Web-Scale Features for Full-Scale Parsing
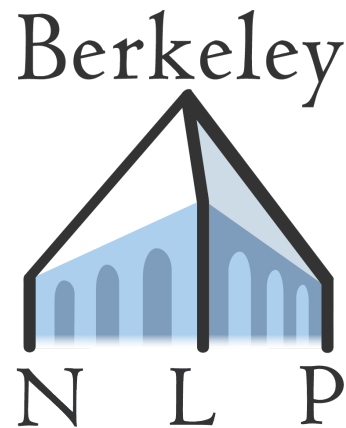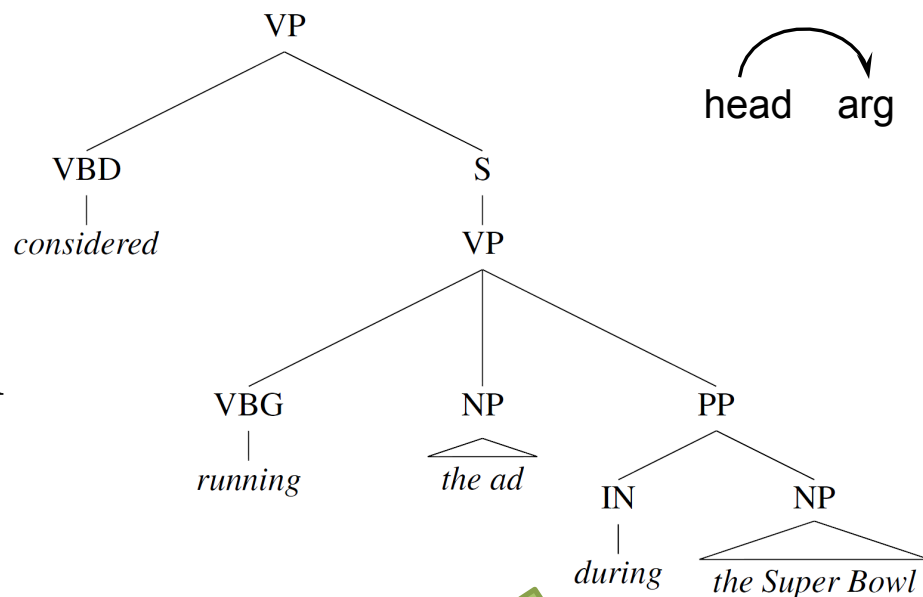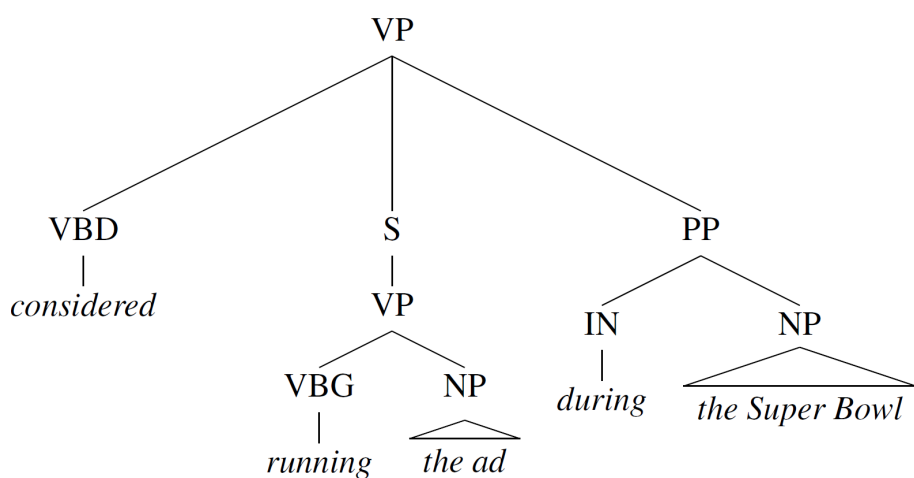
Mohit Bansal and Dan Klein

UC Berkeley

Example

They considered running the ad during the Super Bowl.

head → arg

```
        VP
   /    |    \
 VBD    S     PP
  |     |    /  \
considered VP  IN   NP
       /  \   |    / \
     VBG  NP during the Super Bowl
      |    / \
   running the ad
```

```
        VP
      /    \
    VBD     S
     |      |
 considered VP
          / | \
        VBG NP  PP
         |  |  /  \
      running the ad IN   NP
                   |    / \
                during the Super Bowl
```

Google

✗
*considered * during → 7K*
considered it during → 112
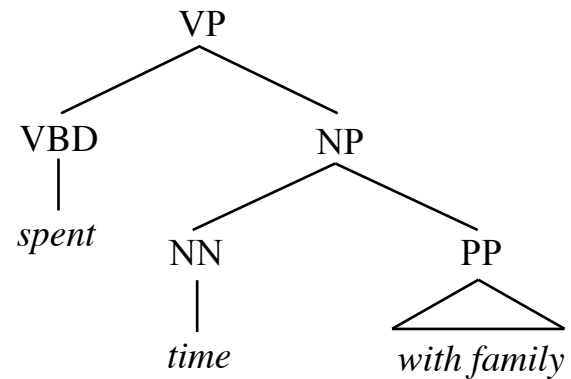
✓
*running * during → 11K*
running it during → 239

[Nakov and Hearst 2005b]

Canonical Ambiguity Type 1

▶ Prepositional phrase (PP) attachment ambiguities (isolated)

VP
VBD   NN   PP
spent time with family

vs.

VP
VBD   NP
spent NN   PP
time with family
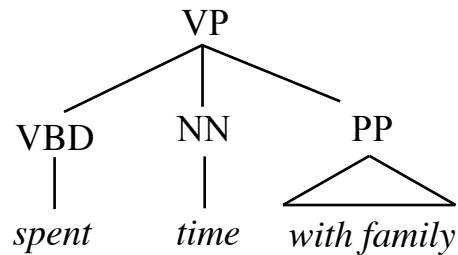
spent with

time with

[Volk 2001; Nakov & Hearst 2005b]

Canonical Ambiguity Type 2

▶ NP coordination ambiguities

NP
├── NP
│   ├── NN — car
│   ├── CC — and
│   └── NN — truck
└── NP
     └── NN — production

vs.

NP
├── NP
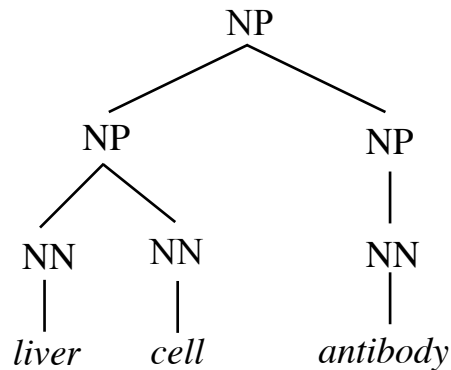│   └── NN — car
├── CC — and
└── NP — truck production
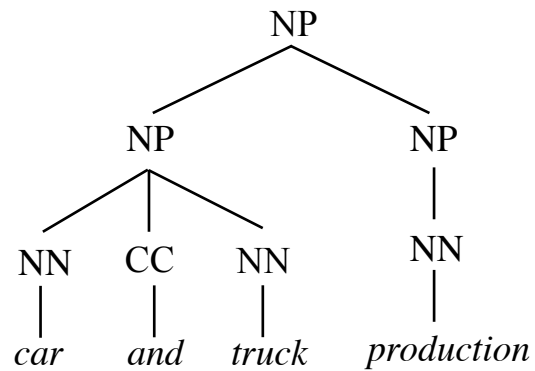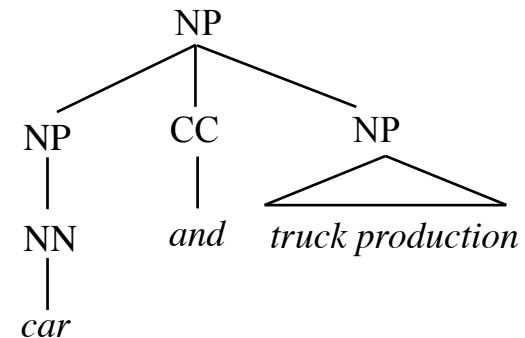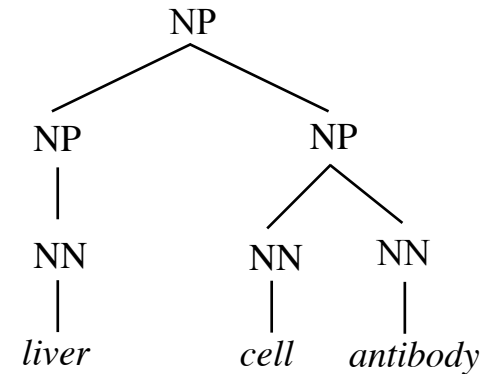
[Nakov & Hearst 2005b; Bergsma *et al*. 2011]

Canonical Ambiguity Type 3

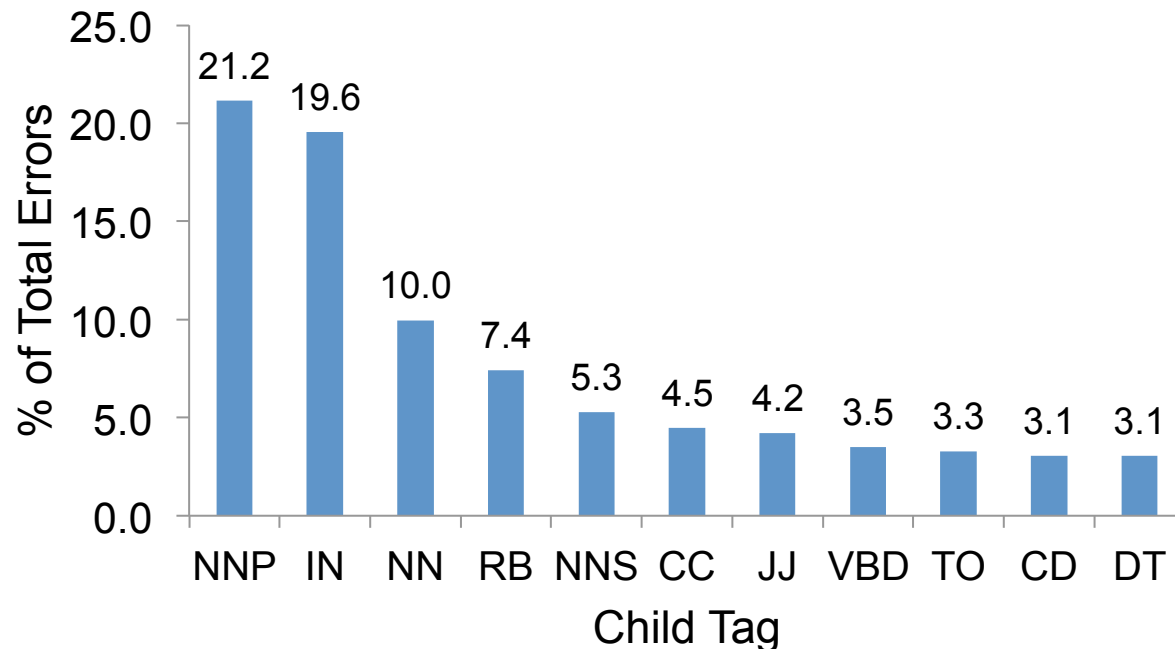▶ Noun compound bracketing ambiguities

[Lapata & Keller 2004; Nakov & Hearst 2005a; Pitler *et al*. 2010]

Parsing Errors
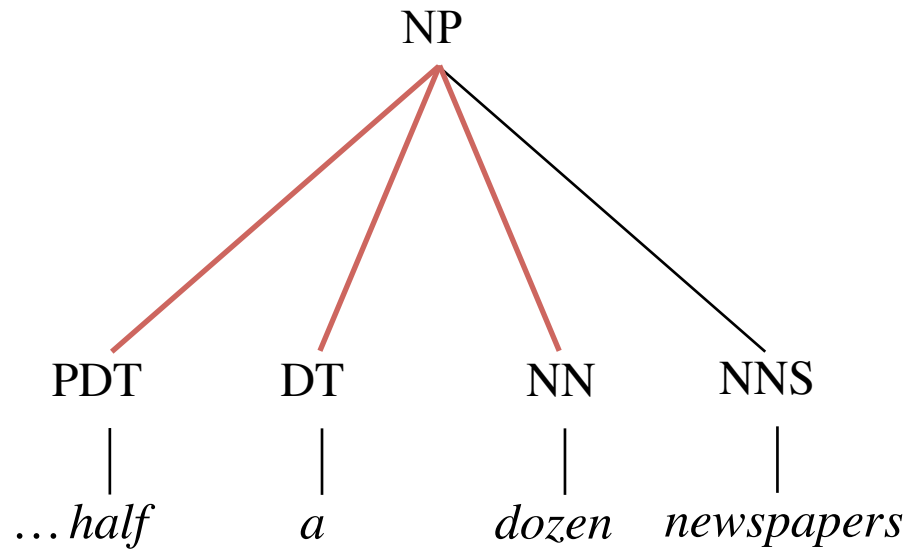
▶ Berkeley parser - errors cast as incorrect dependency attachments

▶ This work - single system that addresses various kinds of ambiguities

... ordered full pages in the Monday editions of half a dozen newspapers .
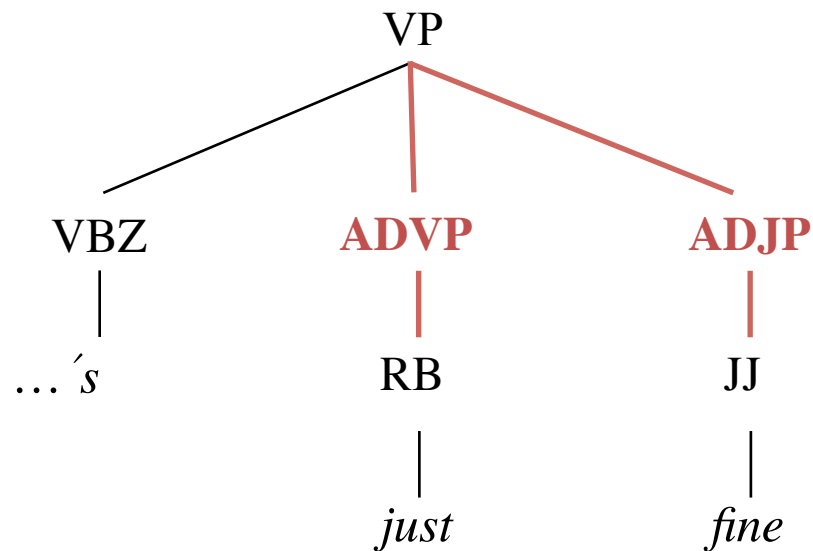

NP

PDT — DT — NN — NNS

...half — a — dozen — newspapers
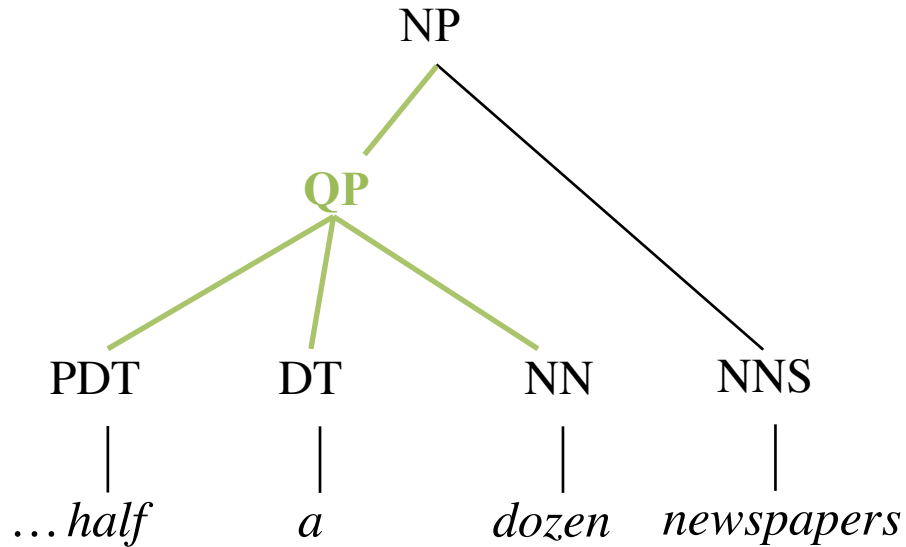
... *ordered full pages in the Monday editions of half a dozen newspapers* .

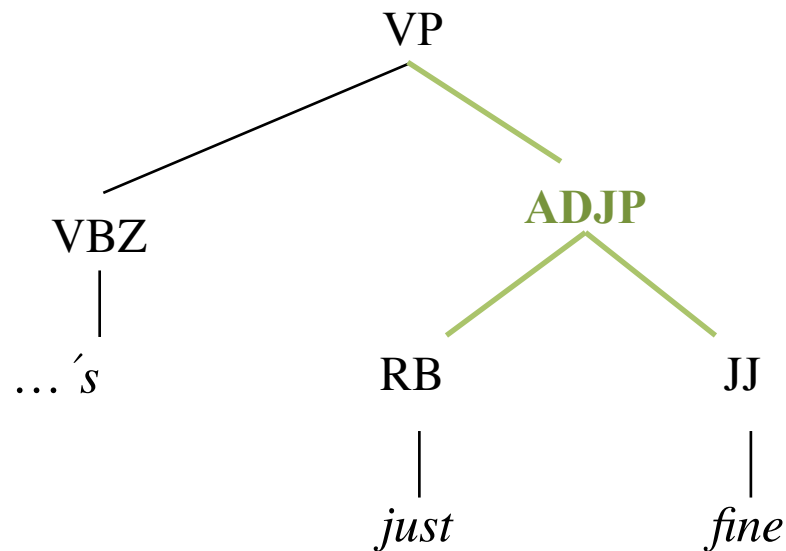*... a familiar message : Keep on investing , the market 's just fine .*

... a familiar message : Keep on investing , the market 's just fine .

```
                    VP
                   /  \
                  /    \
                 /      ADJP
                /       /  \
             VBZ       /    \
              |       RB     JJ
             … ´s     |      |
                     just   fine
```

Using Web-Scale Features

▶ Idea: Edge-factored features that encode web-counts

ϕ(head   arg)

$\boldsymbol{\phi}$(*raising   from*) $\boldsymbol{\phi}$(*$ from*)

raising $ 30   million   from   debt
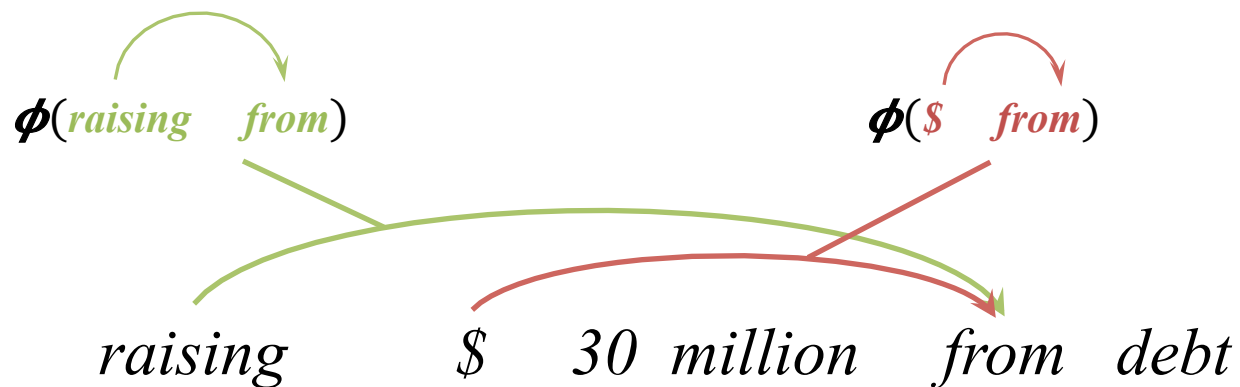
Web-Scale Statistics

▶ Prepositional Phrase (PP) disambiguation

verb / noun attachment ?
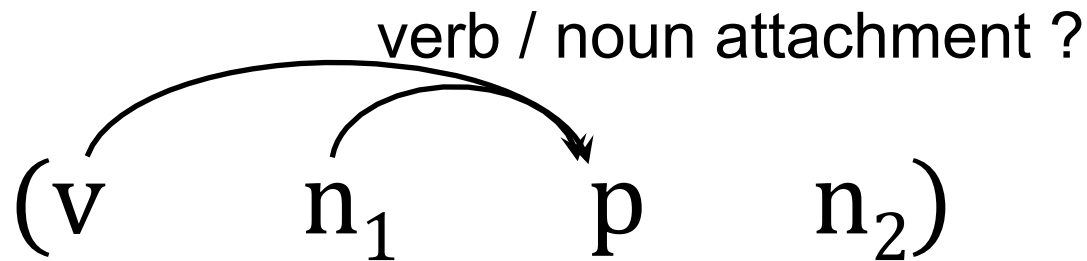
$$(v \quad n_1 \quad p \quad n_2)$$

$$\boxed{\frac{c(v, p, n_2)}{c(v)} > \frac{c(n_1, p, n_2)}{c(n_1)} \; ? \; V : N}$$ (Volk, 2001)

Only 2 competing attachments !

Dependency Features

▶ Discriminative dependency parsing
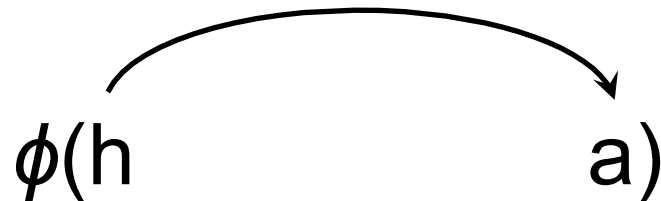
$$\phi(h \quad \overset{\frown}{\longrightarrow} \quad a)$$

1[h, a] (McDonald *et al.*, 2005; *inter alia*)

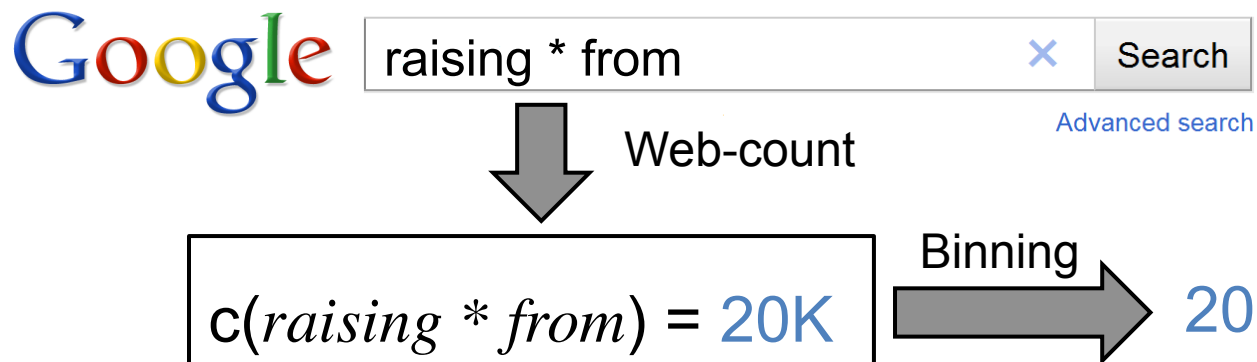**1**[cluster(h), cluster(a)] (Koo *et al.*, 2008; Finkel *et al.*, 2008)

Web-Scale Features

▶ Affinity based Web features

$$\phi(raising \quad +3 \quad from)$$

Google | raising * from | ✕ | Search

Advanced search

Web-count

c(*raising * from*) = 20K → Binning → 20

Web-Scale Features

▶ Affinity based Web features

$$\phi(\textit{raising} \quad +3 \quad \textit{from})$$

$$\mathbf{1}[\text{VBG} - +3 - \text{IN}, \quad 20]$$

Unlexicalized   Lexicalized

Web-Scale Features

▶ Affinity based Web features

$$\phi(h \quad \text{[distance]} \quad a)$$

1[POS(h) — d — POS(a), webcnt]

Web-Scale Features

▶ Paraphrase (context) based Web features

$$\phi(raising \qquad from)$$

Google | raising * from | × | Search

Advanced search

⬇ Top trigrams

raising money from
raising funds from
raising him from
raising it from
raising capital from

....

[Nakov and Hearst 2005b]

Web-Scale Features

- Paraphrase (context) based Web features

$$\phi(\textit{raising} \qquad \textit{from})$$

$$\mathbf{1}[\text{VBG} - \textit{it} - \text{IN}]$$

Web-Scale Features

▶ Paraphrase (context) based Web features

ϕ(h   a)
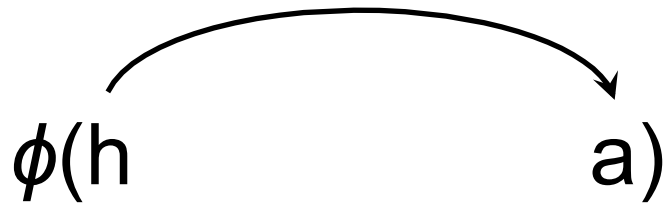
1[POS(h) — context — POS(a)]

▶ Collecting top context words

$$\phi(\text{h} \qquad \text{a})$$
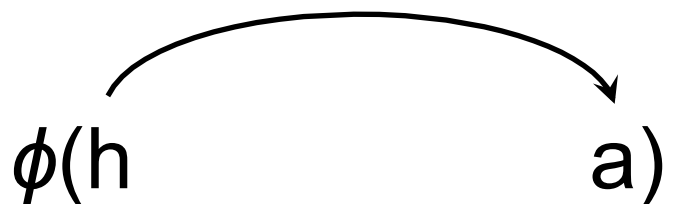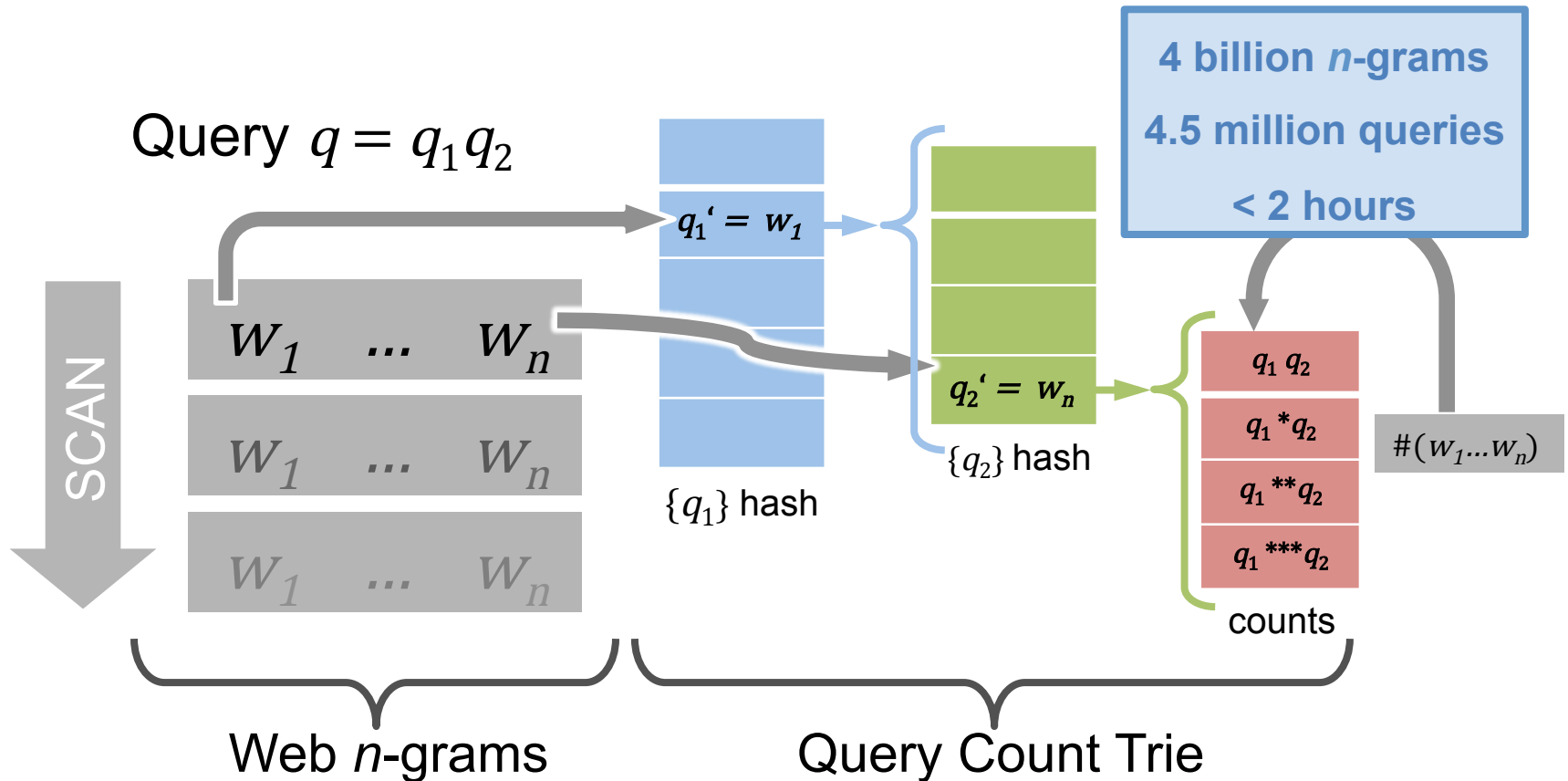
middle : $\text{k-argmax}_{\star} c(h \star a)$

before : $\text{k-argmax}_{\star} c(\star h \, a)$

after : $\text{k-argmax}_{\star} c(h \, a \star)$

Computing Web Statistics Efficiently

▶ Search engines inefficient – use Google *n*-grams (*n* = 1 to 5)

▶ Batch – Collect all queries beforehand, then scan all *n*-grams

Query $q = q_1 q_2$

SCAN

$W_1 \quad \ldots \quad W_n$

$W_1 \quad \ldots \quad W_n$

$W_1 \quad \ldots \quad W_n$

$q_1' = w_1$

$q_2' = w_n$

$\{q_1\}$ hash

$\{q_2\}$ hash

4 billion *n*-grams

4.5 million queries

< 2 hours

$q_1 \, q_2$

$q_1 * q_2$

$q_1 ** q_2$

$q_1 *** q_2$

counts

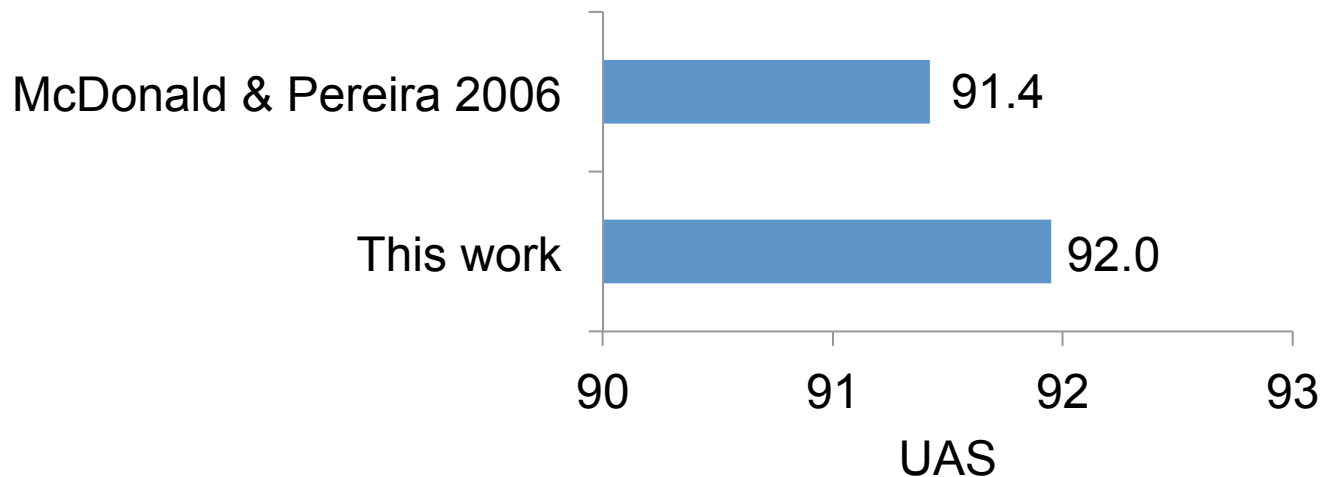$\#(w_1 \ldots w_n)$

Web *n*-grams

Query Count Trie

Parsing Results

▶ Dependency Parsing

  ▶ Web-features integrated into underlying dynamic program

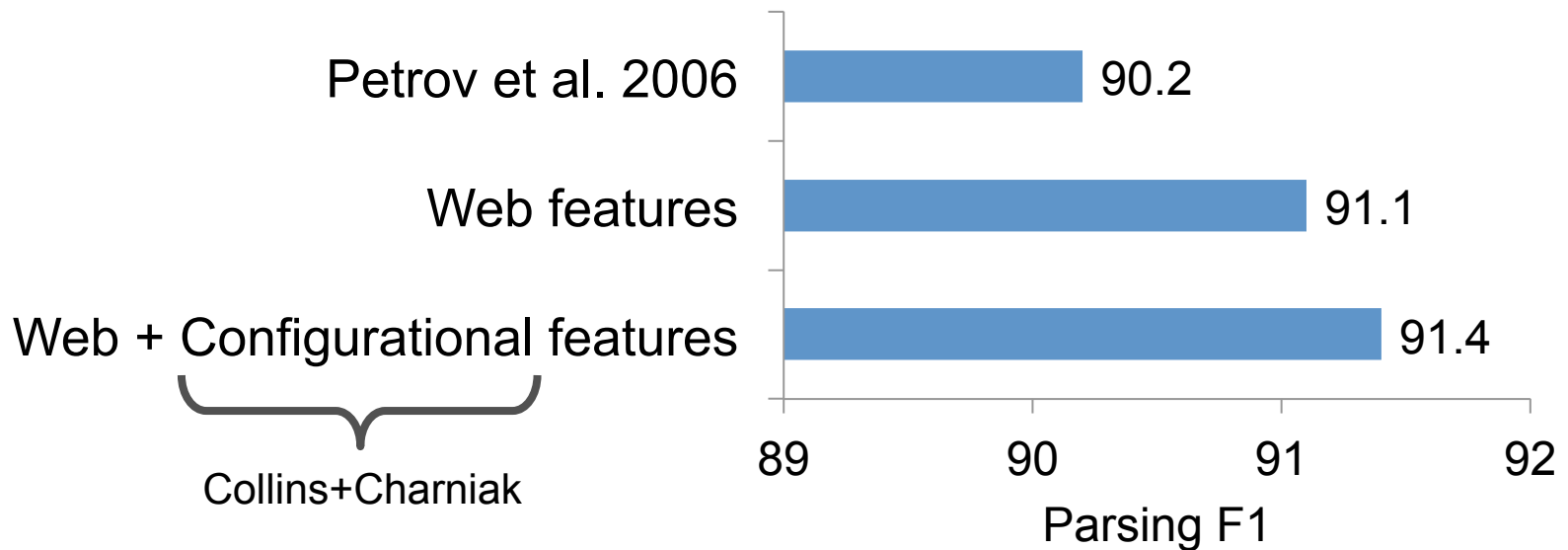 ▶ Error reduction (relative) of 7.0% over order-2 features
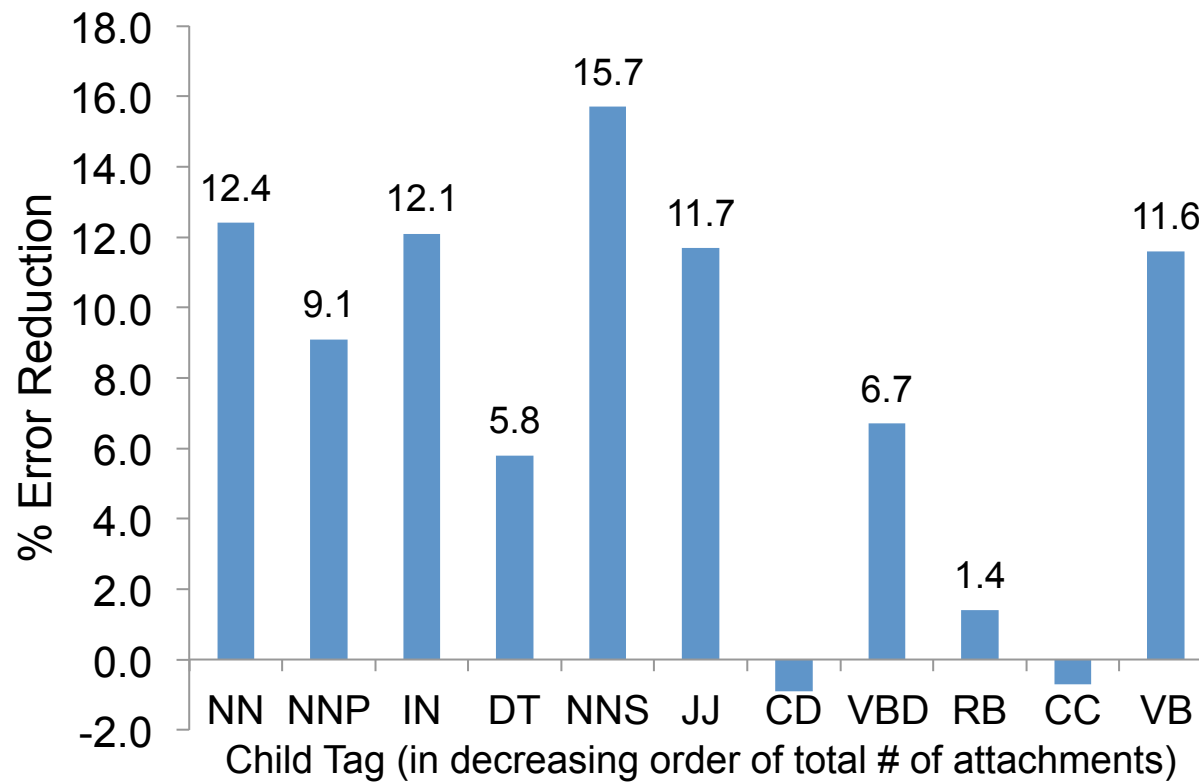
Parsing Results

▶ Constituent Parsing

  ▶ Get *k*-best parses and rerank them discriminatively

  ▶ Error reduction (relative) of 9.2% and 12.2%

Petrov et al. 2006   90.2
Web features   91.1
Web + Configurational features   91.4

Collins+Charniak

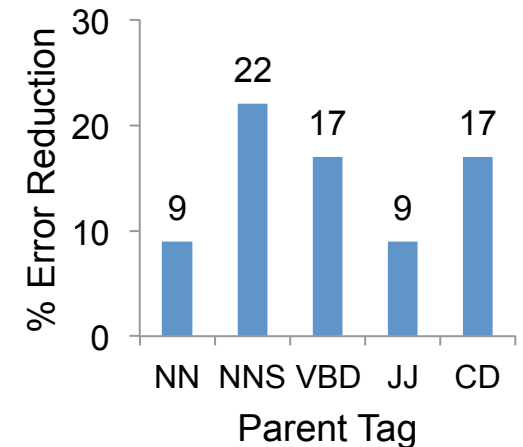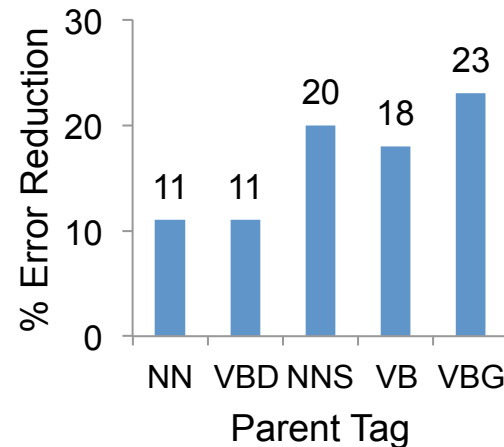89   90   91   92
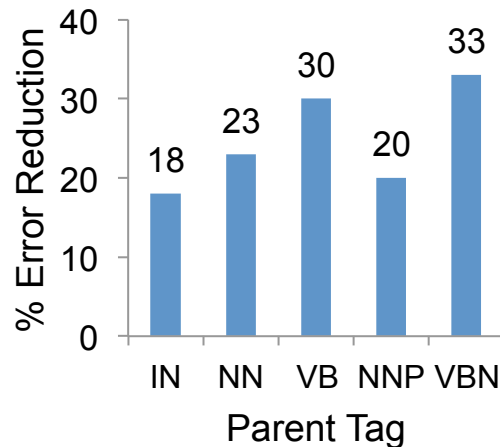
Parsing F1

Error Analysis

▶ Errors reduced for a variety of child (argument) types

Error Analysis

▶ Error reduction for each type of parent attachment for a given child

| Child Tag | NN | IN | JJ |

High-Weight Features

▶ Affinity features

RB → IN   *their bridge back into the big-time*

NN → IN   *an Oct. 19 review of "The, Misanthrope"*

DT → NN   *The new rate will be payable Feb. 15*

▶ Paraphrase features

The   row   by   *sow a row of male-fertile plants*

learned   this   from   *the guile learned from his years in*

purchases and sales   *about stock purchases and sales by*

NN   IN

VBN   DT   IN

NNS   and   NNS

Conclusion

▶ Web-features are powerful disambiguators

▶ Incorporation into end-to-end full-scale parsing

▶ Uniform treatment of all attachment error types

▶ 7-12% relative error reduction in state-of-the-art parsers

▶ Intuitive features surface in the learning setup

Thank you!

Berkeley

N L P

Questions?