Semi-supervised Noun Compound Analysis with Edge and Span Features

11 Dec 2012

MURAWAKI Yūgo KUROHASHI Sadao
Graduate School of Informatics
Kyoto University, Japan

COLING 2012, Mumbai

Semi-supervised Noun Compound Analysis with Edge and Span Features

Semi-supervised

① Noun Compound Analysis

with Edge and Span Features

Semi-supervised

① Noun Compound Analysis

② with Edge and Span Features

③ Semi-supervised

① Noun Compound Analysis

② with Edge and Span Features

③ Semi-supervised

① Noun Compound Analysis

② with Edge and Span Features

Noun Compound Analysis
aka Noun Compound Bracketing

woman aid worker

hydrogen ion exchange

Examples from (Lauer, 1995)

Noun Compound Analysis
aka Noun Compound Bracketing

woman aid worker

N N N

hydrogen ion exchange

N N N

Examples from (Lauer, 1995)

Noun Compound Analysis
aka Noun Compound Bracketing

[woman [aid worker]]

N N N

[[hydrogen ion] exchange]

N N N

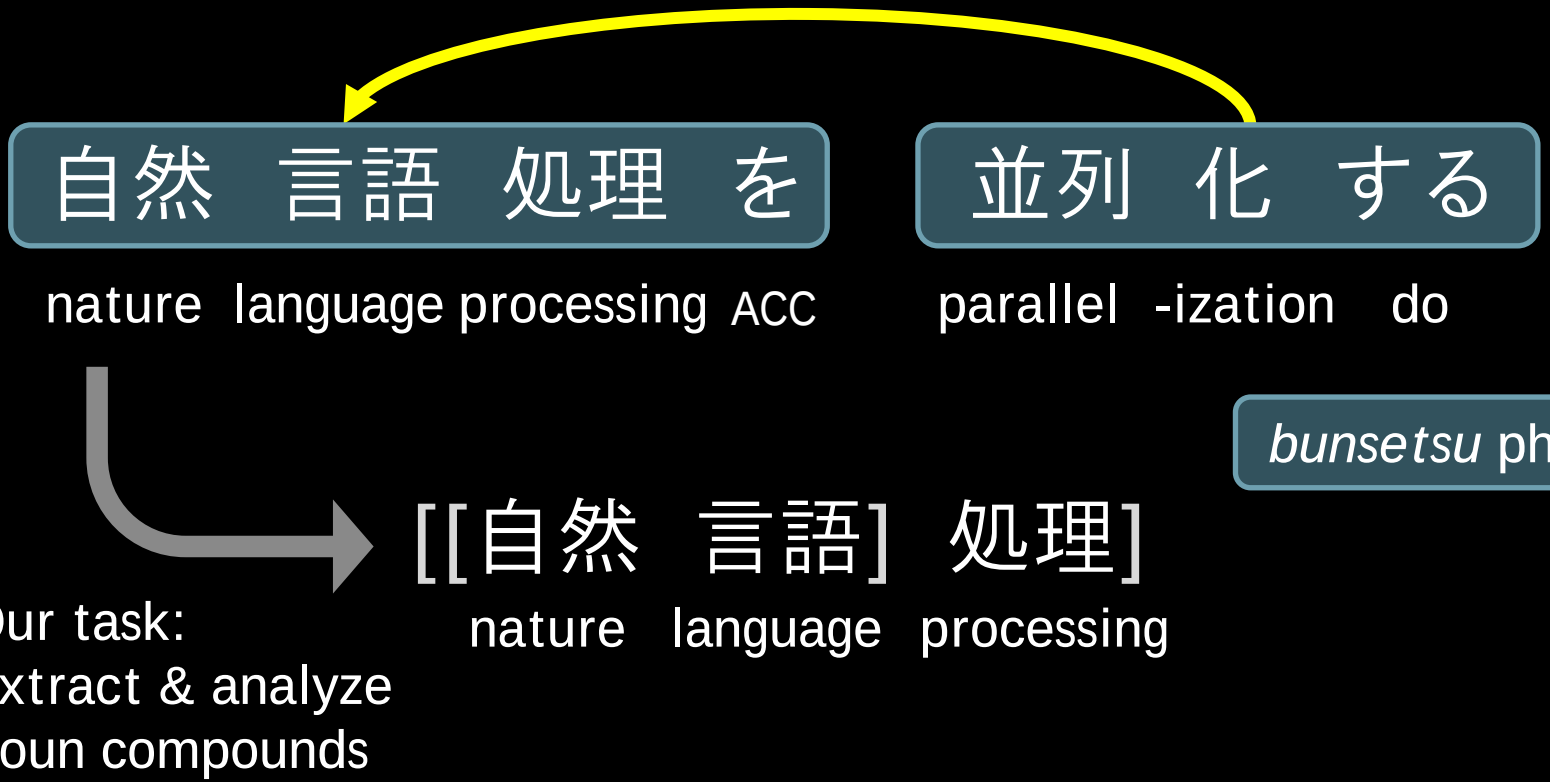
Examples from (Lauer, 1995)

Japanese Noun Compounds 1/2

- Classical *bunsetsu*-based dependency parsing leaves noun compound unanalyzed

Japanese Noun Compounds 1/2

- Classical *bunsetsu*-based dependency parsing leaves noun compound unanalyzed



Japanese Noun Compounds 2/2

自然　言語　処理　を　　並列　化　する

nature　language processing　ACC　　parallel　-ization　do

Japanese Noun Compounds 2/2

- As an agglutinative language, Japanese uses a lot of “glue” markers

Japanese Noun Compounds 2/2

- As an agglutinative language, Japanese uses a lot of “glue” markers
- But none within noun compounds. Why?

nature language processing ACC parallel -ization do

Japanese Noun Compounds 2/2

- As an agglutinative language, Japanese uses a lot of “glue” markers
- But none within noun compounds. Why?

nature language processing ACC parallel -ization do

- And noun compounds can be arbitrarily long

[[[[[第 １７７] 回] 国会] [[国家 [基本 政策]] [委員 会]]] [合同 [審査 会]]]

no. 177 COUNT national diet state base policy member meeting joint review meeting

Representations

Bracketing

[[自然 言語] 処理]

nature language processing

Representations

Bracketing

[[自然 言語] 処理]

nature language processing

Dependency tree

Representations

Bracketing

[[自然 言語] 処理]

nature language processing

One-to-one correspondence

Dependency tree

自然 言語 処理

nature language processing

Representations

Bracketing

[[自然 言語] 処理]

nature language processing

One-to-one correspondence

Dependency tree

We solve this problem

自然 言語 処理

nature language processing

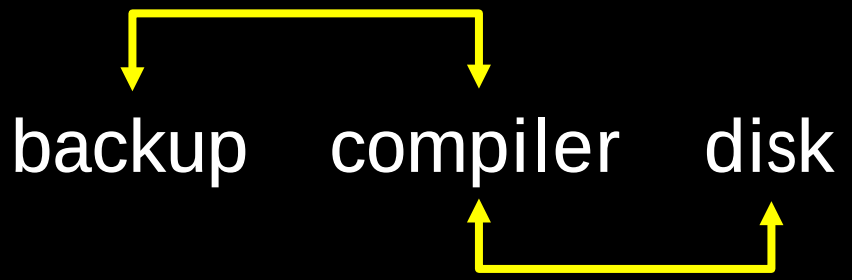
③ Semi-supervised
① Noun Compound Analysis
② with Edge and Span Features

Dependency *beats* Adjacency
for three-word noun compound disambiguation

- Compare two word pairs by lexical association

Adjacency model

(Marcus, 1980)(Liberman+, 1992)
(Pustejovsky+, 1993)(Resnik, 1993)



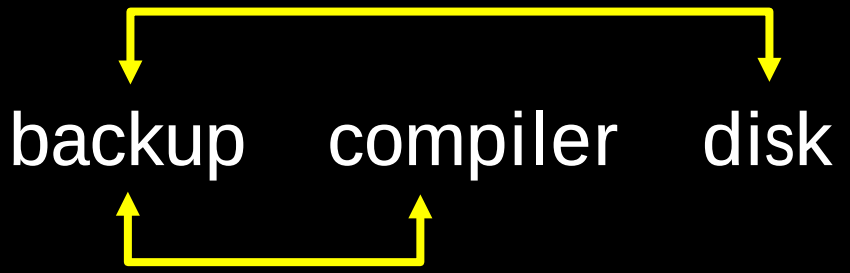
Example from (Lauer, 1995)

Dependency *beats* Adjacency
for three-word noun compound disambiguation

- Compare two word pairs by lexical association

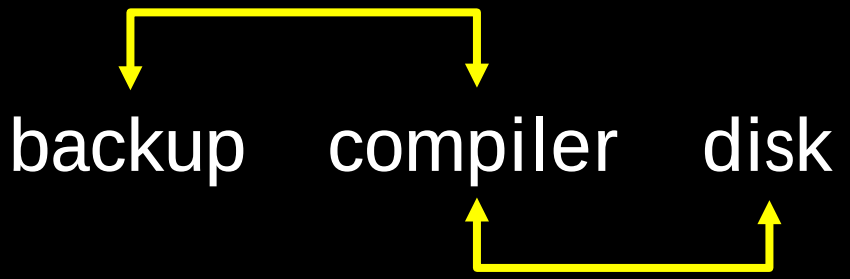
Dependency model

(Lauer, 1995)
and subsequent studies



Adjacency model

(Marcus, 1980)(Liberman+, 1992)
(Pustejovsky+, 1993)(Resnik, 1993)



Example from (Lauer, 1995)

But adjacency seems non-negligible

- Nouns are arranged in a rather fixed order and often, word insertion appears to impair semantic coherence

自然 言語 処理
nature language processing

But adjacency seems non-negligible

- Nouns are arranged in a rather fixed order and often, word insertion appears to impair semantic coherence

Basic idea: Consider *spans* in addition to edges

dependency = edge

adjacency $\xrightarrow[\text{generalize}]{}$ span

Basic idea: Consider *spans* in addition to edges

dependency = edge

adjacency $\xrightarrow[\text{generalize}]{}$ span

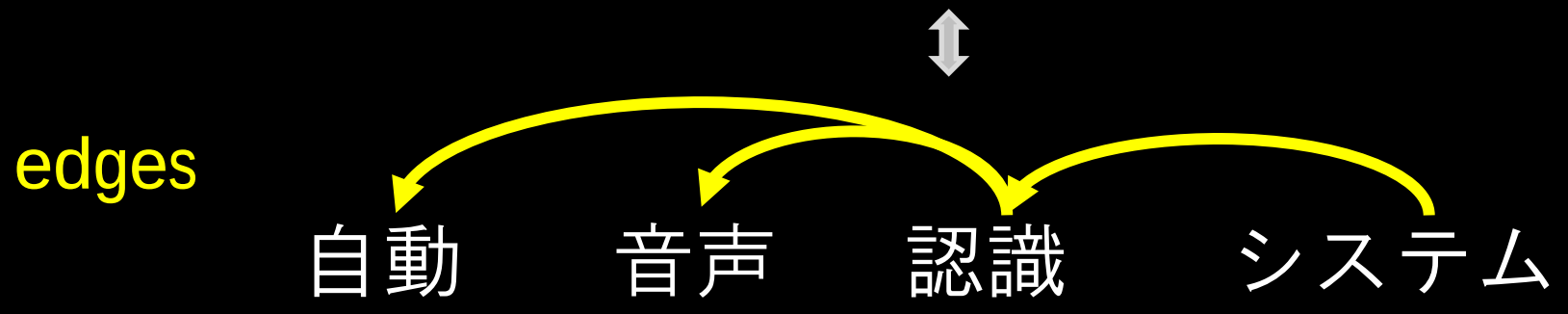
[[自動 [音声 認識]] システム]

automation speech recognition system

Basic idea: Consider *spans* in addition to edges

dependency = edge
adjacency $\xrightarrow[\text{generalize}]{}$ span

[[自動 [音声 認識]] システム]
automation speech recognition system



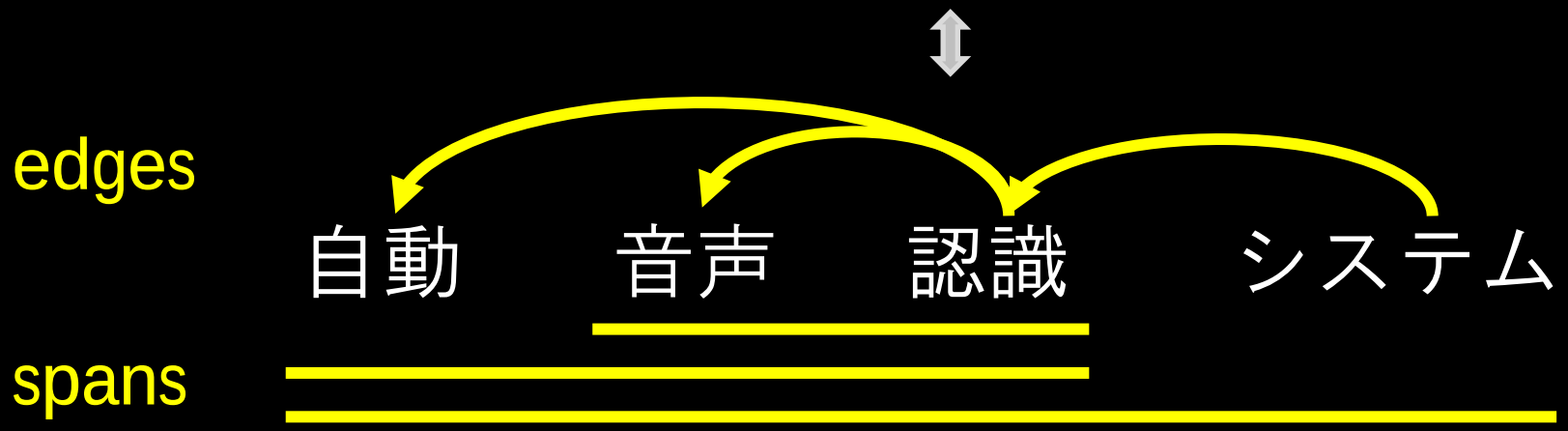
Basic idea: Consider *spans* in addition to edges

dependency = edge

adjacency $\xrightarrow[\text{generalize}]{}$ span

[[自動 [音声 認識]] システム]

automation speech recognition system

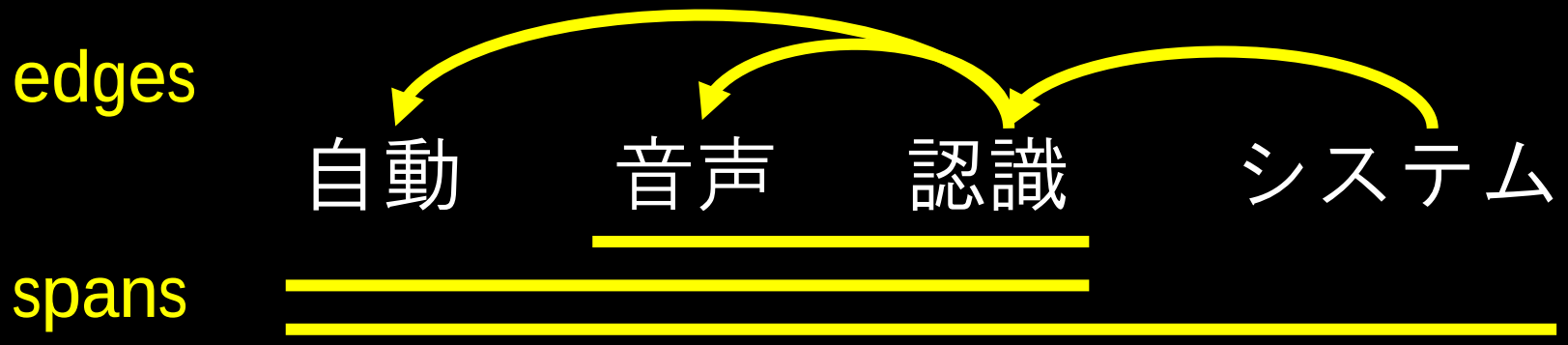


Score-based Parsing

edges

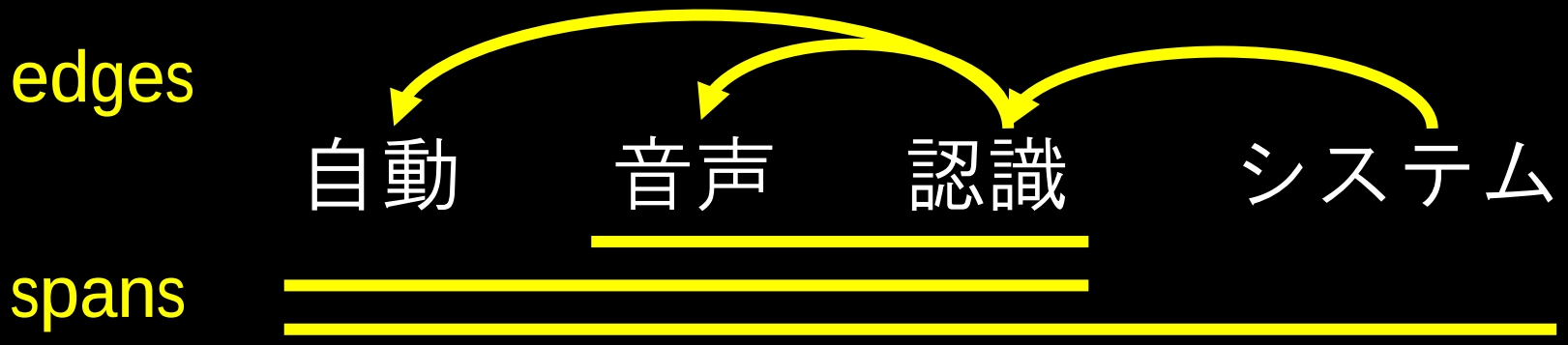
score = edge-score(自動 ← 認識)
+ edge-score(音声 ← 認識)
+ edge-score(認識 ← システム)

Score-based Parsing



score = edge-score(自動 ← 認識)
+ edge-score(音声 ← 認識)
+ edge-score(認識 ← システム)
+ span-score(音声 認識)
+ span-score(自動 音声 認識)
+ span-score(自動 音声 認識 システム)

Score-based Parsing



score = edge-score(自動 ← 認識)
+ edge-score(音声 ← 認識)
+ edge-score(認識 ← システム)
+ span-score(音声 認識)
+ span-score(自動 音声 認識)
+ span-score(自動 音声 認識 システム)

CKY parsing
still applicable

③ Semi-supervised

① Noun Compound Analysis

② with Edge and Span Features

Linear Discriminative Parser

score = edge-score(自動 ← 認識)
+ edge-score(音声 ← 認識)
+ edge-score(認識 ← システム)
+ span-score(音声 認識)
+ span-score(自動 音声 認識)
+ span-score(自動 音声 認識 システム)

Linear Discriminative Parser

$w_{edge} \cdot \phi_{edge}$(自動 ← 認識)

score = edge-score(自動 ← 認識)
+ edge-score(音声 ← 認識)
+ edge-score(認識 ← システム)
+ span-score(音声 認識)
+ span-score(自動 音声 認識)
+ span-score(自動 音声 認識 システム)

Linear Discriminative Parser

$w_{edge} \cdot \phi_{edge}$(自動 ← 認識)

$w_{span} \cdot \phi_{span}$(音声 認識)

score = edge-score(自動 ← 認識)
+ edge-score(音声 ← 認識)
+ edge-score(認識 ← システム)
+ span-score(音声 認識)
+ span-score(自動 音声 認識)
+ span-score(自動 音声 認識 システム)

Linear Discriminative Parser

$w_{edge} \cdot \phi_{edge}$(自動 ← 認識)

$w_{span} \cdot \phi_{span}$(音声 認識)

We need to
1. design the feature set
2. tune the weight using training data

score = edge-score(自動 ← 認識)
+ edge-score(音声 ← 認識)
+ edge-score(認識 ← システム)
+ span-score(音声 認識)
+ span-score(自動 音声 認識)
+ span-score(自動 音声 認識 システム)

Supervised Training

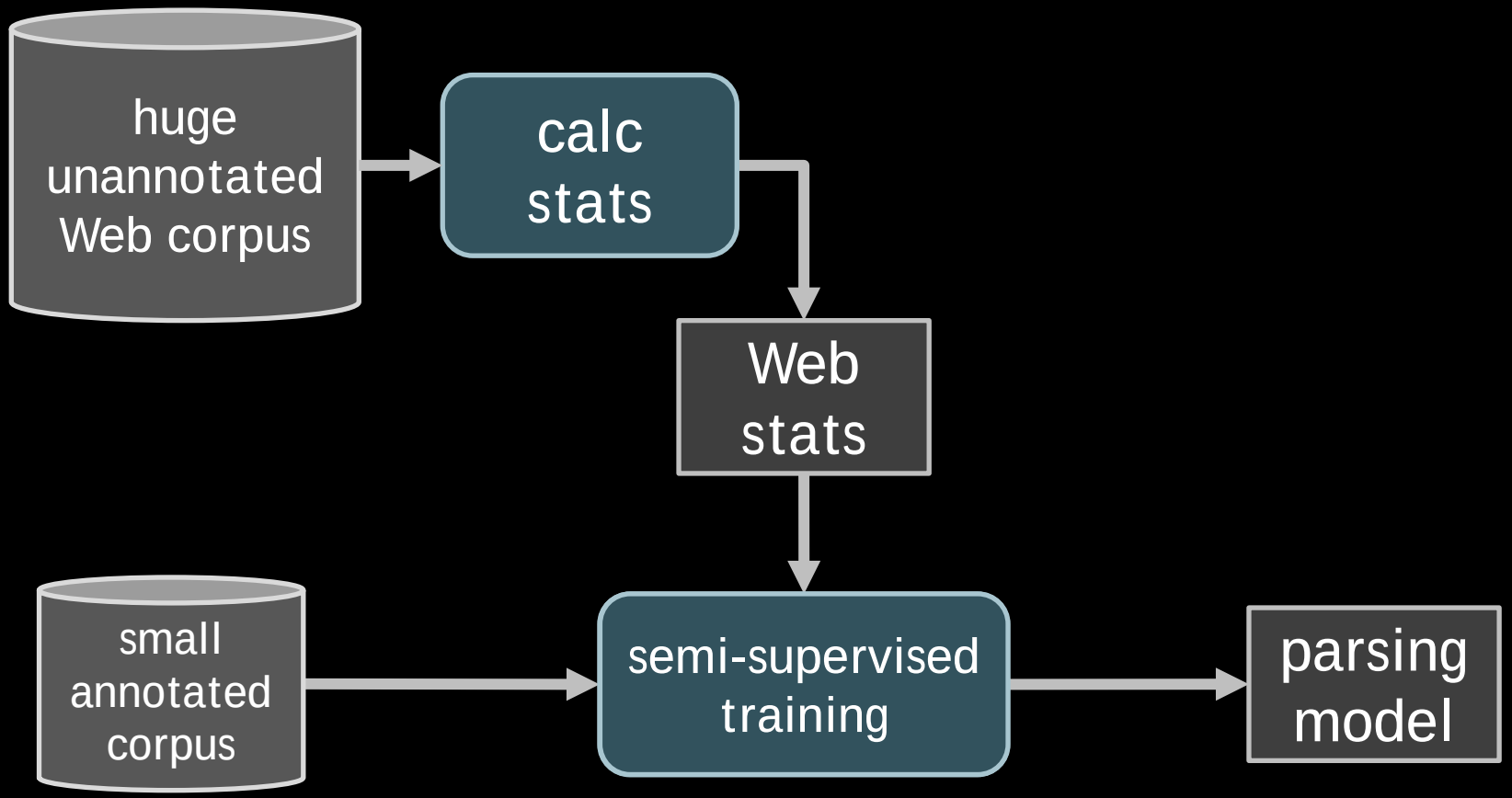
Base Features (Edge)

Feature templates, binary-valued:

- <distance>
- <child, distance>
- <head, distance>
- <child, head>
- <child, head, distance>

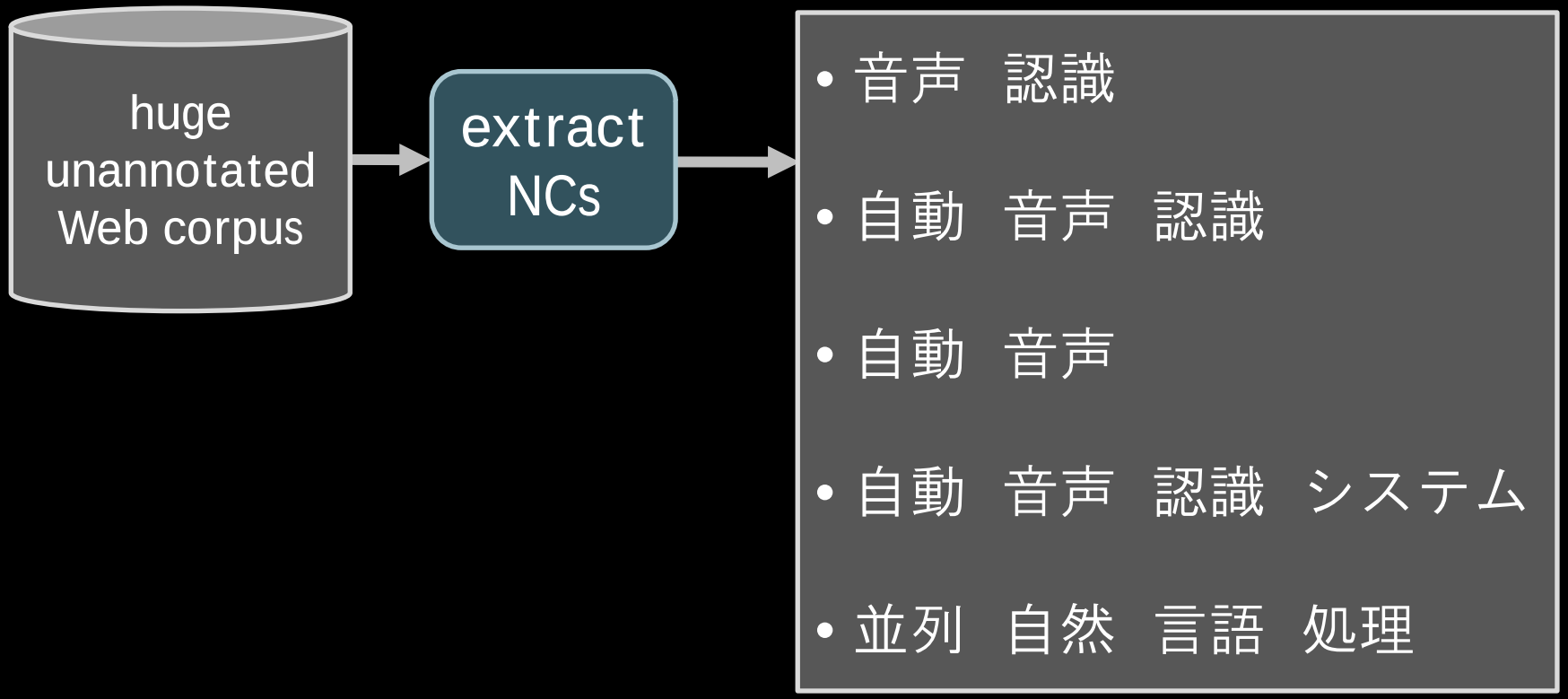
Semi-supervised Training

Semi-supervised Training

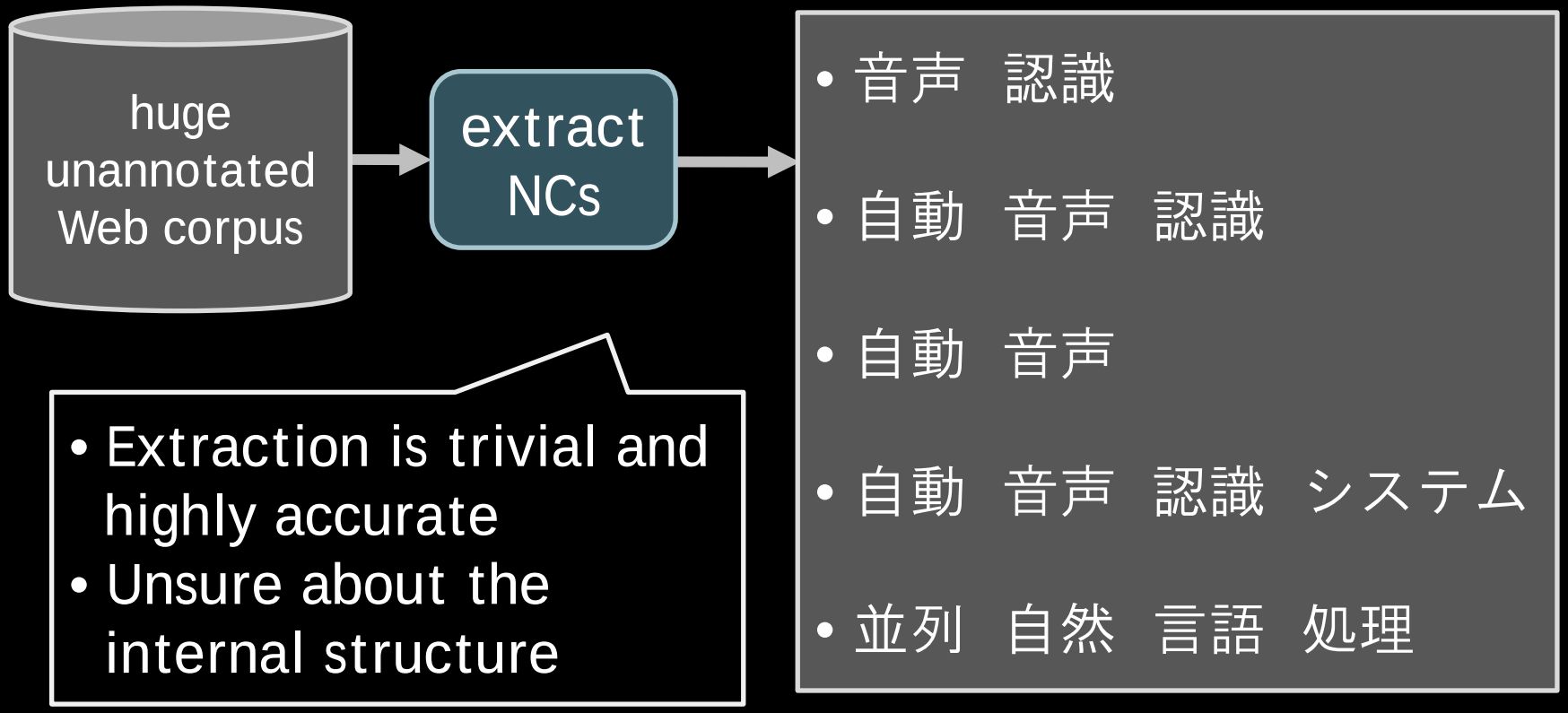


Noun Compound Extraction

Noun Compound Extraction



Noun Compound Extraction



Web Edge Features

- 音声　認識
- 自動　音声　認識
- 自動　音声
- 自動　音声　認識　システム
- 並列　自然　言語　処理

(Lauer, 1995)(Lapata+, 2004)(Nakov+, 2005)(Bergsma+, 2010)(Pitler+, 2010)

Web Edge Features

1. Two-word NCs (log-counts)

2.

- 音声　認識
- 自動　音声　認識
- 自動　音声
- 自動　音声　認識　システム
- 並列　自然　言語　処理

(Lauer, 1995)(Lapata+, 2004)(Nakov+, 2005)(Bergsma+, 2010)(Pitler+, 2010)

Web Edge Features

1. Two-word NCs (log-counts)

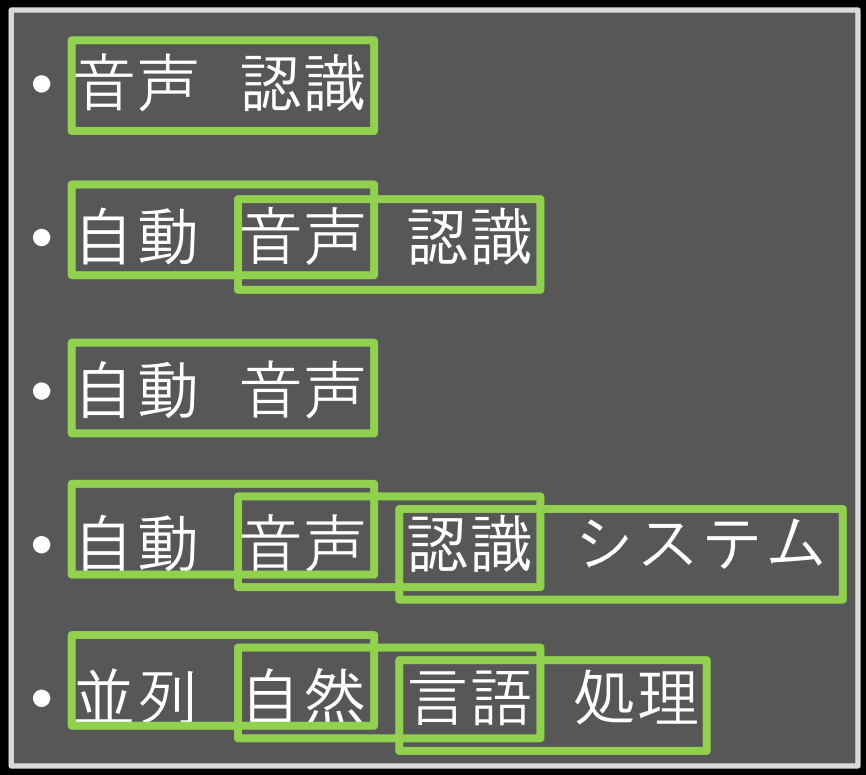
2. Last two words (log counts)

- 音声 認識
- 自動 音声 認識
- 自動 音声
- 自動 音声 認識 システム
- 並列 自然 言語 処理

(Lauer, 1995)(Lapata+, 2004)(Nakov+, 2005)(Bergsma+, 2010)(Pitler+, 2010)

Web Edge Features

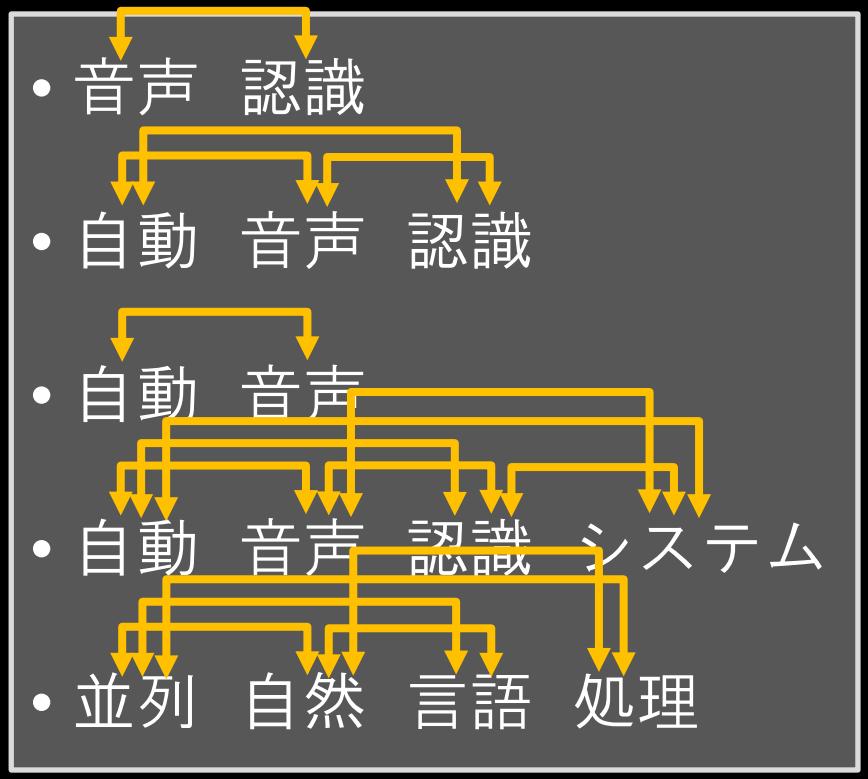
1. Two-word NCs (log-counts)
2. Last two words (log counts)
3. Bigram ($\chi 2$)



(Lauer, 1995)(Lapata+, 2004)(Nakov+, 2005)(Bergsma+, 2010)(Pitler+, 2010)

Web Edge Features

1. Two-word NCs (log-counts)
2. Last two words (log counts)
3. Bigram ($\chi 2$)
4. Co-occurrence (PMI)



(Lauer, 1995)(Lapata+, 2004)(Nakov+, 2005)(Bergsma+, 2010)(Pitler+, 2010)

Web Edge Features

1. Two-word NCs (log-counts)
2. Last two words (log counts)
3. Bigram ($\chi 2$)
4. Co-occurrence (PMI)

Perform best in the experiments

- 自動 音声 認識
- 自動 音声
- 自動 音声 認識 システム
- 並列 自然 言語 処理

(Lauer, 1995)(Lapata+, 2004)(Nakov+, 2005)(Bergsma+, 2010)(Pitler+, 2010)

Web Span Feature

Log-count as a new feature

- 音声　認識
- 自動　音声　認識
- 自動　音声
- 自動　音声　認識　システム
- 並列　自然　言語　処理

Web Span Feature

Log-count as a new feature

- 音声　認識
- 自動　音声　認識
- 自動　音声
- 自動　音声　認識　システム
- 並列　自然　言語　処理

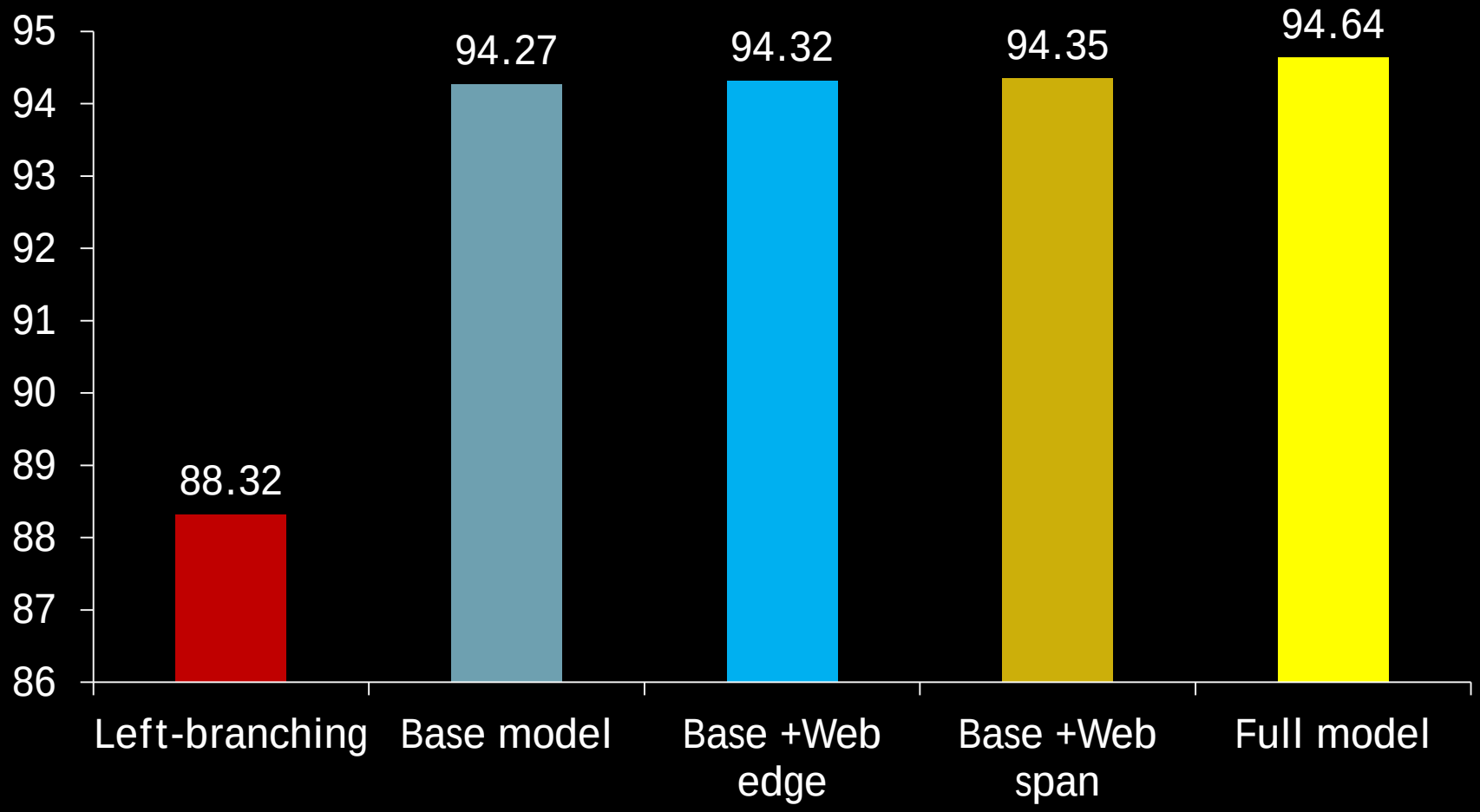
Experimental Settings 1/2

1. In-domain: Abstracts from computer science-related papers
   - Annotated 3,100 noun compounds
   - 5-fold cross-validation
2. Out-of-domain: Abstracts from material science-related papers
   - Annotated 1,000 noun compounds
   - Tested the model trained on in-domain data

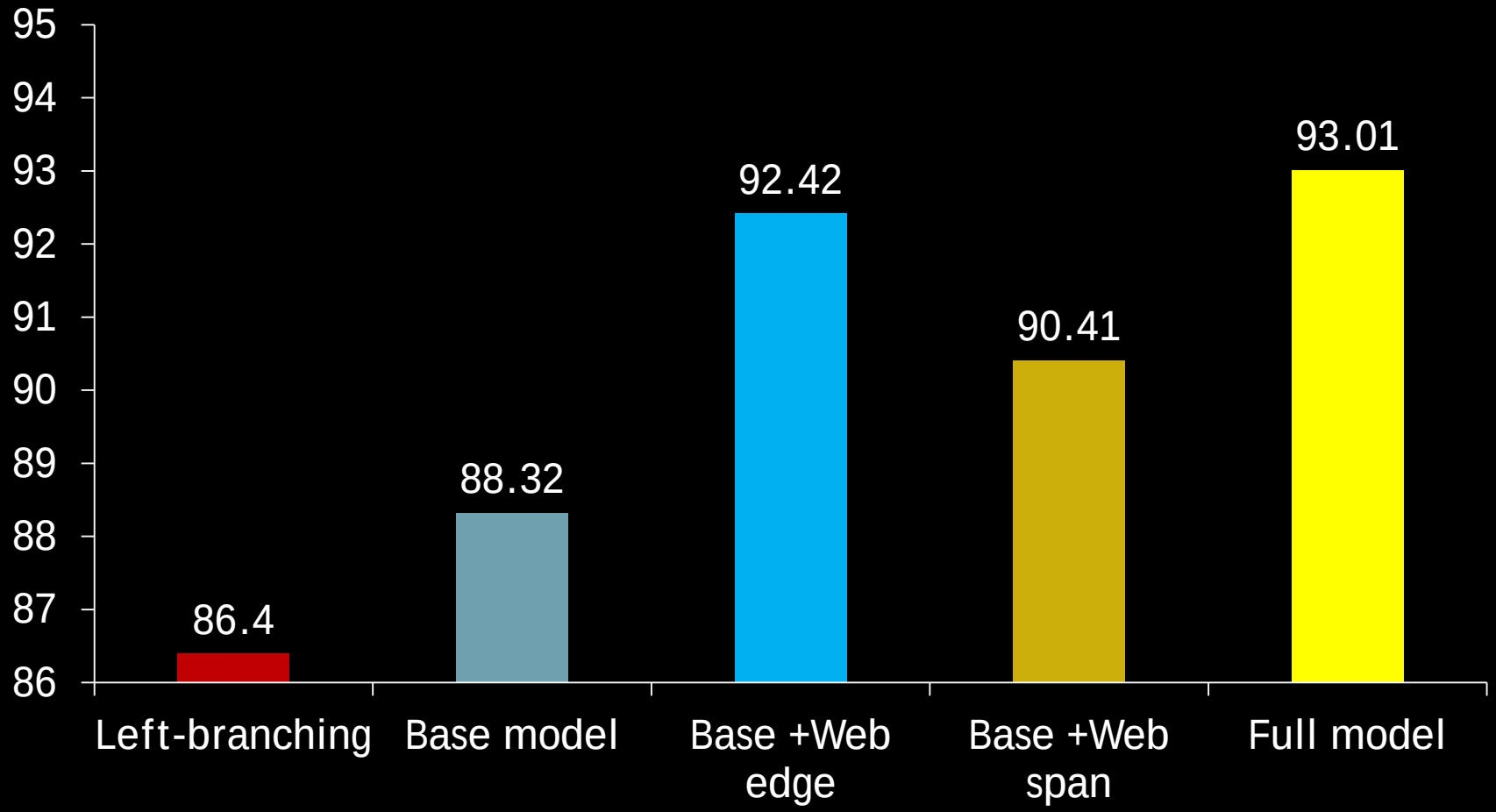
Experimental Settings 2/2

- Web corpus (unannotated)
  - 70 million pages
- Model: Passive-Aggressive training
  - PA-I, prediction-based updates, weight averaging
  - 20 iterations over training data

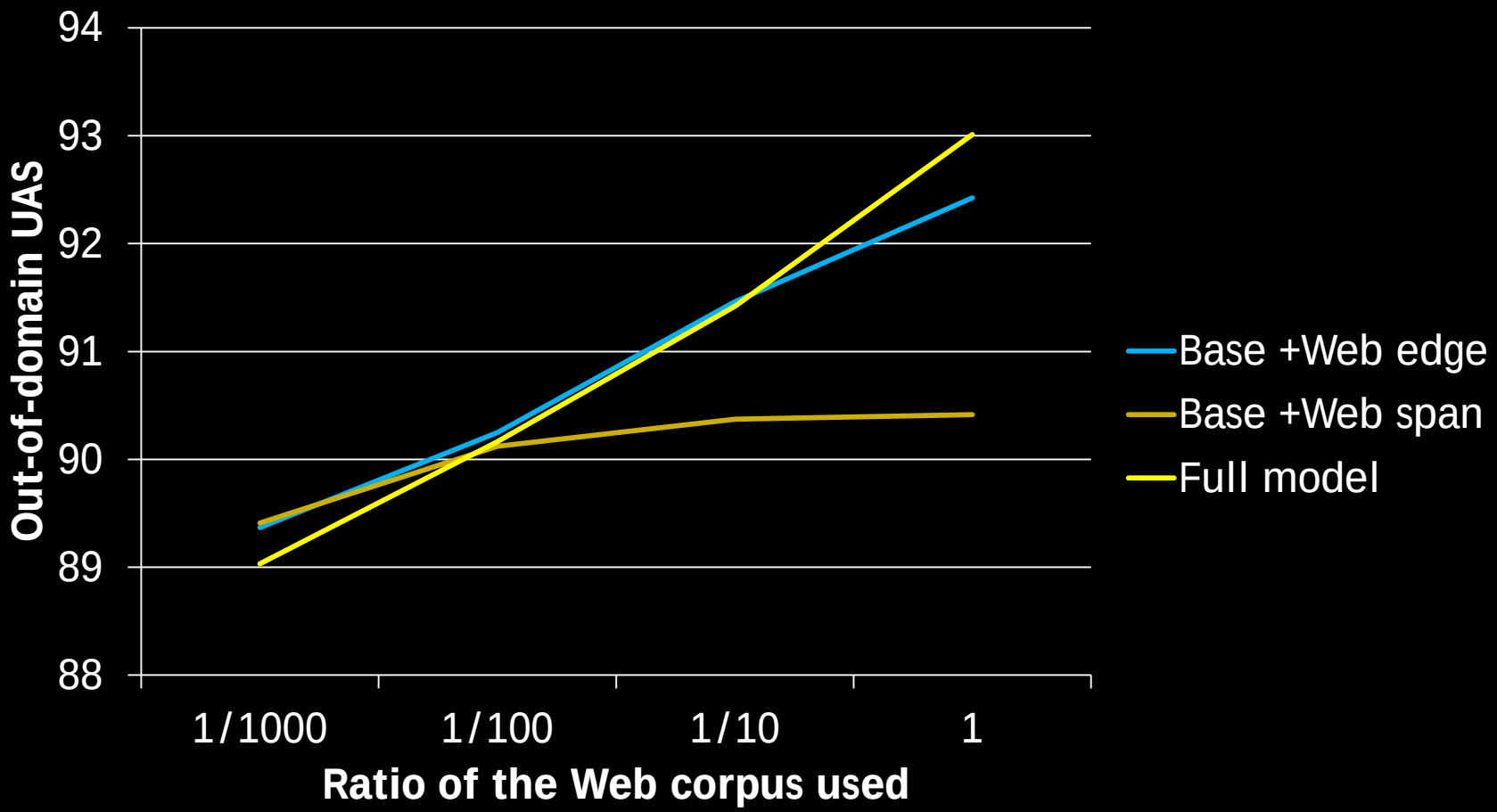
Unlabeled Attachment Scores (in-domain)



Unlabeled Attachment Scores (out-of-domain)



Accuracy in relation to corpus size



Conclusion

- Proposed to use spans in addition to edges in noun compound analysis
- Simple but effective Web statistics incorporated into a discriminative parser
- Spans are shown to have a complementary relationship with edges