Topics, Trends, and Resources in Natural Language Processing (NLP)

Mohit Bansal

TTI-Chicago

(CSC2523, ‘Visual Recognition with Text’, UToronto, Winter 2015 – 01/21/2015)

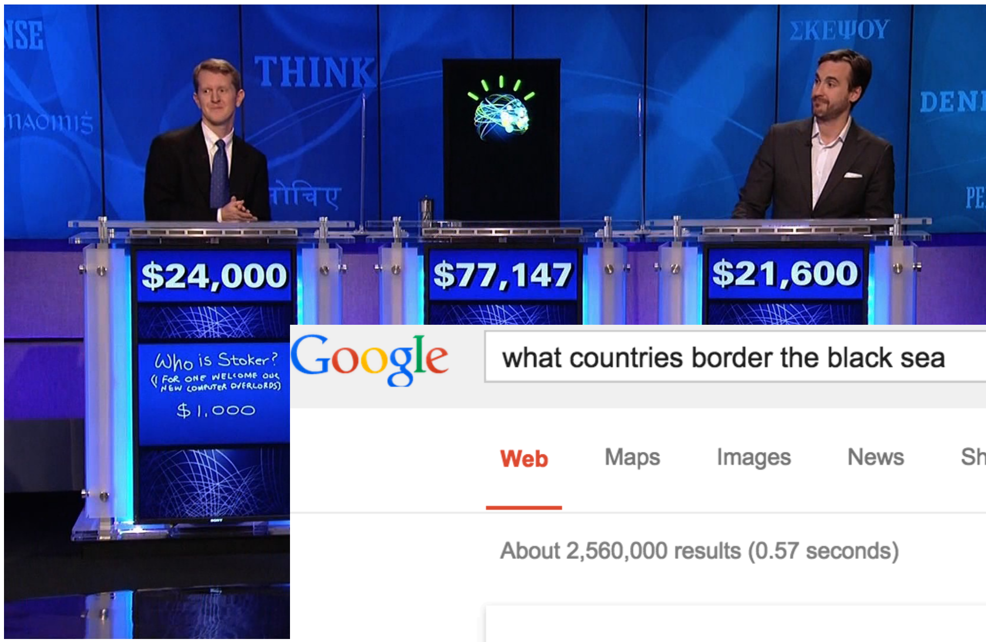
(various slides adapted/borrowed from Dan Klein’s and Chris Manning’s course slides)

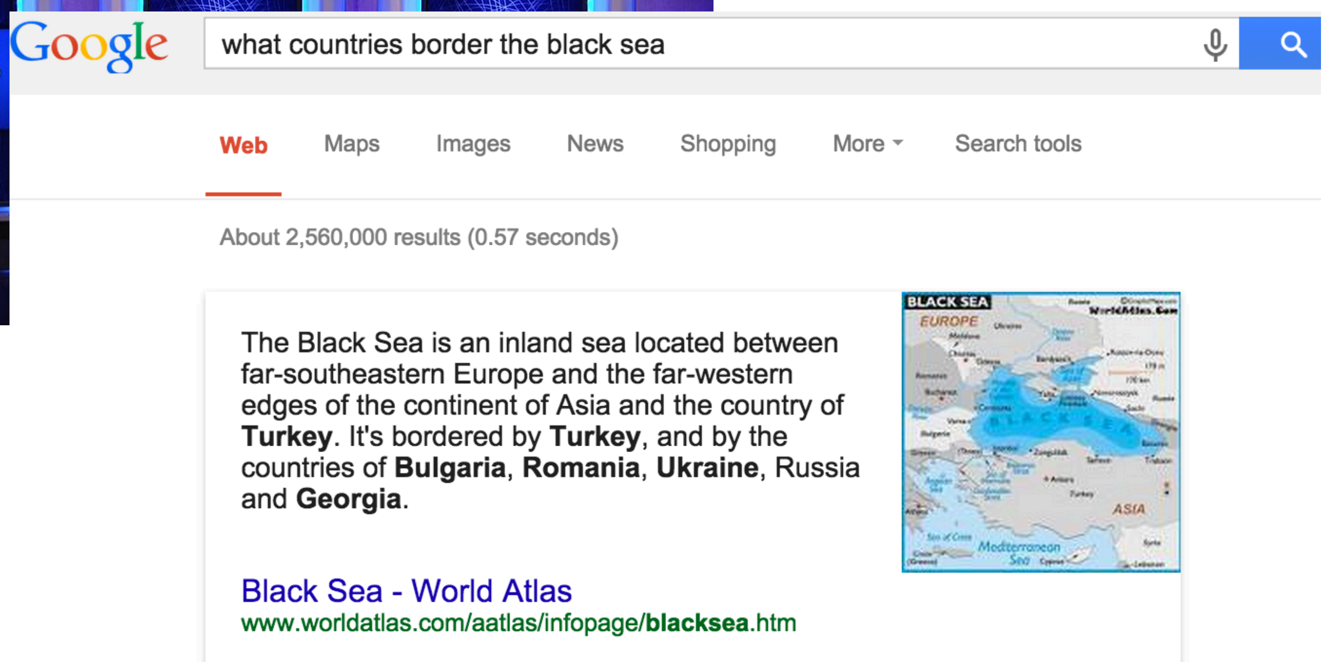
Preface/Disclaimer

- This is meant to be a short (2-3 hours), overview/summary style lecture on some major topics and trends in NLP, with plenty of resource pointers (demos, software, references)
- Hence, it only covers 4-5 topics in some detail, e.g., tagging, parsing, coreference, and semantics (distributional, compositional, semantic parsing, Q&A)
- For some remaining topics, citations and pointers are provided; also, please refer to the full NLP courses and books cited at the end for detailed material
- Inline cites can be matched with full references at the end
- Comments/suggestions welcome: mbansal@ttic.edu

NLP Examples

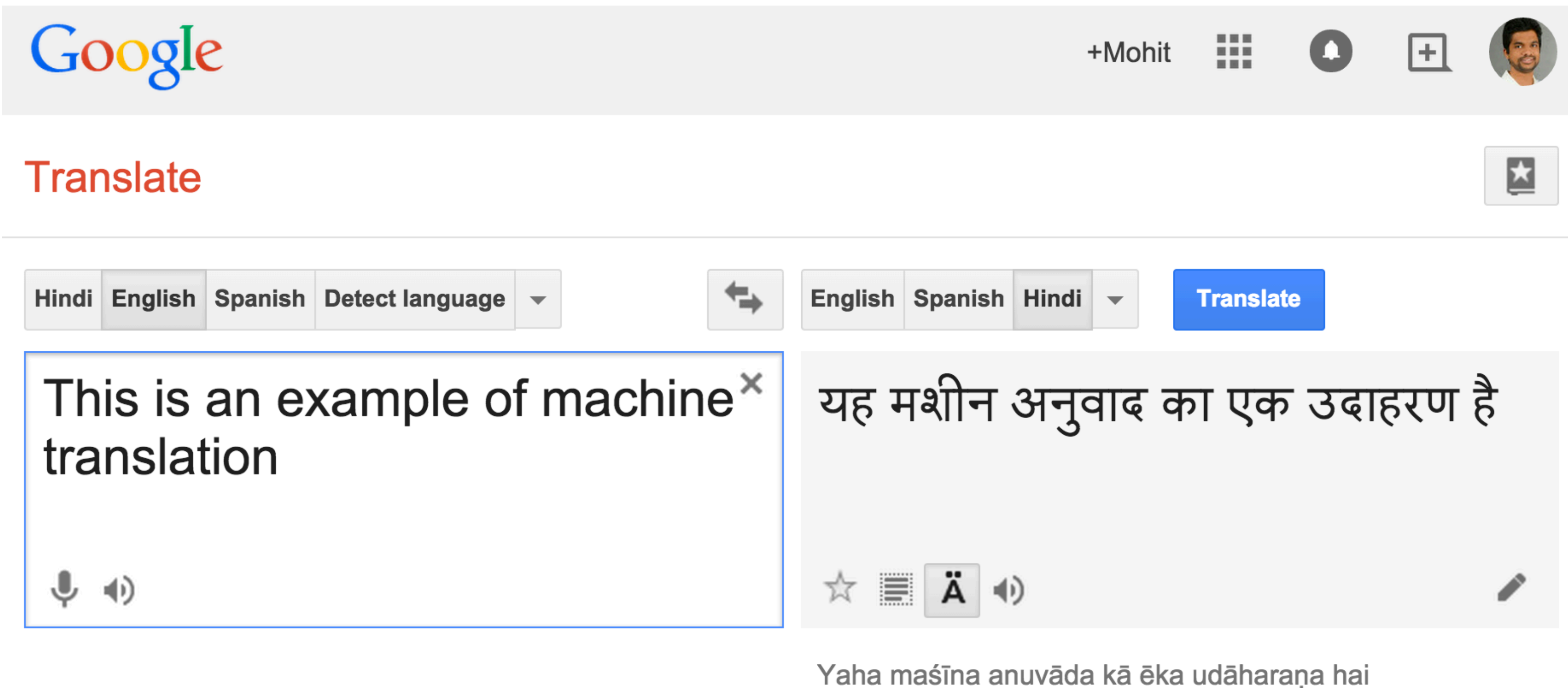
- Question Answering





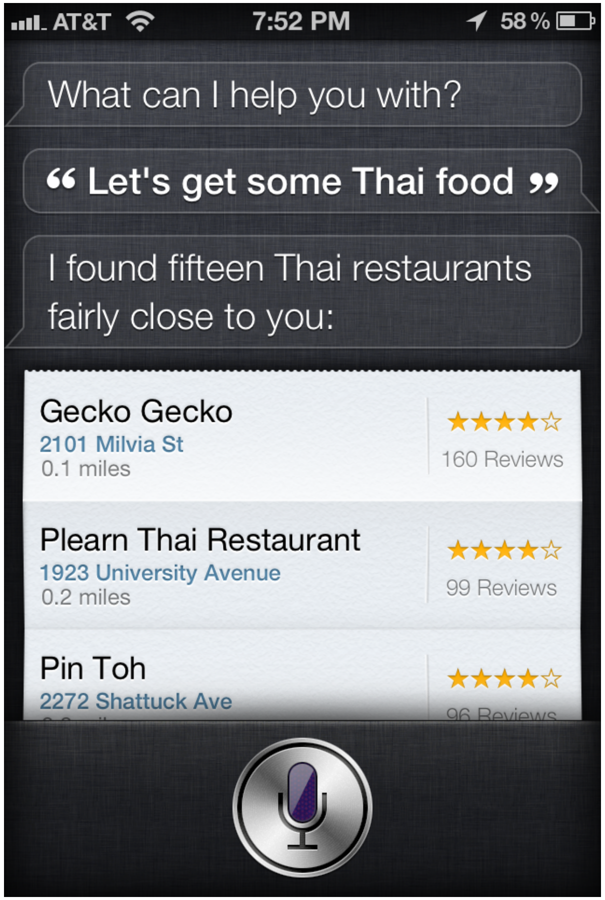
NLP Examples

- Machine Translation



NLP Examples

- Automatic Speech Recognition



Contents

- Part-of-Speech Tagging
- Syntactic Parsing: Constituent, Dependency, CCG, others
- Coreference Resolution
- Distributional Semantics: PMI, NNs, CCA
- Compositional Semantics I: Vector-form, Deep Learning
- Compositional Semantics II: Logic-form, Semantic Parsing, Q&A
- Other Topics: Sentiment Analysis, Machine Translation, Taxonomies, WSI/ WSD, NER, Diachronics, Summarization, Generation, Multimodal, ...
- Some Next Topics: Humor, Sarcasm, Idioms, Human-like Dialog, Poetry

Part-of-Speech Tagging

- Tag sequence of words with syntactic categories (noun, verb, preposition, …)
- Useful in itself:
  - Text-to-speech: *read, lead, record*
  - Lemmatization: *saw*[v] → *see*, *saw*[n] → *saw*
  - Shallow Chunking: grep {JJ | NN}* {NN | NNS}
- Useful for downstream tasks (e.g., in parsing, and as features in various word/text classification tasks)
- Demos: http://nlp.stanford.edu:8080/corenlp/

Penn Treebank Tagset

CC	conjunction, coordinating	and both but either or
CD	numeral, cardinal	mid-1890 nine-thirty 0.5 one
DT	determiner	a all an every no that the
EX	existential there	there
FW	foreign word	gemeinschaft hund ich jeux
IN	preposition or conjunction, subordinating	among whether out on by if
JJ	adjective or numeral, ordinal	third ill-mannered regrettable
JJR	adjective, comparative	braver cheaper taller
JJS	adjective, superlative	bravest cheapest tallest
MD	modal auxiliary	can may might will would
NN	noun, common, singular or mass	cabbage thermostat investment subhumanity
NNP	noun, proper, singular	Motown Cougar Yvette Liverpool
NNPS	noun, proper, plural	Americans Materials States
NNS	noun, common, plural	undergraduates bric-a-brac averages
POS	genitive marker	' 's
PRP	pronoun, personal	hers himself it we them
PRP$	pronoun, possessive	her his mine my our ours their thy your
RB	adverb	occasionally maddeningly adventurously
RBR	adverb, comparative	further gloomier heavier less-perfectly
RBS	adverb, superlative	best biggest nearest worst
RP	particle	aboard away back by on open through
TO	"to" as preposition or infinitive marker	to
UH	interjection	huh howdy uh whammo shucks heck
VB	verb, base form	ask bring fire see take
VBD	verb, past tense	pleaded swiped registered saw
VBG	verb, present participle or gerund	stirring focusing approaching erasing
VBN	verb, past participle	dilapidated imitated reunifed unsettled
VBP	verb, present tense, not 3rd person singular	twist appear comprise mold postpone
VBZ	verb, present tense, 3rd person singular	bases reconstructs marks uses
WDT	WH-determiner	that what whatever which whichever
WP	WH-pronoun	that what whatever which who whom
WP$	WH-pronoun, possessive	whose
WRB	Wh-adverb	however whenever where why

Part-of-Speech Ambiguities

- A word can have multiple parts of speech

VBD		VB			
VBN	VBZ	VBP	VBZ		
NNP	NNS	NN	NNS	CD	NN
Fed	raises	interest	rates	0.5	percent

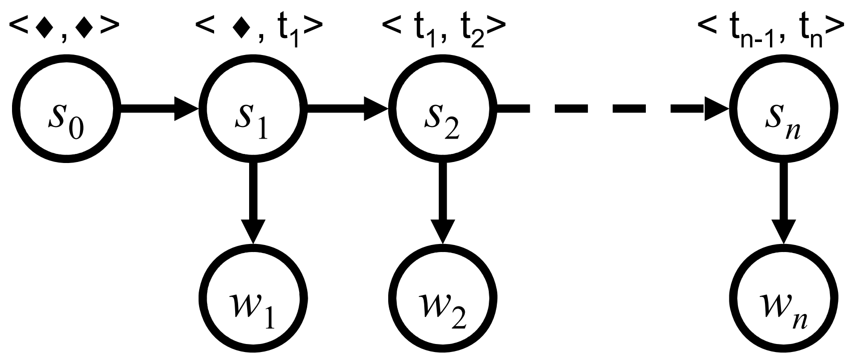
Mrs./NNP Shaefer/NNP never/RB got/VBD **around/RP** to/TO joining/VBG

All/DT we/PRP gotta/VBN do/VB is/VBZ go/VB **around/IN** the/DT corner/NN

Chateau/NNP Petrus/NNP costs/VBZ **around/RB** 250/CD

- Disambiguating features: lexical identity (word), context, morphology (suffixes, prefixes), capitalization, gazetteers (dictionaries), …

Classic Solution: HMMs



$$P(\mathbf{s}, \mathbf{w}) = \prod_i P(s_i|s_{i-1})P(w_i|s_i)$$

- Trigram HMM: states = tag-pairs
- Estimating Transitions: Standard smoothing w/ backoff
- Estimating Emissions: Use unknown word classes (affixes, shapes) and estimate P(t|w) and invert
- Inference: choose most likely (Viterbi) sequence under model

[Brants, 2000]

POS Tagging: Other Models

- Discriminative sequence models with richer features: MEMMs, CRFs (SoA ~= 97%/90% known/unknown)

- Universal POS tagset for multilingual and cross-lingual tagging and parsing [Petrov et al., 2012]

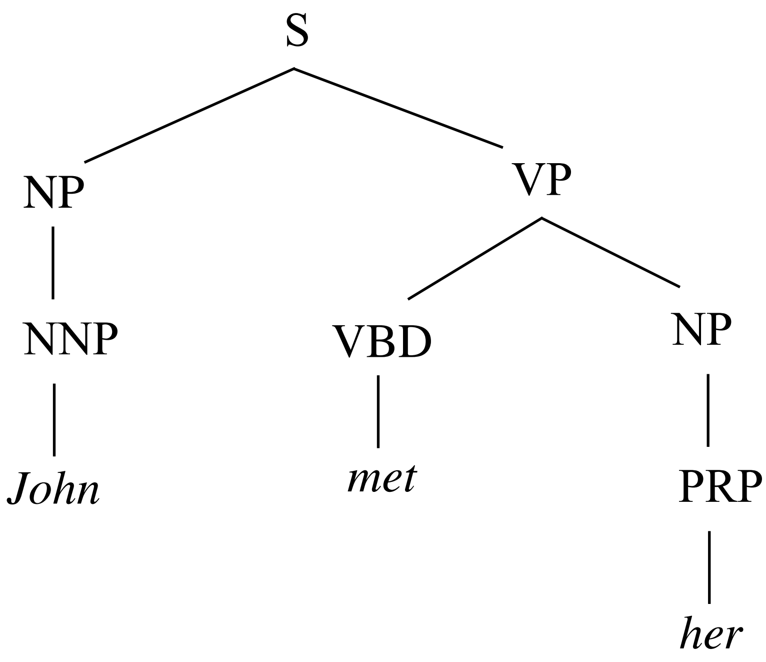
  12 tags: NOUN, VERB, ADJ, ADV, PRON, DET, ADP, NUM, CONJ, PRT, ., X

- Unsupervised tagging also works reasonably well!
  [Yarowsky et al., 2001; Xi and Hwa, 2005; Berg-Kirkpatrick et al., 2010; Christodoulopoulos et al., 2010; Das and Petrov, 2011]

[Brill, 1995; Ratnaparkhi, 1996; Toutanova and Manning, 2000; Toutanova et al., 2003]

Syntactic Parsing -- Constituent

- Phrase-structure parsing or Bracketing



- Demos: http://tomato.banatao.berkeley.edu:8080/parser/parser.html

Probabilistic Context-free Grammars

- A context-free grammar is a tuple <*N, T, S, R*>

N : the set of non-terminals

Phrasal categories: S, NP, VP, ADJP, etc.

Parts-of-speech (pre-terminals): NN, JJ, DT, VB

T : the set of terminals (the words)

S : the start symbol

Often written as ROOT or TOP

Not usually the sentence non-terminal S

R : the set of rules

Of the form $X \rightarrow Y_1\ Y_2 \ldots Y_k$, with $X, Y_i \in N$

Examples: $S \rightarrow NP\ VP$, $VP \rightarrow VP\ CC\ VP$

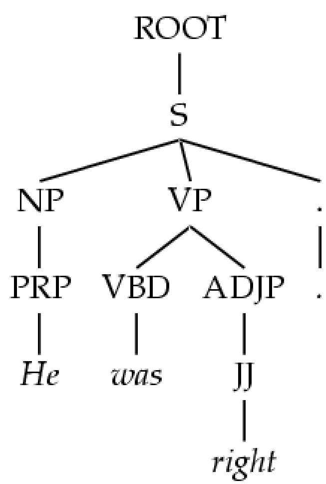
Also called rewrites, productions, or local trees

Probabilistic Context-free Grammars

- A PCFG:
  - Adds a top-down production probability per rule $P(Y_1\ Y_2 \ldots Y_k \mid X)$
  - Allows us to find the 'most probable parse' for a sentence
  - The probability of a parse is just the product of the probabilities of the individual rules

Treebank PCFG

- Need a PCFG for broad coverage parsing
- Extracting a grammar right off the trees is not effective:

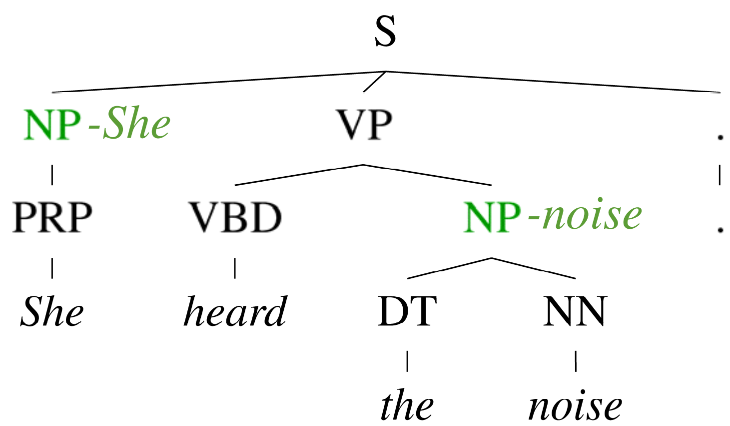


ROOT → S	1
S → NP VP .	1
NP → PRP	1
VP → VBD ADJP	1
…..	

Model	*F1*
Baseline	72.0

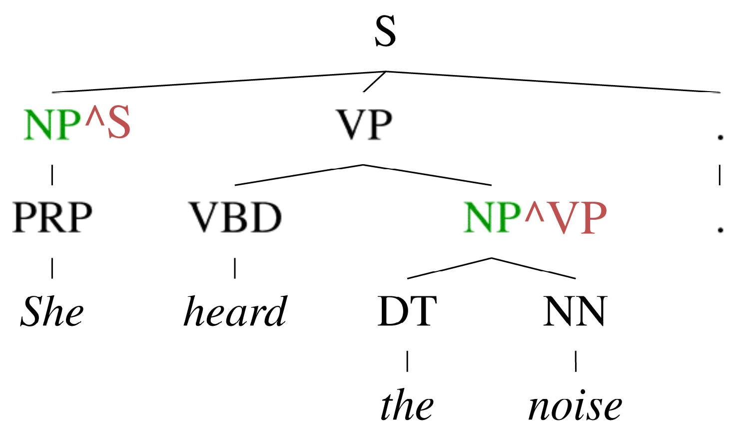
[Charniak, 1996]

Grammar Refinement



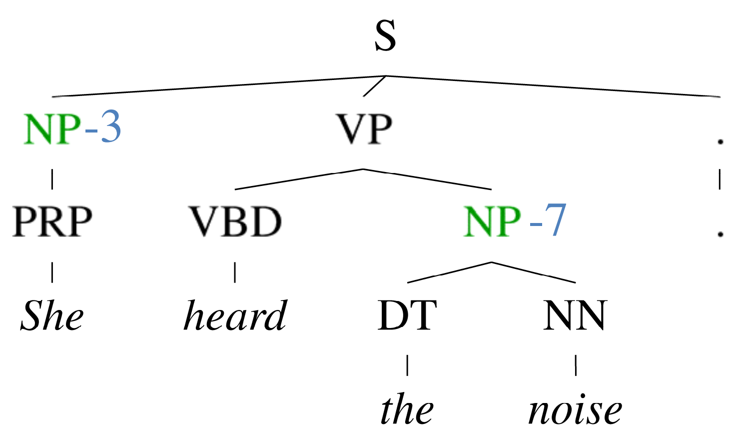
- Conditional independence assumptions often too strong! Not every NP expansion can fill every NP slot
- Better results by enriching the grammar e.g.,
  - Lexicalization [Collins, 1999; Charniak, 2000]

Grammar Refinement



- Conditional independence assumptions often too strong! Not every NP expansion can fill every NP slot
- Better results by enriching the grammar e.g.,
  - Lexicalization [Collins, 1999; Charniak, 2000]
  - Markovization, Manual Tag-splitting [Johnson, 1998; Klein & Manning, 2003]

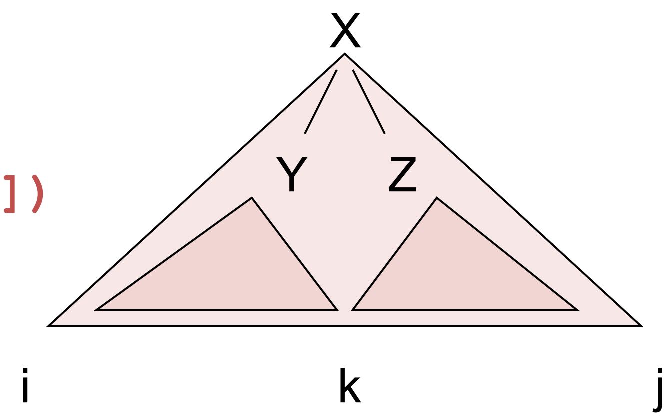
Grammar Refinement



- Conditional independence assumptions often too strong! Not every NP expansion can fill every NP slot
- Better results by enriching the grammar e.g.,
  - Lexicalization [Collins, 1999; Charniak, 2000]
  - Markovization, Manual Tag-splitting [Johnson, 1998; Klein & Manning, 2003]
  - Latent Tag-splitting [Matsuzaki et al., 2005; Petrov et al., 2006]

CKY Parsing Algorithm (Bottom-up)

```
bestScore(s)
 for (i : [0,n-1])
   for (X : tags[s[i]])
     score[X][i][i+1] = tagScore(X,s[i])
 for (diff : [2,n])
   for (i : [0,n-diff])
     j = i + diff
     for (X->YZ : rule)
       for (k : [i+1, j-1])
         score[X][i][j] = max{score[X][i][j], score(X->YZ)
                                               *score[Y][i][k]
                                               *score[Z][k][j]}
```



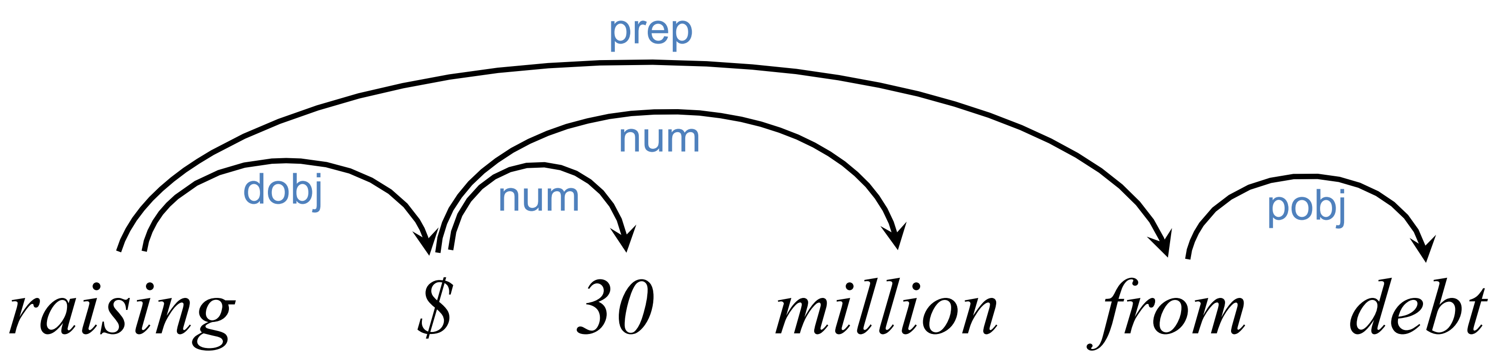
[Cocke, 1970; Kasami, 1965; Younger, 1967]

Some Results

- Collins, 1999 → 88.6 F1 (generative lexical)
- Charniak and Johnson, 2005 → 89.7 / 91.3 F1 (generative lexical / reranking)
- Petrov et al., 2006 → 90.7 F1 (generative unlexical)
- McClosky et al., 2006 – 92.1 F1 (generative + reranking + self-training)

Syntactic Parsing -- Dependency

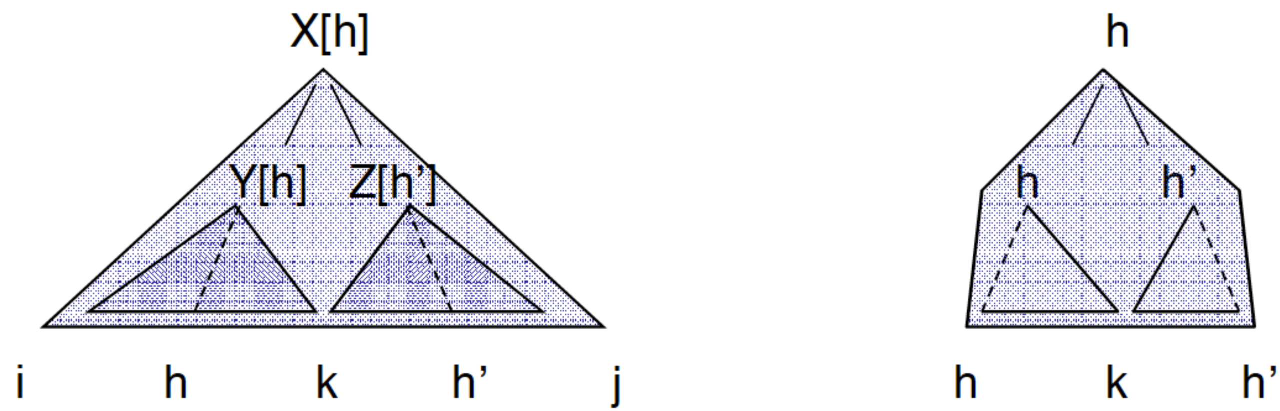
- Predicting directed head-modifier relationship pairs



- Demos: http://nlp.stanford.edu:8080/corenlp/

Syntactic Parsing -- Dependency

- Pure (projective, 1st order) dependency parsing is only cubic [Eisner, 1996]



- Non-projective dependency parsing useful for Czech & other languages – MST algorithms [McDonald et al., 2005]

Parsing: Other Models and Methods

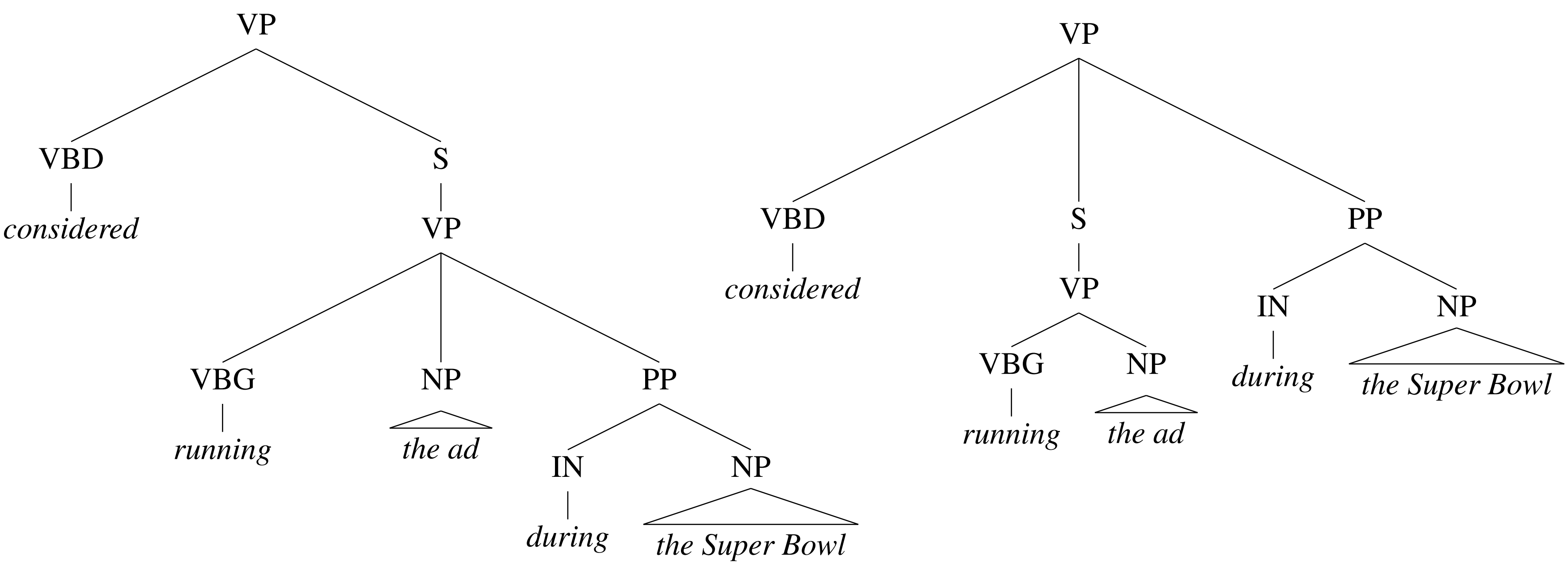
- Combinatory Categorial Grammar [Steedman, 1996, 2000; Clark and Curran, 2004]
- Transition-based Dependency Parsing [Yamada and Matsumoto, 2003; Nivre, 2003]
- Tree-Insertion Grammar, DOP [Schabes and Waters, 1995; Hwa, 1998; Scha, 1990; Bod, 1993; Goodman, 1996; Bansal and Klein, 2010]
- Tree-Adjoining Grammar [Resnik, 1992; Joshi and Schabes, 1998; Chiang, 2000]
- Shift-Reduce Parser [Nivre and Scholz, 2004; Sagae and Lavie, 2005]
- Other: Reranking, A*, K-Best, Self-training, Co-training, System Combination, Cross-lingual Transfer [Sarkar, 2001; Steedman et al., 2003; Charniak and Johnson, 2005; Hwa et al., 2005; Huang and Chiang, 2005; McClosky et al., 2006; Fossum and Knight, 2009; Pauls and Klein, 2009; McDonald et al., 2011]
- Other Demos: http://svn.ask.it.usyd.edu.au/trac/candc/wiki/Demo, http://4.easy-ccg.appspot.com/

World Knowledge is Important

Web Features for Syntactic Parsing

Dependency:

Constituent:



[Nakov and Hearst 2005; Pitler et al., 2010; Bansal and Klein, 2011]

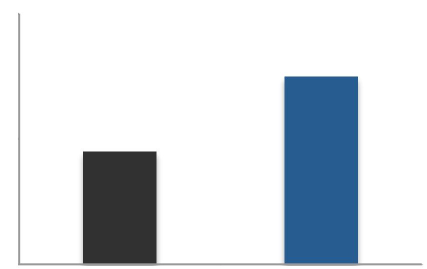
Web Features for Syntactic Parsing

They considered running the ad during the Super Bowl.

Web Ngrams

count(*running it during*) > count(*considered it during*)

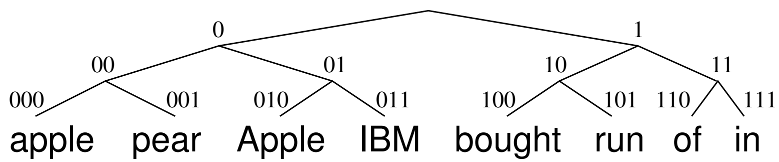
- 7-10% relative error reduction over 90-92% parsers



[Bansal and Klein, 2011]

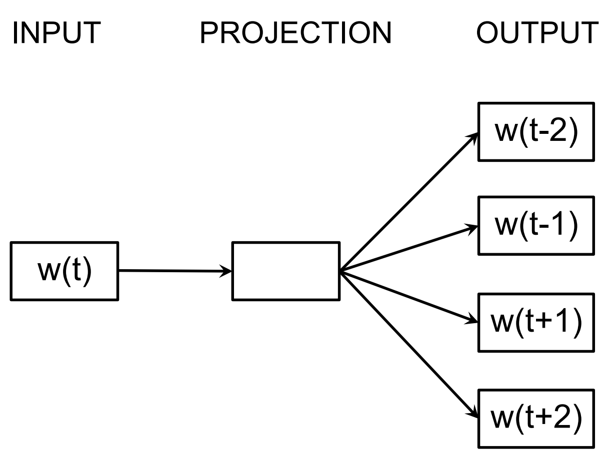
Unsup. Representations for Parsing

- Discrete or continuous, trained on large amounts of context
- BROWN (Brown et al., 1992):



apple → 000
pear → 001
Apple → 010

- SKIPGRAM (Mikolov et al., 2013):



apple → [0.65 0.15 -0.21 0.15 0.70 -0.90]
pear → [0.51 0.05 -0.32 0.20 0.80 -0.95]
Apple → [0.11 0.33 0.51 -0.05 -0.41 0.50]

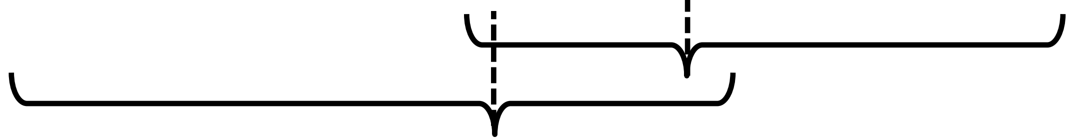
[Koo et al., 2008; Bansal et al., 2014]

Unsup. Representations for Parsing

- Condition on dependency context instead of linear, then convert each dependency to a tuple:

dep label | grandparent | parent | child | dep label

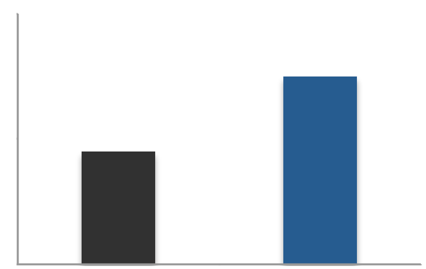
$[PMOD_{<L>} \quad regulation_{<G>} \quad of \quad safety \quad PMOD_{<L>}]$



[*Mr., Mrs., Ms., Prof., III, Jr., Dr.*]
[*Jeffrey, William, Dan, Robert, Stephen, Peter, John, Richard, ...*]
[*Portugal, Iran, Cuba, Ecuador, Greece, Thailand, Indonesia, ...*]

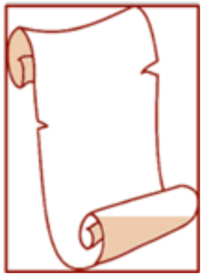
[*his, your, her, its, their, my, our*]
[*Your, Our, Its, My, His, Their, Her*]
[*truly, wildly, politically, financially, completely, potentially, ...*]

- 10% rel. error reduction over 90-92% parsers



[Bansal et al., 2014]

Coreference Resolution

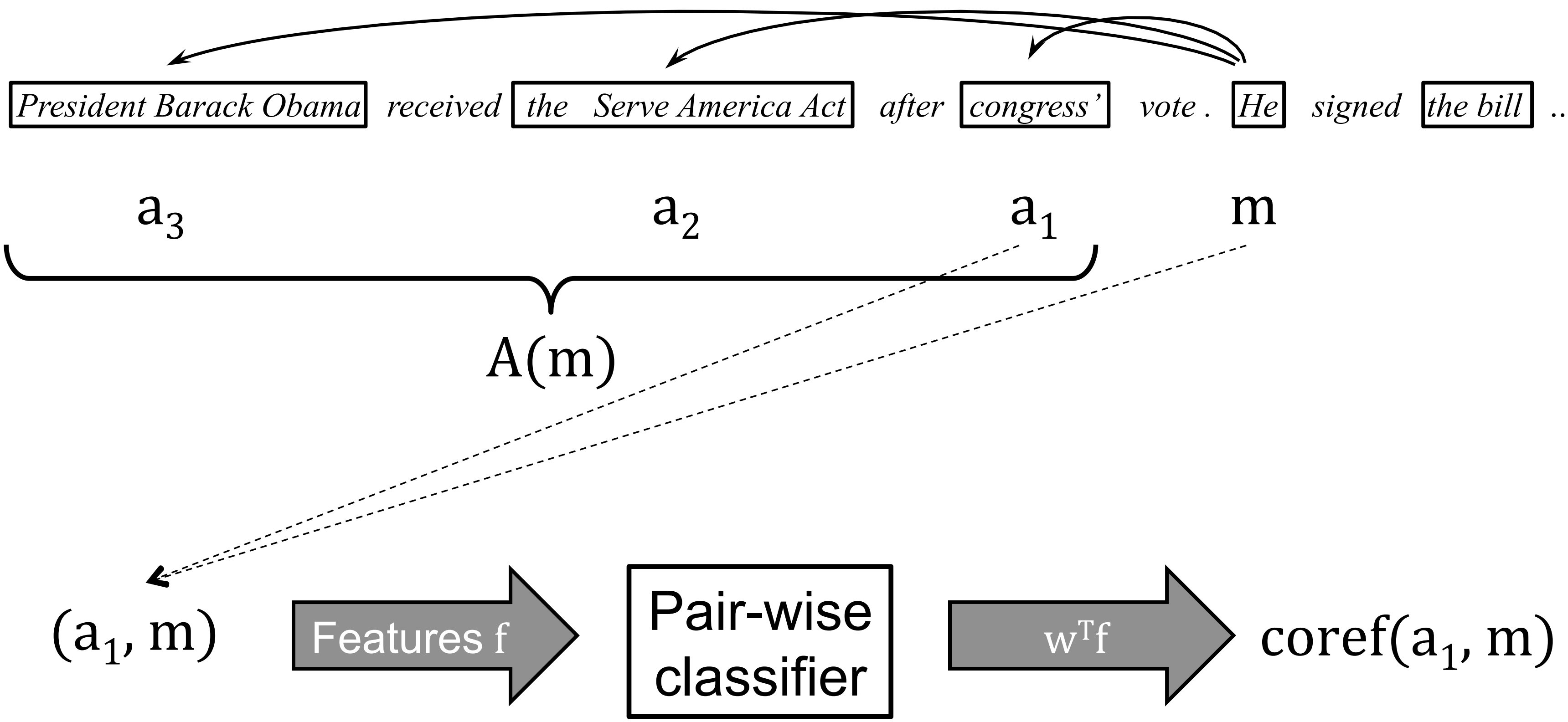
   

President Barack Obama received the Serve America Act after congress' vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.

- Mentions to entity/event clusters
- Demos: http://nlp.stanford.edu:8080/corenlp/process

Mention-pair Models

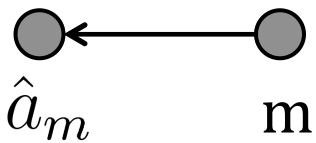
- Pair-wise classification approach:

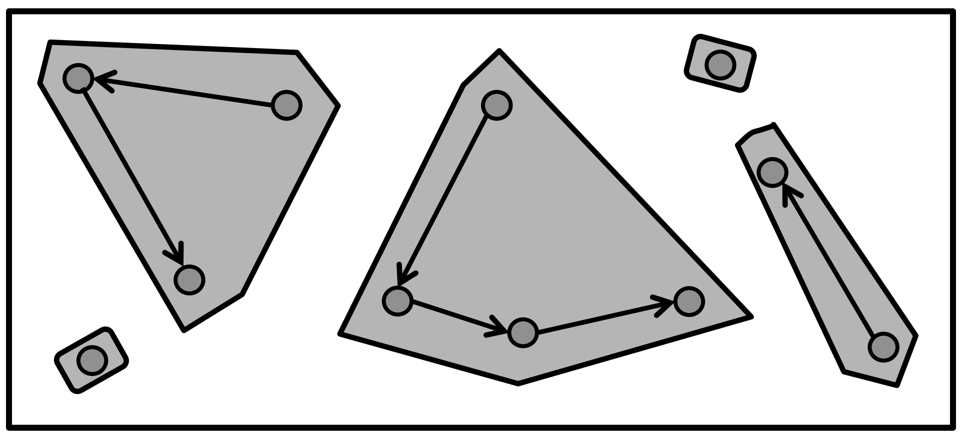


[Soon et al. 2001, Ng and Cardie 2002; Bengtson and Roth, 2008; Stoyanov et al., 2010]

Mention-pair Model

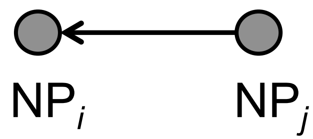
For each mention m, $\hat{a}_m = \underset{a_i \in A(m)}{\operatorname{argmax}}\, coref(a_i, m)$





[Soon et al. 2001, Ng and Cardie 2002; Bengtson and Roth, 2008; Stoyanov et al., 2010]

Standard features



Type	Feature	Description
LEXICAL	SOON_STR	Do the strings match after removing determiners ?
GRAMMATICAL	NUMBER	Do NP_i and NP_j agree in number ?
	GENDER	Do NP_i and NP_j agree in gender ?
	APPOSITIVE	Are the NPs in an appositive relationship ?
SEMANTIC	WORDNET_CLASS	Do NP_i and NP_j have the same WordNet class ?
	ALIAS	Is one NP an alias of the other ?
POSITIONAL	SENTNUM	Distance between the NPs in terms of # of sentences

- Weaknesses: All pairs, Transitivity/Independence errors (*He – Obama – She*), Insufficient information

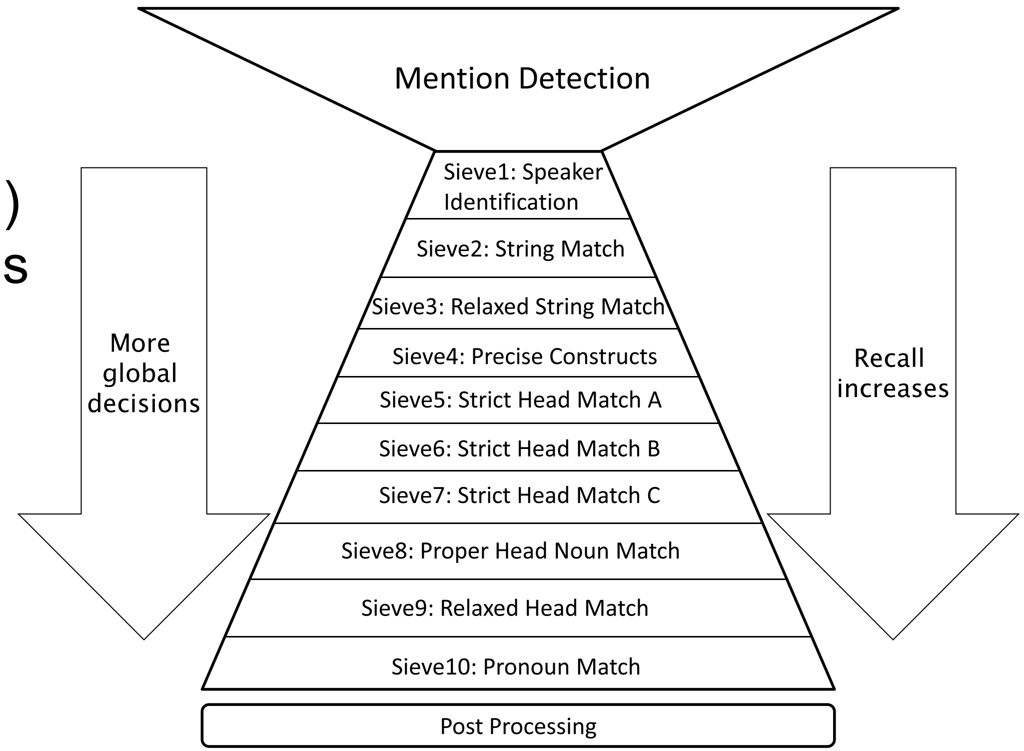
[Soon et al. 2001, Ng and Cardie 2002; Bengtson and Roth, 2008; Stoyanov et al., 2010]

Entity-centric Models

- Each coreference decision is globally informed by previously clustered mentions and their shared attributes

- Lee et al., 2013's deterministic (rule-based) system: multiple, cautious sieves from high to low precision

- Durrett et al., 2013's entity-level model is discriminative, probabilistic using factor graphs and BP

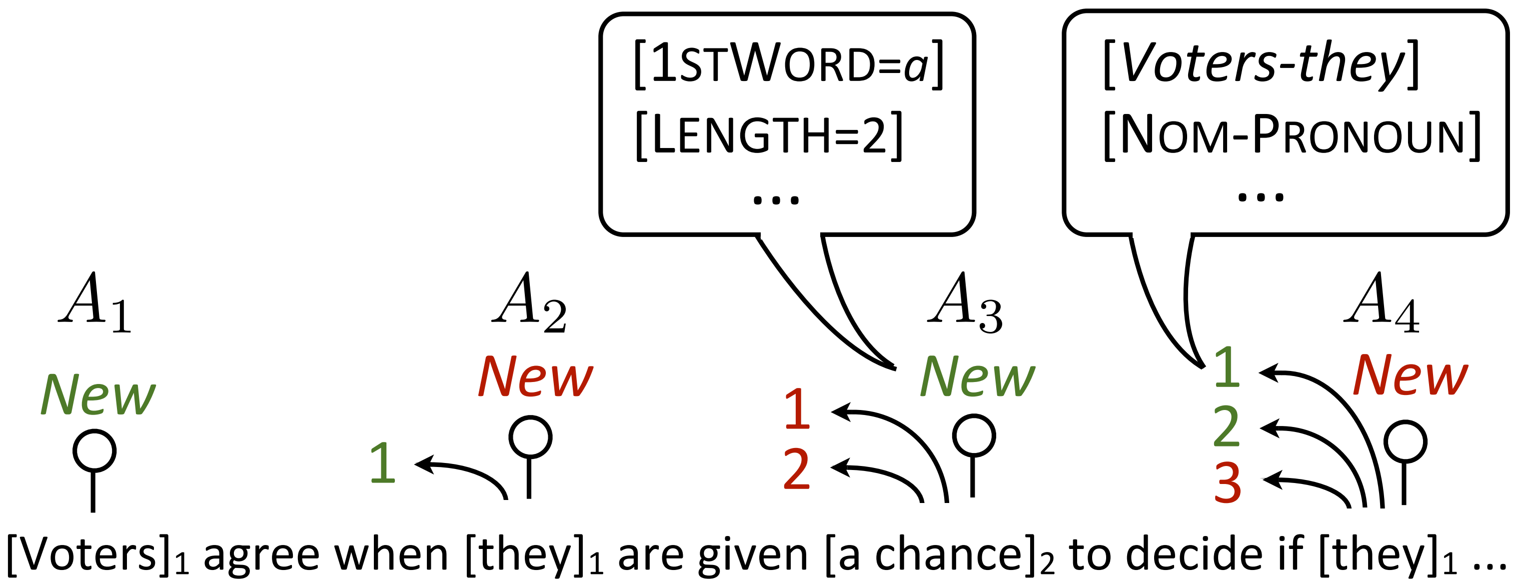


[Haghighi and Klein, 2009; Lee et al., 2013; Durrett et al., 2013]

Mention-Ranking Models (Learned)

- Log-linear model to select at most 1 antecedent for each mention or determine that it begins a new cluster

$$Pr(A_i = a|x) \propto \exp(w^\top f(i,a,x))$$



[Denis and Baldridge, 2008; Durrett and Klein, 2013]

Adding Knowledge to Coref

- External corpora: Web, Wikipedia, YAGO, FrameNet, Gender/Number/Person lists/classifiers, 3D Images, Videos

- Methods:
  - Self-training, Bootstrapping
  - Co-occurrence, Distributional, and Pattern-based Features
  - Entity Linking
  - Visual Cues from 3D Images and Videos

- Daumé III and Marcu, 2005; Markert and Nissim, 2005; Bergsma and Lin, 2006; Ponzetto and Strube, 2006; Haghighi and Klein, 2009; Kobdani et al., 2011; Rahman and Ng, 2011; Bansal and Klein, 2012; Durrett and Klein, 2014; Kong et al., 2014; Ramanathan et al., 2014

Web Features for Coreference

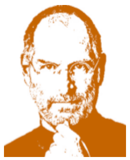
count(*Obama * president*)   vs   count(*Jobs * president*)

When *Obama* *met* *Jobs* *, the president discussed the …*

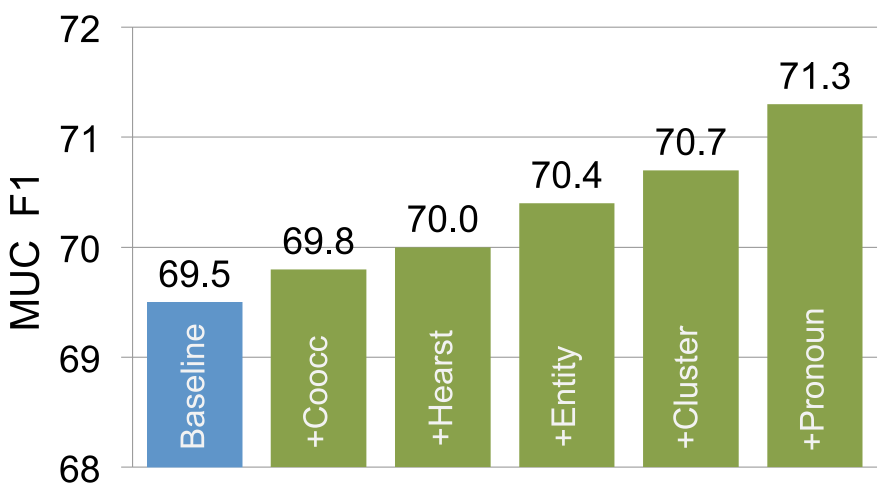
[Bansal and Klein, 2012]

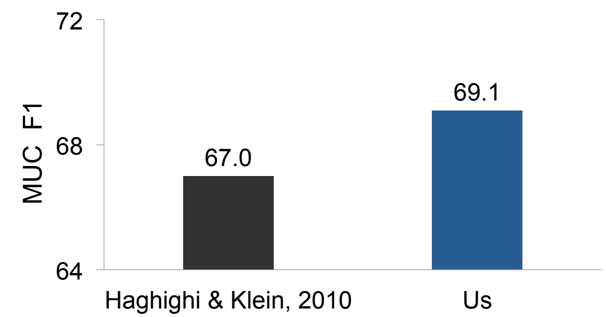
Web Features for Coreference

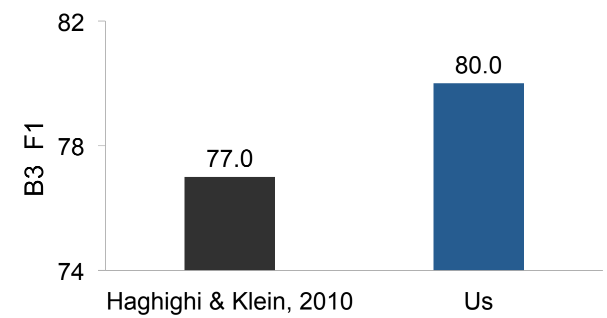
count(*Obama signed bills*) vs count(*Jobs signed bills*)



When Obama met Jobs , the … He signed bills that …



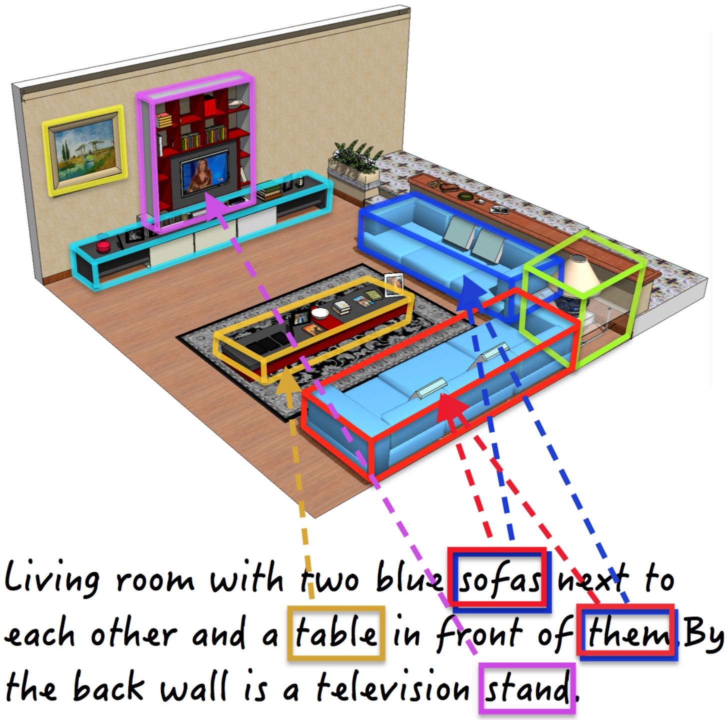




[Bansal and Klein, 2012]

Visual Cues for Coreference

- Joint coreference and 3D image recognition



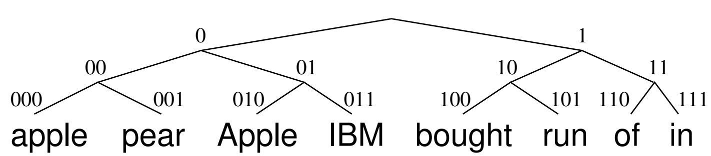
	MUC			B^3		
Method	precision	recall	F1	precision	recall	F1
Stanford	61.56	62.59	62.07	75.05	76.15	75.59
Ours	83.69	51.08	63.44	88.42	70.02	78.15

[Kong, Lin, Bansal, Urtasun, and Fidler, 2014]

Distributional Semantics

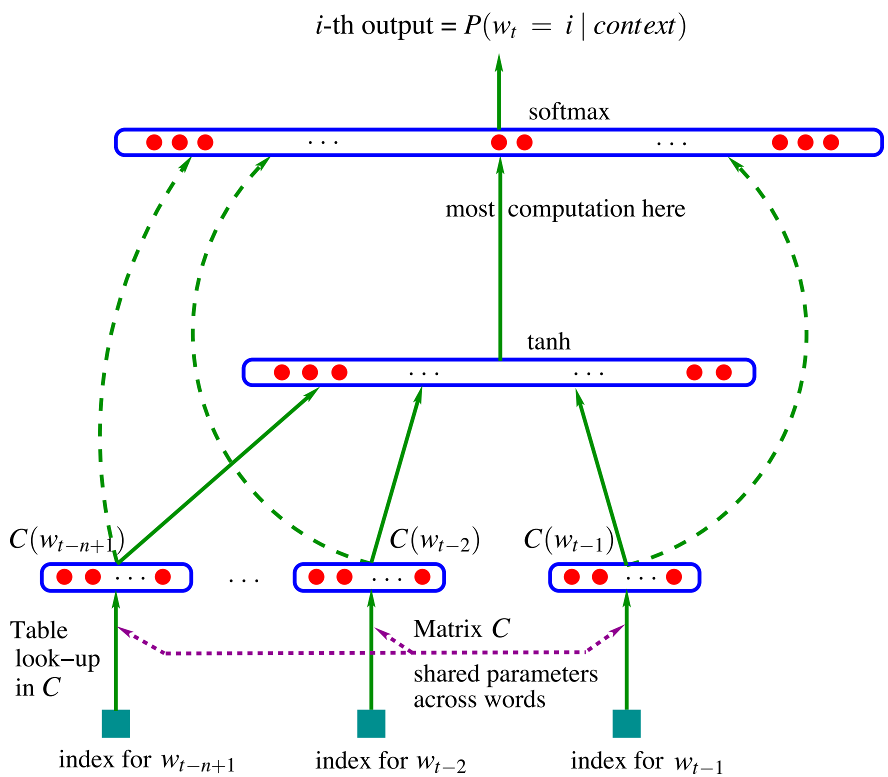
- Words occurring in similar context have similar linguistic behavior (meaning) [Harris, 1954; Firth, 1957]
- Traditional approach: context-counting vectors
  - Count left and right context in window
  - Reweight with PMI or LLR
  - Reduce dimensionality with SVD or NNMF

  [Pereira et al., 1993; Lund & Burgess, 1996; Lin, 1998; Lin and Pantel, 2001; Sahlgren, 2006; Pado & Lapata, 2007; Turney and Pantel, 2010; Baroni and Lenci, 2010]
- More word representations: hierarchical clustering based on bigram LM LL [Brown et al., 1992]



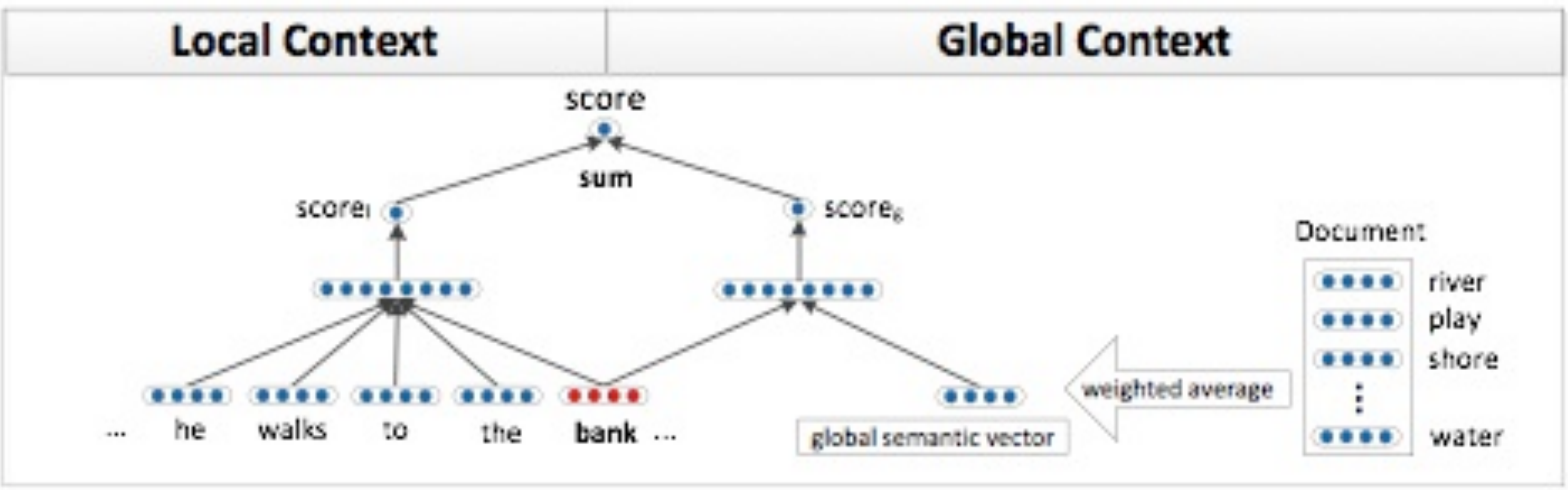
Distributional Semantics -- NNs

- Newer approach: context-predicting vectors (NNs)
  - SENNA [Collobert and Weston, 2008; Collobert et al., 2011]: Multi-layer DNN w/ ranking-loss objective; BoW and sentence-level feature layers, followed by std. NN layers. Similar to [Bengio et al., 2003].



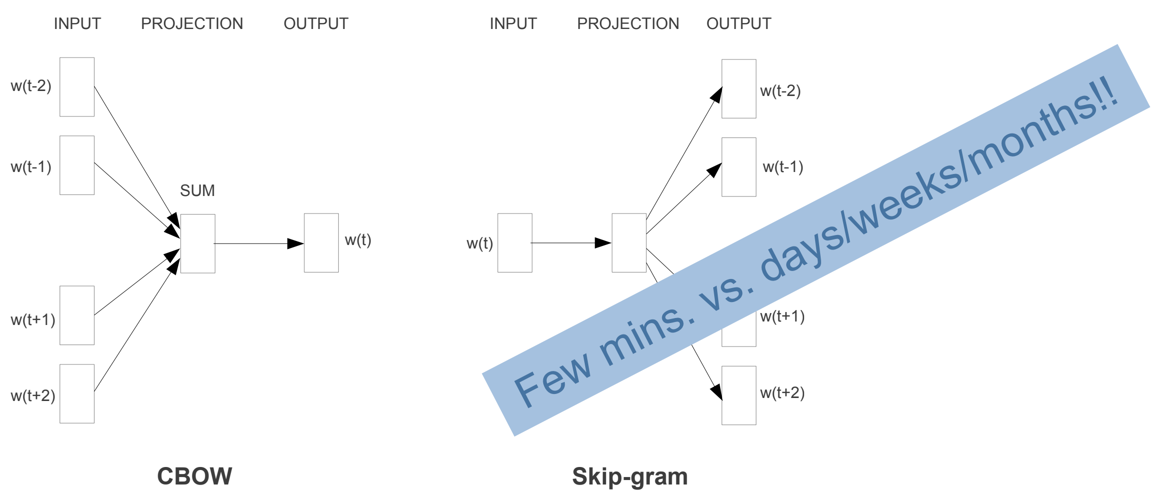
Distributional Semantics -- NNs

- HUANG [Huang et al., 2012]: Add global, document-level context



Distributional Semantics -- NNs

- CBOW, SKIP, word2vec [Mikolov et al., 2013]: Simple, super-fast NN w/ no hidden layer. Continuous BoW model predicts word given context, skip-gram model predicts surrounding words given current word



  - Other: [Mnih and Hinton, 2007; Turian et al., 2010]
- Comparison of count vs. predict (winner) [Baroni et al., 2014]
- Demos: https://code.google.com/p/word2vec, http://metaoptimize.com/projects/wordreprs/, http://ml.nec-labs.com/senna/

Distributional Semantics

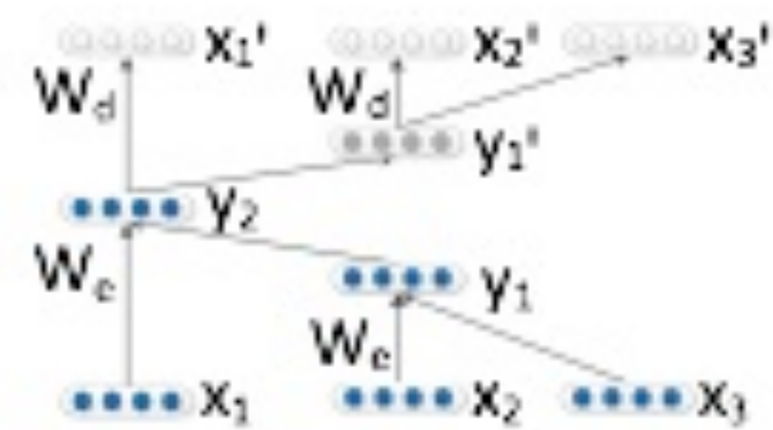
- Other approaches: spectral methods, e.g., CCA
  - Word-context correlation [Dhillon et al., 2011, 2012]
  - Multilingual correlation [Faruqui and Dyer, 2014]
- Some current/next directions: Train task-tailored embeddings to capture specific types of similarity/semantics, e.g.,
  - Dependency context [Bansal et al., 2014, Levy and Goldberg, 2014]
  - Predicate-argument structures [Hashimoto et al., 2014; Madhyastha et al., 2014]
  - Lexicon evidence (PPDB, WordNet, FrameNet) [Xu et al., 2014; Yu and Dredze, 2014; Faruqui et al., 2014]

Compositional Semantics I: NNs

- Composing, combining word vectors to representations for longer units: phrases, sentences, paragraphs, …
- Initial approaches: point-wise sum, multiplication [Mitchell and Lapata, 2010; Blacoe and Lapata, 2012]
- Vector-matrix compositionality [Baroni and Zamparelli, 2010; Zanzotto et al., 2010; Grefenstette and Sadrzadeh, 2011; Socher et al., 2011; Yessenalina and Cardie, 2011]
- Linguistic information added via say parses [Socher et al., 2011b, 2012, 2013a, 2013b, 2014; Hermann and Blunsom, 2013]

Compositional Semantics I: NNs

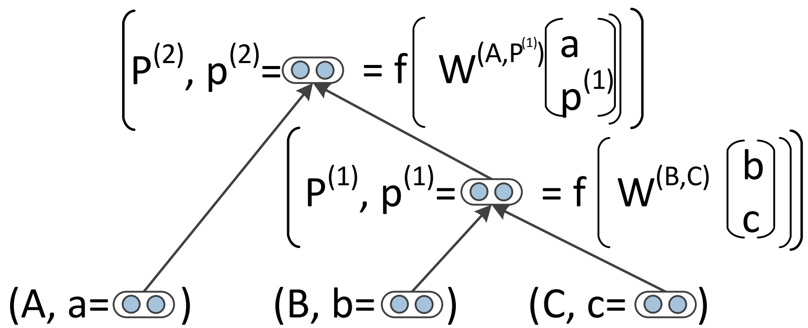
- Socher et al., 2011: Recursive autoencoders (unsupervised) on constituent parse trees

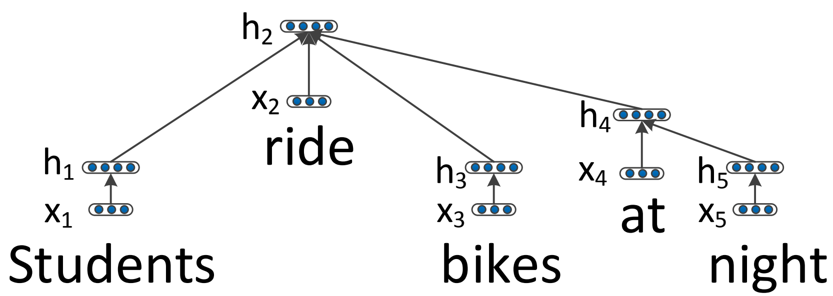


- The unfolding autoencoder which tries to reconstruct all leaf nodes underneath each node.

Compositional Semantics I: NNs

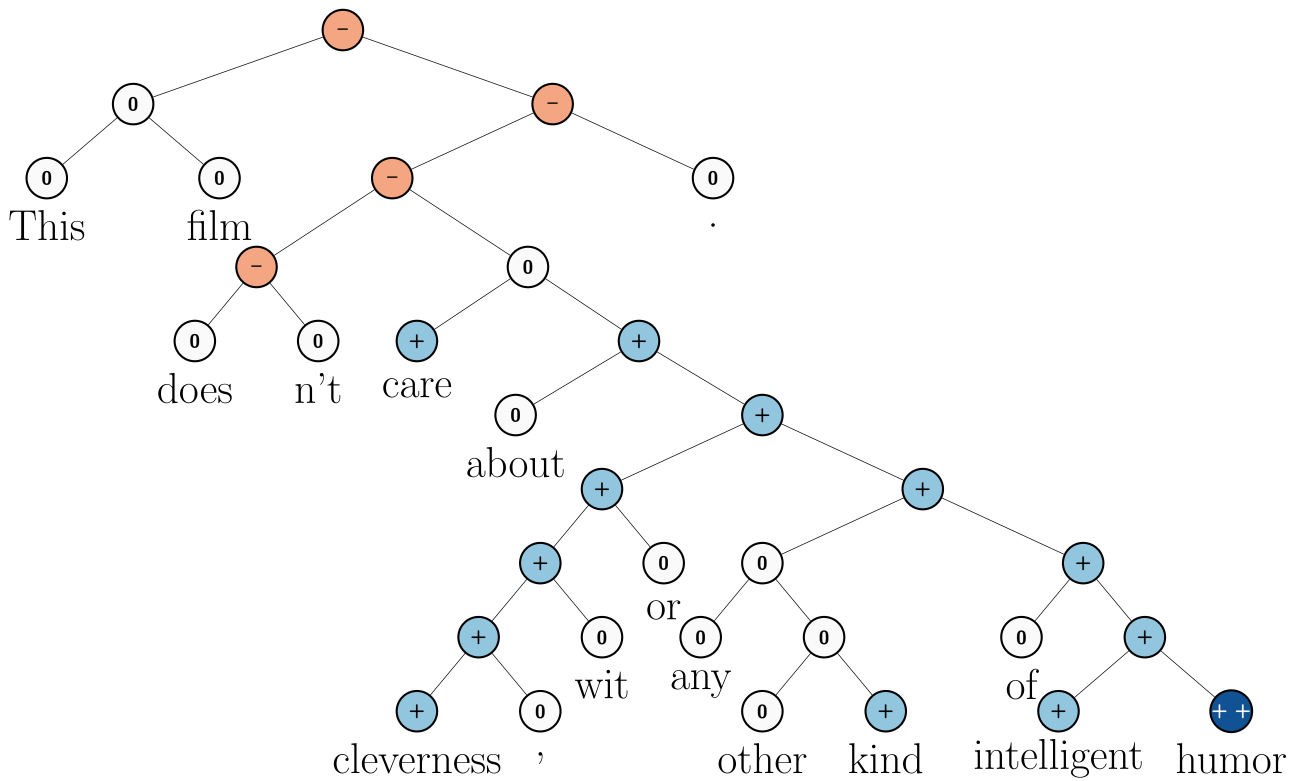
- Socher et al., 2013a, 2014: RNNs on constituent and dependency parse trees





Compositional Semantics I: NNs

- Socher et al., 2013b: Sentiment compositionality



- Demos: http://nlp.stanford.edu:8080/sentiment/rntnDemo.html

[Yessenalina and Cardie, 2011; Socher et al., 2013b]

Compositional Semantics I: NNs

- Various other approaches: [Das and Smith, 2009; Collobert et al., 2011; Grefenstette et al., 2013; Hashimoto et al., 2014; Madhyastha et al., 2014; Chen and Manning, 2014]

- New Deep Learning based Generation: End-to-end MT, Parsing, Caption generation for images, videos [Sutskever et al., 2014; Vinyals et al., 2014a, 2014b; Karpathy and Fei-Fei, 2014; Kiros et al., 2014; Donahue et al., 2014; Fang et al., 2014; Venugopalan et al., 2014]

- Demos: http://deeplearning.net/demos/, http://cs.stanford.edu/people/karpathy/deepimagesent/rankingdemo/, https://www.metamind.io/

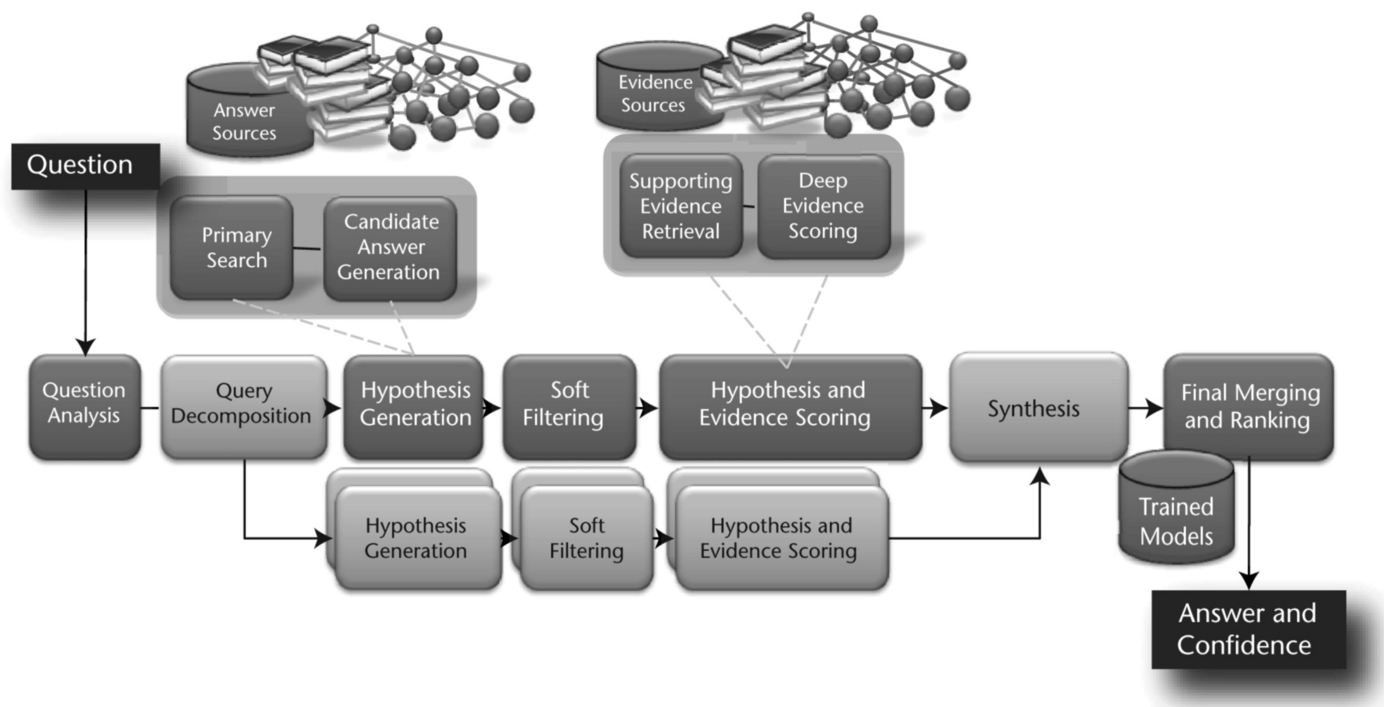
Compositional Semantics II: Logic form

- Logic-based, Semantic Parsing
- Useful for Q&A, IE, grounding, comprehension tasks (summarization, reading tasks)
- A lot of focus on Question Answering

- Demos: http://demo.ark.cs.cmu.edu/parse, www.google.com, Facebook graph search

Question Answering

- Initial approaches to Q&A: pattern matching, pattern learning, query rewriting, information extraction
- Next came a large-scale, open-domain IE system like IBM Watson



[Ferrucci et al., 2010]

Deep Q&A: Semantic Parsing

- Complex, free-form, multi-clause questions

Google | what countries border the black sea

Web · Maps · Images · News · Shopping · More · Search tools

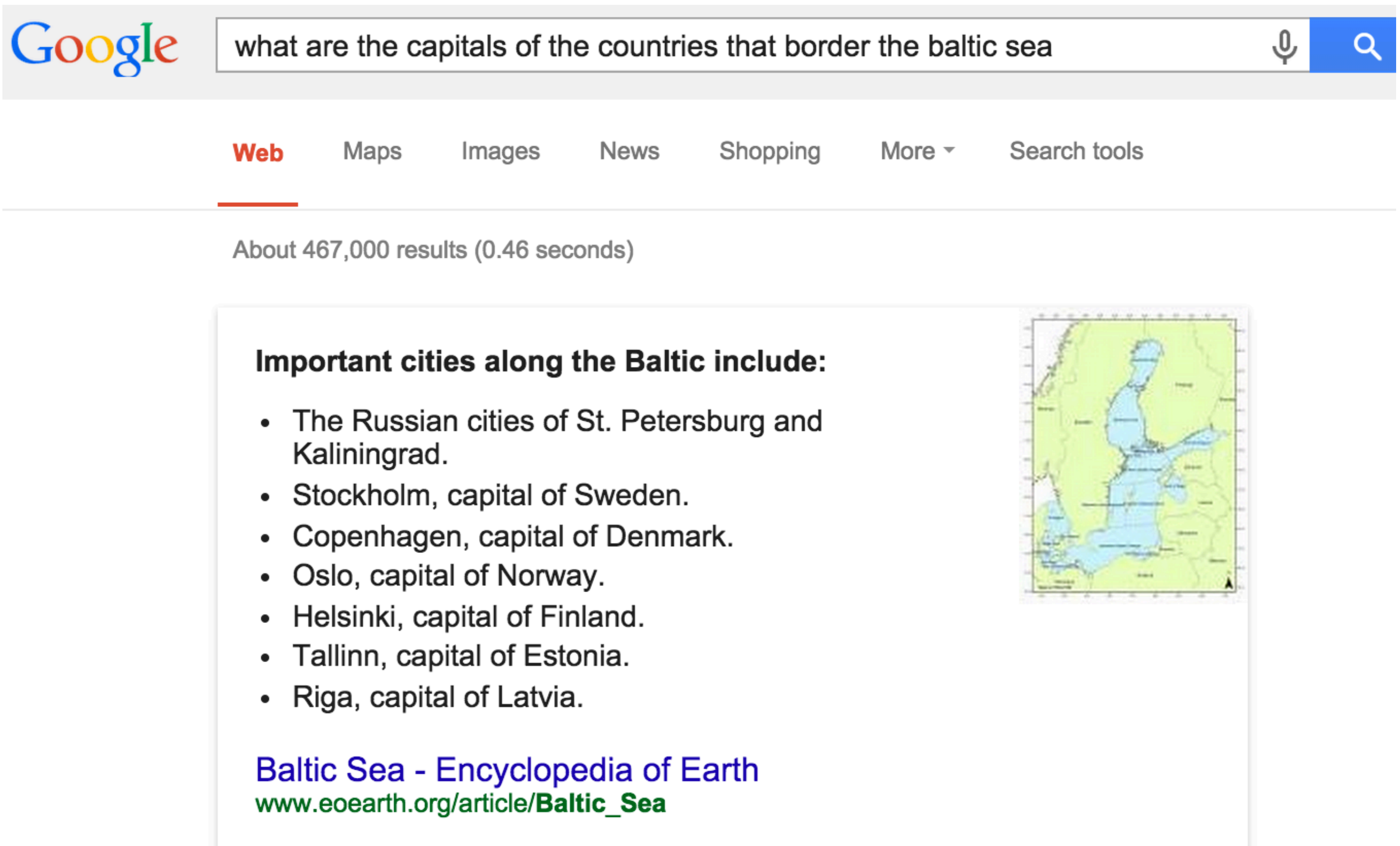
About 2,560,000 results (0.57 seconds)

The Black Sea is an inland sea located between far-southeastern Europe and the far-western edges of the continent of Asia and the country of **Turkey**. It's bordered by **Turkey**, and by the countries of **Bulgaria**, **Romania**, **Ukraine**, Russia and **Georgia**.

Black Sea - World Atlas
www.worldatlas.com/aatlas/infopage/**blacksea**.htm

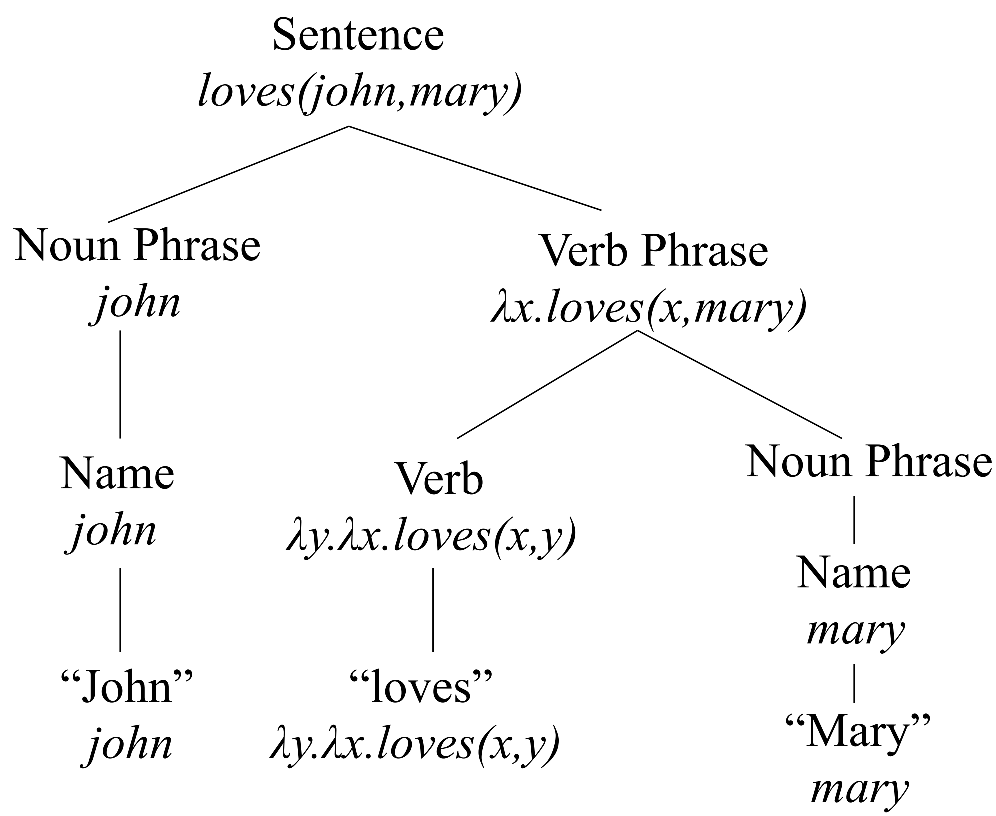
Deep Q&A: Semantic Parsing

- Complex, free-form, multi-clause questions



Semantic Parsing: Logic forms

- Parsing with logic (booleans, individuals, functions) and lambda forms



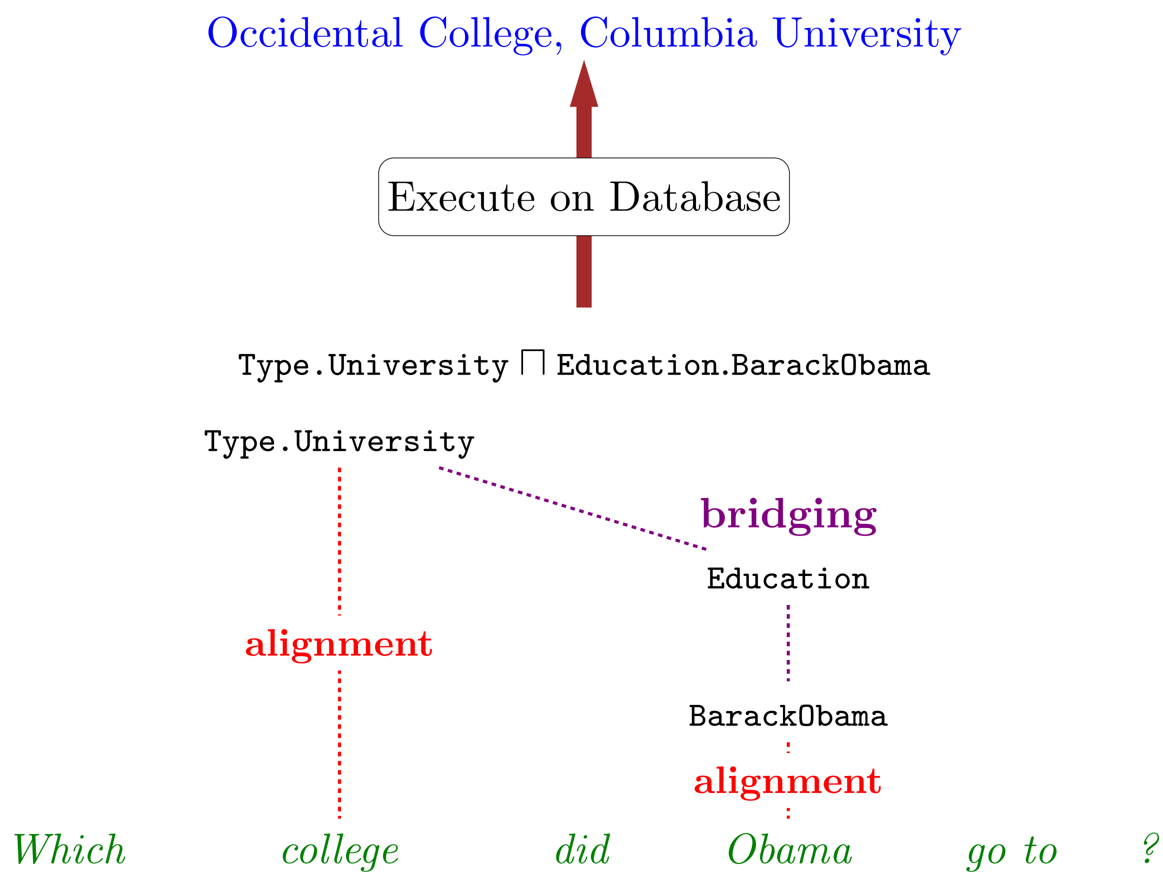
[Wong and Mooney, 2007; Zettlemoyer and Collins, 2007; Poon and Domingos, 2009; Artzi and Zettlemoyer, 2011, 2013; Kwiatkowski et al., 2013; Cai and Yates, 2013; Berant et al., 2013; Poon 2013; Berant and Liang, 2014; Iyyer et al., 2014]

Semantic Parsing Ideas

- Various recent ideas/extensions:
  - unsupervised SP (clustering lambda forms)
  - grounded USP (via databases)
  - Dependency-based compositional semantics (DCS)
  - CCG
  - Bootstrapping w/ conversations
  - On-the-fly ontology matching
  - Question answering on Freebase
  - Paraphrasing
  - RNNs for Q&A
  - Comparison with IE approaches

[Wong and Mooney, 2007; Zettlemoyer and Collins, 2007; Poon and Domingos, 2009; Artzi and Zettlemoyer, 2011, 2013; Kwiatkowski et al., 2013; Cai and Yates, 2013; Berant et al., 2013; Poon 2013; Berant and Liang, 2014; Iyyer et al., 2014; Yao and Van Durne, 2014]

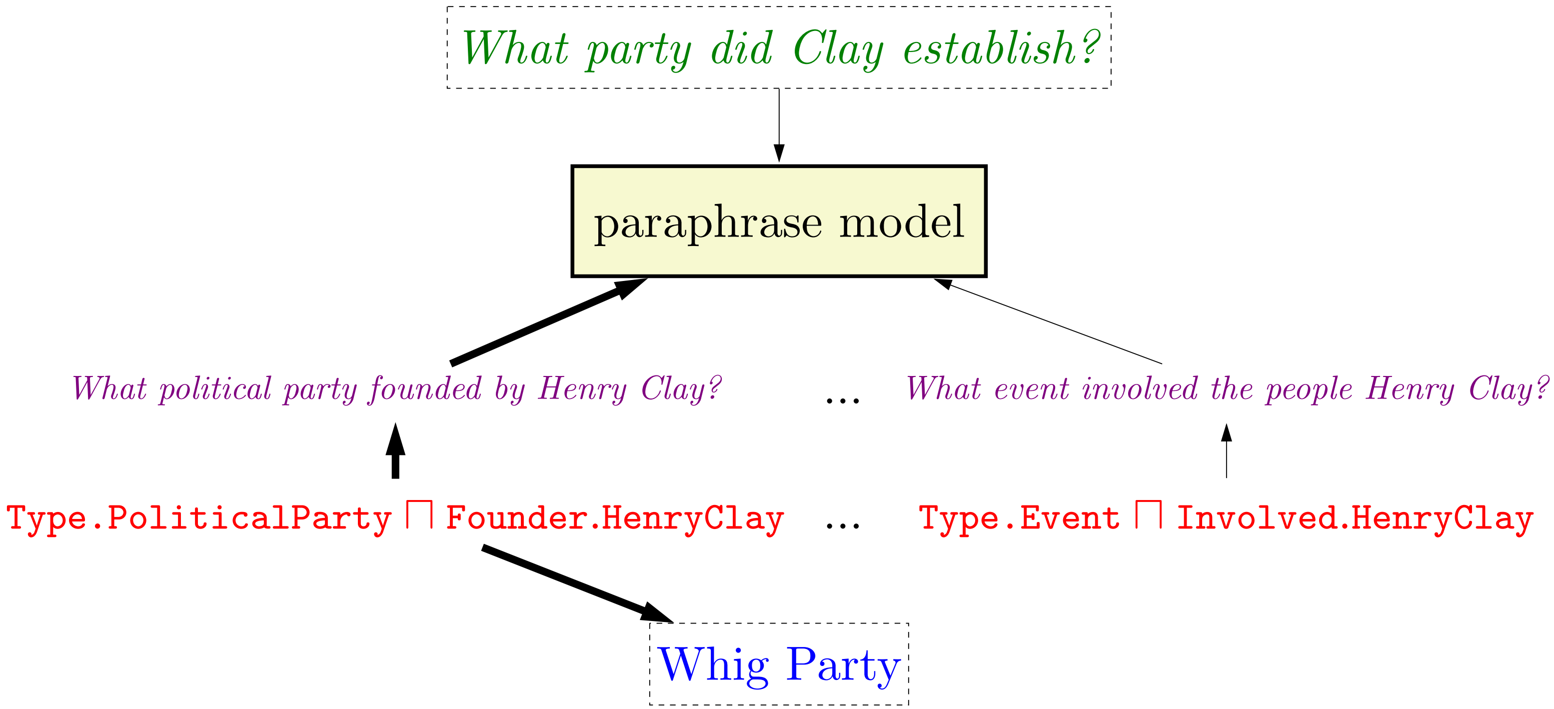
Semantic Parsing on Freebase



Mapping questions to answers via latent logical forms. To narrow down the logical predicate space, they use a (i) coarse *alignment* based on Freebase and a text corpus and (ii) a *bridging* operation that generates predicates compatible with neighboring predicates.

[Berant et al., 2013]

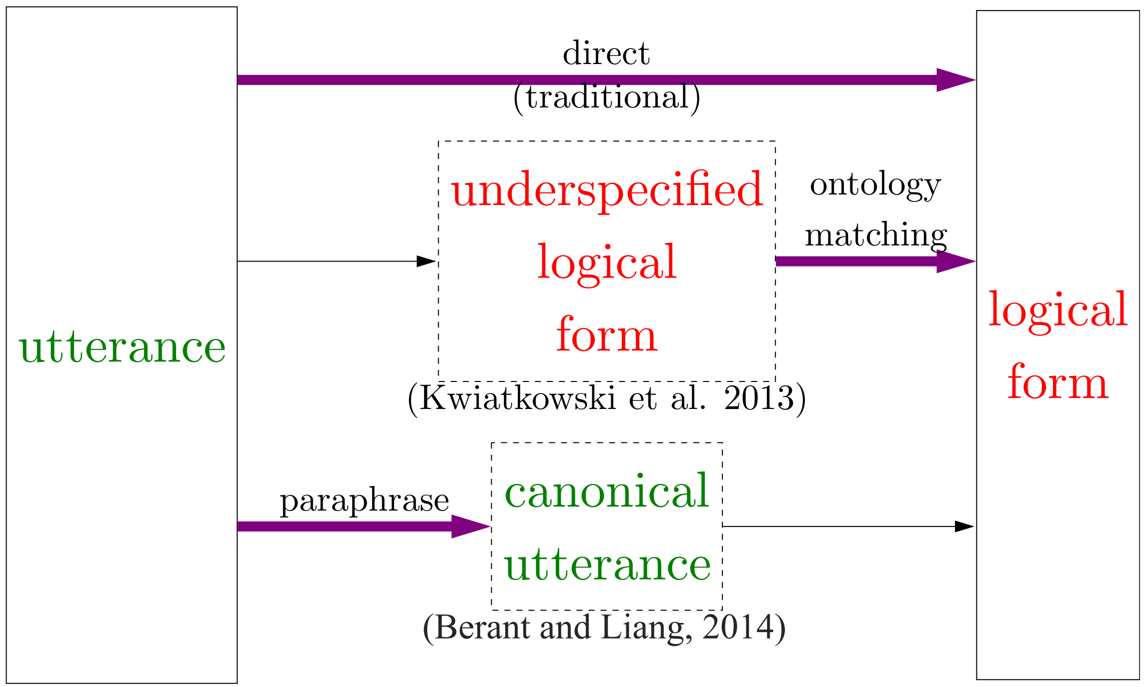
Semantic Parsing via Paraphrasing



For each candidate logical form (red), they generate canonical utterances (purple). The model is trained to paraphrase the input utterance (green) into the canonical utterances associated with the correct denotation (blue).

[Berant and Liang, 2014]

Semantic Parsing via Ontology Matching



The main challenge in semantic parsing is the mismatch between language and the knowledge base. (a) Traditional: map utterances directly to logical forms, (b) Kwiatkowski et al. (2013): map utterance to intermediate, underspecified logical form, then perform ontology matching to handle the mismatch, (c) Berant and Liang (2014): generate intermediate, canonical text utterances for logical forms, then use paraphrase models.

[Kwiatkowski et al., 2013; Berant and Liang, 2014]

Other Topics

- **Machine Translation** [Brown et al., 1990, 1993; Vogel et al., 1996; Wu, 1997; Papineni et al., 2002; Och and Ney, 2002; Och, 2003; Galley et al., 2004; Koehn, 2004; Chiang et al., 2005; Liang et al., 2006a, 2006b; Marcu et al., 2006; Koehn et al., 2007; Gimpel and Smith, 2008; Mi et al., 2008; Chiang, 2010; Galley and Manning, 2010; Bansal et al., 2011; Kalchbrenner and Blunsom, 2013; Vaswani et al., 2013; Auli et al., 2013; Devlin et al., 2014; Sutskever et al., 2014, …many more]
  (Demos: http://www.statmt.org/moses/?n=public.demos, http://lisa.iro.umontreal.ca/mt-demo, https://translate.google.com/)

- **Sentiment Analysis** [Hatzivassiloglou and McKeown, 1997; Das and Chen, 2001; Tong, 2001; Turney, 2002; Pang et al., 2002; Nenkova and Passonneau, 2004; Wiebe et al., 2005; Thomas et al., 2006; Snyder and Barzilay, 2007; Ding et al., 2008; Pang and Lee, 2008; Bansal et al., 2008; Nakagawa et al., 2010; Liu, 2012; Socher et al., 2011, 2013; ...]
  (Demos: http://nlp.stanford.edu:8080/sentiment/rntnDemo.html, http://text-processing.com/demo/sentiment/)

- **Summarization** [Teufel and Moens, 1997; Carbonell and Goldstein, 1998; Knight and Marcu, 2001; White et al., 2001; Lin, 2003, 2004; Daumé III, 2006; Zajic, et al., 2006; Shen et al., 2007; Yih et al., 2007; Schilder and Kondadadi, 2008; Martins and Smith, 2009; Gillick and Favre, 2009; Woodsend and Lapata, 2010; Wang and Cardie, 2012; Hong and Nenkova, 2014; …]
  (Demos: https://semantria.com/demo, http://www.summly.com/)

- **Taxonomy/Ontology Induction** [Widdows, 2003; Snow et al., 2006; Yang and Callan, 2009; Kozareva and Hovy, 2010; Poon and Domingos, 2010; Navigli et al., 2011; Lao et al., 2012; Fountain and Lapata, 2012; Bansal et al., 2014; …]

[*Not exhaustive, various other references]

Many Other Topics …

- Language Modeling
- Word Sense Disambiguation/Induction, NER
- Topic Modeling and Text Classification/Categorization
- Discourse
- Diachronics (Historical Linguistics, Language Reconstruction)
- Decipherment and OCR

Some Next Topics

- Metaphors, Idioms

You:	*I am under the weather today.*
Siri:	*The weather's looking good today …*

- Sarcasm, Insult, Irony, Humor
- Generating realistic stories, poetry, …
- Human-like dialog systems (Turing test)

Resources: Software and Demos

- **POS tagging:** http://nlp.stanford.edu/software/tagger.shtml, https://code.google.com/p/universal-pos-tags/, http://www.ark.cs.cmu.edu/TweetNLP/, …
- **Parsing:** https://code.google.com/p/berkeleyparser/, http://nlp.stanford.edu/software/lex-parser.shtml, https://github.com/BLLIP/bllip-parser, http://www.cs.columbia.edu/~mcollins/code.html, http://www.ark.cs.cmu.edu/TurboParser/
- **Coreference:** http://nlp.stanford.edu/software/dcoref.shtml, http://nlp.cs.berkeley.edu/projects/coref.shtml, http://www.cs.utah.edu/nlp/reconcile/, http://www.bart-coref.org/, http://cogcomp.cs.illinois.edu/page/software_view/Coref
- **Word embeddings:** https://code.google.com/p/word2vec, http://metaoptimize.com/projects/wordreprs/, http://ml.nec-labs.com/senna/, http://nlp.stanford.edu/projects/glove/, http://ttic.uchicago.edu/~mbansal/data/syntacticEmbeddings.zip, http://www.socher.org/index.php/Main/ImprovingWordRepresentationsViaGlobalContextAndMultipleWordPrototypes, http://www.wordvectors.org/web-eacl14-vectors/de-projected-en-512.txt.gz
- **Compositional embeddings**: http://nlp.stanford.edu/sentiment/, http://nal.co/DCNN, http://www.socher.org/index.php/Main/ParsingWithCompositionalVectorGrammars, http://www.socher.org/index.php/Main/DynamicPoolingAndUnfoldingRecursiveAutoencodersForParaphraseDetection
- **Semantic Paring, Q&A (Compositional Semantics II):** http://www-nlp.stanford.edu/software/sempre/, https://bitbucket.org/yoavartzi/spf, https://code.google.com/p/jacana/, http://cs.umd.edu/~miyyer/qblearn/, http://alchemy.cs.washington.edu/usp/, http://www.ark.cs.cmu.edu/SEMAFOR/,
- **Most of the demo links are inline with each topic's slides**

Resources: Courses and Books

- Berkeley NLP course: http://www.cs.berkeley.edu/~klein/cs288/fa14/
- CMU NLP course: www.ark.cs.cmu.edu/NLP
- Stanford NLP course: http://web.stanford.edu/class/cs224n
- Many others: Brown, Columbia, Cornell, JHU, MIT, Maryland, UPenn, …
- Books:
  - Jurafsky and Martin, Speech and Language Processing, 2nd edition, 2009
  - Manning and Shuetze, Foundations of Statistical Natural Language Processing
- Many others references (in the material above) …

References

Artzi, Yoav, and Luke Zettlemoyer. "Bootstrapping semantic parsers from conversations." *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics, 2011.

Artzi, Yoav, and Luke Zettlemoyer. "Weakly Supervised Learning of Semantic Parsers for Mapping Instructions to Actions." *TACL* 1 (2013): 49-62.

Auli, Michael, Michel Galley, Chris Quirk, and Geoffrey Zweig. "Joint Language and Translation Modeling with Recurrent Neural Networks." In *EMNLP*, pp. 1044-1054. 2013.

Bansal, Mohit, Claire Cardie, and Lillian Lee. "The Power of Negative Thinking: Exploiting Label Disagreement in the Min-cut Classification Framework." In*COLING (Posters)*, pp. 15-18. 2008.

Bansal, Mohit, and Dan Klein. "Simple, accurate parsing with an all-fragments grammar." *Proceedings of the 48th annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010.

Bansal, Mohit, and Dan Klein. "Web-scale features for full-scale parsing."*Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

Bansal, Mohit, Chris Quirk, and Robert C. Moore. "Gappy phrasal alignment by agreement." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pp. 1308-1317. Association for Computational Linguistics, 2011.

Bansal, Mohit, and Dan Klein. "Coreference semantics from web features."*Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics, 2012.

Bansal, Mohit, David Burkett, Gerard de Melo, and Dan Klein. "Structured Learning for Taxonomy Induction with Belief Propagation." *ACL*, 2014

Bansal, Mohit, Kevin Gimpel, and Karen Livescu. "Tailoring Continuous Word Representations for Dependency Parsing." *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2014.

Baroni, Marco, and Alessandro Lenci. "Distributional memory: A general framework for corpus-based semantics." *Computational Linguistics* 36.4 (2010): 673-721.

References

Baroni, Marco, and Roberto Zamparelli. "Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space."*Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2010.

Baroni, Marco, Georgiana Dinu, and Germán Kruszewski. "Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors." *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Vol. 1. 2014.

Bengio, Yoshua, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. "A Neural Probabilistic Language Model." *Journal of Machine Learning Research* 3 (2003): 1137-1155.

Bengtson, Eric, and Dan Roth. "Understanding the value of features for coreference resolution." *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2008.

Berant, Jonathan, Andrew Chou, Roy Frostig, and Percy Liang. "Semantic Parsing on Freebase from Question-Answer Pairs." In *EMNLP*, pp. 1533-1544. 2013.

Berant, Jonathan, and Percy Liang. "Semantic parsing via paraphrasing."*Proceedings of ACL*. 2014.

Berg-Kirkpatrick, Taylor, Alexandre Bouchard-Côté, John DeNero, and Dan Klein. "Painless unsupervised learning with features."*Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010.

Bergsma, Shane, and Dekang Lin. "Bootstrapping path-based pronoun resolution." *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2006.

Blacoe, William, and Mirella Lapata. "A comparison of vector-based representations for semantic composition." *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, 2012.

Bod, Rens. "Using an annotated corpus as a stochastic grammar." *Proceedings of the sixth conference on European chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 1993.

References

Brants, Thorsten. "TnT: a statistical part-of-speech tagger." *Proceedings of the sixth conference on Applied natural language processing*. Association for Computational Linguistics, 2000.

Brill, Eric. "Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging." *Computational linguistics*21.4 (1995): 543-565.

Brown, Peter F., Peter V. Desouza, Robert L. Mercer, Vincent J. Della Pietra, and Jenifer C. Lai. "Class-based n-gram models of natural language."*Computational linguistics* 18.4 (1992): 467-479.

Brown, Peter F., Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. "The mathematics of statistical machine translation: Parameter estimation." *Computational linguistics* 19, no. 2 (1993): 263-311.

Cai, Qingqing, and Alexander Yates. "Large-scale Semantic Parsing via Schema Matching and Lexicon Extension." In *ACL (1)*, pp. 423-433. 2013.

Charniak, Eugene. "Tree-bank grammars." *Proceedings of the National Conference on Artificial Intelligence*. 1996.

Charniak, Eugene. "A maximum-entropy-inspired parser." *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*. Association for Computational Linguistics, 2000.

Charniak, Eugene, and Mark Johnson. "Coarse-to-fine n-best parsing and MaxEnt discriminative reranking." *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2005.

Chen, Danqi, and Christopher D. Manning. "A fast and accurate dependency parser using neural networks." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014.

Chiang, David. "Statistical parsing with an automatically-extracted tree adjoining grammar." *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2000.

Chiang, David. "A hierarchical phrase-based model for statistical machine translation." In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 263-270. Association for Computational Linguistics, 2005.

Chiang, David. "Learning to translate with source and target syntax." In*Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1443-1452. Association for Computational Linguistics, 2010.

References

Christodoulopoulos, Christos, Sharon Goldwater, and Mark Steedman. "Two Decades of Unsupervised POS induction: How far have we come?."*Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2010.

Clark, Stephen, and James R. Curran. "Parsing the WSJ using CCG and log-linear models." *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2004.

Cocke, John. "Programming languages and their compilers." *Preliminary notes (Technical report) (2nd revised ed.). CIMS, NYU.,* 1970.

Collins, Michael. "Head-Driven Statistical Models for Natural Language Parsing". Diss. University of Pennsylvania, 1999.

Collobert, Ronan, and Jason Weston. "A unified architecture for natural language processing: Deep neural networks with multitask learning."*Proceedings of the 25th international conference on Machine learning*. ACM, 2008.

Collobert, Ronan, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. "Natural language processing (almost) from scratch."*The Journal of Machine Learning Research* 12 (2011): 2493-2537.

Das, Dipanjan, and Noah A. Smith. "Paraphrase identification as probabilistic quasi-synchronous recognition." *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*. Association for Computational Linguistics, 2009.

Das, Dipanjan, and Slav Petrov. "Unsupervised part-of-speech tagging with bilingual graph-based projections." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

Das, Sanjiv, and Mike Chen. "Yahoo! for Amazon: Extracting market sentiment from stock message boards." In *Proceedings of the Asia Pacific finance association annual conference (APFA)*, vol. 35, p. 43. 2001.

Daumé III, Hal, and Daniel Marcu. "A large-scale exploration of effective global features for a joint entity detection and tracking model." *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2005.

Denis, Pascal, and Jason Baldridge. "Specialized models and ranking for coreference resolution." *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2008.

References

Devlin, Jacob, Rabih Zbib, Zhongqiang Huang, Thomas Lamar, Richard Schwartz, and John Makhoul. "Fast and robust neural network joint models for statistical machine translation." In *52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD, USA, June*. 2014.

Dhillon, Paramveer, Dean P. Foster, and Lyle H. Ungar. "Multi-view learning of word embeddings via cca." *Advances in Neural Information Processing Systems*. 2011.

Dhillon, Paramveer, Jordan Rodu, Dean Foster, and Lyle Ungar. "Two Step CCA: A new spectral method for estimating vector models of words." *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*. 2012.

Ding, Xiaowen, Bing Liu, and Philip S. Yu. "A holistic lexicon-based approach to opinion mining." In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pp. 231-240. ACM, 2008.

Donahue, Jeff, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. "Long-term recurrent convolutional networks for visual recognition and description." *arXiv preprint arXiv:1411.4389* (2014).

Durrett, Greg, David Leo Wright Hall, and Dan Klein. "Decentralized Entity-Level Modeling for Coreference Resolution." *ACL (1)*. 2013.

Durrett, Greg, and Dan Klein. "Easy Victories and Uphill Battles in Coreference Resolution." *EMNLP*. 2013.

Durrett, Greg, and Dan Klein. "A Joint Model for Entity Analysis: Coreference, Typing, and Linking." *Transactions of the Association for Computational Linguistics* 2 (2014): 477-490.

Eisner, Jason M. "Three new probabilistic models for dependency parsing: An exploration." *Proceedings of the 16th conference on Computational linguistics-Volume 1*. Association for Computational Linguistics, 1996.

Fang, Hao, Saurabh Gupta, Forrest Iandola, Rupesh Srivastava, Li Deng, Piotr Dollár, Jianfeng Gao, Xiaodong He, Margaret Mitchell, John C. Platt, Lawrence Zitnick, and Geoffrey Zweig. "From captions to visual concepts and back." *arXiv preprint arXiv:1411.4952* (2014).

Faruqui, Manaal, and Chris Dyer. "Improving vector space word representations using multilingual correlation." *Proc. of EACL. Association for Computational Linguistics* (2014).

Faruqui, Manaal, Jesse Dodge, Sujay K. Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. "Retrofitting Word Vectors to Semantic Lexicons." *arXiv preprint arXiv:1411.4166* (2014).

References

Ferrucci, David, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A. Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John Prager, Nico Schlaefer, and Chris Welty. "Building Watson: An overview of the DeepQA project." *AI magazine* 31.3 (2010): 59-79.

Firth, John R. "A Synopsis of Linguistic Theory 1930–55 (Special Volume of the Philological Society)." (1957).

Fossum, Victoria, and Kevin Knight. "Combining constituent parsers."*Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*. Association for Computational Linguistics, 2009.

Fountain, Trevor, and Mirella Lapata. "Taxonomy induction using hierarchical random graphs." In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 466-476. Association for Computational Linguistics, 2012.

Galley, Michel, Mark Hopkins, Kevin Knight, and Daniel Marcu. *What's in a translation rule*. *NAACL*, 2004.

Galley, Michel, and Christopher D. Manning. "Accurate non-hierarchical phrase-based translation." In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 966-974. Association for Computational Linguistics, 2010.

Gillick, Dan, and Benoit Favre. "A scalable global model for summarization." In*Proceedings of the Workshop on Integer Linear Programming for Natural Langauge Processing*, pp. 10-18. Association for Computational Linguistics, 2009.

Gimpel, Kevin, and Noah A. Smith. "Rich source-side context for statistical machine translation." In *Proceedings of the Third Workshop on Statistical Machine Translation*, pp. 9-17. Association for Computational Linguistics, 2008.

Goodman, Joshua. "Efficient algorithms for parsing the DOP model." *arXiv preprint cmp-lg/9604008* (1996).

Grefenstette, Edward, and Mehrnoosh Sadrzadeh. "Experimental support for a categorical compositional distributional model of meaning." *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011.

Grefenstette, E., G. Dinu, Y. Zhang, M. Sadrzadeh, and M. Baroni. "Multi-Step Regression Learning for Compositional Distributional Semantics." *Proceedings of the 10th International Conference on Computational Semantics* (IWCS 2013). 2013.

References

Haghighi, Aria, and Dan Klein. "Simple coreference resolution with rich syntactic and semantic features." *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*. Association for Computational Linguistics, 2009.

Hashimoto, Kazuma, Pontus Stenetorp, Makoto Miwa, and Yoshimasa Tsuruoka. "Jointly learning word representations and composition functions using predicate-argument structures." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2014.

Hatzivassiloglou, Vasileios, and Kathleen R. McKeown. "Predicting the semantic orientation of adjectives." In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pp. 174-181. Association for Computational Linguistics, 1997.

Harris, Zellig S. "Distributional structure." *Word* (1954).

Hermann, Karl Moritz, and Phil Blunsom. "The Role of Syntax in Vector Space Models of Compositional Semantics." *ACL (1)*. 2013.

Hong, Kai, and Ani Nenkova. "Improving the estimation of word importance for news multi-document summarization." In *Proceedings of EACL*. 2014.

Huang, Eric H., Richard Socher, Christopher D. Manning, and Andrew Y. Ng. "Improving word representations via global context and multiple word prototypes." *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics, 2012.

Huang, Liang, and David Chiang. "Better k-best parsing." *Proceedings of the Ninth International Workshop on Parsing Technology*. Association for Computational Linguistics, 2005.

Hwa, Rebecca. "An empirical evaluation of probabilistic lexicalized tree insertion grammars." *Proceedings of the 17th international conference on Computational linguistics-Volume 1*. Association for Computational Linguistics, 1998.

Hwa, Rebecca, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. "Bootstrapping parsers via syntactic projection across parallel texts." *Natural language engineering* 11.03 (2005): 311-325.

Iyyer, Mohit, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher, and Hal Daumé III. "A neural network for factoid question answering over paragraphs." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 633-644. 2014.

References

Johnson, Mark. "PCFG models of linguistic tree representations."*Computational Linguistics* 24.4 (1998): 613-632.

Kalchbrenner, Nal, and Phil Blunsom. "Recurrent Continuous Translation Models." In *EMNLP*, pp. 1700-1709. 2013.

Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions." *arXiv preprint arXiv:1412.2306* (2014).

Kasami, T. “An efficient recognition and syntax algorithm for context-free languages”. *Scientific Report AFCRL-65-758, Air Force Cambridge Research Laboratory, Bedford MA*, 1965.

Klein, Dan, and Christopher D. Manning. "Accurate unlexicalized parsing."*Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2003.

Kiros, Ryan, Ruslan Salakhutdinov, and Richard S. Zemel. "Unifying visual-semantic embeddings with multimodal neural language models." *arXiv preprint arXiv:1411.2539* (2014).

Knight, Kevin, and Daniel Marcu. "Summarization beyond sentence extraction: A probabilistic approach to sentence compression." *Artificial Intelligence* 139, no. 1 (2002): 91-107.

Kobdani, Hamidreza, Hinrich Schütze, Michael Schiehlen, and Hans Kamp. "Bootstrapping coreference resolution using word associations." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

Koehn, Philipp. "Pharaoh: a beam search decoder for phrase-based statistical machine translation models." In *Machine translation: From real users to research*, pp. 115-124. Springer Berlin Heidelberg, 2004.

Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. "Moses: Open source toolkit for statistical machine translation." In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pp. 177-180. Association for Computational Linguistics, 2007.

Kong, Chen, Dahua Lin, Mohit Bansal, Raquel Urtasun, and Sanja Fidler. "What are you talking about? text-to-image coreference."*Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014.

Koo, Terry, Xavier Carreras, and Michael Collins. "Simple semi-supervised dependency parsing." (2008).

References

Kozareva, Zornitsa, and Eduard Hovy. "A semi-supervised method to learn and construct taxonomies using the web." In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 1110-1118. Association for Computational Linguistics, 2010.

Kwiatkowski, Tom, Eunsol Choi, Yoav Artzi, and Luke Zettlemoyer. "Scaling semantic parsers with on-the-fly ontology matching." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2013.

Lao, Ni, Amarnag Subramanya, Fernando Pereira, and William W. Cohen. "Reading the web with learned syntactic-semantic inference rules." In*Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 1017-1026. Association for Computational Linguistics, 2012.

Lee, Heeyoung, Angel Chang, Yves Peirsman, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. "Deterministic coreference resolution based on entity-centric, precision-ranked rules." *Computational Linguistics* 39.4 (2013): 885-916.

Levy, Omer, and Yoav Goldberg. "Dependency-Based Word Embeddings."*Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. Vol. 2. 2014.

Liang, Percy, Alexandre Bouchard-Côté, Dan Klein, and Ben Taskar. "An end-to-end discriminative approach to machine translation." In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pp. 761-768. Association for Computational Linguistics, 2006.

Liang, Percy, Ben Taskar, and Dan Klein. "Alignment by agreement." In*Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pp. 104-111. Association for Computational Linguistics, 2006.

Lin, Chin-Yew. "Improving summarization performance by sentence compression: a pilot study." In *Proceedings of the sixth international workshop on Information retrieval with Asian languages-Volume 11*, pp. 1-8. Association for Computational Linguistics, 2003.

Lin, Chin-Yew. "Rouge: A package for automatic evaluation of summaries." In*Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pp. 74-81. 2004.

Lin, Dekang. "An information-theoretic definition of similarity." *ICML*. Vol. 98. 1998.

Lin, Dekang, and Patrick Pantel. "Induction of semantic classes from natural language text." *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2001.

References

Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis Lectures on Human Language Technologies* 5, no. 1 (2012): 1-167.

Madhyastha, Pranava S., Xavier Carreras Pérez, and Ariadna Quattoni. "Learning task-specific bilexical embeddings." *COLING*, 2014.

Markert, Katja, and Malvina Nissim. "Comparing knowledge sources for nominal anaphora resolution." *Computational Linguistics* 31.3 (2005): 367-402.

Marcu, Daniel, Wei Wang, Abdessamad Echihabi, and Kevin Knight. "SPMT: Statistical machine translation with syntactified target language phrases." In*Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 44-52. Association for Computational Linguistics, 2006.

Martins, André FT, and Noah A. Smith. "Summarization with a joint model for sentence extraction and compression." *Proceedings of the Workshop on Integer Linear Programming for Natural Langauge Processing*. Association for Computational Linguistics, 2009.

Matsuzaki, Takuya, Yusuke Miyao, and Jun'ichi Tsujii. "Probabilistic CFG with latent annotations." *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2005.

McClosky, David, Eugene Charniak, and Mark Johnson. "Effective self-training for parsing." *Proceedings of the main conference on human language technology conference of the North American Chapter of the Association of Computational Linguistics*. Association for Computational Linguistics, 2006.

McDonald, Ryan, Fernando Pereira, Kiril Ribarov, and Jan Hajič. "Non-projective dependency parsing using spanning tree algorithms." *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2005.

McDonald, Ryan, Slav Petrov, and Keith Hall. "Multi-source transfer of delexicalized dependency parsers." *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011.

Mi, Haitao, Liang Huang, and Qun Liu. "Forest-Based Translation." In *ACL*, pp. 192-199. 2008.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed representations of words and phrases and their compositionality." *Advances in Neural Information Processing Systems*. 2013.

Mitchell, Jeff, and Mirella Lapata. "Composition in distributional models of semantics." *Cognitive science* 34.8 (2010): 1388-1429.

References

Mnih, Andriy, and Geoffrey Hinton. "Three new graphical models for statistical language modelling." *Proceedings of the 24th international conference on Machine learning*. ACM, 2007.

Nakagawa, Tetsuji, Kentaro Inui, and Sadao Kurohashi. "Dependency tree-based sentiment classification using CRFs with hidden variables." In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 786-794. Association for Computational Linguistics, 2010.

Nakov, Preslav, and Marti Hearst. "Using the web as an implicit training set: application to structural ambiguity resolution." *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2005.

Navigli, Roberto, Paola Velardi, and Stefano Faralli. "A graph-based algorithm for inducing lexical taxonomies from scratch." In *IJCAI*, pp. 1872-1877. 2011.

Ng, Vincent, and Claire Cardie. "Improving machine learning approaches to coreference resolution." *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2002.

Nivre, Joakim. "An efficient algorithm for projective dependency parsing."*Proceedings of the 8th International Workshop on Parsing Technologies (IWPT*. 2003.

Nivre, Joakim, and Mario Scholz. "Deterministic dependency parsing of English text." *Proceedings of the 20th international conference on Computational Linguistics*. Association for Computational Linguistics, 2004.

Och, Franz Josef, and Hermann Ney. "Discriminative training and maximum entropy models for statistical machine translation." In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 295-302. Association for Computational Linguistics, 2002.

Och, Franz Josef. "Minimum error rate training in statistical machine translation." In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pp. 160-167. Association for Computational Linguistics, 2003.

Padó, Sebastian, and Mirella Lapata. "Dependency-based construction of semantic space models." *Computational Linguistics* 33.2 (2007): 161-199.

References

Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pp. 79-86. Association for Computational Linguistics, 2002.

Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis."*Foundations and trends in information retrieval* 2, no. 1-2 (2008): 1-135.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. "BLEU: a method for automatic evaluation of machine translation." In *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 311-318. Association for Computational Linguistics, 2002.

Pauls, Adam, and Dan Klein. "k-best A* parsing." *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*. Association for Computational Linguistics, 2009.

Pereira, Fernando, Naftali Tishby, and Lillian Lee. "Distributional clustering of English words." *Proceedings of the 31st annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 1993.

Petrov, Slav, Leon Barrett, Romain Thibaux, and Dan Klein. "Learning accurate, compact, and interpretable tree annotation." *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2006.

Petrov, Slav, Dipanjan Das, and Ryan McDonald. "A universal part-of-speech tagset." *LREC 2012*.

Pitler, Emily, Shane Bergsma, Dekang Lin, and Kenneth Church. "Using web-scale N-grams to improve base NP parsing performance." *Proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics, 2010.

Ponzetto, Simone Paolo, and Michael Strube. "Exploiting semantic role labeling, WordNet and Wikipedia for coreference resolution." *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*. Association for Computational Linguistics, 2006.

Poon, Hoifung, and Pedro Domingos. "Unsupervised semantic parsing." *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. Association for Computational Linguistics, 2009.

References

Poon, Hoifung, and Pedro Domingos. "Unsupervised ontology induction from text." In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics*, pp. 296-305. Association for Computational Linguistics, 2010.

Poon, Hoifung. "Grounded Unsupervised Semantic Parsing." *ACL (1)*. 2013.

Rahman, Altaf, and Vincent Ng. "Coreference resolution with world knowledge."*Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011.

Ramanathan, Vignesh, Armand Joulin, Percy Liang, and Li Fei-Fei. "Linking people in videos with “their” names using coreference resolution." *Computer Vision–ECCV 2014*. Springer International Publishing, 2014. 95-110.

Ratnaparkhi, Adwait. "A maximum entropy model for part-of-speech tagging."*Proceedings of the conference on empirical methods in natural language processing*. Vol. 1. 1996.

Resnik, Philip. "Probabilistic tree-adjoining grammar as a framework for statistical natural language processing." *Proceedings of the 14th conference on Computational linguistics-Volume 2*. Association for Computational Linguistics, 1992.

Sagae, Kenji, and Alon Lavie. "A best-first probabilistic shift-reduce parser."*Proceedings of the COLING/ACL on Main conference poster sessions*. Association for Computational Linguistics, 2006.

Sahlgren, Magnus. "The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces." (2006).

Sarkar, Anoop. "Applying co-training methods to statistical parsing."*Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*. Association for Computational Linguistics, 2001.

Schabes, Yves, and Richard C. Waters. "Tree insertion grammar: cubic-time, parsable formalism that lexicalizes context-free grammar without changing the trees produced." *Computational Linguistics* 21.4 (1995): 479-513.

Schilder, Frank, and Ravikumar Kondadadi. "FastSum: fast and accurate query-based multi-document summarization." In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pp. 205-208. Association for Computational Linguistics, 2008.

References

Shen, Dou, Jian-Tao Sun, Hua Li, Qiang Yang, and Zheng Chen. "Document Summarization Using Conditional Random Fields." In *IJCAI*, vol. 7, pp. 2862-2867. 2007.

Snow, Rion, Daniel Jurafsky, and Andrew Y. Ng. "Semantic taxonomy induction from heterogenous evidence." In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pp. 801-808. Association for Computational Linguistics, 2006.

Snyder, Benjamin, and Regina Barzilay. "Multiple Aspect Ranking Using the Good Grief Algorithm." In *HLT-NAACL*, pp. 300-307. 2007.

Socher, Richard, Eric H. Huang, Jeffrey Pennin, Christopher D. Manning, and Andrew Y. Ng. "Dynamic pooling and unfolding recursive autoencoders for paraphrase detection." *Advances in Neural Information Processing Systems*. 2011.

Socher, Richard, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, and Christopher D. Manning. "Semi-supervised recursive autoencoders for predicting sentiment distributions." *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011.

Socher, Richard, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. "Semantic compositionality through recursive matrix-vector spaces." *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, 2012.

Socher, Richard, John Bauer, Christopher D. Manning, and Andrew Y. Ng. "Parsing with compositional vector grammars." *In Proceedings of the ACL conference*. 2013.

Socher, Richard, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. "Recursive deep models for semantic compositionality over a sentiment treebank." *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2013.

Socher, Richard, Q. Le, C. Manning, and A. Ng. "Grounded Compositional Semantics for Finding and Describing Images with Sentences." *Transactions of the Association for Computational Linguistics*, 2014.

Soon, Wee Meng, Hwee Tou Ng, and Daniel Chung Yong Lim. "A machine learning approach to coreference resolution of noun phrases." *Computational linguistics* 27.4 (2001): 521-544.

Steedman, Mark. "Surface structure and interpretation." (1996).

References

Steedman, Mark. *The syntactic process*. Vol. 35. Cambridge: MIT press, 2000.

Steedman, Mark, Miles Osborne, Anoop Sarkar, Stephen Clark, Rebecca Hwa, Julia Hockenmaier, Paul Ruhlen, Steven Baker, and Jeremiah Crim. "Bootstrapping statistical parsers from small datasets."*Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics-Volume 1*. Association for Computational Linguistics, 2003.

Stoyanov, Veselin, Claire Cardie, Nathan Gilbert, Ellen Riloff, David Buttler, and David Hysom. "Coreference resolution with reconcile." *Proceedings of the ACL 2010 Conference Short Papers*. Association for Computational Linguistics, 2010.

Sutskever, Ilya, Oriol Vinyals, and Quoc VV Le. "Sequence to sequence learning with neural networks." *Advances in Neural Information Processing Systems*. 2014.

Teufel, Simone, and Marc Moens. "Summarizing scientific articles: experiments with relevance and rhetorical status." *Computational linguistics* 28, no. 4 (2002): 409-445.

Thomas, Matt, Bo Pang, and Lillian Lee. "Get out the vote: Determining support or opposition from Congressional floor-debate transcripts." In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pp. 327-335. Association for Computational Linguistics, 2006.

Tong, Richard M. "An operational system for detecting and tracking opinions in on-line discussion." In *Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification*, vol. 1, p. 6. 2001.

Toutanova, Kristina, Dan Klein, Christopher D. Manning, and Yoram Singer. "Feature-rich part-of-speech tagging with a cyclic dependency network." *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. Association for Computational Linguistics, 2003.

Toutanova, Kristina, and Christopher D. Manning. "Enriching the knowledge sources used in a maximum entropy part-of-speech tagger." *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13*. Association for Computational Linguistics, 2000.

Turian, Joseph, Lev Ratinov, and Yoshua Bengio. "Word representations: a simple and general method for semi-supervised learning." *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010.

References

Turney, Peter D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." In *Proceedings of the 40th annual meeting on association for computational linguistics*, pp. 417-424. Association for Computational Linguistics, 2002.

Turney, Peter D., and Patrick Pantel. "From frequency to meaning: Vector space models of semantics." *Journal of artificial intelligence research* 37.1 (2010): 141-188.

Vaswani, Ashish, Yinggong Zhao, Victoria Fossum, and David Chiang. "Decoding with Large-Scale Neural Language Models Improves Translation." In*EMNLP*, pp. 1387-1392. 2013.

Venugopalan, Subhashini, Huijuan Xu, Jeff Donahue, Marcus Rohrbach, Raymond Mooney, and Kate Saenko. "Translating Videos to Natural Language Using Deep Recurrent Neural Networks." arXiv preprint arXiv:1412.4729 (2014).

Vinyals, Oriol, Alexander Toshev, Samy Bengio, and Dumitru Erhan. "Show and tell: A neural image caption generator." *arXiv preprint arXiv:1411.4555* (2014).

Vinyals, Oriol, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton. "Grammar as a Foreign Language." *arXiv preprint arXiv:1412.7449* (2014).

Vogel, Stephan, Hermann Ney, and Christoph Tillmann. "HMM-based word alignment in statistical translation." In *Proceedings of the 16th conference on Computational linguistics-Volume 2*, pp. 836-841. Association for Computational Linguistics, 1996.

Wang, Lu, and Claire Cardie. "Focused meeting summarization via unsupervised relation extraction." In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 304-313. Association for Computational Linguistics, 2012.

White, Michael, Tanya Korelsky, Claire Cardie, Vincent Ng, David Pierce, and Kiri Wagstaff. "Multidocument summarization via information extraction." In*Proceedings of the first international conference on Human language technology research*, pp. 1-7. Association for Computational Linguistics, 2001.

Widdows, Dominic. "Unsupervised methods for developing taxonomies by combining syntactic and statistical information." In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pp. 197-204. Association for Computational Linguistics, 2003.

Wiebe, Janyce, Theresa Wilson, and Claire Cardie. "Annotating expressions of opinions and emotions in language." *Language resources and evaluation* 39, no. 2-3 (2005): 165-210.

References

Woodsend, Kristian, and Mirella Lapata. "Automatic generation of story highlights." In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 565-574. Association for Computational Linguistics, 2010.

Wong, Yuk Wah, and Raymond J. Mooney. "Learning synchronous grammars for semantic parsing with lambda calculus." *Annual Meeting-Association for computational Linguistics*. Vol. 45. No. 1. 2007.

Wu, Dekai. "Stochastic inversion transduction grammars and bilingual parsing of parallel corpora." *Computational linguistics* 23, no. 3 (1997): 377-403.

Xi, Chenhai, and Rebecca Hwa. "A backoff model for bootstrapping resources for non-English languages." *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2005.

Xu, Chang, Yalong Bai, Jiang Bian, Bin Gao, Gang Wang, Xiaoguang Liu, and Tie-Yan Liu. "RC-NET: A General Framework for Incorporating Knowledge into Word Representations." *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 2014.

Yamada, Hiroyasu, and Yuji Matsumoto. "Statistical dependency analysis with support vector machines." *Proceedings of IWPT*. Vol. 3. 2003.

Yang, Hui, and Jamie Callan. "A metric-based framework for automatic taxonomy induction." In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pp. 271-279. Association for Computational Linguistics, 2009.

Yao, Xuchen, and Benjamin Van Durme. "Information extraction over structured data: Question answering with freebase." In Proceedings of ACL. 2014.

Yarowsky, David, Grace Ngai, and Richard Wicentowski. "Inducing multilingual text analysis tools via robust projection across aligned corpora." *Proceedings of the first international conference on Human language technology research*. Association for Computational Linguistics, 2001.

Yessenalina, Ainur, and Claire Cardie. "Compositional matrix-space models for sentiment analysis." *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011.

Yih, Wen-tau, Joshua Goodman, Lucy Vanderwende, and Hisami Suzuki. "Multi-Document Summarization by Maximizing Informative Content-Words." In*IJCAI*, vol. 2007, p. 20th. 2007.

References

Younger, Daniel H. "Recognition and parsing of context-free languages in time n^3." *Information and control* 10.2 (1967): 189-208.

Yu, Mo and Dredze, Mark. “Improving lexical embeddings with semantic knowledge.” *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 2014.

Zajic, David M., Bonnie Dorr, Jimmy Lin, and Richard Schwartz. "Sentence compression as a component of a multi-document summarization system." In*Proceedings of the 2006 Document Understanding Workshop, New York*. 2006.

Zanzotto, Fabio Massimo, Ioannis Korkontzelos, Francesca Fallucchi, and Suresh Manandhar. "Estimating linear models for compositional distributional semantics." *Proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics, 2010.

Zettlemoyer, Luke S., and Michael Collins. "Online learning of relaxed CCG grammars for parsing to logical form." *In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL-2007*. 2007.

Thank you!

http://ttic.uchicago.edu/~mbansal/