Creating Robust Supervised Classifiers via Web-Scale N-gram Data

Shane Bergsma
University of Alberta

Emily Pitler
University of Pennsylvania

Dekang Lin
Google, Inc.

ACL 2010

New Web-Scale N-gram Data

- Details in: [Lin et al., LREC 2010]
  - Same source as Google N-grams Version 1
  - More pre-processing: duplicate sentence removal, length+alphabetical constraints
- Includes part-of-speech tags!

flies	**1643568**	**NNS|611646 VBZ|1031922**
caught the flies ,	**11**	**VBD|DT|NNS|,|11**
plane flies really well	**10**	**NN|VBZ|RB|RB|10**

Overview

- Features from web-scale N-gram data:
  - Count(***some N-gram***) in web corpus
- Open questions:
  1. How well do web-scale N-gram features work when combined with conventional features?
  2. How well do classifiers with web-scale N-gram features perform on new *domains*?
- **Conclusion**: N-gram features are *essential*

Feature Classes

- Lex (lexical features): $\mathbf{x}_{\mathrm{Lex}}$
  - Many thousands of **binary** features indicating a property of the strings to be classified
- N-gm (N-gram count features): $\mathbf{x}_{\mathrm{Ngm}}$
  - A few dozen **real-valued** features for the **logarithmic** counts of various things
- The classifier:

$$h(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x}$$

$$\mathbf{x} = (\mathbf{x}_{\mathrm{Lex}}, \mathbf{x}_{\mathrm{Ngm}})$$

Uses of New N-gram Data

- Applications:
  1. **Adjective Ordering**
  2. **Real-Word Spelling Correction**
  3. **Noun Compound Bracketing**
  4. Verb Part-of-Speech Tagging
     - benefits of N-grams not so clear cut (see paper)
- All experiments: linear SVM classifier, report *Accuracy* (%)

1. Adjective Ordering

- “**green big** truck” or
  “**big green** truck”?

- Used in translation, generation, etc.
- Not a syntactic issue but a semantic issue:
  - size precedes colour, etc.

Adjective Ordering

- As a classification problem:
  - Take adjectives in alphabetical order
  - Decision: is alphabetical order correct or not?
- Why not just most frequent order on web?
  - 87% for web order but 94% for classifier

Adjective Ordering Features

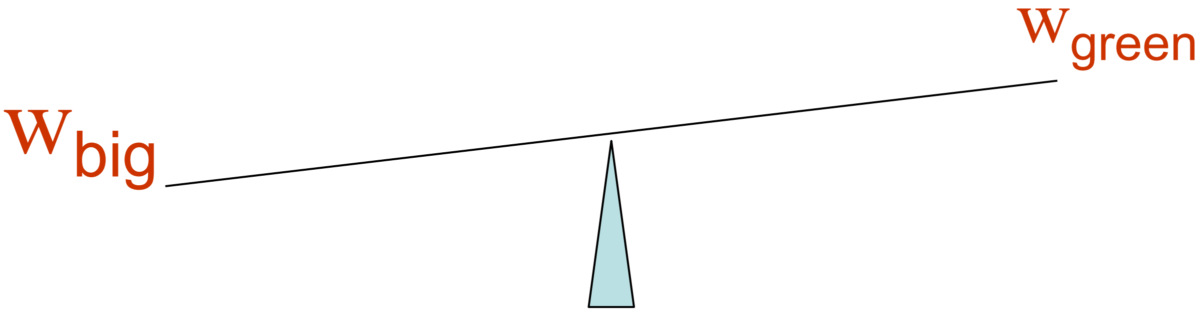
- Lex features: indicators for the adjectives
  - adj$_1$ indicated with +1, adj$_2$ indicated with -1
  - E.g. “big green”

$$\mathbf{x}_{Lex} = (\ldots,0,0,0,0,0,0,0,\mathbf{+1},0,0,0,0,\ldots,0,0,0,\mathbf{-1},0,0,0,0,0,0,0,0,0,0,0,\ldots)$$

Decision: $h_{Lex}(\mathbf{x}_{Lex}) = \mathbf{w}_{Lex} \cdot \mathbf{x}_{Lex}$

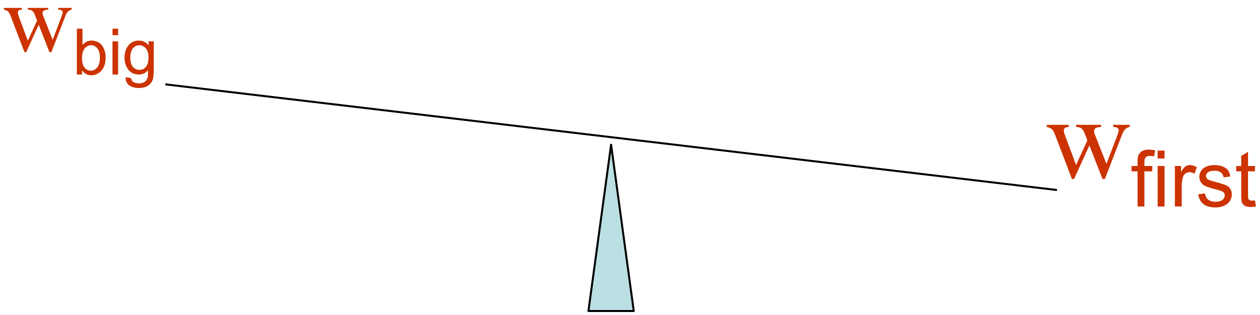
$h_{Lex}(\mathbf{x}_{Lex}) = w_{big} - w_{green}$

Adjective Ordering Features



big green truck

Adjective Ordering Features



first big storm

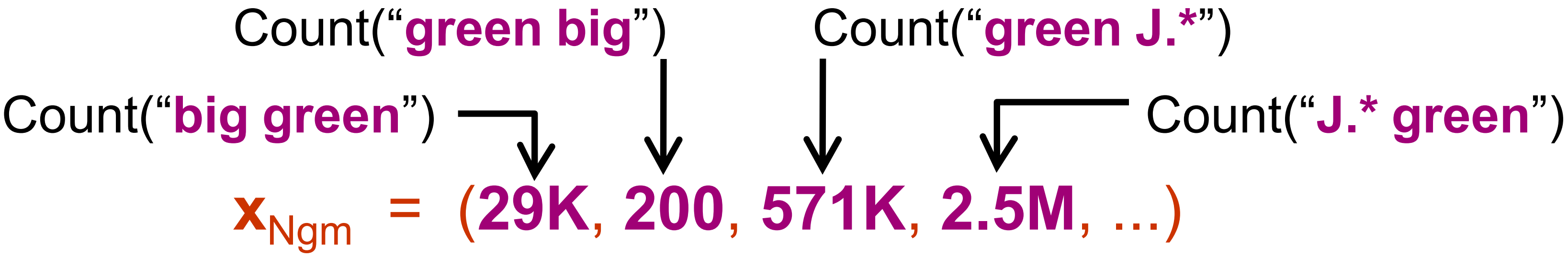
Adjective Ordering Features

w_{first} w_{big} w_{young} w_{green} w_{Canadian}

Adjective Ordering Features

- N-gm features:

Count("**big green**")	Count("**green big**")
Count("**big J.***")	Count("**green J.***")
Count("**J.* big**")	Count("**J.* green**")...



Count("**green big**")

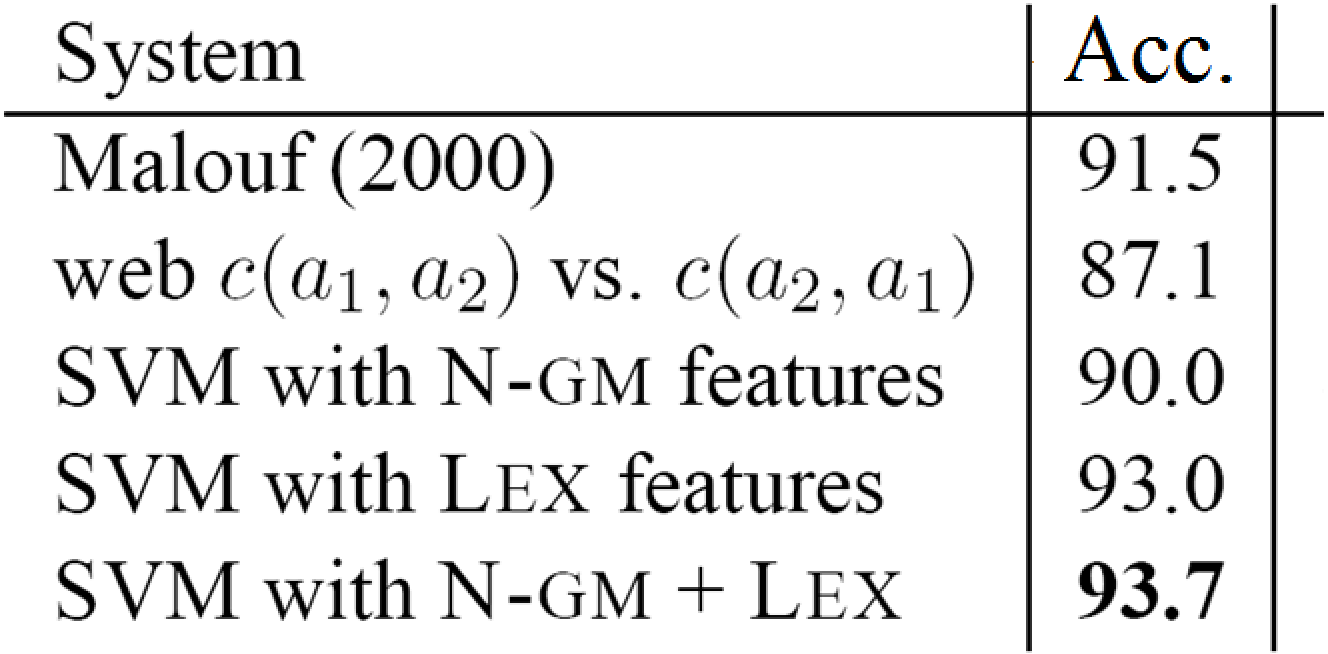
Count("**green J.***")

Count("**big green**")

Count("**J.* green**")

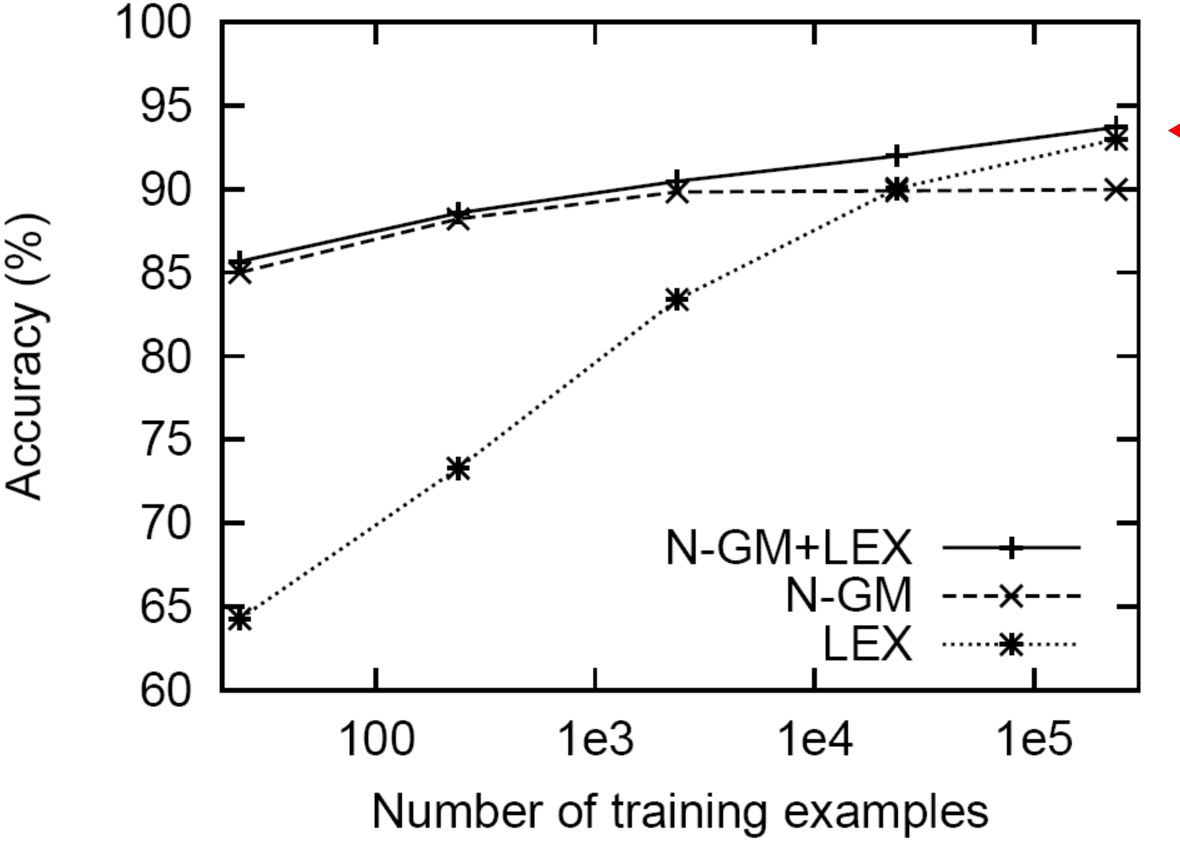
x_{Ngm} = (**29K**, **200**, **571K**, **2.5M**, …)

Adjective Ordering Results



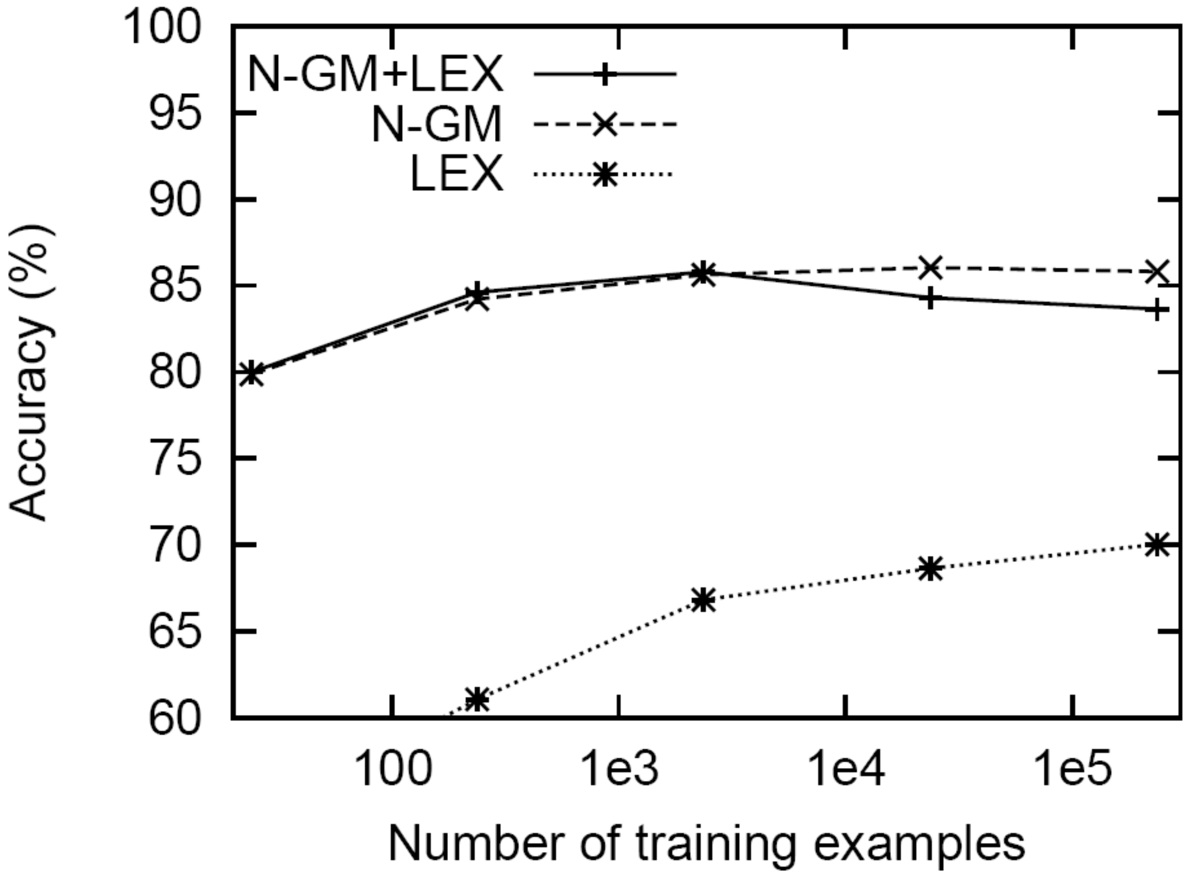
System	Acc.
Malouf (2000)	91.5
web $c(a_1, a_2)$ vs. $c(a_2, a_1)$	87.1
SVM with N-GM features	90.0
SVM with LEX features	93.0
SVM with N-GM + LEX	**93.7**

In-Domain Learning Curve



← 93.7%

Out-of-Domain Learning Curve



2. Real-Word Spelling Correction

- Classifier predicts correct word in context:

  “Let me know **weather** you like it.”

  “**weather**” or “**whether**”

Spelling Correction

- Lex features:
  - Presence of particular words (and phrases) preceding or following the confusable word

Spelling Correction

- N-gm feats: Leverage multiple relevant contexts:

Bergsma et al., 2009

```
Let me know _
    me know _ you
       know _ you like
            _ you like it
```

  - Five 5-grams, four 4-grams, three 3-grams and two 2-grams span the confusable word

Spelling Correction

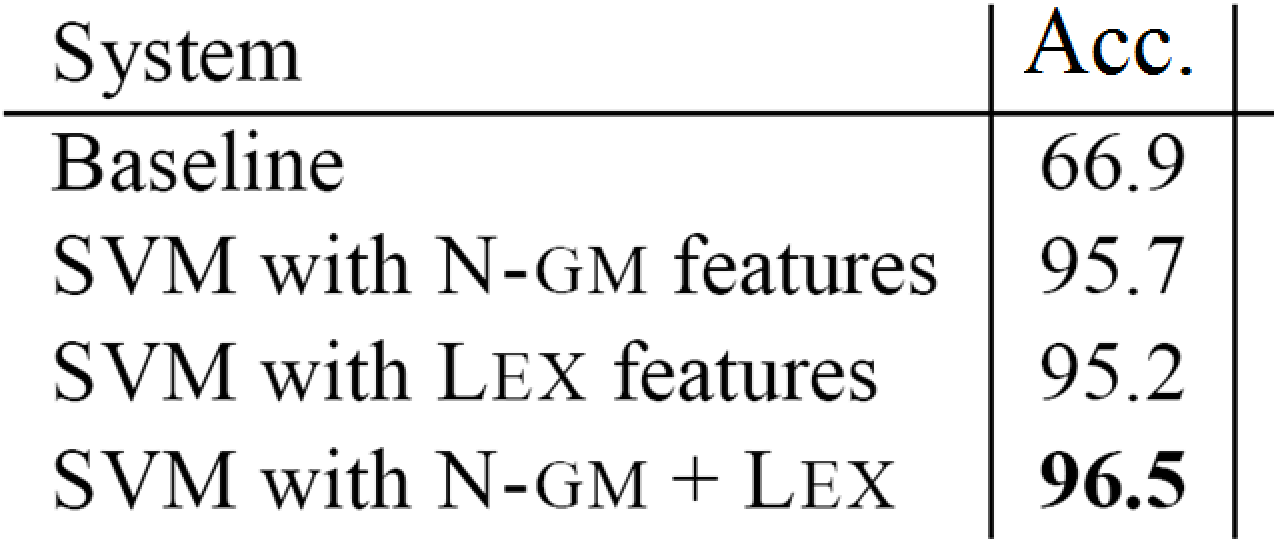
- N-gm features:
  - Count("let me know **weather** you") 5-grams
  - Count("me know **weather** you like")

    …
  - Count("let me know **weather**") 4-grams
  - Count("me know **weather** you")
  - Count("know **weather** you like")

    …
  - Count("let me know **whether** you") 5-grams

    …

Spelling Correction Results



System	Acc.
Baseline	66.9
SVM with N-GM features	95.7
SVM with LEX features	95.2
SVM with N-GM + LEX	**96.5**

3. Noun Compound Bracketing

- "… bus driver"
  - female (bus driver)
  - *(female bus) driver
  - (school bus) driver

3-word case is a binary classification: **right** or **left** bracketing

Noun Compound Bracketing

- Lex features:
  - binary features for all words, pairs, and the triple, plus capitalization pattern

Vadas & Curran, 2007

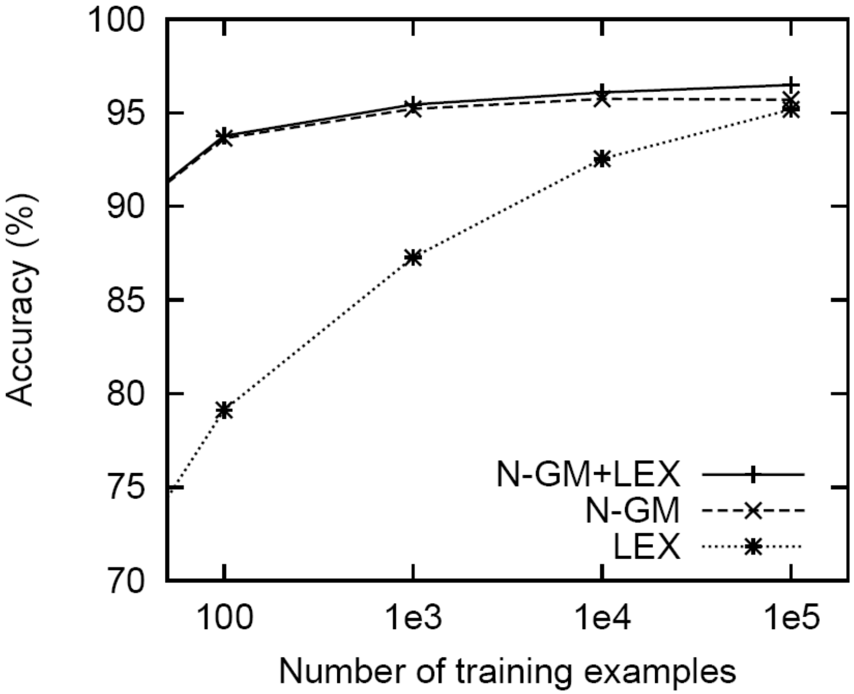
Noun Compound Bracketing

- N-gm features, e.g. "female bus driver"
  - Count("female bus") → predicts left
  - Count("female driver") → predicts right
  - Count("bus driver") → predicts right
  - Count("femalebus")
  - Count("busdriver")
  - etc.

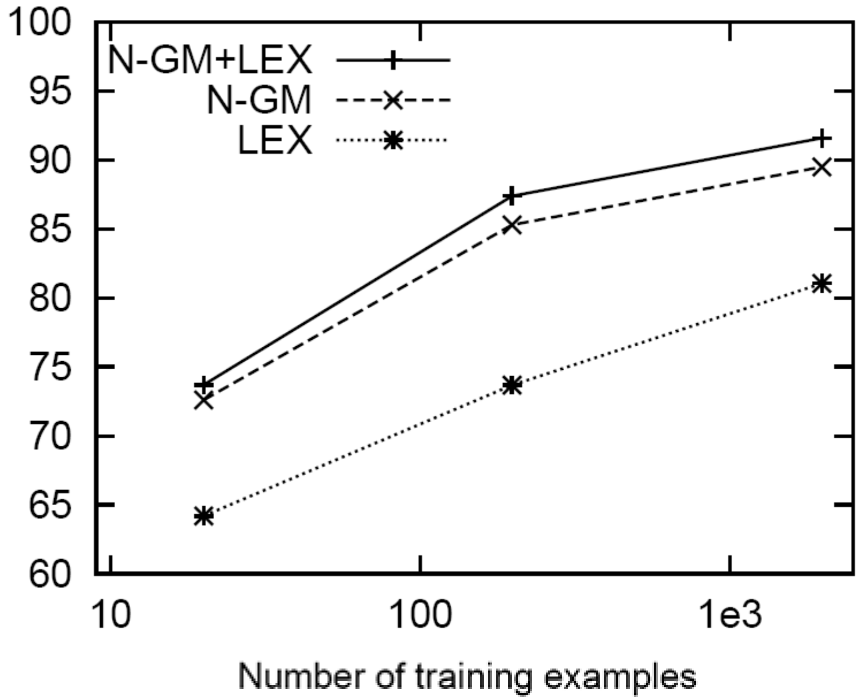
Nakov & Hearst, 2005

In-Domain Learning Curves

Spelling Correction



Noun Compound Bracketing



Out of Domain

Spelling Correction

Errors nearly double when you remove N-gram features

Noun Compound Bracketing

No N-gram features = BAD

Conclusion

- It's good to mix standard lexical features with N-gram count features
- Domain sensitivity of NLP

Thanks

- Google, Inc.
- Johns Hopkins University
- Liam:

(getting one good picture took 45 minutes)