Ph.D. Thesis Oral Defense
Stanford Artificial Intelligence Laboratory (SAIL)

Valentin I. Spitkovsky

Grammar Induction and Parsing with Dependency-and-Boundary Models
**Problem:** Unsupervised Parsing (and Grammar Induction)
Problem: Unsupervised Parsing (and Grammar Induction)

- **Input**: Raw Text

  ... By most measures, the nation's industrial sector is now growing very slowly — if at all. Factory payrolls fell in September. So did the Federal Reserve ...
The Problem

Problem: Unsupervised Parsing (and Grammar Induction)

- **Input**: Raw Text *(Sentences, Tokens and POS-tags)*

  ... By most measures, the nation's industrial sector is now growing very slowly — if at all. Factory payrolls fell in September. So did the Federal Reserve ...

- **Sentences**: Factory payrolls fell in September.
- **Tokens**: Factory, payrolls, fell, in, September.
- **POS-tags**: NN, NNS, VBD, IN, NN.
Problem: Unsupervised Parsing (and Grammar Induction)

- **Input:** Raw Text (Sentences, Tokens and POS-tags)
  
  ... By most measures, the nation's industrial sector is now growing very slowly — if at all. Factory payrolls fell in September. So did the Federal Reserve ...

- **Output:** Syntactic Structures (and a Probabilistic Grammar)
Motivation: Unsupervised (Dependency) Parsing
Motivation: Unsupervised (Dependency) Parsing

- Insert your favorite reason(s) why you’d like to parse anything in the first place...
Motivation: Unsupervised (Dependency) Parsing

- Insert your favorite reason(s) why you’d like to parse anything in the first place...

- ... adjust for any data without reference tree banks: — i.e., understudied languages or genres (e.g., legal).
Motivation: Unsupervised (Dependency) Parsing

- Insert your favorite reason(s) why you’d like to parse anything in the first place...

- ... adjust for any data without reference tree banks: — i.e., understudied languages or genres (e.g., legal).

- Potential applications:
  - machine translation
    — word alignment, phrase extraction, reordering;
  - web search
    — retrieval, query refinement;
  - question answering, speech recognition, etc.
Evaluation: Directed Dependency Accuracy
Evaluation: Directed Dependency Accuracy

Scoring example:

```
Factory payrolls fell in September .
```

Directed Score: \( \frac{3}{5} = 60\% \)  
(left/right-branching baselines: \( \frac{2}{5} = 40\% \))
Evaluation: Directed Dependency Accuracy

Scoring example:

Factory payrolls fell in September.

Directed Score: $\frac{3}{5} = 60\%$ (left/right-branching baselines: $\frac{2}{5} = 40\%$);

Undirected Score: $\frac{4}{5} = 80\%$ (left/right-branching baselines: $\frac{4}{5} = 80\%$).
Prior Art: A Brief History

- 1992 — word classes  
  (Carroll and Charniak)
- 1998 — greedy linkage via mutual information  
  (Yuret)
- 2001 — iterative re-estimation with EM  
  (Paskin)
Prior Art: A Brief History

- 1992 — word classes (Carroll and Charniak)
- 1998 — greedy linkage via mutual information (Yuret)
- 2001 — iterative re-estimation with EM (Paskin)
- 2004 — right-branching baseline — valence (DMV) (Klein and Manning)
Prior Art: A Brief History

- 1992 — word classes (Carroll and Charniak)
- 1998 — greedy linkage via mutual information (Yuret)
- 2001 — iterative re-estimation with EM (Paskin)

- 2004 — right-branching baseline
  — valence (DMV) (Klein and Manning)
Prior Art: A Brief History

- 1992 — word classes (Carroll and Charniak)
- 1998 — greedy linkage via mutual information (Yuret)
- 2001 — iterative re-estimation with EM (Paskin)
- 2004 — right-branching baseline
  — valence (DMV) (Klein and Manning)
- 2004 — annealing techniques (Smith and Eisner)
- 2005 — contrastive estimation (Smith and Eisner)
- 2006 — structural biasing (Smith and Eisner)
- 2007 — common cover link representation (Seginer)
- 2008 — logistic normal priors (Cohen et al.)
- 2009 — lexicalization and smoothing (Headden et al.)
- 2009 — soft parameter tying (Cohen and Smith)
Prior Art: Dependency Model with Valence
Prior Art: Dependency Model with Valence

- a head-outward model, with word classes and valence/adjacency (Klein and Manning, 2004)
Prior Art: Dependency Model with Valence

- a head-outward model, with word classes and valence/adjacency (Klein and Manning, 2004)

\[ h \]

e.g.: verb
Prior Art: Dependency Model with Valence

- a head-outward model, with word classes and valence/adjacency (Klein and Manning, 2004)

  ![Dependency Diagram]

  e.g.: verb
Prior Art: Dependency Model with Valence

- a head-outward model, with word classes and valence/adjacency (Klein and Manning, 2004)

\[ a_1 \rightarrow h \]

e.g.: verb noun
Prior Art: Dependency Model with Valence

- a head-outward model, with word classes and valence/adjacency (Klein and Manning, 2004)

![Diagram of dependency model with valence/adjacency]

e.g.: verb noun phrase

\[ a_1 \]

\[ h \]
Prior Art: Dependency Model with Valence

- a head-outward model, with word classes and valence/adjacency (Klein and Manning, 2004)

![Diagram of dependency model with valence]

e.g.: verb noun phrase etc...
Prior Art: Dependency **Model** with Valence

- a **head-outward model**, with **word classes** and **valence/adjacency** (Klein and Manning, 2004)

![Diagram](image)

- e.g.: verb
- noun
- phrase
- etc...
Prior Art: Dependency Model with Valence

- a head-outward model, with word classes and valence/adjacency (Klein and Manning, 2004)

\[ h \]

\[ a_1 \quad a_2 \]

\[ \text{e.g.: verb \quad noun \quad phrase \quad etc...} \]
Prior Art: Dependency Model with Valence

- a head-outward model, with word classes and valence/adjacency (Klein and Manning, 2004)

Diagram:

- $h$ (head)
- $a_1$, $a_2$
- e.g.: verb, noun, phrase, etc...
- STOP

Diagram shows the dependency model with valence/adjacency.
Prior Art: Dependency Model with Valence

- a head-outward model, with word classes and valence/adjacency

(Klein and Manning, 2004)

\[
\mathbb{P}(t_h) = \prod_{\text{dir} \in \{L,R\}} \left[ \mathbb{P}_{\text{STOP}}(c_h, \text{dir}, 1_{n=0}) \prod_{i=1}^{n} \mathbb{P}(t_{a_i}) \mathbb{P}_{\text{ATTACH}}(c_h, \text{dir}, c_{a_i}) 
\right]
\]

\[
(1 - \mathbb{P}_{\text{STOP}}(c_h, \text{dir}, 1_{i=1}))
\]

\[
\text{n} = |\text{args}(h, \text{dir})|
\]
Prior Art: Unsupervised Learning Engine
**Prior Art: Unsupervised Learning Engine**

- parameter fitting via expectation-maximization (EM)
Prior Art: Unsupervised Learning Engine

- Parameter fitting via expectation-maximization (EM)
Prior Art: Unsupervised Learning Engine

- parameter fitting via expectation-maximization (EM)
  - by means of inside-outside re-estimation (Baker, 1979)

\[ w_1 w_{p-1} w_p \cdots w_q w_{q+1} \cdots w_m \]

(Manning and Schütze, 1999)
Prior Art: Unsupervised Learning Engine

- parameter fitting via expectation-maximization (EM)
Prior Art: The Standard Corpus — WSJ

![Graph showing sentences and tokens distribution over WSJk]
Prior Art: The Standard Corpus — WSJ10

Sentences (1,000s)

Tokens (1,000s)
Outline

- Introduction to the Grammar Induction Problem
  - The Inputs and Outputs (text $\rightarrow$ dependency parses)
  - A Probabilistic Model (DMV)
  - Parameter Fitting (EM)
Outline

- Introduction to the Grammar Induction Problem
  - The Inputs and Outputs (text $\rightarrow$ dependency parses)
  - A Probabilistic Model (DMV)
  - Parameter Fitting (EM)

- Part I: Non-Convex Optimization
Outline

- Introduction to the Grammar Induction Problem
  - The Inputs and Outputs (text $\rightarrow$ dependency parses)
  - A Probabilistic Model (DMV)
  - Parameter Fitting (EM)

- Part I: Non-Convex Optimization

- Part II: Parsing Constraints
Outline

- Introduction to the Grammar Induction Problem
  - The Inputs and Outputs (text $\rightarrow$ dependency parses)
  - A Probabilistic Model (DMV)
  - Parameter Fitting (EM)

- Part I: Non-Convex Optimization

- Part II: Parsing Constraints

- Part III: Generative Models
Outline

- Introduction to the Grammar Induction Problem
  - The Inputs and Outputs (text → dependency parses)
  - A Probabilistic Model (DMV)
  - Parameter Fitting (EM)

- Part I: Non-Convex Optimization

- Part II: Parsing Constraints

- Part III: Generative Models

- State-of-the-Art Integrated Solution
Outline

- Introduction to the Grammar Induction Problem
  - The Inputs and Outputs \( \text{(text} \rightarrow \text{dependency parses)} \)
  - A Probabilistic Model \( \text{(DMV)} \)
  - Parameter Fitting \( \text{(EM)} \)

- Part I: Non-Convex Optimization

- Part II: Parsing Constraints

- Part III: Generative Models

- State-of-the-Art Integrated Solution \( \text{(in submission)} \)
Optimization

Viterbi Training

Baby Steps

Lateen EM
Issue I: Why so little data?

Directed Dependency Accuracy (%) on WSJ40
**Issue I: Why so little data?**

Directed Dependency Accuracy (%) on WSJ40

- **Uninformed**
Issue I: Why so little data?

Directed Dependency Accuracy (%) on WSJ40

Uninformed
**Issue I: Why so little data?**

Directed Dependency Accuracy (%) on WSJ40

- **Oracle**
- **Uninformed**
**Issue I: Why so little data?**

![Directed Dependency Accuracy (%) on WSJ40](image)

- **Oracle**
- **Uninformed**

5 10 15 20 25 30 35 40
Issue II: Non-convex objectives...

- maximizing the probability of data (sentence strings):

\[
\hat{\theta}_{\text{UNS}} = \arg \max_{\theta} \sum_{s} \log \sum_{t \in T(s)} \mathbb{P}_\theta(t) \mathbb{P}_\theta(s)
\]
**Issue II: Non-convex objectives...**

- maximizing the probability of data (sentence strings):

\[
\hat{\theta}_{\text{UNS}} = \arg \max_{\theta} \sum_s \log \sum_{t \in T(s)} \mathbb{P}_\theta(t) \mathbb{P}_\theta(s)
\]

- supervised objective would be convex:

\[
\hat{\theta}_{\text{SUP}} = \arg \max_{\theta} \sum_s \log \mathbb{P}_\theta(t^*(s))
\]
**Issue II: Non-convex objectives...**

- maximizing the probability of data (sentence strings):
  
  \[
  \hat{\theta}_{\text{UNS}} = \arg \max_{\theta} \sum_s \log \sum_{t \in T(s)} P_\theta(t) P_\theta(s)
  \]

- supervised objective would be convex:
  
  \[
  \hat{\theta}_{\text{SUP}} = \arg \max_{\theta} \sum_s \log P_\theta(t^*(s))
  \]

- could optimize likeliest parse trees (also non-convex):
  
  \[
  \hat{\theta}_{\text{VIT}} = \arg \max_{\theta} \sum_s \max_{t \in T(s)} \log P_\theta(t)
  \]
Classic EM:

Directed Dependency Accuracy (%) on WSJ40

Oracle

Uninformed
Viterbi EM:

Directed Dependency Accuracy (%) on WSJ40

Oracle

Uninformed
Hard vs. Soft EM:

- **Viterbi EM**: zoom in on likeliest tree
Hard vs. Soft EM:

- Viterbi EM: zoom in on likeliest tree
  - does not degrade with input length
  - better retains supervised solutions
  - actually learns from scratch
    — see also Cohen and Smith (2010)
Hard vs. Soft EM:

- **Viterbi EM**: zoom in on likeliest tree
  - does not degrade with input length
  - better retains supervised solutions
  - actually learns from scratch
    — see also Cohen and Smith (2010)

- **Classic EM**: “focus across the board”
Hard vs. Soft EM:

- **Viterbi EM**: zoom in on likeliest tree
  - does not degrade with input length
  - better retains supervised solutions
  - actually learns from scratch
    — see also Cohen and Smith (2010)

- **Classic EM**: “focus across the board”
  - hard to see the trees for the forest...
  - ... but known to work with shorter inputs!
Idea I: Baby Steps ... as Non-convex Optimization
Idea I: Baby Steps ... as Non-convex Optimization

- start with an easy (convex) case
Idea I: Baby Steps ... as Non-convex Optimization

- start with an easy (convex) case
- slowly extend it to the fully complex target task
Idea I: Baby Steps ... as Non-convex Optimization

- start with an easy (convex) case
- slowly extend it to the fully complex target task
- take tiny (cautious) steps in the problem space
Idea 1: Baby Steps ... as Non-convex Optimization

- start with an easy (convex) case
- slowly extend it to the fully complex target task
- take tiny (cautious) steps in the problem space
- ... try not to stray far from relevant neighborhoods in the solution space
Idea I: Baby Steps ... as Non-convex Optimization

- start with an easy (convex) case
- slowly extend it to the fully complex target task
- take tiny (cautious) steps in the problem space
- ... try not to stray far from relevant neighborhoods in the solution space

**base case**: sentences of length one (trivial — no init)
Idea I: Baby Steps ... as Non-convex Optimization

- start with an easy (convex) case
- slowly extend it to the fully complex target task
- take tiny (cautious) steps in the problem space
- ... try not to stray far from relevant neighborhoods in the solution space

- **base case**: sentences of length one (trivial — no init)
- **incremental step**: smooth $WSJ^k$; re-init $WSJ(k + 1)$
Idea I: Baby Steps ... as Graduated Learning
Idea I: Baby Steps ... as Graduated Learning

- WSJ1 — Atone (verbs!)
Idea I: Baby Steps ... as Graduated Learning

- **WSJ1** — Atone (verbs!)

- **WSJ2** — It is. (pronouns!)
  Darkness fell.
Idea I: Baby Steps ... as Graduated Learning

- **WSJ1** — Atone (verbs!)

- **WSJ2** — It is. (pronouns!)
  Darkness fell.

- **WSJ3** — But many have. (determiners!)
  They didn’t.

Become a Lobbyist
Idea I: Baby Steps ... and Related Notions
Idea I: Baby Steps ... and Related Notions

- shaping

(Skinner, 1938)
Idea I: Baby Steps ... and Related Notions

- shaping  Psychology / Cognitive Science  (Skinner, 1938)
- less is more  (Kail, 1984; Newport, 1988; 1990)
- starting small  (Elman, 1993)
  - scaffold on model complexity  [restrict memory]
  - scaffold on data complexity  [restrict input]
Idea I: Baby Steps ... and Related Notions

- shaping Psychology / Cognitive Science (Skinner, 1938)
- less is more (Kail, 1984; Newport, 1988; 1990)
- starting small (Elman, 1993)
  - scaffold on model complexity [restrict memory]
  - scaffold on data complexity [restrict input]
- stepping stones (Brown et al., 1993)
- coarse-to-fine NLP / AI (Charniak and Johnson, 2005)
- curriculum learning (Bengio et al., 2009)
Idea I: Baby Steps ... and Related Notions

- shaping
  Psychology / Cognitive Science (Skinner, 1938)
- less is more
  (Kail, 1984; Newport, 1988; 1990)
- starting small
  (Elman, 1993)

  ▶ scaffold on model complexity [restrict memory]
  ▶ scaffold on data complexity [restrict input]

- stepping stones
  (Brown et al., 1993)
- coarse-to-fine NLP / AI (Charniak and Johnson, 2005)
- curriculum learning
  (Bengio et al., 2009)

- continuation methods
  (Allgower and Georg, 1990)
- deterministic annealing
  Math / OR (Rose, 1998)
Idea I: Baby Steps ... and Related Notions

- shaping Psychology / Cognitive Science (Skinner, 1938)
- less is more (Kail, 1984; Newport, 1988; 1990)
- starting small (Elman, 1993)
  - scaffold on model complexity [restrict memory]
  - scaffold on data complexity [restrict input]
- stepping stones (Brown et al., 1993)
- coarse-to-fine NLP / AI (Charniak and Johnson, 2005)
- curriculum learning (Bengio et al., 2009)
- continuation methods (Allgower and Georg, 1990)
- deterministic annealing Math / OR (Rose, 1998)

successive approximations!
Aside I: Speaking of successive approximations...
Idea I: Baby Steps ... Observations & Results!

Directed Dependency Accuracy (%) on WSJ40

Oracle

Uninformed
Idea I: Baby Steps ... Observations & Results!

Directed Dependency Accuracy (%) on WSJ40

- Oracle
- Baby Steps
- Uninformed
Idea I: Baby Steps ... Observations & Results!

Directed Dependency Accuracy (%) on WSJ40

- Oracle
- Baby Steps
- Uninformed
Idea I: Baby Steps ... Observations & Results!

Directed Dependency Accuracy (%) on WSJ40

- Oracle
- Less is More
- Baby Steps
- Uninformed
- K&M*

5 10 15 20 25 30 35 40 45 50 55 60 65 70

Idea II: Less is More & Leapfrog ... a Hack!

Directed Dependency Accuracy (%) on WSJ$k$

- Oracle
- K&M
- Baby Steps
- Uninformed
Idea II: Less is More & Leapfrog ... a Hack!

Directed Dependency Accuracy (%) on WSJ\textsubscript{k}

- Oracle
- Baby Steps
- K&M\textsuperscript{*}
- Uninformed
Idea II: Less is More & Leapfrog ... a Hack!

Directed Dependency Accuracy (%) on WSJk

Oracle
Baby Steps
K&M*
Uninformed
Idea II: Less is More & Leapfrog ... a Hack!

Directed Dependency Accuracy (%) on WSJk

- Oracle
- Baby Steps
- K&M* (K&M Asterisk)
- Uninformed
**Idea II: Less is More & Leapfrog ... a Hack!**

Directed Dependency Accuracy (%) on WSJk

- **Oracle**
- **Baby Steps**
- **K&M**
- **Uninformed**
Idea II: Less is More & Leapfrog ... a Hack!

Directed Dependency Accuracy (%) on WSJ$k$

- Oracle
- Leapfrog
- Baby Steps
- K&M*
- Uninformed
Aside II: Also, to be fair...

Undirected Dependency Accuracy (%) on WSJk
Lateen EM: Intuition

- Use both objectives (a primary and a secondary).
Lateen EM: Intuition

- Use **both** objectives (a primary and a secondary).

As a captain can’t count on favorable winds, so an unsupervised learner can’t rely on co-operative gradients.
Lateen EM: Intuition

- Use both objectives (a primary and a secondary).

As a captain can’t count on favorable winds, so an unsupervised learner can’t rely on co-operative gradients.

Lateen strategies **de-emphasize fixed points**, e.g., by tacking around local attractors, in a zig-zag fashion.
Lateen EM: Simple Algorithm
Lateen EM: Simple Algorithm

- Alternate ordinary soft and hard EM algorithms: switching when stuck helps escape local optima.
**Lateen EM: Simple Algorithm**

- **Alternate** ordinary soft and hard EM algorithms: switching when **stuck** helps **escape** local optima.

E.g., Italian grammar induction improves from 41.8% to 56.2% after three lateen alternations:

**Pumping action:**
- hard EM pushes down the top curve (primary objective);
- soft EM pushes down the bottom curve (the secondary objective), often at the expense of the primary.
Lateen EM: Early-Stopping
Lateen EM: Early-Stopping

- Use one objective to validate moves proposed by the other: stop if the secondary objective gets worse.
Lateen EM: Early-Stopping

- Use one objective to validate moves proposed by the other: stop if the secondary objective gets worse.

- 30% faster, on average, than either standard EM.
Optimization: Summary
Optimization: Summary

- Take *guesswork out* of tuning EM:
Optimization: Summary

- Take **guesswork out** of tuning EM:
  - Lateen EM ties **termination** to a sign change
**Optimization: Summary**

- Take **guesswork out** of tuning EM:
  - Lateen EM ties **termination** to a sign change;
  - Viterbi EM and Baby Steps don’t require **initializers**.
Optimization: Summary

- Take **guesswork out** of tuning EM:
  - Lateen EM ties termination to a sign change;
  - Viterbi EM and Baby Steps don’t require initializers.

- Exploit **multiple views and iterate** (Blum and Mitchell, 1998):
Optimization: Summary

- Take **guesswork out** of tuning EM:
  - Lateen EM ties termination to a sign change;
  - Viterbi EM and Baby Steps don’t require initializers.

- **Exploit multiple views and iterate** (Blum and Mitchell, 1998):
  - Lateen EM optimizes for Viterbi parse trees and forests;
Optimization: Summary

- Take **guesswork out** of tuning EM:
  - Lateen EM ties termination to a sign change;
  - Viterbi EM and Baby Steps don’t require initializers.

- Exploit **multiple views and iterate** (Blum and Mitchell, 1998):
  - Lateen EM optimizes for Viterbi parse **trees and forests**;
  - Baby Steps focuses on **simpler sentences** in the data.
Optimization: Summary

- Take **guesswork out** of tuning EM:
  - Lateen EM ties termination to a sign change;
  - Viterbi EM and Baby Steps don’t require initializers.

- Exploit **multiple views and iterate** (Blum and Mitchell, 1998):
  - Lateen EM optimizes for Viterbi parse trees *and* forests;
  - Baby Steps focuses on simpler sentences in the data.

- Capitalize on EM’s **advantages**:
Optimization: Summary

- Take **guesswork out** of tuning EM:
  - Lateen EM ties termination to a sign change;
  - Viterbi EM and Baby Steps don’t require initializers.

- Exploit **multiple views and iterate** (Blum and Mitchell, 1998):
  - Lateen EM optimizes for Viterbi parse trees *and* forests;
  - Baby Steps focuses on simpler sentences in the data.

- Capitalize on EM’s **advantages**:
  - guarantee to **not harm** a primary likelihood
Optimization: Summary

- Take **guesswork out** of tuning EM:
  - Lateen EM ties termination to a sign change;
  - Viterbi EM and Baby Steps don’t require initializers.

- Exploit **multiple views and iterate** (Blum and Mitchell, 1998):
  - Lateen EM optimizes for Viterbi parse trees *and* forests;
  - Baby Steps focuses on simpler sentences in the data.

- Capitalize on EM’s **advantages**:
  - guarantee to **not harm** a primary likelihood;
  - begin with **large steps** in a parameter space.
Optimization: Summary

- Take **guesswork out** of tuning EM:
  - Lateen EM ties termination to a sign change;
  - Viterbi EM and Baby Steps don’t require initializers.

- Exploit **multiple views and iterate** (Blum and Mitchell, 1998):
  - Lateen EM optimizes for Viterbi parse trees *and* forests;
  - Baby Steps focuses on simpler sentences in the data.

- Capitalize on EM’s **advantages**:
  - guarantee to not harm a primary likelihood;
  - begin with large steps in a parameter space.

- Ameliorate EM’s **disadvantages**:  

**Optimization: Summary**

- **Take guesswork out of tuning EM:**
  - Lateen EM ties termination to a sign change;
  - Viterbi EM and Baby Steps don’t require initializers.

- **Exploit multiple views and iterate** (Blum and Mitchell, 1998):
  - Lateen EM optimizes for Viterbi parse trees *and* forests;
  - Baby Steps focuses on simpler sentences in the data.

- **Capitalize on EM’s advantages:**
  - guarantee to not harm a primary likelihood;
  - begin with large steps in a parameter space.

- **Ameliorate EM’s disadvantages:**
  - *avoid* taking disproportionately many (and ever-smaller) steps to approach a likelihood’s fixed point.
Optimization: Summary

- Take **guesswork out** of tuning EM:
  - Lateen EM ties termination to a sign change;
  - Viterbi EM and Baby Steps don’t require initializers.

- Exploit **multiple views and iterate** (Blum and Mitchell, 1998):
  - Lateen EM optimizes for Viterbi parse trees *and* forests;
  - Baby Steps focuses on simpler sentences in the data.

- Capitalize on EM’s **advantages**:
  - guarantee to not harm a primary likelihood;
  - begin with large steps in a parameter space.

- Ameliorate EM’s **disadvantages**:
  - **avoid** taking disproportionately many (and ever-smaller) steps to approach a likelihood’s fixed point;
  - **escape** by changing objectives and mixing solutions!
Constraints

Web Markup

Punctuation

Capitalization
Constraints: Supervised and Unsupervised
Constraints: Supervised and Unsupervised

- compact summaries of high-level insights into a domain
Constraints: **Supervised** and **Unsupervised**

- compact summaries of high-level insights into a domain — can significantly reduce the search space
Constraints: **Supervised** and **Unsupervised**

- **compact summaries** of high-level insights into a domain
  — can significantly **reduce** the **search space**
  — often easier to enforce than to model
Constraints: Supervised and Unsupervised

- **compact summaries** of high-level insights into a domain — can significantly **reduce** the search space
  — often easier to enforce than to model

- relevant to **unsupervised learning**  (less rope to hang self)
Constraints: Supervised and Unsupervised

- compact summaries of high-level insights into a domain — can significantly reduce the search space — often easier to enforce than to model

- relevant to unsupervised learning (less rope to hang self) — in general, steer at the “right” regularities in data
Constraints: Supervised and Unsupervised

- compact summaries of high-level insights into a domain
  — can significantly reduce the search space
  — often easier to enforce than to model

- relevant to unsupervised learning (less rope to hang self)
  — in general, steer at the “right” regularities in data
  — linguistic structure underdetermined by raw text
Constraints: Supervised and Unsupervised

- compact summaries of high-level insights into a domain
  — can significantly reduce the search space
  — often easier to enforce than to model

- relevant to unsupervised learning (less rope to hang self)
  — in general, steer at the “right” regularities in data
  — linguistic structure underdetermined by raw text

- partial bracketings (Pereira and Schabes, 1992)
Constraints: Supervised and Unsupervised

- compact summaries of high-level insights into a domain — can significantly reduce the search space — often easier to enforce than to model

- relevant to unsupervised learning (less rope to hang self) — in general, steer at the “right” regularities in data — linguistic structure underdetermined by raw text

- partial bracketings (Pereira and Schabes, 1992)

- synchronous grammars (Alshawi and Douglas, 2000)

- linear-time parsing, skewness of trees, etc. (Seginer, 2007)

- sparse posterior regularization (Ganchev et al., 2009)
Syntax of Markup: Common Constituents

..., but $[S [NP the <a>Toronto Star] [VP reports [NP this] [PP in the softest possible way]]]$
Syntax of Markup: Common Constituents

..., but \[ S [NP \text{the <a>Toronto Star</a>} [VP \text{reports [NP this]} [PP \text{in the softest possible way}] </a>]] , [S stating ...]]

\[ S \rightarrow NP \_VP \]
... but \[ S \ [NP \ \text{the} \ [a] \textit{Toronto Star} \ [VP \ \text{reports} \ [NP \ \text{this}] \ [PP \ \text{in the softest possible way}] </a> \], \[ S \ \text{stating} \ldots \]]

\[ S \rightarrow NP \_VP \_ DT \ NNP \_ NNP \_ VBZ \_ NP \_ PP \_ S \]
### Syntax of Markup: Constituent Productions

<table>
<thead>
<tr>
<th>Production</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (\rightarrow) NNP NNP</td>
<td>9.6%</td>
</tr>
<tr>
<td>NP (\rightarrow) NNP</td>
<td>4.6%</td>
</tr>
<tr>
<td>NP (\rightarrow) NP PP</td>
<td>3.4%</td>
</tr>
<tr>
<td>NP (\rightarrow) NNP NNP NNP</td>
<td>2.4%</td>
</tr>
<tr>
<td>NP (\rightarrow) DT NNP NNP</td>
<td>2.1%</td>
</tr>
<tr>
<td>NP (\rightarrow) NN</td>
<td>1.8%</td>
</tr>
<tr>
<td>NP (\rightarrow) DT NNP NNP NNP</td>
<td>1.7%</td>
</tr>
<tr>
<td>NP (\rightarrow) DT NN</td>
<td>1.7%</td>
</tr>
<tr>
<td>NP (\rightarrow) DT NNP NNP</td>
<td>1.6%</td>
</tr>
<tr>
<td>S (\rightarrow) NP VP</td>
<td>1.4%</td>
</tr>
<tr>
<td>NP (\rightarrow) DT NNP NNP NNP</td>
<td>1.2%</td>
</tr>
<tr>
<td>NP (\rightarrow) DT JJ NN</td>
<td>1.1%</td>
</tr>
<tr>
<td>NP (\rightarrow) NNS</td>
<td>1.0%</td>
</tr>
<tr>
<td>NP (\rightarrow) JJ NN</td>
<td>0.8%</td>
</tr>
<tr>
<td>NP (\rightarrow) NP NP</td>
<td>0.8%</td>
</tr>
</tbody>
</table>

Total: 35.3%
Syntax of Markup: Constituent Productions

<table>
<thead>
<tr>
<th>Production</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP → NNP NNP</td>
<td>9.6</td>
</tr>
<tr>
<td>NP → NNP</td>
<td>4.6</td>
</tr>
<tr>
<td>NP → NP PP</td>
<td>3.4</td>
</tr>
<tr>
<td>NP → NNP NNP NNP</td>
<td>2.4</td>
</tr>
<tr>
<td>NP → DT NNP NNP</td>
<td>2.1</td>
</tr>
<tr>
<td>NP → NN</td>
<td>1.8</td>
</tr>
<tr>
<td>NP → DT NNP NNP NNP</td>
<td>1.7</td>
</tr>
<tr>
<td>NP → DT NN</td>
<td>1.7</td>
</tr>
<tr>
<td>NP → DT NNP NNP</td>
<td>1.6</td>
</tr>
<tr>
<td>S → NP VP</td>
<td>1.4</td>
</tr>
<tr>
<td>NP → DT NNP NNP NNP</td>
<td>1.2</td>
</tr>
<tr>
<td>NP → DT JJ NN</td>
<td>1.1</td>
</tr>
<tr>
<td>NP → NNS</td>
<td>1.0</td>
</tr>
<tr>
<td>NP → JJ NN</td>
<td>0.8</td>
</tr>
<tr>
<td>NP → NP NP</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>35.3</td>
</tr>
</tbody>
</table>
# Syntax of Markup: Constituent Productions

<table>
<thead>
<tr>
<th>Rule</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$NP \rightarrow NNP _ NNP$</td>
<td>9.6</td>
</tr>
<tr>
<td>$NP \rightarrow NNP$</td>
<td>4.6</td>
</tr>
<tr>
<td>$NP \rightarrow NP _ PP$</td>
<td>3.4</td>
</tr>
<tr>
<td>$NP \rightarrow NNP _ NNP _ NNP$</td>
<td>2.4</td>
</tr>
<tr>
<td>$NP \rightarrow DT _ NNP _ NNP$</td>
<td>2.1</td>
</tr>
<tr>
<td>$NP \rightarrow NN$</td>
<td>1.8</td>
</tr>
<tr>
<td>$NP \rightarrow DT _ NNP _ NNP _ NNP$</td>
<td>1.7</td>
</tr>
<tr>
<td>$NP \rightarrow DT _ NN$</td>
<td>1.7</td>
</tr>
<tr>
<td>$NP \rightarrow DT _ NNP _ NNP$</td>
<td>1.6</td>
</tr>
<tr>
<td>$S \rightarrow NP _ VP$</td>
<td>1.4</td>
</tr>
<tr>
<td>$NP \rightarrow DT _ NNP _ NNP _ NNP$</td>
<td>1.2</td>
</tr>
<tr>
<td>$NP \rightarrow DT _ JJ _ NN$</td>
<td>1.1</td>
</tr>
<tr>
<td>$NP \rightarrow NNS$</td>
<td>1.0</td>
</tr>
<tr>
<td>$NP \rightarrow JJ _ NN$</td>
<td>0.8</td>
</tr>
<tr>
<td>$NP \rightarrow NP _ NP$</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>35.3</strong></td>
</tr>
</tbody>
</table>
Syntax of Markup: Common Dependencies

..., but [S [NP the <a>Toronto Star][VP reports [NP this]]
[PP in the softest possible way]<a>][S stating ...]]

DT NNP NNP VBZ DT IN DT JJS JJ NN
**Syntax of Markup: Common Dependencies**

..., but [S [NP the <a>Toronto Star</a>] [VP reports [NP this] [PP in the softest possible way]] <a>,</a> [S stating ...]]

![Diagram of dependencies]

```
DT    NNP    NNP    VBZ    DT    IN    DT    JJS    JJ    NN
```

```
DT    NNP    VBZ
```
Syntax of Markup: Common Dependencies

..., but $[S [NP \text{the } \text{Toronto Star}] [VP \text{reports } [NP \text{this}] [PP \text{in the softest possible way}] </a> [S \text{stating ...}]])$
**Syntax of Markup: Head-Outward Spawns**

<table>
<thead>
<tr>
<th>Part of Speech</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNP</td>
<td>24.4</td>
</tr>
<tr>
<td>NN</td>
<td>8.1</td>
</tr>
<tr>
<td>DT</td>
<td>6.1</td>
</tr>
<tr>
<td>NNP</td>
<td>5.9</td>
</tr>
<tr>
<td>NNS</td>
<td>4.5</td>
</tr>
<tr>
<td>NNPS</td>
<td>1.4</td>
</tr>
<tr>
<td>VBG</td>
<td>1.3</td>
</tr>
<tr>
<td>NNP</td>
<td>1.2</td>
</tr>
<tr>
<td>NNP</td>
<td>1.2</td>
</tr>
<tr>
<td>NN</td>
<td>1.0</td>
</tr>
<tr>
<td>IN</td>
<td>1.0</td>
</tr>
<tr>
<td>VBN</td>
<td>1.0</td>
</tr>
<tr>
<td>DT</td>
<td>0.9</td>
</tr>
<tr>
<td>JJ</td>
<td>0.9</td>
</tr>
<tr>
<td>POS</td>
<td>0.9</td>
</tr>
<tr>
<td>JJ</td>
<td>0.8</td>
</tr>
</tbody>
</table>

Total: 59.4%
### Syntax of Markup: Head-Outward Spawns

<table>
<thead>
<tr>
<th>Tag</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNP</td>
<td>24.4</td>
</tr>
<tr>
<td>NN</td>
<td>8.1</td>
</tr>
<tr>
<td>DT</td>
<td>6.1</td>
</tr>
<tr>
<td>NN</td>
<td>5.9</td>
</tr>
<tr>
<td>NNS</td>
<td>4.5</td>
</tr>
<tr>
<td>NNPS</td>
<td>1.4</td>
</tr>
<tr>
<td>VBG</td>
<td>1.3</td>
</tr>
<tr>
<td>NNP</td>
<td>1.2</td>
</tr>
<tr>
<td>VBD</td>
<td>1.0</td>
</tr>
<tr>
<td>IN</td>
<td>1.0</td>
</tr>
<tr>
<td>VBN</td>
<td>1.0</td>
</tr>
<tr>
<td>DT</td>
<td>0.9</td>
</tr>
<tr>
<td>JJ</td>
<td>0.9</td>
</tr>
<tr>
<td>VBZ</td>
<td>0.9</td>
</tr>
<tr>
<td>POS</td>
<td>0.9</td>
</tr>
<tr>
<td>JJ</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>59.4</strong></td>
</tr>
</tbody>
</table>
Syntax of Markup: Head-Outward Spawns

<table>
<thead>
<tr>
<th></th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNP</td>
<td>24.4</td>
</tr>
<tr>
<td>NN</td>
<td>8.1</td>
</tr>
<tr>
<td>DT NNP</td>
<td>6.1</td>
</tr>
<tr>
<td>DT NN</td>
<td>5.9</td>
</tr>
<tr>
<td>NNS</td>
<td>4.5</td>
</tr>
<tr>
<td>NNPS</td>
<td>1.4</td>
</tr>
<tr>
<td>VBG</td>
<td>1.3</td>
</tr>
<tr>
<td>NNP NNP NN</td>
<td>1.2</td>
</tr>
<tr>
<td>VBD</td>
<td>1.0</td>
</tr>
<tr>
<td>IN</td>
<td>1.0</td>
</tr>
<tr>
<td>VBN</td>
<td>1.0</td>
</tr>
<tr>
<td>DT JJ</td>
<td>0.9</td>
</tr>
<tr>
<td>VBZ</td>
<td>0.9</td>
</tr>
<tr>
<td>POS NNP</td>
<td>0.9</td>
</tr>
<tr>
<td>JJ</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>59.4</td>
</tr>
</tbody>
</table>
Formating:

- strong connection between markup and syntax
Formatting:

- strong connection between markup and syntax
- yields a suite of **accurate** parsing **constraints**
Formatting:

- strong connection between markup and syntax
- yields a suite of accurate parsing constraints
- improves grammar induction over web data
Formatting:

- strong connection between markup and syntax
- yields a suite of accurate parsing constraints
- improves grammar induction over web data, but...
Formatting:

- strong connection between markup and syntax
- yields a suite of accurate parsing constraints
- improves grammar induction over web data, but...
  - works better with more natural language-resources!
Formatting:

- strong connection between markup and syntax
- yields a suite of accurate parsing constraints
- improves grammar induction over web data, but...
  - works better with more natural language-resources!

- missing structural cues:
Formatting:

- strong connection between markup and syntax
- yields a suite of accurate parsing constraints
- improves grammar induction over web data, but...
  - works better with more natural language-resources!

- missing structural cues:
  - e.g., punctuation (will make heavy use)
Formatting:

- strong connection between markup and syntax
- yields a suite of accurate parsing constraints
- improves grammar induction over web data, but...
  - works better with more natural language-resources!

- missing structural cues:
  - e.g., **punctuation** (will make heavy use)
  - and capitalization (will not discuss today)
Formatting:

- strong connection between markup and syntax
- yields a suite of accurate parsing constraints
- improves grammar induction over web data, but...
  - works better with more natural language-resources!

- missing structural cues:
  - e.g., punctuation (will make heavy use)
  - and capitalization (will not discuss today)

- raw word streams often difficult even for humans
  - e.g., transcribed utterances (Kim and Woodland, 2002)
ALTHOUGH IT PROBABLY HAS REDUCED THE LEVEL OF EXPENDITURES FOR SOME PURCHASERS UTILIZATION MANAGEMENT LIKE MOST OTHER COST CONTAINMENT STRATEGIES DOESN’T APPEAR TO HAVE ALTERED THE LONG-TERM RATE OF INCREASE IN HEALTH-CARE COSTS THE INSTITUTE OF MEDICINE AN AFFILIATE OF THE NATIONAL ACADEMY OF SCIENCES CONCLUDED AFTER A TWO-YEAR STUDY.
Example:

---

Example:

Although it probably has reduced the level of expenditures for some purchasers,
Example: 

Although it probably has reduced the level of expenditures for some purchasers, utilization management —
Example:

\[\text{Although it probably has reduced the level of expenditures for some purchasers, } \text{utilization management} \quad \text{— like most other cost containment strategies} \quad \]
Example: 

\[ \text{Although it probably has reduced the level of expenditures for some purchasers, utilization management — like most other cost containment strategies} \text{ — doesn’t appear to have altered the long-term rate of increase in health-care costs,} \]
Example:

Although it probably has reduced the level of expenditures for some purchasers, utilization management — like most other cost containment strategies — doesn't appear to have altered the long-term rate of increase in health-care costs, the Institute of Medicine,
Example: 

Although it probably has reduced the level of expenditures for some purchasers, utilization management — like most other cost containment strategies — doesn’t appear to have altered the long-term rate of increase in health-care costs, the Institute of Medicine, an affiliate of the National Academy of Sciences,
Example:

Although it probably has reduced the level of expenditures for some purchasers, utilization management — like most other cost containment strategies — doesn’t appear to have altered the long-term rate of increase in health-care costs, the Institute of Medicine, an affiliate of the National Academy of Sciences, concluded after a two-year study.
**Intuition:**

- **strong constraint**
Intuition:

- strong constraint: \((\text{head} \leftarrow \text{head})\) in training
Intuition:

- strong constraint: \((\text{head} \leftarrow \text{head})\) in training

\[
\begin{align*}
\text{word head,} & \quad \text{head word word,} \\
\text{head word word word word word word word word word word,}
\end{align*}
\]
**Intuition:**

- **strong constraint:** \((\text{head} \leftarrow \text{head})\) in training

```
word head, head word word, head word word word word word word word word.
```
**Intuition:**

- **strong constraint:** \((\text{head} \leftarrow \text{head})\) in training
Intuition:

- **strong constraint:** \((\text{head} \leftarrow \text{head})\) in training

Other countries, including West Germany, may have a hard time justifying continued membership.
Intuition:

- weak constraint
Intuition:

- **weak** constraint: (head ← external word) in inference
**Intuition:**

- **weak constraint:** \((\text{head} \leftarrow \text{external word})\) in inference
Intuition:

- **weak constraint**: \((\text{head} \leftarrow \text{external word})\) in inference
Intuition:

- **weak constraint:** \((\text{head} \leftarrow \text{external word})\) in inference
**Intuition:**

- **weak constraint:** (head ← external word) in inference

IFI also **has** nonvoting preferred shares,

which are quoted on the Milan stock exchange.
Linguistic Analysis:

- punctuation and syntax are related
  (Nunberg, 1990; Briscoe, 1994; Jones 1994; Doran, 1998, *inter alia*)
Linguistic Analysis:

- punctuation and syntax are related (Nunberg, 1990; Briscoe, 1994; Jones 1994; Doran, 1998, *inter alia*)

- 49.4% of inter-punctuation fragments are constituents
Linguistic Analysis:

- punctuation and syntax are related (Nunberg, 1990; Briscoe, 1994; Jones 1994; Doran, 1998, *inter alia*)

- 49.4% of inter-punctuation fragments are constituents

- many fragments derive precisely themselves:

<table>
<thead>
<tr>
<th></th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IN</strong></td>
<td>9.6</td>
</tr>
<tr>
<td><strong>NN</strong></td>
<td>7.2</td>
</tr>
<tr>
<td><strong>NNP</strong></td>
<td>6.3</td>
</tr>
<tr>
<td><strong>CD</strong></td>
<td>3.8</td>
</tr>
<tr>
<td><strong>VBD</strong></td>
<td>3.2</td>
</tr>
<tr>
<td><strong>VBZ</strong></td>
<td>3.0</td>
</tr>
<tr>
<td><strong>RB</strong></td>
<td>2.8</td>
</tr>
<tr>
<td><strong>VBG</strong></td>
<td>2.2</td>
</tr>
<tr>
<td><strong>VBP</strong></td>
<td>1.9</td>
</tr>
<tr>
<td><strong>NNS</strong></td>
<td>1.8</td>
</tr>
<tr>
<td><strong>WDT</strong></td>
<td>1.6</td>
</tr>
<tr>
<td><strong>MD</strong></td>
<td>1.1</td>
</tr>
<tr>
<td><strong>VBN</strong></td>
<td>1.1</td>
</tr>
<tr>
<td><strong>IN</strong></td>
<td>1.0</td>
</tr>
<tr>
<td><strong>VBD</strong></td>
<td>0.7</td>
</tr>
<tr>
<td><strong>JJ</strong></td>
<td>52.8</td>
</tr>
</tbody>
</table>
Training:

- **strong** constraint, e.g.,

  ... arrests followed a “Snake Day” at Utrecht ...

— already 74.0% agreement with head-percolated trees.
Inference:

- weak constraint, e.g.,

  Maryland Club also distributes tea, which ...

  — now 92.9% agreement with head-percolated trees!
Modeling

Context-Sensitive Unsupervised Tags
Dependency-and-Boundary Models
Reduced Models of Grammar
Example: Actually, state-of-the-art uses gold tags!

```
IN  PRP  RB  VBZ  VBN  DT  NN  IN  NNS  IN  DT  
NNS  NN  NN  IN  RBS  JJ  NN  NN  NNS  VBZ  RB  
VB  TO  VB  VBN  DT  JJ  NN  IN  NN  IN  JJ  NNS  
DT  NNP  IN  NNP  DT  NN  IN  DT  NNP  NNP  IN  
NNPS  VBD  IN  DT  JJ  NN
```
**Word Categories**: Two benefits of tag-sets...

- **Grouping**: pooling statistics of words that play similar syntactic roles improves generalization (reduces sparsity):
**Word Categories**: Two benefits of tag-sets...

- **Grouping**: pooling statistics of words that play similar syntactic roles improves generalization (reduces sparsity):
  - indeed, working directly with words does not do well.
Word Categories: Two benefits of tag-sets...

- **Grouping**: pooling statistics of words that play similar syntactic roles improves generalization (reduces sparsity):
  - indeed, working directly with words does not do well.

- **Disambiguation**: for words that take on multiple parts of speech, knowing gold tags limits the parsing search space:
Word Categories: Two benefits of tag-sets...

- **Grouping**: pooling statistics of words that play similar syntactic roles improves generalization (reduces sparsity):
  - indeed, working directly with words does not do well.

- **Disambiguation**: for words that take on multiple parts of speech, knowing gold tags limits the parsing search space:
**Word Categories:** Two benefits of tag-sets...

- **Grouping:** pooling statistics of words that play similar syntactic roles improves generalization (reduces sparsity):
  - indeed, working directly with words does not do well.

- **Disambiguation:** for words that take on multiple parts of speech, knowing gold tags limits the parsing search space:

Can improve with **context-sensitive** unsupervised clusters!
Word Categories: Two benefits of tag-sets...

- **Grouping**: pooling statistics of words that play similar syntactic roles improves generalization (reduces sparsity):
  - indeed, working directly with words does not do well.

- **Disambiguation**: for words that take on multiple parts of speech, knowing gold tags limits the parsing search space:

Can improve with context-sensitive unsupervised clusters!

- start with a hard assignment; (standard word clustering)
**Word Categories**: Two benefits of tag-sets...

- **Grouping**: pooling statistics of words that play similar syntactic roles improves generalization (reduces sparsity):
  - indeed, working directly with words does not do well.

- **Disambiguation**: for words that take on multiple parts of speech, knowing gold tags limits the parsing search space:

Can improve with **context-sensitive** unsupervised clusters!

1. start with a hard assignment; (standard word clustering)
2. inject context-colored noise; (get out of the local optimum)
**Word Categories:** Two benefits of tag-sets...

- **Grouping:** pooling statistics of words that play similar syntactic roles improves generalization (reduces sparsity):
  - indeed, working directly with words does not do well.

- **Disambiguation:** for words that take on multiple parts of speech, knowing gold tags limits the parsing search space:

Can improve with **context-sensitive** unsupervised clusters!

1. start with a hard assignment; (standard word clustering)
2. inject context-colored noise; (get out of the local optimum)
3. Viterbi-train a bitag HMM. (the unTagger)
Word Categories: Breaking the barrier...

Swapped out gold tags for unsupervised word categories:

<table>
<thead>
<tr>
<th>System Description</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>“punctuation” with gold tags</td>
<td>58.4</td>
</tr>
</tbody>
</table>
Word Categories: Breaking the barrier...

- Swapped out gold tags for unsupervised word categories:

<table>
<thead>
<tr>
<th>System Description</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>“punctuation” with gold tags</td>
<td>58.4</td>
</tr>
<tr>
<td>“punctuation” with monosemous induced tags</td>
<td>58.2 (−0.2)</td>
</tr>
</tbody>
</table>
Word Categories: Breaking the barrier...

- Swapped out gold tags for unsupervised word categories:
  
<table>
<thead>
<tr>
<th>System Description</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>“punctuation” with <strong>gold</strong> tags</td>
<td><strong>58.4</strong></td>
</tr>
<tr>
<td>“punctuation” with <strong>monosemous</strong> induced tags</td>
<td><strong>58.2</strong> (−0.2)</td>
</tr>
<tr>
<td>“punctuation” with <strong>context-sensitive</strong> induced tags</td>
<td><strong>59.1</strong> (:+0.7)</td>
</tr>
</tbody>
</table>
Word Categories: Breaking the barrier...

- Swapped out gold tags for unsupervised word categories:

<table>
<thead>
<tr>
<th>System Description</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>“punctuation” with gold tags</td>
<td>58.4</td>
</tr>
<tr>
<td>“punctuation” with monosemous induced tags</td>
<td>58.2 (−0.2)</td>
</tr>
<tr>
<td>“punctuation” with context-sensitive induced tags</td>
<td>59.1 (+0.7)</td>
</tr>
</tbody>
</table>

- only a small drop from switching to (monosemous) unsupervised clusters — previous systems lost ∼ 5pts
Word Categories: Breaking the barrier...

- Swapped out gold tags for unsupervised word categories:

  
<table>
<thead>
<tr>
<th>System Description</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>“punctuation” with gold tags</td>
<td>58.4</td>
</tr>
<tr>
<td>“punctuation” with monosemous induced tags</td>
<td>58.2 (−0.2)</td>
</tr>
<tr>
<td>“punctuation” with context-sensitive induced tags</td>
<td>59.1 (+0.7)</td>
</tr>
</tbody>
</table>

- only a small drop from switching to (monosemous) unsupervised clusters — previous systems lost ∼ 5pts;

- first state-of-the-art system to improve with (context-sensitive) unsupervised clusters!
Dependency-and-Boundary Models (DBMs):

- Use boundary cues in head-driven dependency grammars.
Dependency-and-Boundary Models (DBMs):

- Use boundary cues in head-driven dependency grammars.
- E.g., induce structure by working inwards from edges.
Dependency-and-Boundary Models (DBMs):

- Use boundary cues in head-driven dependency grammars.
- E.g., induce structure by working inwards from edges:

```
DT  NN  VBZ  IN  DT  NN
[The check] is in [the mail].
```

Subject NP  Object NP
**Dependency-and-Boundary Models (DBMs):**

- Use boundary cues in head-driven dependency grammars.
- E.g., induce structure by working inwards from edges:

```
DT    NN  VBZ      IN    DT    NN
[The check] is in [the mail].

Subject NP  Object NP
```

▶ learn from left fringe (determiner DT) to parse object NP
Dependency-and-Boundary Models (DBMs):

- Use boundary cues in head-driven dependency grammars.
- E.g., induce structure by working inwards from edges:

  ![Diagram]

  - [The check] is in [the mail].

  Subject $\text{NP}$
  
  Object $\text{NP}$

  - learn from left fringe (determiner DT) to parse object NP
  - based on right fringe (noun NN), correctly parse subject NP
Dependency-and-Boundary Models (DBMs):

- Use boundary cues in head-driven dependency grammars.
- E.g., induce structure by working inwards from edges:

```
DT  NN  VBZ  IN  DT  NN
[The check] is in [the mail].
Subject NP  Object NP
```

- learn from left fringe (determiner DT) to parse object NP
- based on right fringe (noun NN), correctly parse subject NP
- between them, glean make-up of larger phrases (e.g., VP)
Dependency-and-Boundary Models: Highlights
**Dependency-and-Boundary Models**

- **DBM-1**: Use words at the fringes.
Dependency-and-Boundary Models: Highlights

- DBM-1: Use words at the fringes.
  - truly head-outward model

(Alshawi, 1996)
Dependency-and-Boundary Models: Highlights

- DBM-1: Use words at the fringes.
  - truly head-outward model
  - conditions on what can more often be seen!

(Alishawi, 1996)
Dependency-and-Boundary Models: Highlights

- DBM-1: Use words at the fringes.
  - truly head-outward model
  - conditions on what can more often be seen!

- DBM-2: Fragments differ from complete sentences.
Dependency-and-Boundary Models: Highlights

- **DBM-1: Use words at the fringes.**
  - truly head-outward model
  - conditions on what can more often be seen!

- **DBM-2: Fragments differ from complete sentences.**
  - incomplete fragments are uncharacteristically short
Dependency-and-Boundary Models: Highlights

- **DBM-1**: Use words at the fringes.
  - truly head-outward model
  - conditions on what can more often be seen!

- **DBM-2**: Fragments differ from complete sentences.
  - incomplete fragments are uncharacteristically short
  - roots of fragments are generally not verbs or modals
Dependency-and-Boundary Models: Highlights

- **DBM-1**: Use words at the fringes.
  - truly head-outward model
  - conditions on what can more often be seen!

- **DBM-2**: Fragments differ from complete sentences.
  - incomplete fragments are uncharacteristically short
  - roots of fragments are generally not verbs or modals

- **DBM-3**: Learn to stitch together fragments.
Class-based, head-outward generation (Alshawi, 1996)
Class-based, head-outward generation

(Alshawi, 1996)

\[ \begin{align*}
C_h \\
C_e \\
dir &= R \\
adj &= T
\end{align*} \]
Class-based, head-outward generation

$\text{dir} = \text{R} \quad \text{adj} = \text{T}$
Class-based, head-outward generation

$ch = R$

$adj = T$

(Alshawi, 1996)
Class-based, head-outward generation

\[ \text{dir} = R \quad \text{adj} = F \]

(Alshawi, 1996)
Class-based, head-outward generation

(Alshawi, 1996)

dir = R

adj = F
Class-based, head-outward generation

(dir = R)

(adj = F)

(Alshawi, 1996)
Class-based, head-outward generation

\[ \text{dir} = R \quad \text{adj} = F \]
Class-based, head-outward generation

(Alshawi, 1996)
Class-based, head-outward generation (Alshawi, 1996)

\[ \text{dir} = \mathbb{R} \]

\[ \text{adj} = \mathbb{F} \]

\[ \mathbb{P}_{\text{ROOT}}(c_h \mid \text{comp}) \]
Class-based, head-outward generation

\[ ch_{dir} = R \]

\[ cd_1 \]

\[ cd_2 \]

\[ ce \]

\[ STOP \]

\[ adj = F \]

\[ P_{\text{ROOT}}(ch \mid \text{comp}) \]

\[ P_{\text{ATTACH}}(cd \mid ch, dir, cross) \]
Class-based, head-outward generation

(Alshawi, 1996)

\[ \Pr_{\text{ROOT}}(c_h \mid \text{comp}) \Pr_{\text{ATTACH}}(c_d \mid c_h, \text{dir}, \text{cross}) \Pr_{\text{STOP}}(\text{dir}, \text{adj}, c_e, \text{comp}) \]
How to better exploit more data better?

- more text in long sentences, but those can be hard...
How to better exploit more data better?

- more text in long sentences, but those can be hard...
  - require Viterbi training, punctuation constraints, etc.
How to better exploit more data better?

- more text in long sentences, but those can be hard...
  - require Viterbi training, punctuation constraints, etc.

- could we “start small” \textit{and still use} more data?
How to better exploit more data better?

- more text in long sentences, but those can be hard...
  - require Viterbi training, punctuation constraints, etc.

- could we “start small” and still use more data?
  - ... and wouldn’t it be nice if we could just split things up!
What if we chopped up input at punctuation?

- impact on *quantity* of data (with a 15-token threshold):
  - more and simpler word sequences incorporated earlier
What if we chopped up input at punctuation?

- impact on quantity of data (with a 15-token threshold):
  - more and simpler word sequences incorporated earlier
  - much more dense coverage of available data:
    - number of training inputs goes up 2.2x
    - number of tokens increases 4.3x
What if we chopped up input at punctuation?

- **impact on quantity of data (with a 15-token threshold):**
  - more and simpler word sequences incorporated earlier
  - much more dense coverage of available data:
    - number of training inputs goes up 2.2x
    - number of tokens increases 4.3x

- **but, also impact on quality of data**
What if we chopped up input at punctuation?

- **impact on quantity of data** (with a 15-token threshold):
  - more and simpler word sequences incorporated earlier
  - much more dense coverage of available data:
    - number of training inputs goes up 2.2x
    - number of tokens increases 4.3x

- **but, also impact on quality of data:**
  - many fewer complete sentences exhibiting full structure
What if we chopped up input at punctuation?

- **impact on quantity of data (with a 15-token threshold):**
  - more and simpler word sequences incorporated earlier
  - much more dense coverage of available data:
    - number of training inputs goes up 2.2x
    - number of tokens increases 4.3x

- **but, also impact on quality of data:**
  - many fewer complete sentences exhibiting full structure
  - even less representative than short inputs:
    - but mostly phrases and clauses...
What if we chopped up input at punctuation?

- **impact on quantity of data (with a 15-token threshold):**
  - more and simpler word sequences incorporated earlier
  - much *more dense* coverage of available data:
    - number of training inputs goes up 2.2x
    - number of tokens increases 4.3x

- **but, also impact on quality of data:**
  - many fewer complete sentences exhibiting full structure
  - even *less representative* than short inputs:
    - but mostly phrases and clauses...

... however, we have an appropriate model family!
Example (cont’d)
### Example (cont’d)

<table>
<thead>
<tr>
<th>complete</th>
<th>length &amp; type</th>
<th>left &amp; right</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>51</td>
<td>S</td>
</tr>
<tr>
<td>incomplete</td>
<td>12</td>
<td>SBAR</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>PP</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>VP</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>NP</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>VP</td>
</tr>
</tbody>
</table>
**Example (cont’d)**

<table>
<thead>
<tr>
<th></th>
<th>DBM-2</th>
<th>left &amp; right</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>complete</strong></td>
<td><strong>length &amp; type</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>51</strong></td>
<td><strong>S</strong></td>
</tr>
<tr>
<td><strong>incomplete</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>12</strong></td>
<td><strong>SBAR</strong></td>
</tr>
<tr>
<td></td>
<td><strong>2</strong></td>
<td><strong>NP</strong></td>
</tr>
<tr>
<td></td>
<td><strong>6</strong></td>
<td><strong>PP</strong></td>
</tr>
<tr>
<td></td>
<td><strong>14</strong></td>
<td><strong>VP</strong></td>
</tr>
<tr>
<td></td>
<td><strong>4</strong></td>
<td><strong>NP</strong></td>
</tr>
<tr>
<td></td>
<td><strong>8</strong></td>
<td><strong>NP</strong></td>
</tr>
<tr>
<td></td>
<td><strong>5</strong></td>
<td><strong>VP</strong></td>
</tr>
</tbody>
</table>
## Example (cont’d)

<table>
<thead>
<tr>
<th>Length &amp; Type</th>
<th>DBM-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>left &amp; right</td>
</tr>
<tr>
<td>51 S</td>
<td>IN</td>
</tr>
<tr>
<td>12 SBAR</td>
<td>IN</td>
</tr>
<tr>
<td>2 NP</td>
<td>NN</td>
</tr>
<tr>
<td>6 PP</td>
<td>IN</td>
</tr>
<tr>
<td>14 VP</td>
<td>VBZ</td>
</tr>
<tr>
<td>4 NP</td>
<td>DT</td>
</tr>
<tr>
<td>8 NP</td>
<td>DT</td>
</tr>
<tr>
<td>5 VP</td>
<td>VBD</td>
</tr>
<tr>
<td>Incomplete</td>
<td></td>
</tr>
<tr>
<td>12 SBAR</td>
<td>IN</td>
</tr>
<tr>
<td>2 NP</td>
<td>NN</td>
</tr>
<tr>
<td>6 PP</td>
<td>IN</td>
</tr>
<tr>
<td>14 VP</td>
<td>VBZ</td>
</tr>
<tr>
<td>4 NP</td>
<td>DT</td>
</tr>
<tr>
<td>8 NP</td>
<td>DT</td>
</tr>
<tr>
<td>5 VP</td>
<td>VBD</td>
</tr>
</tbody>
</table>
### Example (cont’d)

<table>
<thead>
<tr>
<th></th>
<th>length &amp; type</th>
<th>left &amp; right</th>
</tr>
</thead>
<tbody>
<tr>
<td>complete</td>
<td>51 S</td>
<td>IN NN</td>
</tr>
<tr>
<td>incomplete</td>
<td>12 SBAR</td>
<td>IN NNS</td>
</tr>
<tr>
<td></td>
<td>2 NP</td>
<td>NN NN</td>
</tr>
<tr>
<td></td>
<td>6 PP</td>
<td>IN NNS</td>
</tr>
<tr>
<td></td>
<td>14 VP</td>
<td>VBZ NNS</td>
</tr>
<tr>
<td></td>
<td>4 NP</td>
<td>DT NNP</td>
</tr>
<tr>
<td></td>
<td>8 NP</td>
<td>DT NNPS</td>
</tr>
<tr>
<td></td>
<td>5 VP</td>
<td>VBD NN</td>
</tr>
</tbody>
</table>

**DBM-3**

**partial parse forests**

“easy-first” (Goldberg and Elhadad, 2010)
### Example (cont’d)

<table>
<thead>
<tr>
<th></th>
<th>length &amp; type</th>
<th>left &amp; right</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>complete</strong></td>
<td>51</td>
<td>IN</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>NN</td>
</tr>
<tr>
<td><strong>incomplete</strong></td>
<td>12</td>
<td>IN</td>
</tr>
<tr>
<td></td>
<td>SBAR</td>
<td>NNS</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>NN</td>
</tr>
<tr>
<td></td>
<td>NP</td>
<td>NN</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>IN</td>
</tr>
<tr>
<td></td>
<td>PP</td>
<td>NNS</td>
</tr>
<tr>
<td><strong>reduced model</strong></td>
<td>14</td>
<td>VBZ</td>
</tr>
<tr>
<td></td>
<td>VP</td>
<td>NNS</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>DT</td>
</tr>
<tr>
<td></td>
<td>NP</td>
<td>NNP</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>DT</td>
</tr>
<tr>
<td></td>
<td>NP</td>
<td>NNPS</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>VBD</td>
</tr>
<tr>
<td></td>
<td>VP</td>
<td>NN</td>
</tr>
</tbody>
</table>

**partial parse forests**  

“easy-first” (Goldberg and Elhadad, 2010)
Integrated System
System: Combiners...

(from Leapfrog)
**System: Combiners...**

(from Leapfrog)

\[ C_1 \rightarrow L_D \]
\[ C_1^* = L(C_1) \]
\[ + \]
\[ C_1^* + C_2^* = C_+ \]
\[ L_D \]
\[ C_2 \rightarrow L_D \]
\[ C_2^* = L(C_2) \]

\[ \text{argmax } L_D \]
System: Combiners...

(from Leapfrog)
System: Inductors... (from Baby Steps and Lateen EM)
**System:** Inductors...

(from Baby Steps and Lateen EM)
**System**: Inductors…

(from Baby Steps and Lateen EM)
System: Iterating... (from Baby Steps, Less is More, etc.)
System: Iterating...

(from Baby Steps, Less is More, etc.)
**System:** Iterating...

(from Baby Steps, Less is More, etc.)
System: Iterating...

(from Baby Steps, Less is More, etc.)
System: Results ... on Section 23 of WSJ
System: Results ... on Section 23 of WSJ

Right-Branching (Klein and Manning, 2004) 31.7%
System: Results ... on Section 23 of WSJ

Right-Branching (Klein and Manning, 2004) 31.7%
DMV @10 34.2%
System: Results ... on Section 23 of WSJ

Right-Branching (Klein and Manning, 2004) 31.7%
DMV @10 34.2%
Baby Steps @15 39.2%
Baby Steps @45 39.4%
System: Results

... on Section 23 of WSJ

- Right-Branching (Klein and Manning, 2004) 31.7%
- DMV @10 34.2%
- Baby Steps @15 39.2%
- Baby Steps @45 39.4%
- Soft Parameter Tying (Cohen and Smith, 2009) 42.2%
Reduced Models

Example Continued

**System: Results** ... on Section 23 of WSJ

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy @10</th>
<th>Accuracy @15</th>
<th>Accuracy @45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right-Branching (Klein and Manning, 2004)</td>
<td>31.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMV</td>
<td></td>
<td>34.2%</td>
<td></td>
</tr>
<tr>
<td>Baby Steps</td>
<td></td>
<td>39.2%</td>
<td></td>
</tr>
<tr>
<td>Baby Steps</td>
<td></td>
<td>39.4%</td>
<td></td>
</tr>
<tr>
<td>Soft Parameter Tying (Cohen and Smith, 2009)</td>
<td>42.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less is More</td>
<td></td>
<td>44.1%</td>
<td></td>
</tr>
</tbody>
</table>
### System: Results on Section 23 of WSJ

<table>
<thead>
<tr>
<th>Model</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right-Branching</td>
<td>31.7%</td>
</tr>
<tr>
<td>DMV</td>
<td>34.2%</td>
</tr>
<tr>
<td>Baby Steps</td>
<td>39.2%</td>
</tr>
<tr>
<td>Soft Parameter Tying</td>
<td>42.2%</td>
</tr>
<tr>
<td>Less is More</td>
<td>44.1%</td>
</tr>
<tr>
<td>Leapfrog</td>
<td>45.0%</td>
</tr>
</tbody>
</table>

(Klein and Manning, 2004)

(Cohen and Smith, 2009)
System: Results ... on Section 23 of WSJ

Right-Branching (Klein and Manning, 2004) 31.7%
DMV @10 34.2%
Baby Steps @15 39.2%
Baby Steps @45 39.4%
Soft Parameter Tying (Cohen and Smith, 2009) 42.2%
Less is More @15 44.1%
Leapfrog @45 45.0%

(Gimpel and Smith, 2012) 53.1%
(Gillenwater et al., 2010) 53.3%
(Bisk and Hockenmaier, 2012) 53.3%
(Blunsom and Cohn, 2010) 55.7%
(Tu and Honavar, 2012) 57.0%
**System: Results** ... on Section 23 of WSJ

<table>
<thead>
<tr>
<th>Method</th>
<th>Reference</th>
<th>@10</th>
<th>@15</th>
<th>@45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right-Branching</td>
<td>(Klein and Manning, 2004)</td>
<td>31.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DMV</td>
<td></td>
<td>34.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baby Steps</td>
<td></td>
<td>39.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baby Steps</td>
<td></td>
<td>39.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soft Parameter Tying</td>
<td>(Cohen and Smith, 2009)</td>
<td>42.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less is More</td>
<td></td>
<td>44.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leapfrog</td>
<td>(Gimpel and Smith, 2012)</td>
<td>53.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Gillenwater et al., 2010)</td>
<td>53.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Bisk and Hockenmaier, 2012)</td>
<td>53.3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Blunsom and Cohn, 2010)</td>
<td>55.7%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Tu and Honavar, 2012)</td>
<td>57.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated System</td>
<td>— no initializers or gold tags —</td>
<td>64.4%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reduced Models

**System: Results**

Right-Branching  
(Klein and Manning, 2004) 31.7%

DMV  
@10 34.2%

Baby Steps  
@15 39.2%

Soft Parameter Tying  
(Cohen and Smith, 2009) 42.2%

Less is More  
@15 44.1%

Leapfrog  
(Gimpel and Smith, 2012) 53.1%

(Gillenwater et al., 2010) 53.3%

(Bisk and Hockenmaier, 2012) 53.3%

(Blunsom and Cohn, 2010) 55.7%

(Tu and Honavar, 2012) 57.0%

**Integrated System**

— no initializers or gold tags — 64.4%

**Supervised DBM**  
76.3%
**System**: Results ... on Section 23 of WSJ

- **Right-Branching** (Klein and Manning, 2004) 31.7%
  - @10 34.2%
- **DMV**
  - @15 39.2%
  - @45 39.4%
- **Baby Steps**
  - @15 44.1%
  - @45 45.0%
- **Soft Parameter Tying** (Cohen and Smith, 2009) 42.2%
- **Less is More**
  - @15 44.1%
  - @45 45.0%
- **Leapfrog**

During the same time, *supervised* (constituency) parsing advanced from $91.8 F_1$ (Petrov, 2010) to $92.4 F_1$ (Shindo et al., 2012).

- **Integrated System** — no initializers or gold tags — 64.4%
  - **Supervised DBM** 76.3%
System: Results ... on 23 CoNLL sets

- Also state-of-the-art for multi-lingual evaluation, across 19 languages from disparate families:

  English  +  Arabic
  Basque   Greek
  Bulgarian Hungarian
  Catalan   Italian
  Chinese   Japanese
  Czech     Portuguese
  Danish    Slovenian
  Dutch     Spanish
  German    Swedish
  Turkish

Thanks!

Omri Abend, Eneko Agirre, Hiyan Alshawi, Serafim Batzoglou, John Bauer, Thorsten Brants, Dan Cer, Nate Chambers, Angel Chang, Pi-Chuan Chang, Wanxiang Che, Johnny Chen, Jenny Finkel, Andy Golding, Spence Green, Sonal Gupta, David Hall, Dan Jurafsky, Eisar Lipkovitz, Ting Liu, Chris Manning, Marie de Marneffe, David McClosky, Ryan McDonald, Liz Morin, Andrew Ng, Peter Norvig, Art Owen, Marius Paşca, Fernando Pereira, Slav Petrov, Daniel Pipes, Agnieszka Purves, Daniel Ramage, Marta Recasens, Roi Reichart, Roy Schwartz, Richard Socher, Mihai Surdeanu, Julie Tibshirani, Mengqiu Wang, Eric Yeh, anonymous reviewers, and many others of Stanford NLP, Fannie & John Hertz Foundation, and Google Inc.
Thanks!

Omri Abend, Eneko Agirre, Hiyan Alshawi, Serafim Batzoglou, John Bauer, Thorsten Brants, Dan Cer, Nate Chambers, Angel Chang, Pi-Chuan Chang, Wanxiang Che, Johnny Chen, Jenny Finkel, Andy Golding, Spence Green, Sonal Gupta, David Hall, Dan Jurafsky, Eisar Lipkovitz, Ting Liu, Chris Manning, Marie de Marneffe, David McClosky, Ryan McDonald, Liz Morin, Andrew Ng, Peter Norvig, Art Owen, Marius Pașca, Fernando Pereira, Slav Petrov, Daniel Pipes, Agnieszka Purves, Daniel Ramage, Marta Recasens, Roi Reichart, Roy Schwartz, Richard Socher, Mihai Surdeanu, Julie Tibshirani, Mengqiu Wang, Eric Yeh, anonymous reviewers, and many others of Stanford NLP, Fannie & John Hertz Foundation, and Google Inc.
Thanks!

Omri Abend, Eneko Agirre, Hiyan Alshawi, Serafim Batzoglou, John Bauer, Thorsten Brants, Dan Cer, Nate Chambers, Angel Chang, Pi-Chuan Chang, Wanxiang Che, Johnny Chen, Jenny Finkel, Andy Golding, Spence Green, Sonal Gupta, David Hall, Dan Jurafsky, Eisar Lipkovitz, Ting Liu, Chris Manning, Marie de Marneffe, David McClosky, Ryan McDonald, Liz Morin, Andrew Ng, Peter Norvig, Art Owen, Marius Paşca, Fernando Pereira, Slav Petrov, Daniel Pipes, Agnieszka Purves, Daniel Ramage, Marta Recasens, Roi Reichart, Roy Schwartz, Richard Socher, Mihai Surdeanu, Julie Tibshirani, Mengqiu Wang, Eric Yeh, anonymous reviewers, and many others of Stanford NLP, Fannie & John Hertz Foundation, and Google Inc.
Thanks!

Omri Abend, Eneko Agirre, Hiyan Alshawi, Serafim Batzoglou, John Bauer, Thorsten Brants, Dan Cer, Nate Chambers, Angel Chang, Pi-Chuan Chang, Wanxiang Che, Johnny Chen, Jenny Finkel, Andy Golding, Spence Green, Sonal Gupta, David Hall, Dan Jurafsky, Eisar Lipkovitz, Ting Liu, Chris Manning, Marie de Marneffe, David McClosky, Ryan McDonald, Liz Morin, Andrew Ng, Peter Norvig, Art Owen, Marius Pašca, Fernando Pereira, Slav Petrov, Daniel Pipes, Agnieszka Purves, Daniel Ramage, Marta Recasens, Roi Reichart, Roy Schwartz, Richard Socher, Mihai Surdeanu, Julie Tibshirani, Mengqiu Wang, Eric Yeh, anonymous reviewers, and many others of Stanford NLP, Fannie & John Hertz Foundation, and Google Inc.
Thanks!

Omri Abend, Eneko Agirre, Hiyan Alshawi, Serafim Batzoglou, John Bauer, Thorsten Brants, Dan Cer, Nate Chambers, Angel Chang, Pi-Chuan Chang, Wanxiang Che, Johnny Chen, Jenny Finkel, Andy Golding, Spence Green, Sonal Gupta, David Hall, Dan Jurafsky, Eisar Lipkovitz, Ting Liu, Chris Manning, Marie de Marneffe, David McClosky, Ryan McDonald, Liz Morin, Andrew Ng, Peter Norvig, Art Owen, Marius Paşca, Fernando Pereira, Slav Petrov, Daniel Pipes, Agnieszka Purves, Daniel Ramage, Marta Recasens, Roi Reichart, Roy Schwartz, Richard Socher, Mihai Surdeanu, Julie Tibshirani, Mengqiu Wang, Eric Yeh, anonymous reviewers, and many others of Stanford NLP, Fannie & John Hertz Foundation, and Google Inc.
Dramatization: http://www.youtube.com/embed/ncFCdCjBqcE?start=4&end=66.5
“... the real challenge is to make simple things look beautiful.”
— Glenn Corteza, Tanguero Argentino
Thanks again!

Questions?