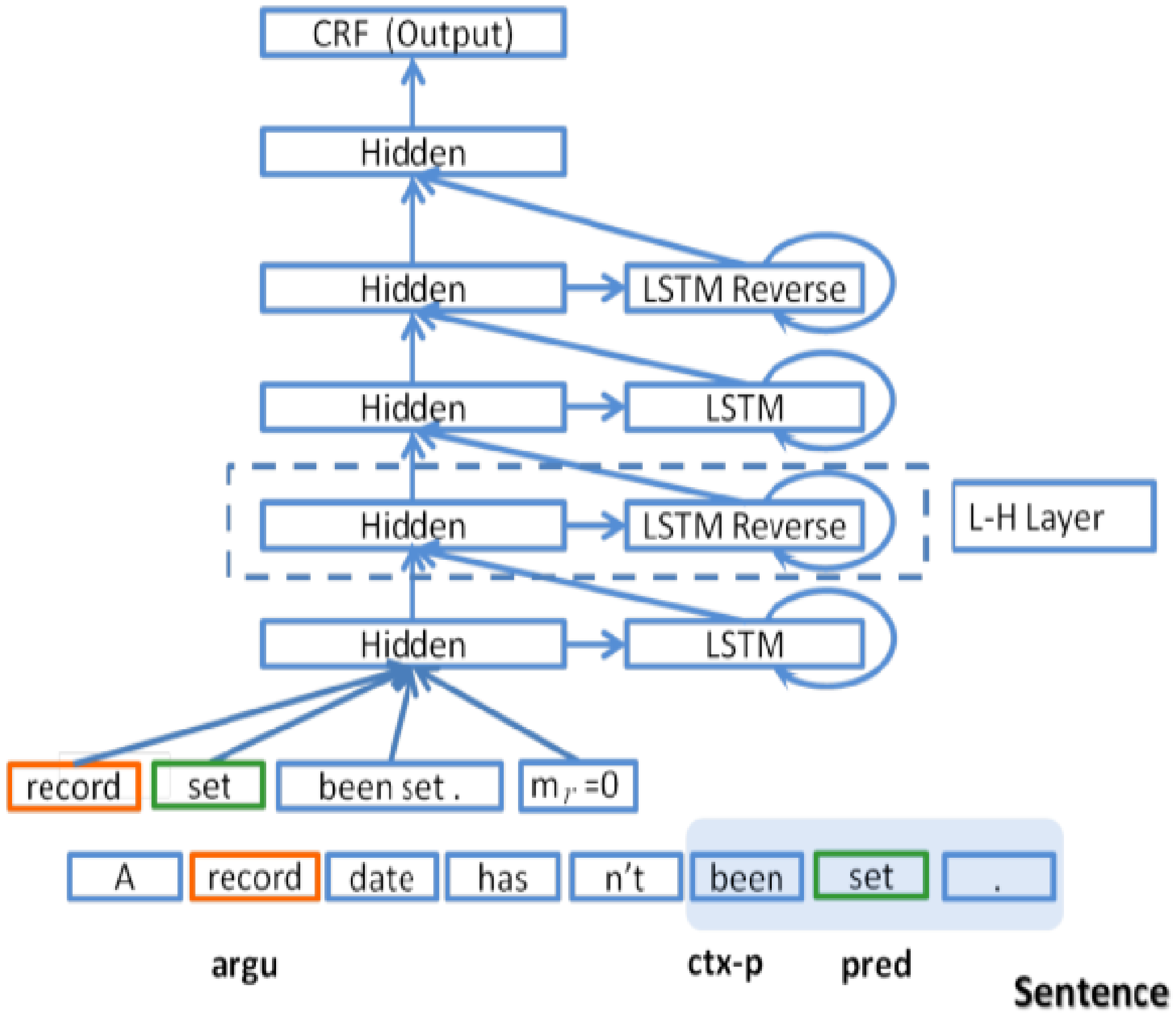
Recurrent
Network :
Latent vars from
(t-1) are fed into
time t;
Recursively encode
Past data



Four Features:

- Predicate
- Arguments
- Context
- Region

- F-score: 81



DB-LSTM Network

zhou-xu-15_end-to-end-semantic-role-labeling-w-RNN

Semantic Role Labeling

FrameNet

Capturing descriptions of the same event by different nouns/verbs

[$_{\text{Arg1}}$ The price of bananas] increased [$_{\text{Arg2}}$ 5%].

[$_{\text{Arg1}}$ The price of bananas] rose [$_{\text{Arg2}}$ 5%].

There has been a [$_{\text{Arg2}}$ 5%] rise [$_{\text{Arg1}}$ in the price of bananas].

FrameNet

- Baker et al. 1998, Fillmore et al. 2003, Fillmore and Baker 2009, Ruppenhofer et al. 2006
- Roles in PropBank are specific to a verb
- Role in FrameNet are specific to a **frame: a** background knowledge structure that defines a set of frame-specific semantic roles, called **frame elements**,
    - includes a set of pred cates that use these roles
    - each word evokes a frame and profiles some aspect of the frame

The “Change position on a scale” Frame

This frame consists of words that indicate the change of an ITEM’s position on a scale (the ATTRIBUTE) from a starting point (INITIAL VALUE) to an end point (FINAL VALUE)

[$_{\text{ITEM}}$ Oil] *rose* [$_{\text{ATTRIBUTE}}$ in price] [$_{\text{DIFFERENCE}}$ by 2%].

[$_{\text{ITEM}}$ It] has *increased* [$_{\text{FINAL\_STATE}}$ to having them 1 day a month].

[$_{\text{ITEM}}$ Microsoft shares] *fell* [$_{\text{FINAL\_VALUE}}$ to 7 5/8].

[$_{\text{ITEM}}$ Colon cancer incidence] *fell* [$_{\text{DIFFERENCE}}$ by 50%] [$_{\text{GROUP}}$ among men].

a steady *increase* [$_{\text{INITIAL\_VALUE}}$ from 9.5] [$_{\text{FINAL\_VALUE}}$ to 14.3] [$_{\text{ITEM}}$ in dividends]

a [$_{\text{DIFFERENCE}}$ 5%] [$_{\text{ITEM}}$ dividend] *increase*...

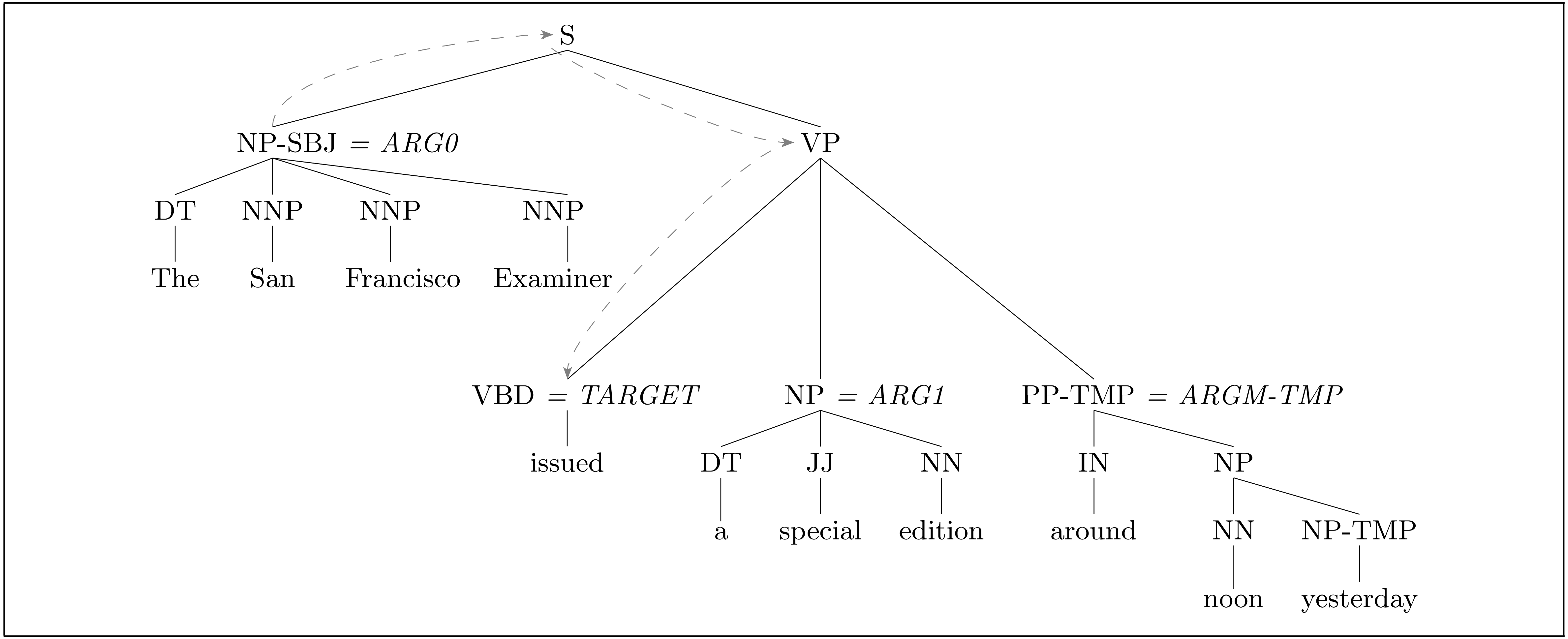
The “Change position on a scale” Frame

VERBS:
advance
climb
decline
decrease
diminish
dip
double
drop
dwindle
edge
explode
fall
fluctuate
gain
grow
increase
jump
move
mushroom
plummet
reach
rise
rocket
shift
skyrocket
slide
soar
swell
swing
triple
tumble

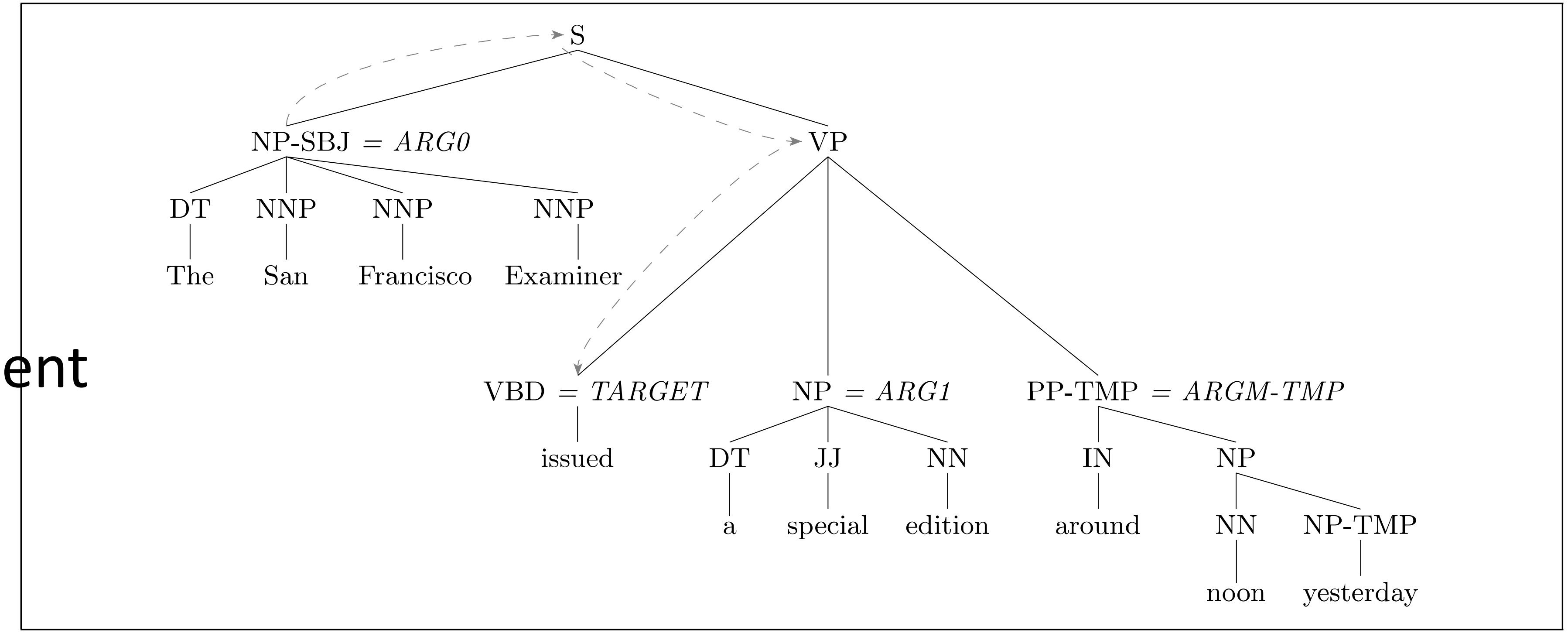
NOUNS:
decline
decrease
escalation
explosion
fall
fluctuation
gain
growth
hike
increase
rise
shift
tumble

ADVERBS:
increasingly

Syntactic path constraints from Training set



Features



Headword of constituent

Examiner

Headword POS

NNP

Voice of the clause

Active

Subcategorization of pred

VP -> VBD NP PP

Named Entity type of constit

ORGANIZATION

First and last words of constit

The, Examiner

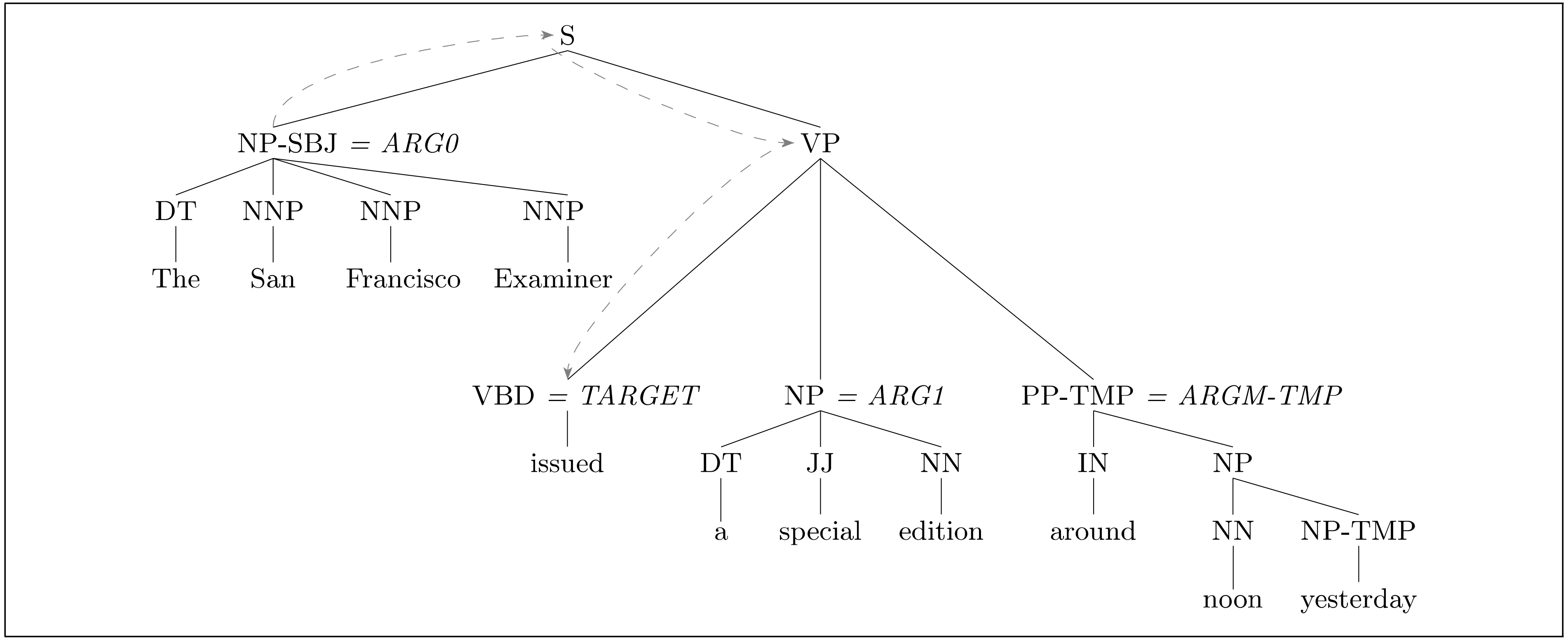
Linear position,clause re: predicate

before

Path Features

Path in the parse tree from the constituent to the predicate

$$NP \uparrow S \downarrow VP \downarrow VBD$$



Frequent path features

Frequency	Path	Description
14.2%	VB↑VP↓PP	PP argument/adjunct
11.8	VB↑VP↑S↓NP	subject
10.1	VB↑VP↓NP	object
7.9	VB↑VP↑VP↑S↓NP	subject (embedded VP)
4.1	VB↑VP↓ADVP	adverbial adjunct
3.0	NN↑NP↑NP↓PP	prepositional complement of noun
1.7	VB↑VP↓PRT	adverbial particle
1.6	VB↑VP↑VP↑VP↑S↓NP	subject (embedded VP)
14.2		no matching parse constituent
31.4	Other	

From Palmer, Gildea, Xue 2010

Final feature vector

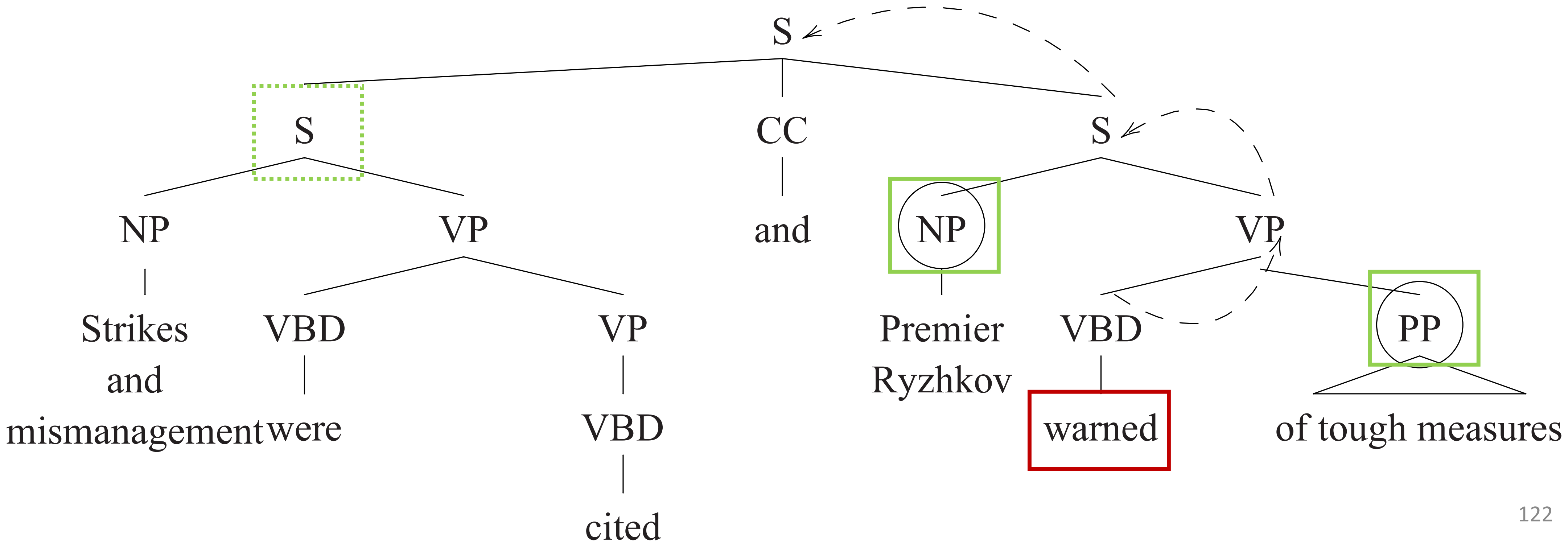
- For “The San Francisco Examiner”,
- Arg0, [issued, NP, Examiner, NNP, active, before, VP$\rightarrow$NP PP, ORG, The, Examiner, $NP\uparrow S\downarrow VP\downarrow VBD$]
- Other features could be used as well
  - sets of n-grams inside the constituent
  - other path features
    - the upward or downward halves
    - whether particular nodes occur in the path

3-step version of SRL algorithm

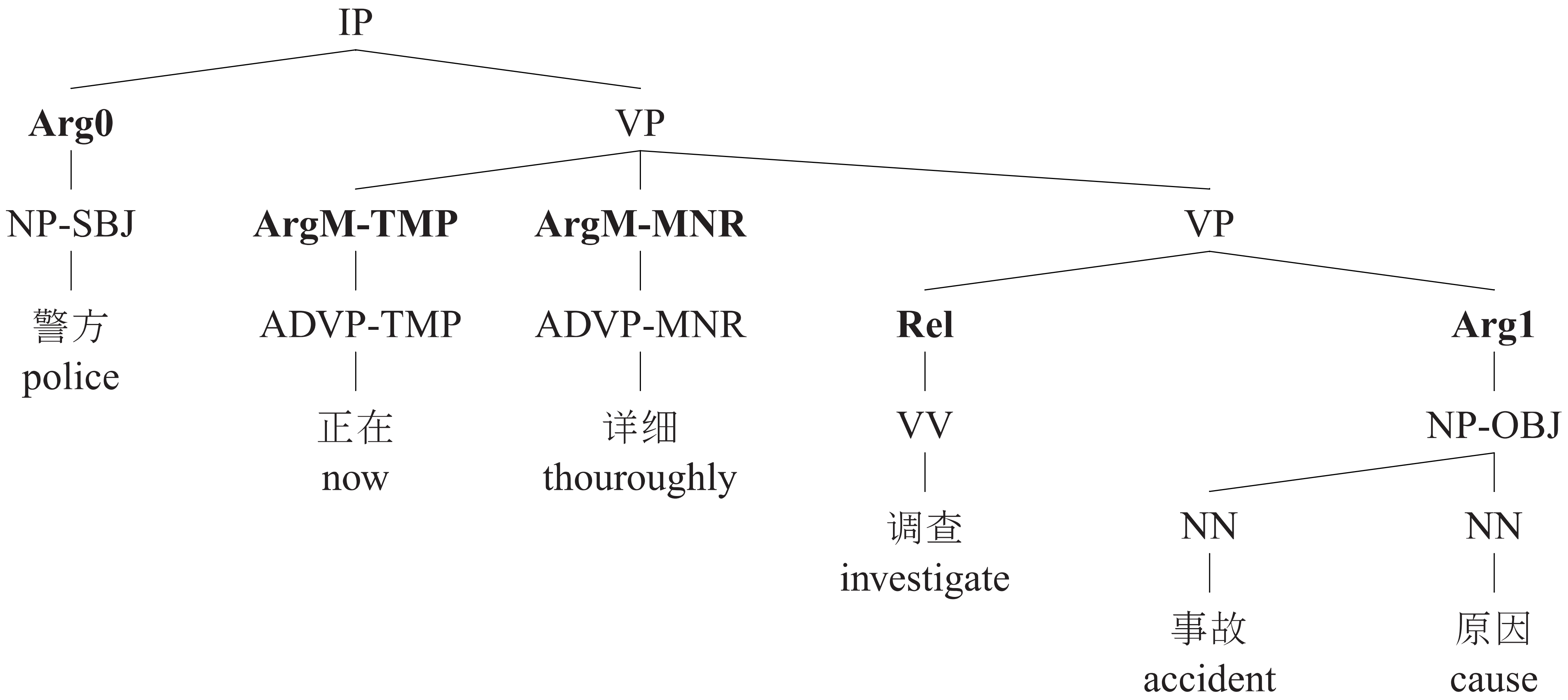
1. **Pruning**: use simple heuristics to prune unlikely constituents.
2. **Identification**: a binary classification of each node as an argument to be labeled or a NONE.
3. **Classification**: a 1-of-*N* classification of all the constituents that were labeled as arguments by the previous stage

Pruning heuristics – Xue and Palmer (2004)

- Add sisters of the predicate, then aunts, then great-aunts, etc
  - But ignoring anything in a coordination structure



Not just English



"The police are thoroughly investigating the cause of the accident."

Not just verbs: NomBank

Meyers et al. 2004

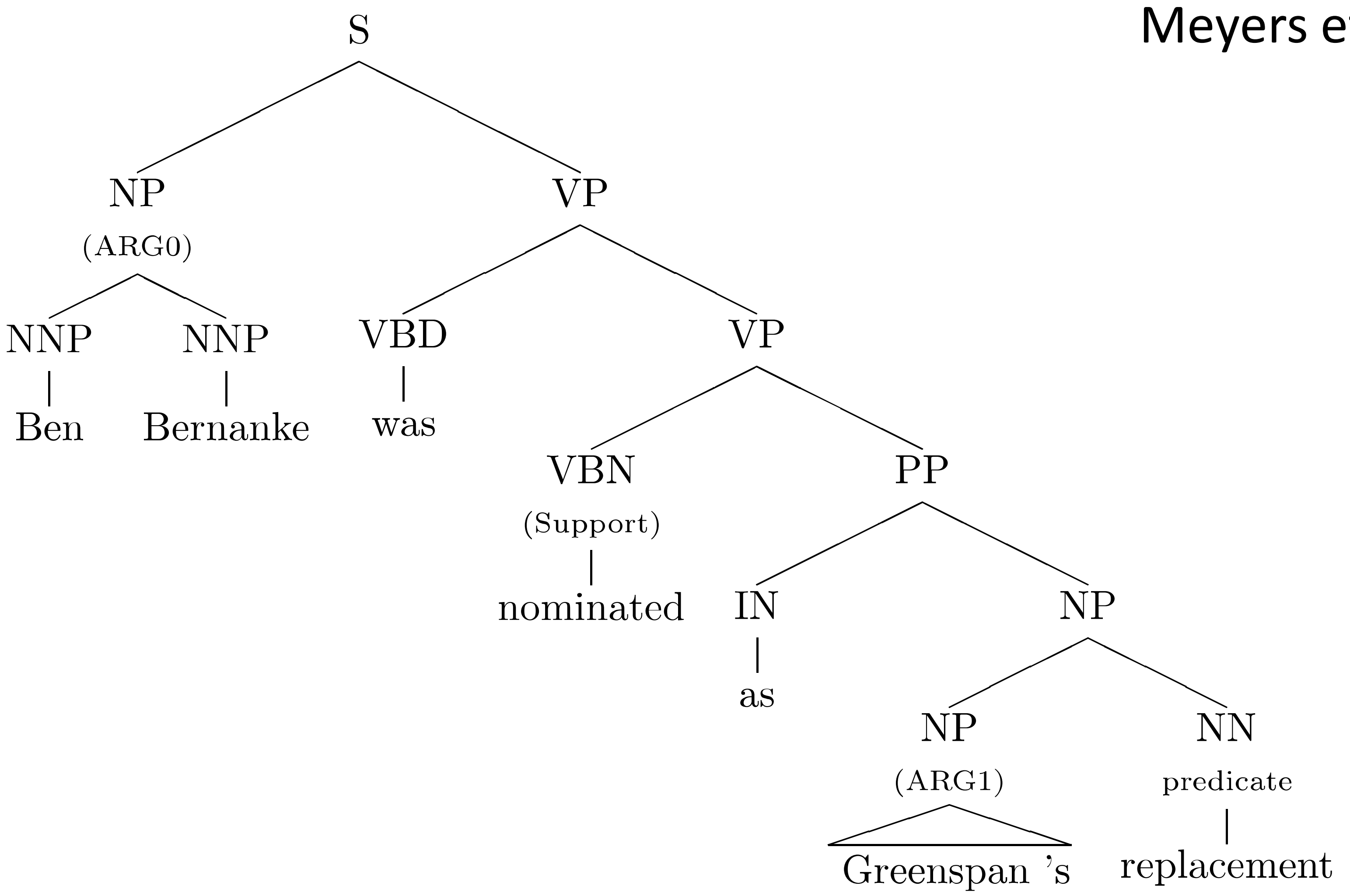


Figure from Jiang and Ng 2006

Additional Issues for nouns

- Features:
  - Nominalization lexicon (employment→ employ)
  - Morphological stem
    - Healthcare, Medicate → care
- Different positions
  - Most arguments of nominal predicates occur inside the NP
  - Others are introduced by support verbs
  - Especially light verbs "X made an argument", "Y took a nap"

Semantic Role Labeling

- A level of shallow semantics for representing events and their participants
  - Intermediate between parses and full semantics
- Two common architectures, for various languages
  - FrameNet: frame-specific roles
  - PropBank: Proto-roles
- Current systems extract by
  - parsing sentence
  - Finding predicates in the sentence
    - For each one, classify each parse tree constituent

Other Semantic Models

Generative Lexicon

a. The newspaper fired the journalist after the fiasco. (organization)

b. Mary spilled coffee on the newspaper. (physical object)

c. John read the newspaper at leisure. (information)

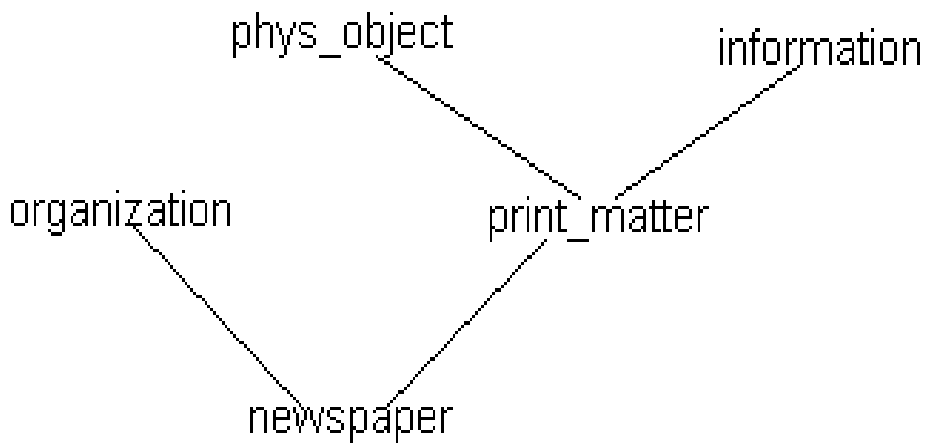
Lexeme Properties

- Newspaper

  = print_matter.org_lcp

- Print_matter

  = phys_object.info_lcp



[pustejovsky 95 : Generative Lexicon] : lcp = Lexical Conceptual Paradigm

Generative Lexicon : Semantic Parameters

I. Qualia stucture in the Generative Lexicon:

1. Constitutive qualia

    dictionary(x): CONST = lexical_entry(y)

2. Formal qualia

    dictionary(x): FORMAL = book(x)

3. Telic qualia:

    dictionary(x): TELIC = consult(y,x)

4. Agentive qualia

    dictionary(x): AGENT = compile(z, x)

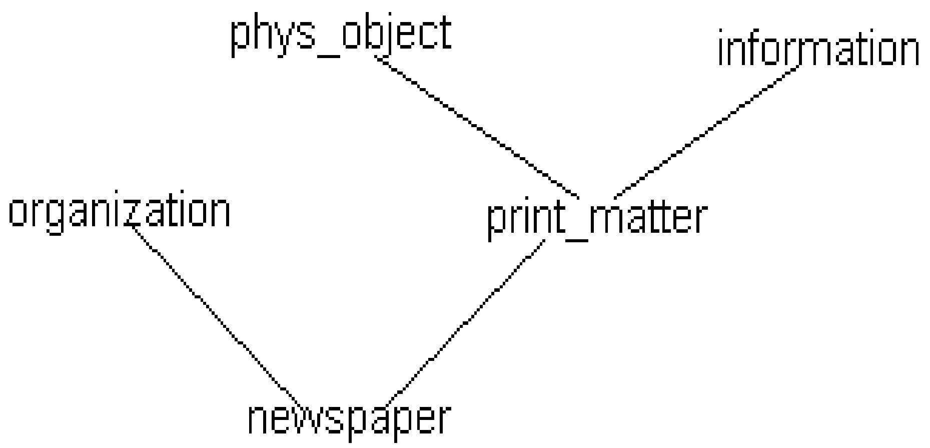
Lexical conceptual paradigm: lcp

a. The newspaper fired the journalist after the fiasco. (organization)

b. Mary spilled coffee on the newspaper. (physical object)

c. John read the newspaper at leisure. (information)

- Newspaper
- = print_matter.org_lcp
- Print_matter
- = phys_object.info_lcp



UNL (Universal Networking Language)

- Universal Words (UWs) – List of Senses

    water(icl>liquid>thing)

- UNL Dictionary – map to Natural Languages

- Relations – ontologies (icl<), modifiers. . . (39)

    mod(water(icl>liquid), safe(mod>thing));

- Attributes

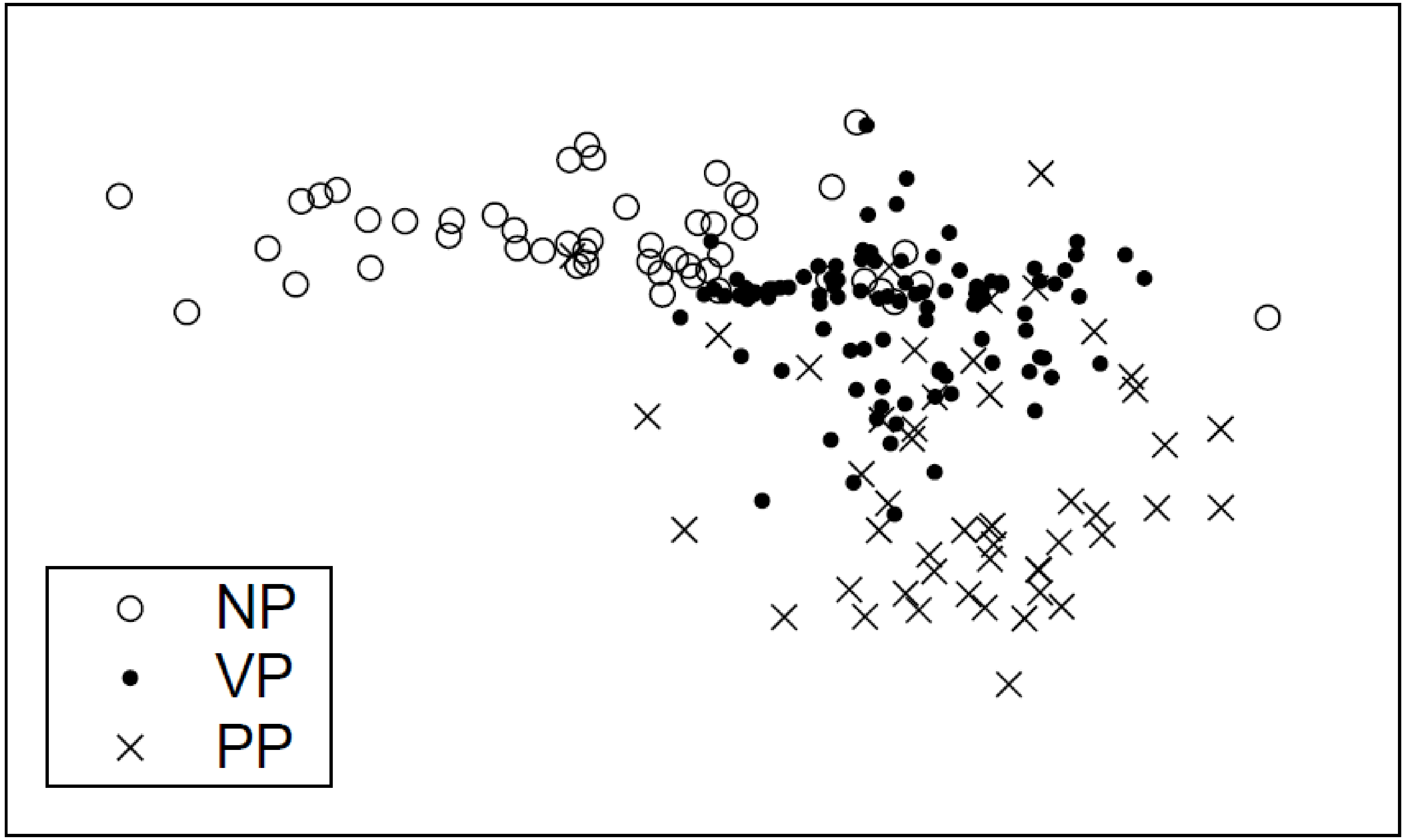
    mineral.@pl

- Knowledge Base (KB) : Relations between UW’s

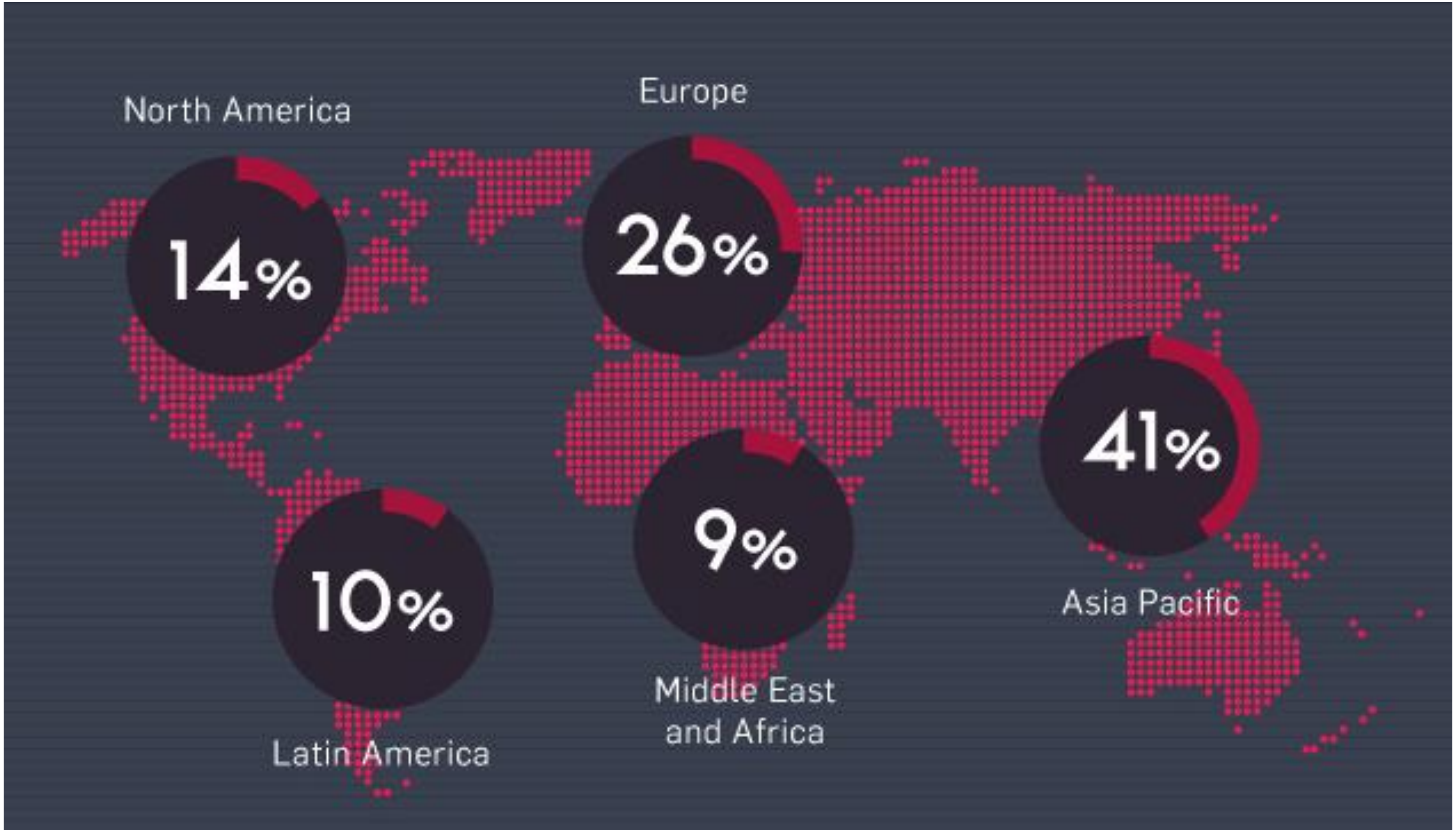
Can’t ignore punctuation

Syntax as Dimensionality Reduction

context vectors for three types of phrases
→ PCA → space of first two principal components



Web Users Map- 2014



- http://www.statista.com