# Discriminative Learning of Selectional Preference from Unlabeled Text

**Shane Bergsma** — University of Alberta
**Dekang Lin** — Google, Inc.
**Randy Goebel** — University of Alberta

**EMNLP 2008**

# Selectional Preferences

- Compatibility of word pairs:
  - "eat bacon" vs. "eat Reliance Industries Ltd."
- Input: verb-object pair  Output: plausibility
- Method:
  - Train SVM to distinguish pairs observed in a corpus from other (unobserved) pairings
  - DSP: Discriminative Selectional Preference

# Outline

1. Introduction and Motivation
2. Learning from Unlabeled Text
3. Features
4. Experiments and Results

# Selectional Preferences

- Which *arguments* can go with which *predicates*?
- Typically, the argument is a noun, and the predicate is a verb or adjective:
- This paper: verbs and object nouns
- Typical approach: look at a corpus
- Unfortunately, observed data is never sufficient – can we use it to *generalize*?

# Motivation

- Pronoun Resolution:

  "My dog ate my homework so I couldn't finish it"

- Parsing:

  "Later we ate their signature *Choucroute garnie*"

- Model of human acquisition

# Human Acquistion

# Selectional Preferences

| Verb | Noun |
|---|---|
| eat | quail |
| eat | *Choucroute garnie* |
| eat | Reliance Industries Ltd. |
| eat | Harry Whittington |

# Selectional Preferences

| Verb | Noun |
|---|---|
| shoot | quail |
| shoot | *Choucroute garnie* |
| shoot | Reliance Industries Ltd. |
| shoot | Harry Whittington |

# Selectional Preference Features

- Did the noun and verb occur together before or not?
- What other words does the noun occur with (its distribution)?
- Number of tokens
- Capitalized, upper or lower-case
- Contains tokens like "Harry" or "Ltd."

# Combining Feature Information

- Plausibility score of $(v,n)$ is sum of a weighted linear combination of arbitrary and potentially interdependent features:

$$y = \boldsymbol{\lambda} \cdot \boldsymbol{\Phi}(v, n)$$

- Use discriminative training to set feature weights: Discriminative Selectional Preference (DSP)

# Training Examples

- Parse a corpus and collect verb-object co-occurrence statistics:

$$\mathrm{MI}(v, n) = \log \frac{\Pr(v, n)}{\Pr(v)\Pr(n)} = \log \frac{\Pr(n|v)}{\Pr(n)}$$

- Positives: MI greater than some threshold
- Negatives: for each positive, match with nouns of similar freq. that are not positive

# Training

- For efficiency, DSP trains a separate classifier for each verb
- Features are for noun only:

$$y^v = \boldsymbol{\lambda}^v \cdot \boldsymbol{\Phi}^v(n)$$

- 57K features, 6.5 million training instances

# Similarity Smoothing

- Other approaches generalize from similar predicates
- Dagan, Lee, & Pereira (1999):

$$\Pr_{\mathrm{SIM}}(n|v) = \sum_{v' \in \textsc{Sims}(v)} \mathit{Sim}(v', v) \Pr(n|v')$$

- E.g. More likely to believe “shoot quail” if we’ve seen “hunt quail” or “stab quail”

# Co-occurrence Features

- For $(v,n)$, DSP has features for probability of $n$ occurring as object of other verbs, $v'$.
  - E.g. for $\mathrm{SP}(shoot\text{-}n)$, feature[10] = $\Pr(n|hunt)$

$$y^v = \sum_{v'} \lambda^v_{v'} \Pr(n|v')$$

- Also features for number of tokens, case, capitalization, semantic-class, etc.

# Implementation

- Parse 3 GB AQUAINT corpus using Minipar
- Use MI>0 as threshold, have 2 negatives for every positive
- 95% for training, 2.5% for development, 2.5% for testing
- Use SVM-light, set C-parameter and j-parameter on development set.

# Feature Weights: $\Pr_{\text{object}}(n|join)$

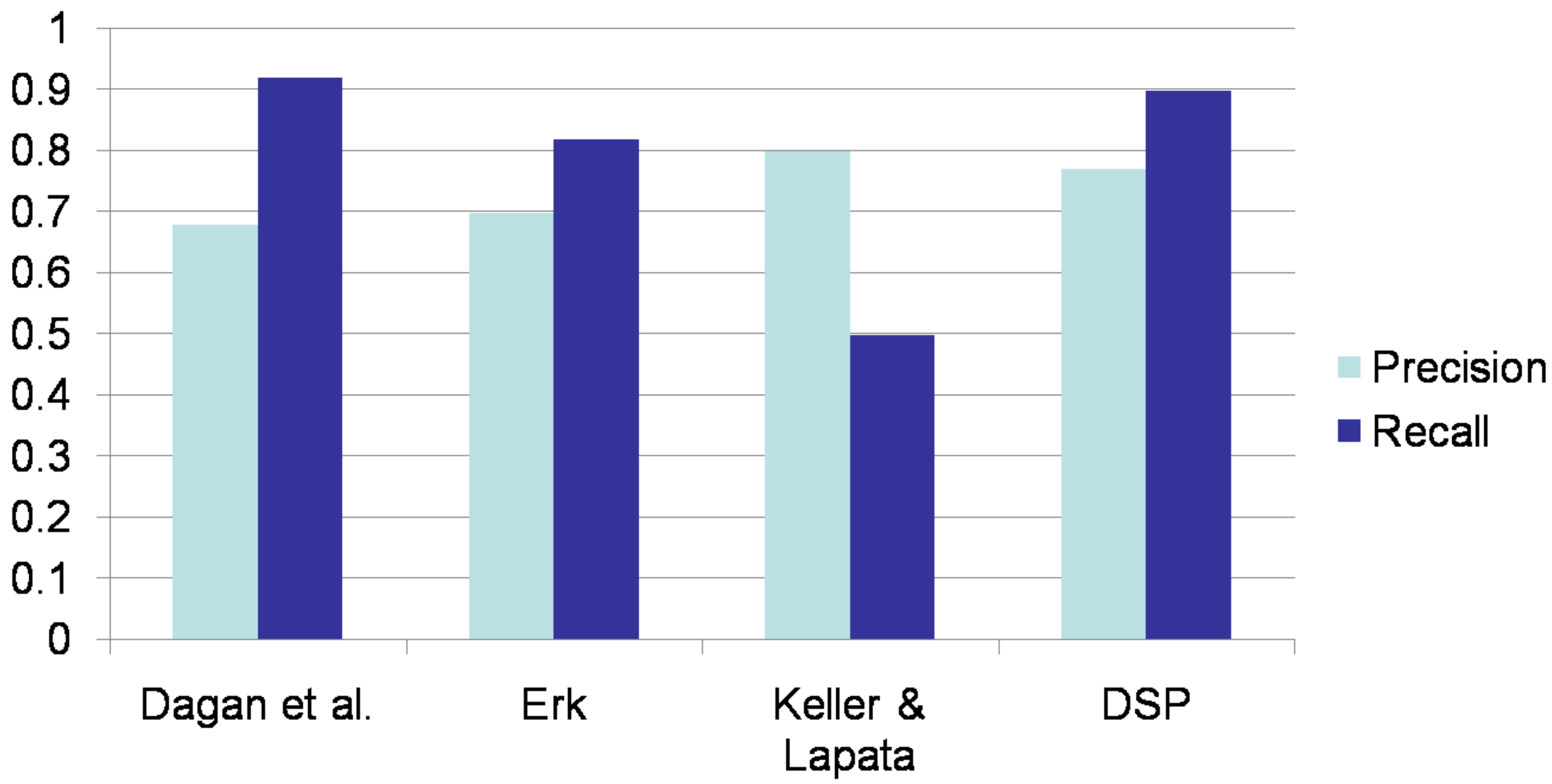
| Learned Weights | | Lin (1998) Similarity | |
|---|---|---|---|
| lead-$n$ | 1.42 | *participate-n* | *0.164* |
| rejoin-$n$ | 1.39 | lead-$n$ | 0.150 |
| form-$n$ | 1.34 | return to-$n$ | 0.148 |
| belong to-$n$ | 1.31 | *say-n* | *0.143* |
| found-$n$ | 1.31 | rejoin-$n$ | 0.142 |
| quit-$n$ | 1.29 | *sign-n* | *0.142* |
| guide-$n$ | 1.19 | meet-$n$ | 0.142 |
| induct-$n$ | 1.19 | *include-n* | *0.141* |
| $n$-launch | 1.18 | leave-$n$ | 0.140 |

# String-based Feature Weights

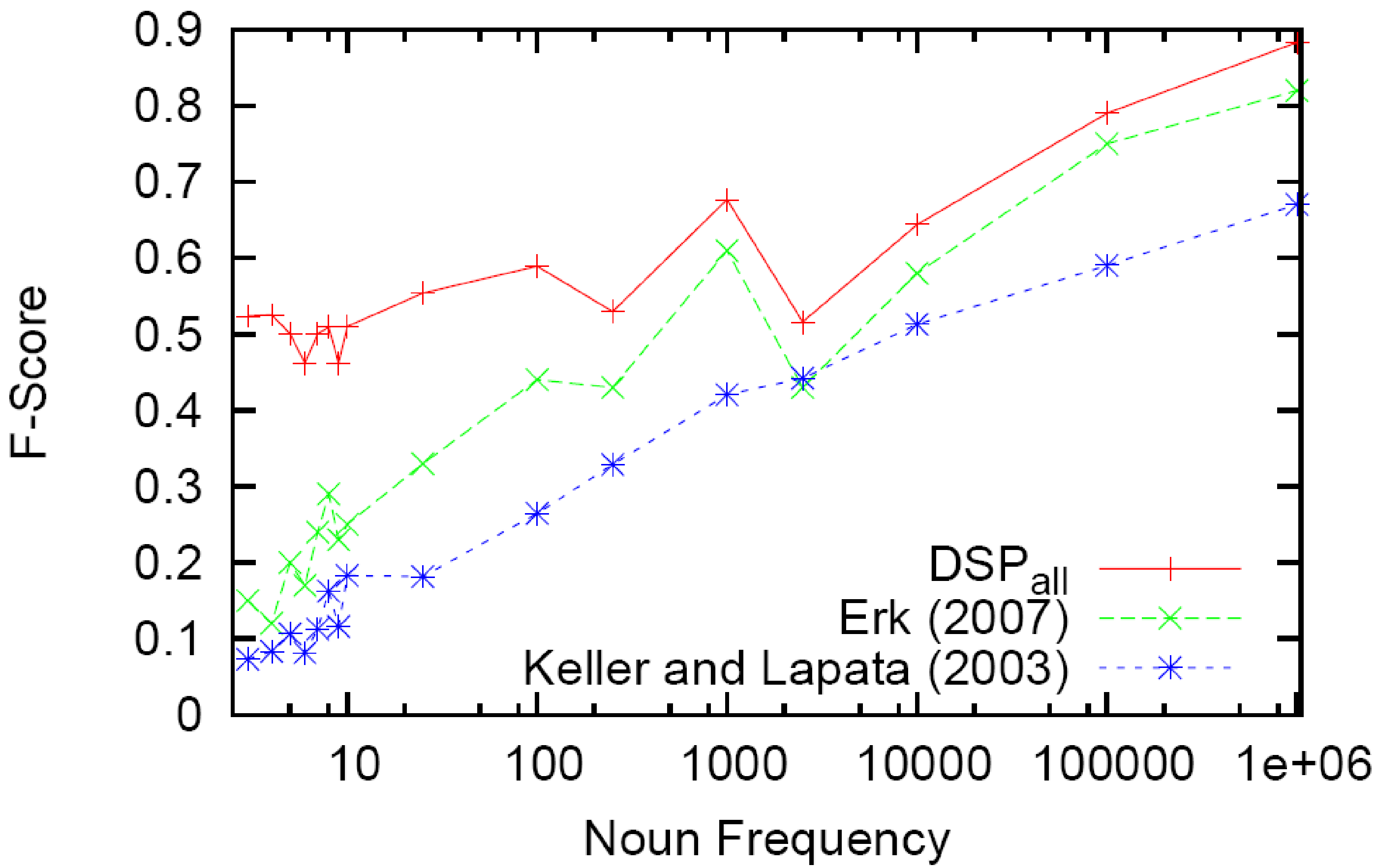
- E.g. Is noun lower-case?

| Verb | Weight |
|---|---|
| become | 0.972 |
| eat | 0.505 |
| embroil | -0.573 |
| accuse | -0.675 |

# Disambiguation Results



# Results by Noun Frequency

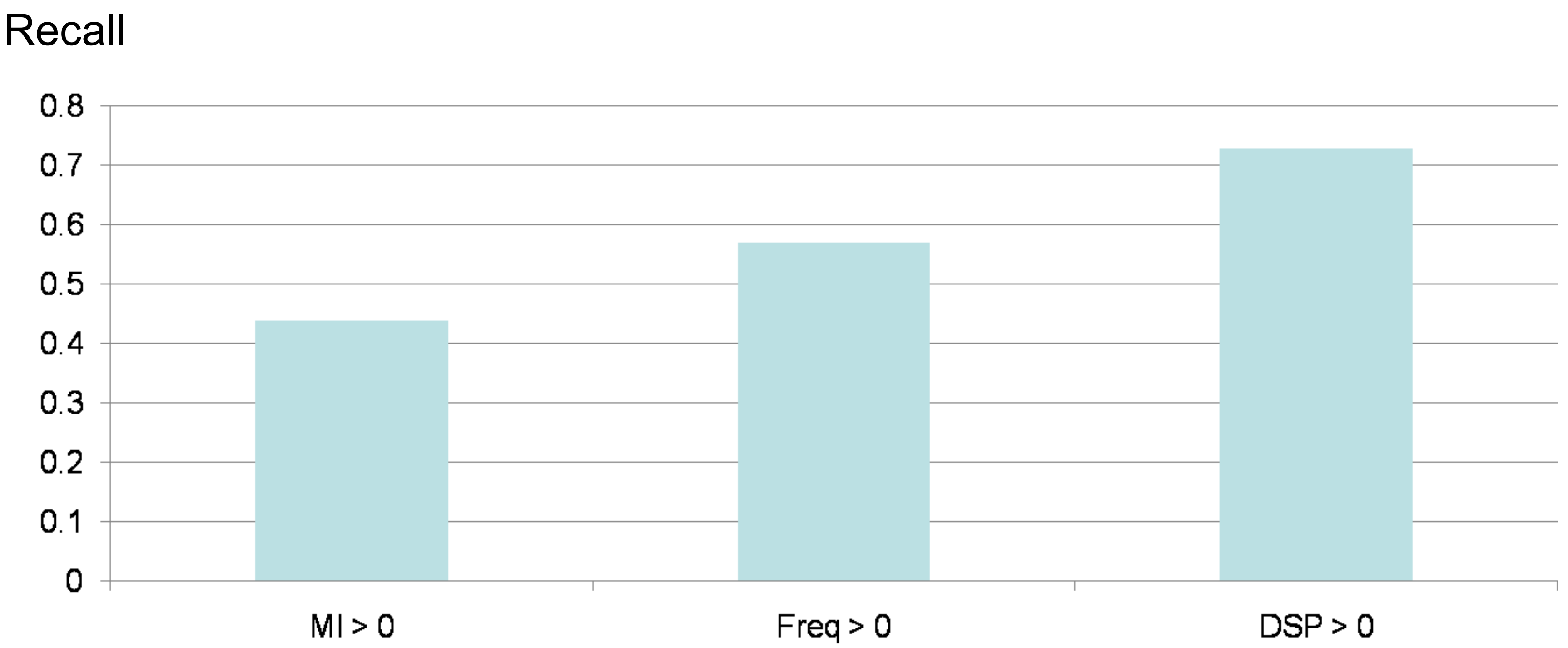


# Human Plausibility

- Human Plausibility – 16 $(v,n)$-pairs used in Resnik (1996, Journal of Cognition)
  - e.g. "repeat comment" / "repeat journal"
- DSP scores plausible ahead of implausible in each case
- MI also scores plausibles highly, but undefined for most negatives

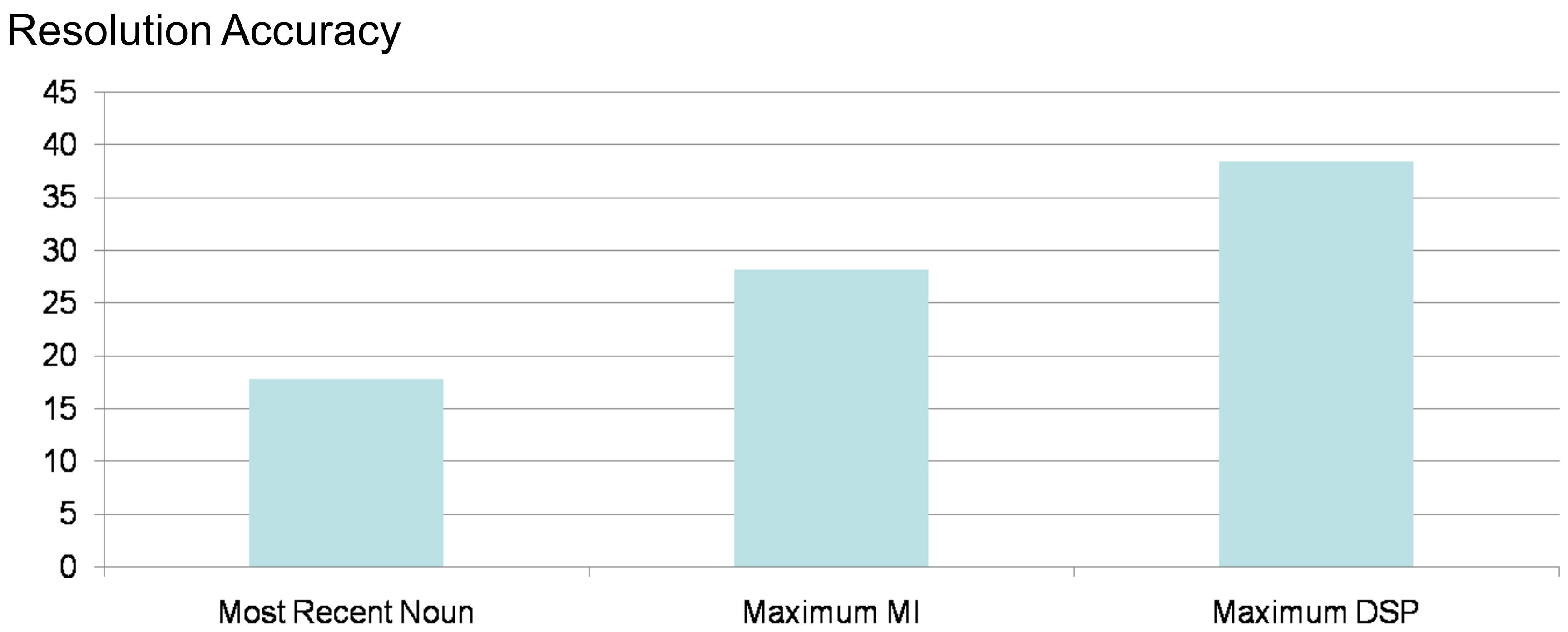
# Unseen Corpus Experiment

- Proportion of accepted $(v,n)$ pairs in San Jose Mercury News Corpus



# Pronoun Resolution

- MUC-7 Corpus
- Object pronouns, e.g. "study it"



# Conclusion

- Discriminative training for selectional preference improves performance
- Unlike most supervised approaches, no labeling or annotation cost
- Allows for combination of arbitrary features
- Yields similar-word list as latent information
- Lots more details in paper

# Future Work

- DSP uses no direct co-occurrence information, could use counts from other corpora (e.g. the web) as a feature in DSP
- Lots of other potential features:
  - *e.g. "Choucroute garnie*" listed as French cuisine on Wikipedia

# Thanks!