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

• 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.
  b. He got the rope on Tuesday.
Sentences and Context

• a. The window broke
  b. Ram fell through it

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

CAUSE
CONSEQUENCE

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

Pragmatics
| Discourse Semantics
| Lexical Semantics
| Syntax
| Morphology

Complexity Increases

More World Knowledge
More Linguistic knowledge
Specification of Meaning

Many Meanings

Words Inside a Sentence

Words Inside a Discourse

Under specified

Single Meaning

Words

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 $\rightarrow$ semantics for complex expressions
  • Model-Theoretic: Map phrases / words $\rightarrow$ 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) $\rightarrow$ Denotata

• Modern versions of Montague grammar – avoid “translation”
Montagovian Translation [1973]

A student sleeps

Lexicon:

student, N: \( \lambda u. \text{stud}(u) \)
sleep, V: \( \lambda x. \text{sl}(x) \)
a, DET: \( \lambda P. \lambda Q. \exists x_i. (P(x_i) \land Q(x_i)) \)
Montagovian Translation [1973]

\[
S: \exists x_i. (\text{stud}(x_i) \land \text{sl}(x_i)) \\
NP: \lambda Q. \exists x_i. (\text{stud}(x_i) \land Q(x_i)) \\
\text{DET: } \lambda P. \lambda Q. \exists x_i. (P(x_i) \land Q(x_i)) \\
\text{N: } \lambda u. \text{stud}(u) \\
\text{V: } \lambda x. \text{sl}(x) \\
A \quad \text{student} \quad \text{sleeps}
\]
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

<table>
<thead>
<tr>
<th>Frame Elements</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Ingestibles</td>
<td>Core</td>
</tr>
<tr>
<td>Ingestor</td>
<td>Core</td>
</tr>
<tr>
<td>Instrument</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Manner</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Means</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Place</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Source</td>
<td>Peripheral</td>
</tr>
<tr>
<td>Time</td>
<td>Peripheral</td>
</tr>
</tbody>
</table>
Lexical Units in: *Ingestion*

<table>
<thead>
<tr>
<th>English</th>
<th>Hindi</th>
<th>Bangla</th>
</tr>
</thead>
<tbody>
<tr>
<td>breakfast.v</td>
<td>नाश्ता</td>
<td>prAtarAsh v</td>
</tr>
<tr>
<td>Consume.v</td>
<td>भोग करना</td>
<td>bhog k.v</td>
</tr>
<tr>
<td>drink.v</td>
<td>पी</td>
<td>khA.v</td>
</tr>
<tr>
<td>eat.v</td>
<td>खा</td>
<td>khA.v</td>
</tr>
<tr>
<td>feast.v</td>
<td>भोज करना</td>
<td>bhoj k.v</td>
</tr>
<tr>
<td>feed.v</td>
<td>खिला</td>
<td>khAoyA.v</td>
</tr>
<tr>
<td>gulp.v</td>
<td>निगल</td>
<td>gelA.v</td>
</tr>
<tr>
<td>have.v</td>
<td>ले</td>
<td>Neo.v</td>
</tr>
<tr>
<td>munch.v</td>
<td>चबा</td>
<td>chebA.v</td>
</tr>
<tr>
<td>nibble.v</td>
<td>कुत्तर</td>
<td>ThokrA.v</td>
</tr>
<tr>
<td>sip.n</td>
<td>घूंट</td>
<td>chumuk.n</td>
</tr>
<tr>
<td>sip.v</td>
<td>घूंट लेना</td>
<td>Chumuk de.v</td>
</tr>
</tbody>
</table>
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.
### Reviews

**Summary** - Based on 377 reviews

<table>
<thead>
<tr>
<th>Rating</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 star</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4 stars</td>
</tr>
</tbody>
</table>

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."
HP Officejet 6500A E710N Multifunction Printer

$121.53 - $242.39 (14 stores)

Average rating: 4.57 stars (144)

Most mentioned:
- Performance (57)
- Ease of Use (43)
- Print Speed (39)
- Connectivity (31)

Show reviews by source:
- Best Buy (140)
- CNET (5)
- Amazon.com (3)
Twitter sentiment versus Gallup Poll of Consumer Confidence

Twitter sentiment:

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

Type in a word and we'll highlight the good and the bad

"united airlines"

**Sentiment analysis for "united airlines"**

Sentiment by Percent

- Negative (68%)
- Positive (32%)

Sentiment by Count

- Positive (11)
- Negative (23)

**Tweets:**

- **lljacobson:** OMG... Could @United airlines have worse customer service? W8g now 15 minutes on hold 4 questions about a flight 2DAY that need a human. Posted 2 hours ago
- **12345clumsy6789:** I hate United Airlines Ceiling!!! Fukn impossible to get my conduit in this damn mess! ? Posted 2 hours ago
- **EMLandPRGbelgju:** EML/PRG fly with Q8 united airlines and 24seven to an exotic destination. http://t.co/Z9QloA7F Posted 2 hours ago
- **CountAdam:** FANTASTIC customer service from United Airlines at XNA today. Is tweet more, but cell phones off now! Posted 4 hours ago
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


• Polarity detection:
  • Is an IMDB movie review positive or negative?

• Data: *Polarity Data 2.0*:
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
[\{\}\\((\[dDpP/:\]\){@\\}]) [\-o\*\']?  # optional nose
[;:=8]  # eyes
[<>]?:  # optional hat/brow
```

# optional hat/brow
# eyes
# optional nose
# mouth
### reverse orientation
# mouth
# optional nose
# eyes
# optional hat/brow
Extracting Features for Sentiment Classification

• How to handle negation
  • I didn’t like this movie
  • 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
didn’t NOT_like NOT_this NOT_movie but I
Reminder: Naïve Bayes

\[ c_{NB} = \arg \max_{c_j \in C} P(c_j) \prod_{i \in \text{positions}} P(w_i \mid c_j) \]

\[ \hat{P}(w \mid c) = \frac{\text{count}(w, c) + 1}{\text{count}(c) + |V|} \]
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
    
    $doc_j \leftarrow$ all docs with class $= c_j$

    $$P(c_j) \leftarrow \frac{|doc_j|}{|\text{total \#\ documents}|}$$

• 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 $doc_j$
  • For each word $w_k$ in *Vocabulary*
    
    $n_k \leftarrow$ # of occurrences of $w_k$ in $Text_j$

    $$P(w_k \mid c_j) \leftarrow \frac{n_k +}{n + |\text{Vocabulary}|}$$
Boolean Multinomial Naïve Bayes on a test document $d$

• First remove all duplicate words from $d$
• Then compute NB using the same equation:

\[
c_{NB} = \arg\max_{c_j \in C} P(c_j) \prod_{i \in \text{positions}} P(w_i | c_j)
\]
## Normal vs. Boolean Multinomial NB

<table>
<thead>
<tr>
<th>Normal</th>
<th>Doc</th>
<th>Words</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>1</td>
<td>Chinese Beijing Chinese</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Chinese Chinese Shanghai</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Chinese Macao</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Tokyo Japan Chinese</td>
<td>j</td>
</tr>
<tr>
<td>Test</td>
<td>5</td>
<td>Chinese Chinese Chinese Tokyo Japan</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Boolean</th>
<th>Doc</th>
<th>Words</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>1</td>
<td>Chinese Beijing</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Chinese Shanghai</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Chinese Macao</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Tokyo Japan Chinese</td>
<td>j</td>
</tr>
<tr>
<td>Test</td>
<td>5</td>
<td>Chinese Tokyo Japan</td>
<td>?</td>
</tr>
</tbody>
</table>
Binarized (Boolean feature) Multinomial Naïve Bayes


• 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

<table>
<thead>
<tr>
<th></th>
<th>Opinion Lexicon</th>
<th>General Inquirer</th>
<th>SentiWordNet</th>
<th>LIWC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPQA</td>
<td>33/5402 (0.6%)</td>
<td>49/2867 (2%)</td>
<td>1127/4214 (27%)</td>
<td>12/363 (3%)</td>
</tr>
<tr>
<td>Opinion Lexicon</td>
<td>32/2411 (1%)</td>
<td></td>
<td>1004/3994 (25%)</td>
<td>9/403 (2%)</td>
</tr>
<tr>
<td>General Inquirer</td>
<td></td>
<td></td>
<td>520/2306 (23%)</td>
<td>1/204 (0.5%)</td>
</tr>
<tr>
<td>SentiWordNet</td>
<td></td>
<td></td>
<td></td>
<td>174/694 (25%)</td>
</tr>
<tr>
<td>LIWC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sentiment via bag of Words
Analyzing the polarity of each word in IMDB

• 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:

  \[ P(w \mid c) = \frac{f(w, c)}{\sum_c f(w, c)} \]

• Make them comparable between words
  • Scaled likelihood:

  \[ \frac{P(w \mid c)}{P(w)} \]

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 $v_{d_i} \in \mathbb{R}^{V_i}$
- Each element in $v_{d_i}$ is the product of term frequency and inverse document frequency: $\text{tfidf}(t, d) = \text{tf}(t, d) \times \log\left(\frac{||D||}{df(t)}\right)$
- 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

Other sentiment feature: Logical negation

• 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

IMDB (4,073,228 tokens)  Five-star reviews (846,444 tokens)

Scaled likelihood

\[ \frac{P(w|c)}{P(w)} \]
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 a sentence
    • “Given this sentence, is the aspect food, décor, service, value, or NONE”
Putting it all together: Finding sentiment for aspects

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।

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

<table>
<thead>
<tr>
<th>Composition</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>88.42</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>88.41</td>
</tr>
<tr>
<td>Multiplication</td>
<td>50.30</td>
</tr>
</tbody>
</table>
Paragraph Vector Models

Classifier

Average/Concatenate

Word Matrix

[Le and Mikolov 14]
Paragraph Vector Models

[Le and Mikolov 14]
Word2vec variants

![Bar chart comparing SkipGram and CBOW accuracy in various distributed semantic models.](chart.png)
Weighted average vs other models

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maas et al. (2011)</td>
<td>88.89</td>
</tr>
<tr>
<td>NBSVM-bi (Wang &amp; Manning, 2012)</td>
<td>91.22</td>
</tr>
<tr>
<td>NBSVM-uni (Wang &amp; Manning, 2012)</td>
<td>88.29</td>
</tr>
<tr>
<td>SVM-uni (Wang &amp; Manning, 2012)</td>
<td>89.16</td>
</tr>
<tr>
<td>Paragraph Vector (Le and Mikolov (2014))</td>
<td>92.58</td>
</tr>
<tr>
<td>WordVector+Wiki (Our Method)</td>
<td>88.60</td>
</tr>
<tr>
<td>WordVector+TfIdf (Our Method)</td>
<td>89.03</td>
</tr>
<tr>
<td>WordVector Averaging+TfIdf+Document Vector</td>
<td>93.91</td>
</tr>
</tbody>
</table>

Table 6.1: Results on IMDB Movie Review Dataset

[singh & mukerjee 15]
Semantic Role Labelling
The police officer detained the suspect at the scene of the crime.
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

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

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

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:

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

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

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)
agree.01
Arg0: Agreer
Arg1: Proposition
Arg2: Other entity agreeing

Ex1: [Arg0 The group] agreed [Arg1 it wouldn’t make an offer].
Ex2: [ArgM-TMP Usually] [Arg0 John] agrees [Arg2 with Mary] [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: [Arg1 Sales] fell [Arg4 to $25 million] [Arg3 from $27 million].
Ex2: [Arg1 The average junk bond] fell [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:

[Arg0 Big Fruit Co. ] increased [Arg1 the price of bananas].
[Arg1 The price of bananas] was increased again [Arg0 by Big Fruit Co. ]
[Arg1 The price of bananas] increased [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      |
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

Analysts

have been expecting

* T*-1

a GM-Jaguar pact

Arg0

Arg1

Arg2

that would give company)))))))))))

an eventual 30% stake in the

British company

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

(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

Martha Palmer 2013
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)

<table>
<thead>
<tr>
<th>Language</th>
<th>Final Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>10,615*</td>
</tr>
<tr>
<td>Chinese</td>
<td>24,642</td>
</tr>
<tr>
<td>Arabic</td>
<td>7,015</td>
</tr>
</tbody>
</table>

From Martha Palmer 2013 Tutorial
Plus nouns and light verbs

Example Noun: *Decision*

← Roleset: Arg0: decider, Arg1: decision…

← “[your_{ARG_0}] [decision_{REL}]
[to say look I don't want to go through this anymore_{ARG_1}]”

Example within an LVC: *Make a decision*

← “[the President_{ARG_0}] [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| x |C|
Word-Word matrix (raw counts)

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

<table>
<thead>
<tr>
<th></th>
<th>aardvark</th>
<th>computer</th>
<th>data</th>
<th>pinch</th>
<th>result</th>
<th>sugar</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>apricot</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>pineapple</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>digital</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>information</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Co-occurrence vectors: Weighting

\[ tTest(w_i, c_j) = \frac{p(w_i, c_j) - p(w_i)p(c_j)}{\sqrt{p(w_i)p(c_j)}} \]

Values: \([-1, 1]\)
(often \(\sim 0\))

\[ PPMI(w_i, c_j) = p(w_i, c_j) \log \left( \frac{p(w_i, c_j)}{p(w_i)p(c_j)} \right) \]

: \([0, \infty]\)

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

[polajnar & clark 14]
Co-occurrence vectors: Weighting

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 revealed, 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
Context Selection (CS)

• Keep only the N highest-weighted context words (sparsify)

• Select these $c_j$ to maximize correlation across all words in the evaluation dataset
Word Vectors via SVD

\[ M = U \Sigma V' \]
\[ l \times c \quad l \times l \times c \times c \]

Keep top \( k \) eigenvectors: \( U_k \Sigma_k V'_k = [l \times k] [k \times k] [k \times c] \)

\( k \)-Word vectors: eigenvectors of \( U_k \Sigma_k \)
Evaluating Word Vector models

Word-pair similarity – gold standards

MEN [Bruni et al. 2012]: 3000 word pairs
WS-353 [Finkelstein + et al. 2002]: 353 pairs
WS-Sim [Agirre et al. 2009]: small
SimLex-999 [Hill et al. 2014]: distinguish semantic similarity from association

Turney 12: two different WVs for similarity vs. association
Evaluating Word Vector models

Blue = tuned for sparseness

[polajnar & clark 14]
Word Vector Composition Operators

**Sum** \( \vec{x} + \vec{y} = \{ x_i + y_i \}_i \)

**Prod** \( \vec{x} \odot \vec{y} = \{ x_i \cdot y_i \}_i \)

**Kron** \( \vec{x} \otimes \vec{y} = \{ x_i \cdot y_j \}_{i,j} \)

**Conv** \( \vec{x} \boxtimes \vec{y} = \left\{ \sum_{j=0}^{n} (x)_j \% n \cdot (y)_{(i-j)\% n} \right\}_i \)
Evaluating Composition : (t-test)

• Phrasal similarity dataset : mitchell / lapata 2010

<table>
<thead>
<tr>
<th>Oper</th>
<th>N=140</th>
<th>N=3300</th>
<th>N=10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>sum</td>
<td>ttest</td>
<td>0.40 (0.41)</td>
<td>0.40 (0.40)</td>
</tr>
<tr>
<td></td>
<td>SVD_{100}</td>
<td>0.37 (0.42)</td>
<td>0.35 (0.41)</td>
</tr>
<tr>
<td>prod</td>
<td>ttest</td>
<td>0.32 (0.32)</td>
<td>0.40 (0.40)</td>
</tr>
<tr>
<td></td>
<td>SVD_{100}</td>
<td>0.25 (0.23)</td>
<td>0.23 (0.23)</td>
</tr>
<tr>
<td>kron</td>
<td>SVD_{100}</td>
<td>0.31 (0.34)</td>
<td>0.34 (0.38)</td>
</tr>
<tr>
<td></td>
<td>SVD_{700}</td>
<td>0.39 (0.39)</td>
<td>0.37 (0.37)</td>
</tr>
<tr>
<td>conv</td>
<td>RI_{512}</td>
<td>0.10 (0.12)</td>
<td>0.26 (0.21)</td>
</tr>
<tr>
<td></td>
<td>RI_{1024}</td>
<td>0.22 (0.15)</td>
<td>0.29 (0.27)</td>
</tr>
<tr>
<td></td>
<td>RI_{4096}</td>
<td>0.16 (0.19)</td>
<td>0.33 (0.34)</td>
</tr>
</tbody>
</table>

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

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

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
Semantic Role Labeling

FrameNet
Capturing descriptions of the same event by different nouns/verbs

[Arg1 The price of bananas] increased [Arg2 5%].
[Arg1 The price of bananas] rose [Arg2 5%].
There has been a [Arg2 5%] rise [Arg1 in the price of bananas].
FrameNet


• 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 predicates 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)

[ITEM Oil] rose [ATTRIBUTE in price] [DIFFERENCE by 2%].

[ITEM It] has increased [FINAL_STATE to having them 1 day a month].

[ITEM Microsoft shares] fell [FINAL_VALUE to 7 5/8].

[ITEM Colon cancer incidence] fell [DIFFERENCE by 50%] [GROUP among men].

a steady increase [INITIAL_VALUE from 9.5] [FINAL_VALUE to 14.3] [ITEM in dividends]

a [DIFFERENCE 5%] [ITEM dividend] increase...
# The “Change position on a scale” Frame

<table>
<thead>
<tr>
<th><strong>VERBS:</strong></th>
<th>dwindle</th>
<th>move</th>
<th>soar</th>
<th>escalation</th>
<th>shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>advance</td>
<td>edge</td>
<td>mushroom</td>
<td>swell</td>
<td>explosion</td>
<td>tumble</td>
</tr>
<tr>
<td>climb</td>
<td>explode</td>
<td>plummet</td>
<td>swing</td>
<td>fall</td>
<td></td>
</tr>
<tr>
<td>decline</td>
<td>fall</td>
<td>reach</td>
<td>triple</td>
<td>fluctuation</td>
<td></td>
</tr>
<tr>
<td>decrease</td>
<td>fluctuate</td>
<td>rise</td>
<td>tumble</td>
<td>gain</td>
<td></td>
</tr>
<tr>
<td>diminish</td>
<td>gain</td>
<td>rocket</td>
<td></td>
<td>growth</td>
<td></td>
</tr>
<tr>
<td>dip</td>
<td>grow</td>
<td>shift</td>
<td><strong>NOUNS:</strong></td>
<td>hike</td>
<td></td>
</tr>
<tr>
<td>double</td>
<td>increase</td>
<td>skyrocket</td>
<td>decline</td>
<td>increase</td>
<td></td>
</tr>
<tr>
<td>drop</td>
<td>jump</td>
<td>slide</td>
<td>decrease</td>
<td>rise</td>
<td></td>
</tr>
</tbody>
</table>

**ADVERBS:** increasingly
Syntactic path constraints from Training set

```
S
  NP-SBJ = ARG0
     DT  NNP  NNP  NNP
       The  San  Francisco  Examiner
     VBD = TARGET
        issued
     NP = ARG1
        DT  JJ  NN
           a  special  edition
     PP-TMP = ARGM-TMP
        IN  NP
           around  NN  NP-TMP
             noon  yesterday
```
Features

Headword of constituent
Examiner

Headword POS
NNP

Voice of the clause
Active

Subcategorization of pred
VP -> VBD NP PP

Named Entity type of consti
ORGANIZATION

First and last words of consti
The, Examiner

Linear position, clause re: predicate
before
Path Features

Path in the parse tree from the constituent to the predicate

\[ \text{NP} \uparrow \text{S} \downarrow \text{VP} \downarrow \text{VBD} \]
# Frequent path features

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

From Palmer, Gildea, Xue 2010
Final feature vector

• For “The San Francisco Examiner”,
• Arg0, [issued, NP, Examiner, NNP, active, before, VP→NP PP, ORG, The, Examiner, ]
  \[ \text{NP} \uparrow \text{S} \downarrow \text{VP} \downarrow \text{VBD} \]

• 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

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

- North America: 14%
- Europe: 26%
- Latin America: 10%
- Middle East and Africa: 9%
- Asia Pacific: 41%

• http://www.statista.com