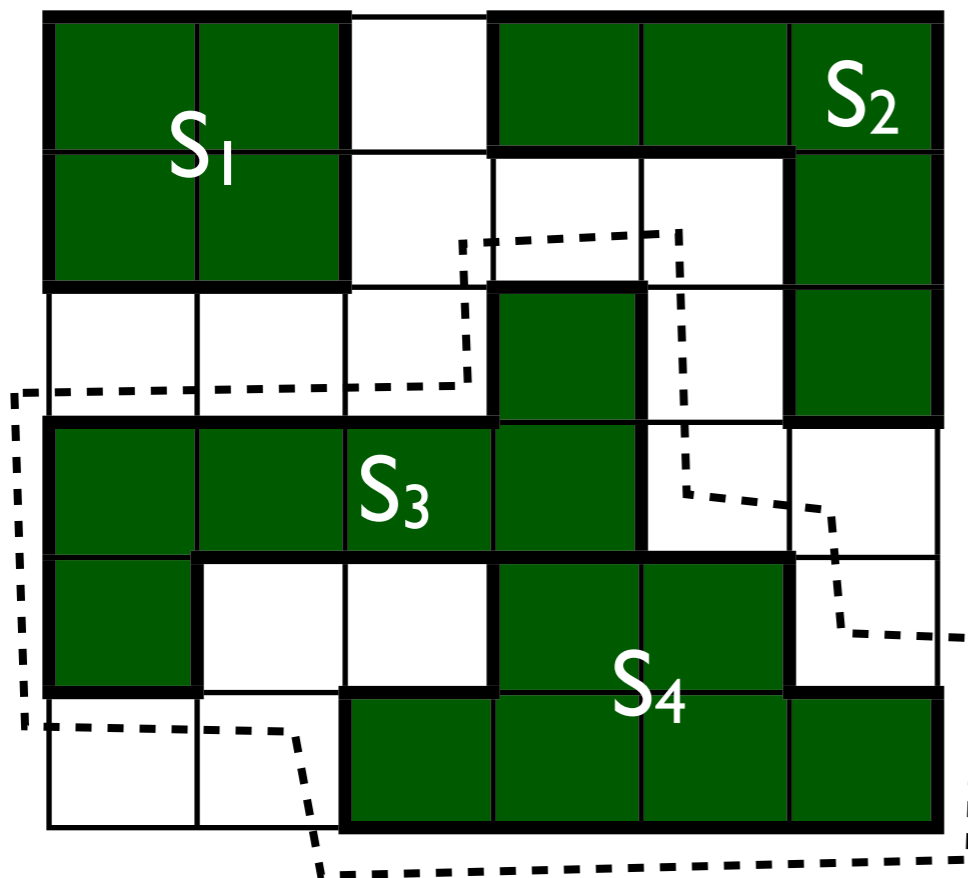


ICSI and UTD at TAC 2009

Dan Gillick
Benoit Favre
Dilek Hakkani-Tur
Berndt Bohnet
Yang Liu
Shasha Xie



Maximum coverage model



Each grid square is a concept

Each sentence contains some concepts

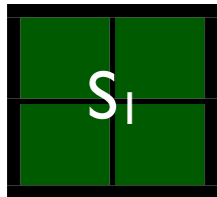
Objective: maximum concept coverage

Subject to: summary length constraint

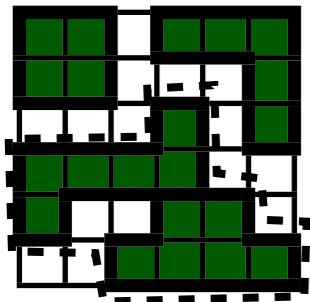
TAC 2008 recap



Concepts are word bigrams valued by frequency



Sentences contain concepts



Summaries are valued by the sum of their unique concept values

$$\begin{aligned}
 &\text{Maximize: } \sum_i w_i c_i \\
 &\text{Subject to: } \sum_j l_j s_j \leq L \\
 &\quad s_j \text{Occ}_{ij} \leq c_i, \quad \forall i, j \\
 &\quad \sum_j s_j \text{Occ}_{ij} \geq c_i \quad \forall i \\
 &\quad c_i \in \{0, 1\} \quad \forall i \\
 &\quad s_j \in \{0, 1\} \quad \forall j
 \end{aligned}$$

Integer Linear Programming solution

Outline

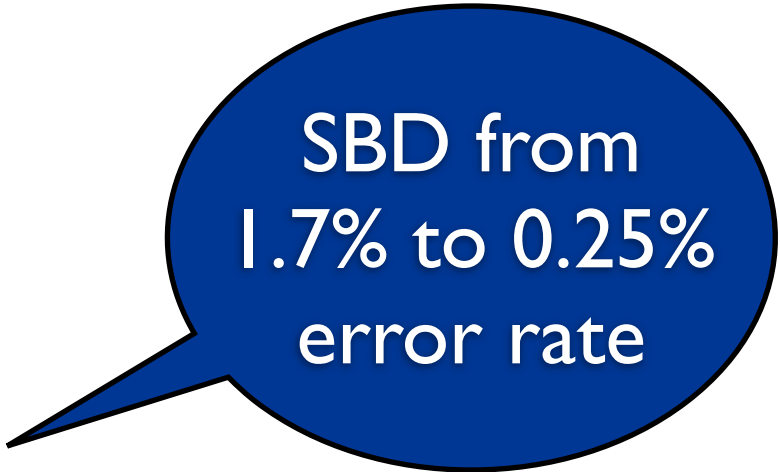
- **Improve linguistic quality**
 - Fix sentence boundary detection
 - Fix unclear references
 - Fix relative time/date

- **Improve content**
 - Use sentence position

Outline

- Improve linguistic quality
 - Fix sentence boundary detection
 - Fix unclear references
 - Fix relative time/date

- Improve content
 - Use sentence position



SBD from
1.7% to 0.25%
error rate

Outline

- Improve linguistic quality
 - Fix sentence boundary detection
 - Fix unclear references
 - Fix relative time/date
- Improve content
 - Use sentence position

SBD from
1.7% to 0.25%
error rate

Safely prune
30% of
sentences

Outline

- Improve linguistic quality
 - Fix sentence boundary detection
 - Fix unclear references
 - Fix relative time/date
- Improve content
 - Use sentence position

SBD from
1.7% to 0.25%
error rate

Safely prune
30% of
sentences

Making
sentence
compression
work

Outline

- Improve linguistic quality
 - Fix sentence boundary detection
 - Fix unclear references
 - Fix relative time/date
- Improve content
 - Use sentence position

SBD from
1.7% to 0.25%
error rate

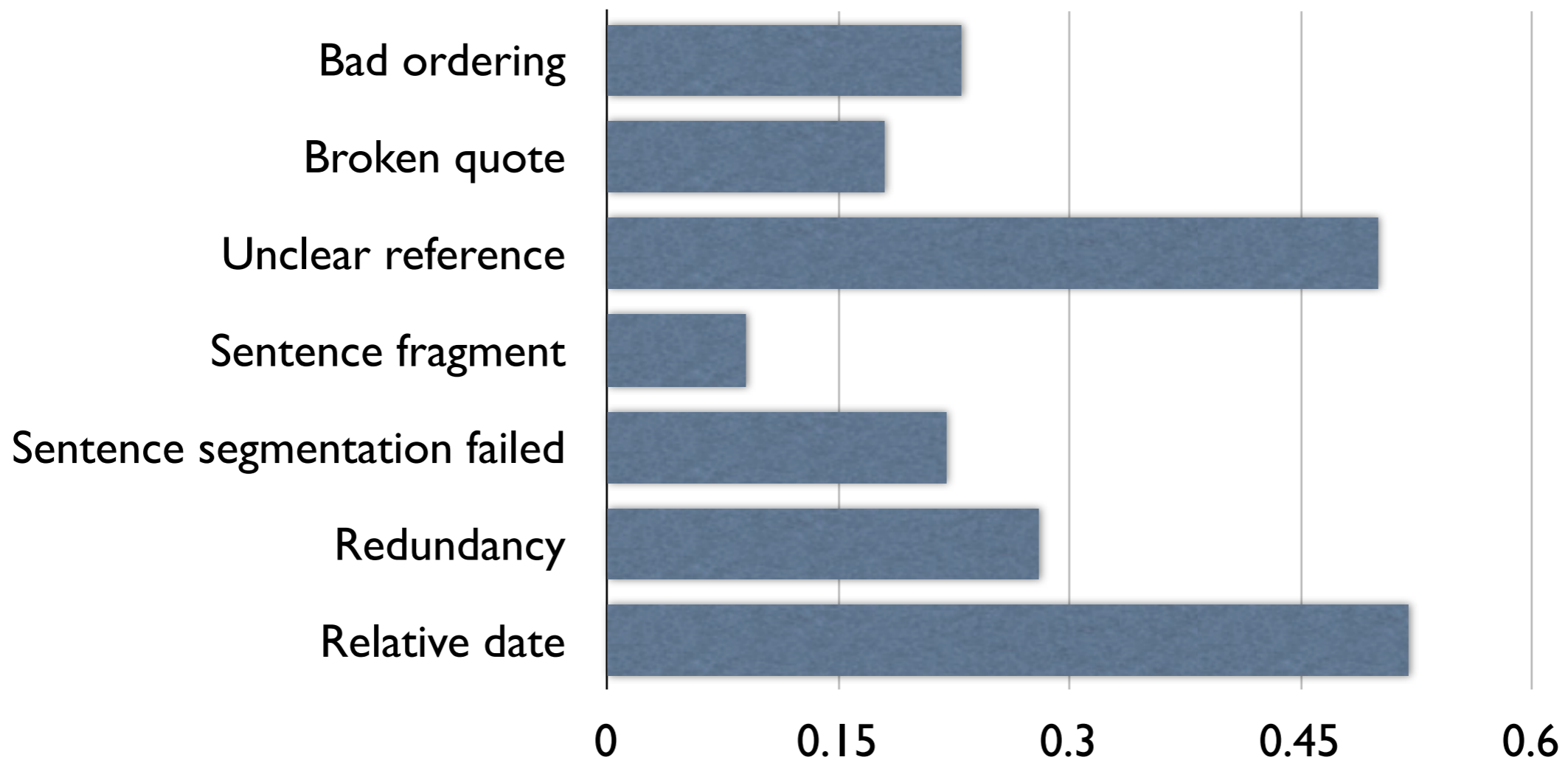
Safely prune
30% of
sentences

Making
sentence
compression
work

Improve
ROUGE by 20%
with one line of
code

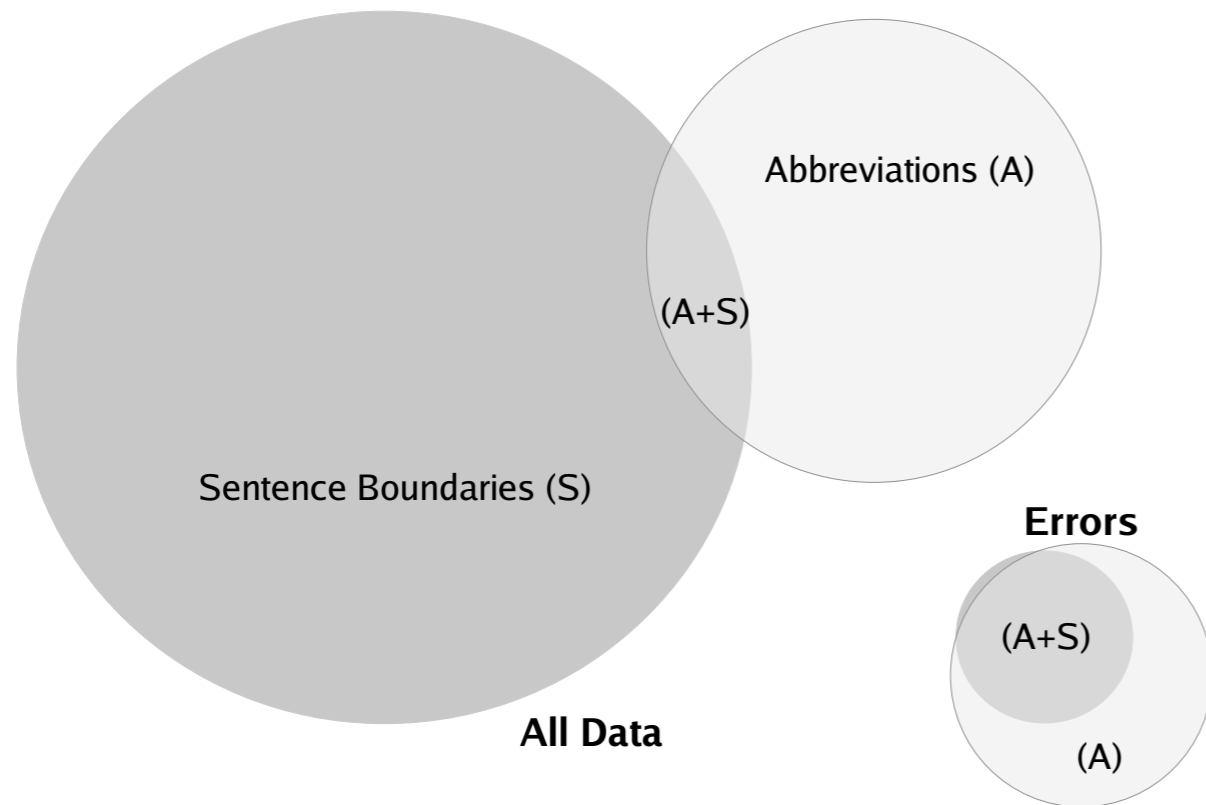
TAC 2008 error analysis

culprits in low-scoring summaries



Fixing sentence segmentation

Wall Street Journal Corpus

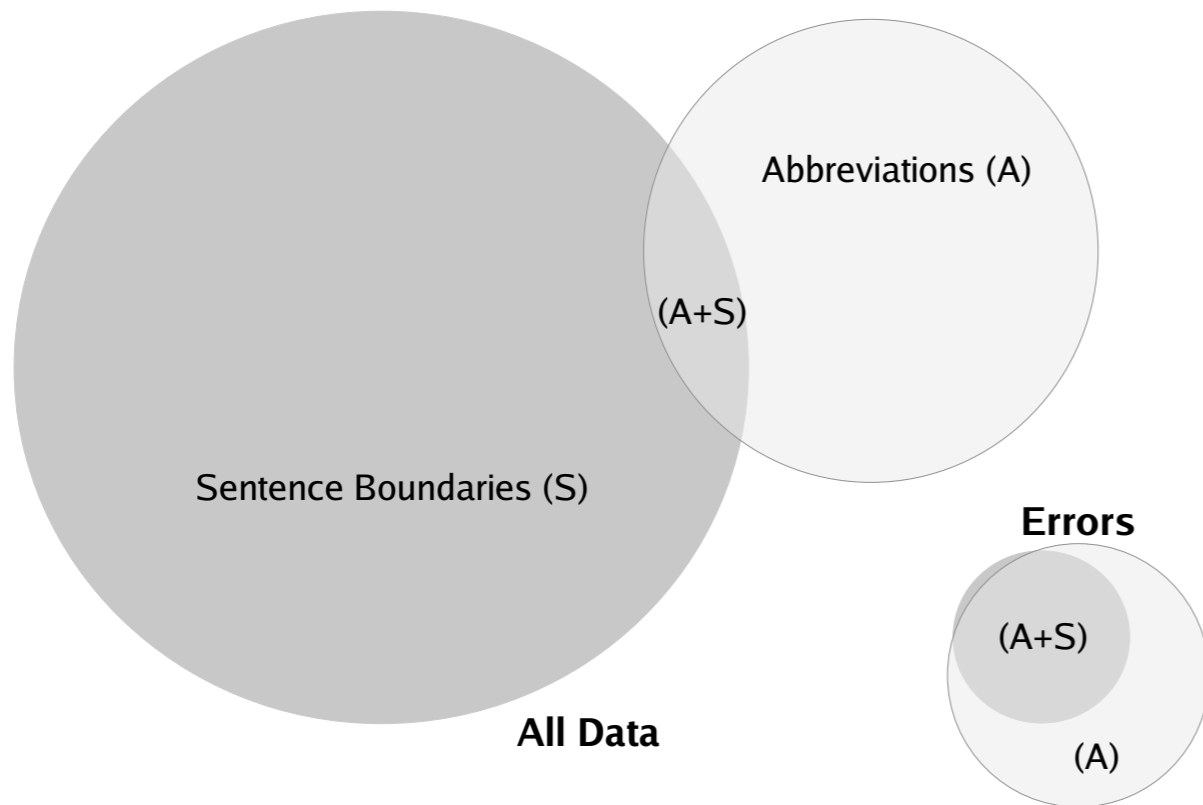


... outside the **U.S.**
Amoco already has
shed such operations ...

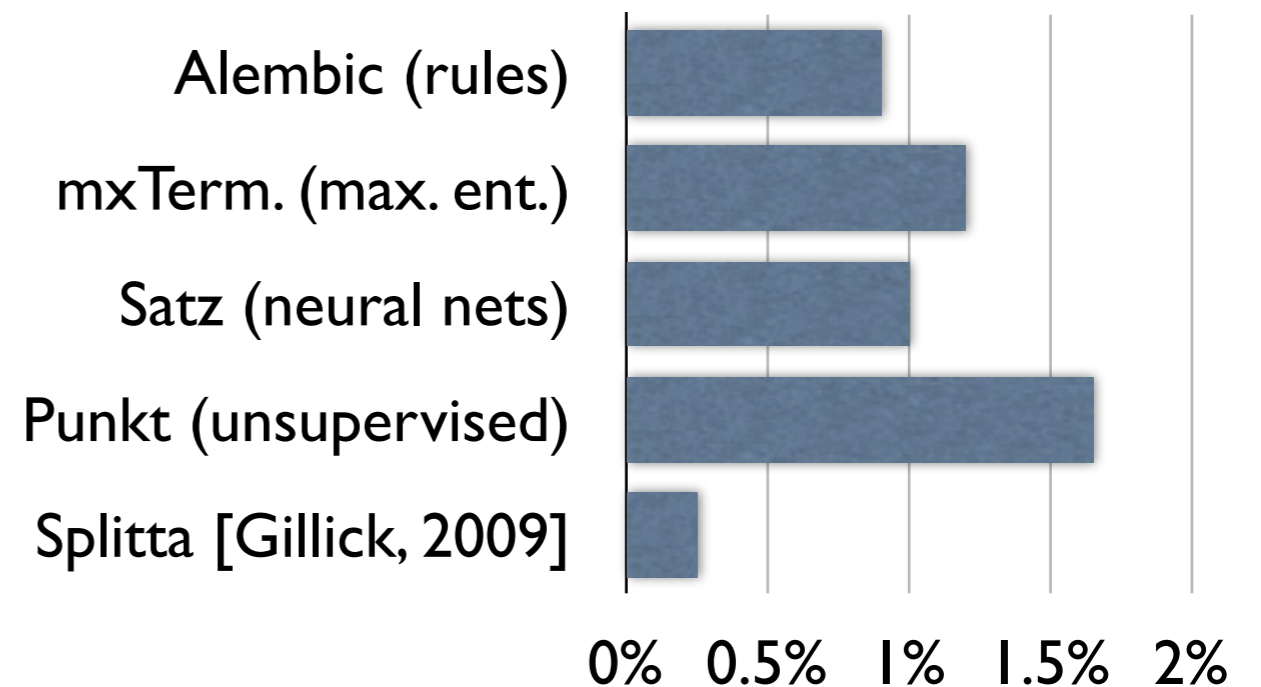
... news that the **U.K.**
Panel on Takeover and
Mergers ...

Fixing sentence segmentation

Wall Street Journal Corpus



SBD error rates on WSJ



... outside the **U.S.**
Amoco already has
 shed such operations ...

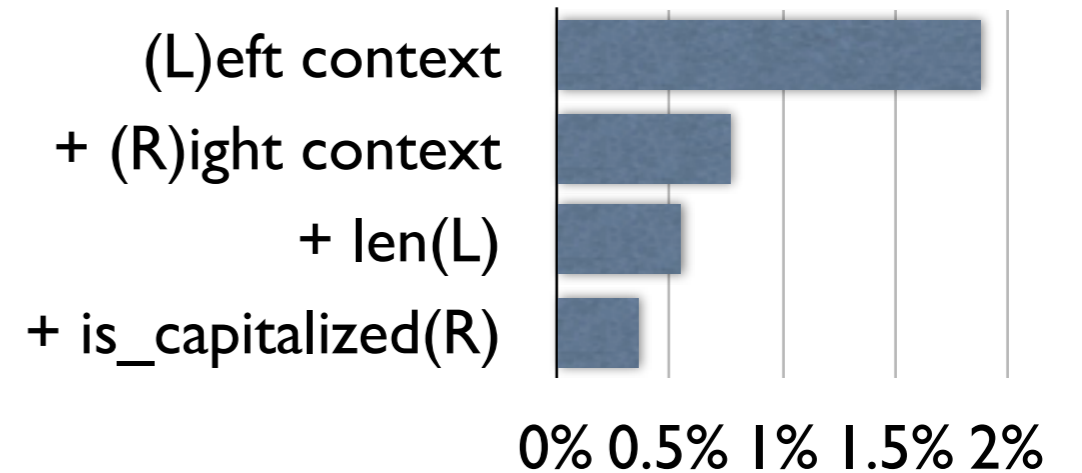
... news that the **U.K.**
Panel on Takeover and
 Mergers ...

Fixing sentence segmentation

Key innovations:

- Local context (one word on each side of the period) is sufficient.

WSJ test error rates

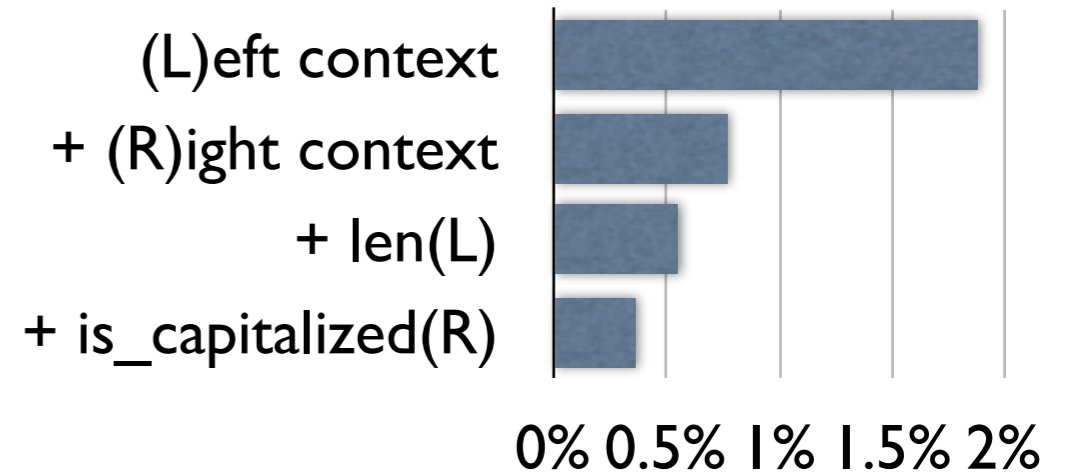


Fixing sentence segmentation

Key innovations:

- Local context (one word on each side of the period) is sufficient.
- Using an abbreviation list is too coarse.

WSJ test error rates



Abbr.	Ends sentence	Total	Ratio
Inc.	109	683	0.16
Co.	80	566	0.14
Corp.	67	699	0.10
U.S.	45	800	0.06
Calif.	24	86	0.28
Ltd.	23	112	0.21

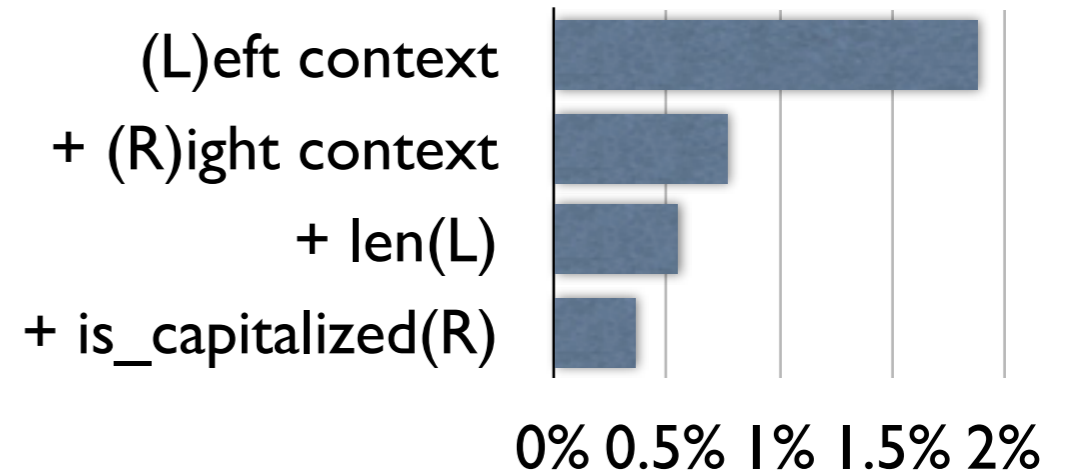
80% of sentence-ending abbrs. in the WSJ test set

Fixing sentence segmentation

Key innovations:

- Local context (one word on each side of the period) is sufficient.
- Using an abbreviation list is too coarse.

WSJ test error rates



Abbr.	Ends sentence	Total	Ratio
Inc.	109	683	0.16
Co.	80	566	0.14
Corp.	67	699	0.10
U.S.	45	800	0.06
Calif.	24	86	0.28
Ltd.	23	112	0.21

80% of sentence-ending abbrs. in the WSJ test set

Download: code.google.com/p/splitta

Fixing unclear references

Idea: prune sentences with unresolved pronouns

1. Tag text to find pronouns
2. Extract features from a parser
3. Train an AdaBoost classifier using OntoNotes data
4. Identify posterior threshold with maximum F-score
5. Prune sentence if one pronoun is classified as unresolved

Fixing unclear references

Idea: prune sentences with unresolved pronouns

1. Tag text to find pronouns
2. Extract features from a parser
3. Train an AdaBoost classifier using OntoNotes data
4. Identify posterior threshold with maximum F-score
5. Prune sentence if one pronoun is classified as unresolved

- 89% F-score on our test set
- **30%** of sentences pruned
- Marginal decline in ROUGE

Input	R-2 (A)	R-2 (B)
unfiltered	0.1990	0.1987
filtered	0.1942	0.1960

Max. ROUGE experiment

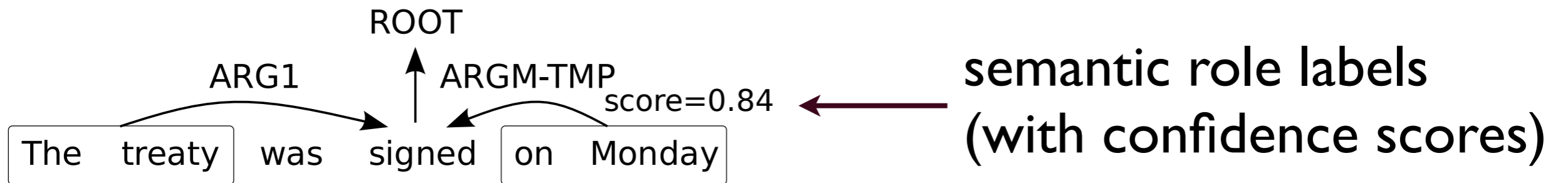
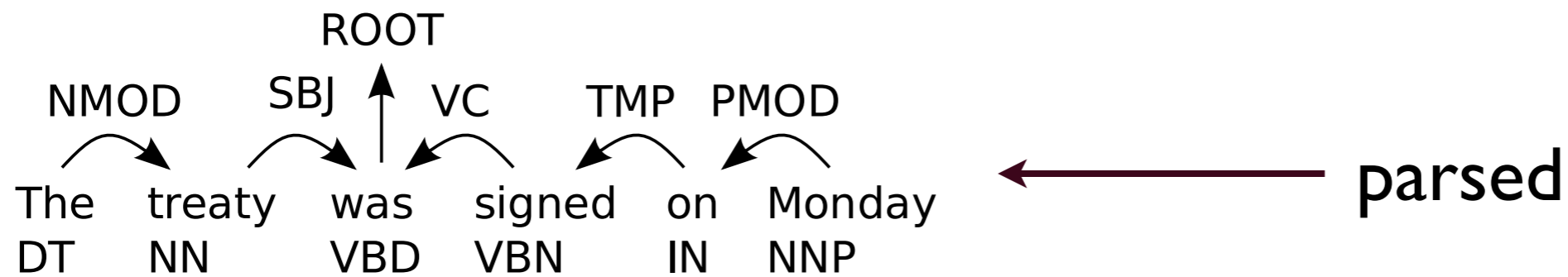
Fixing relative dates

- The treaty was signed on Monday.
- The treaty was signed Monday morning.
- The treaty was signed Monday before noon.
- The treaty signing took all of Monday.
- Last Monday, the treaty was signed.
- It was Monday when the treaty was finally signed.

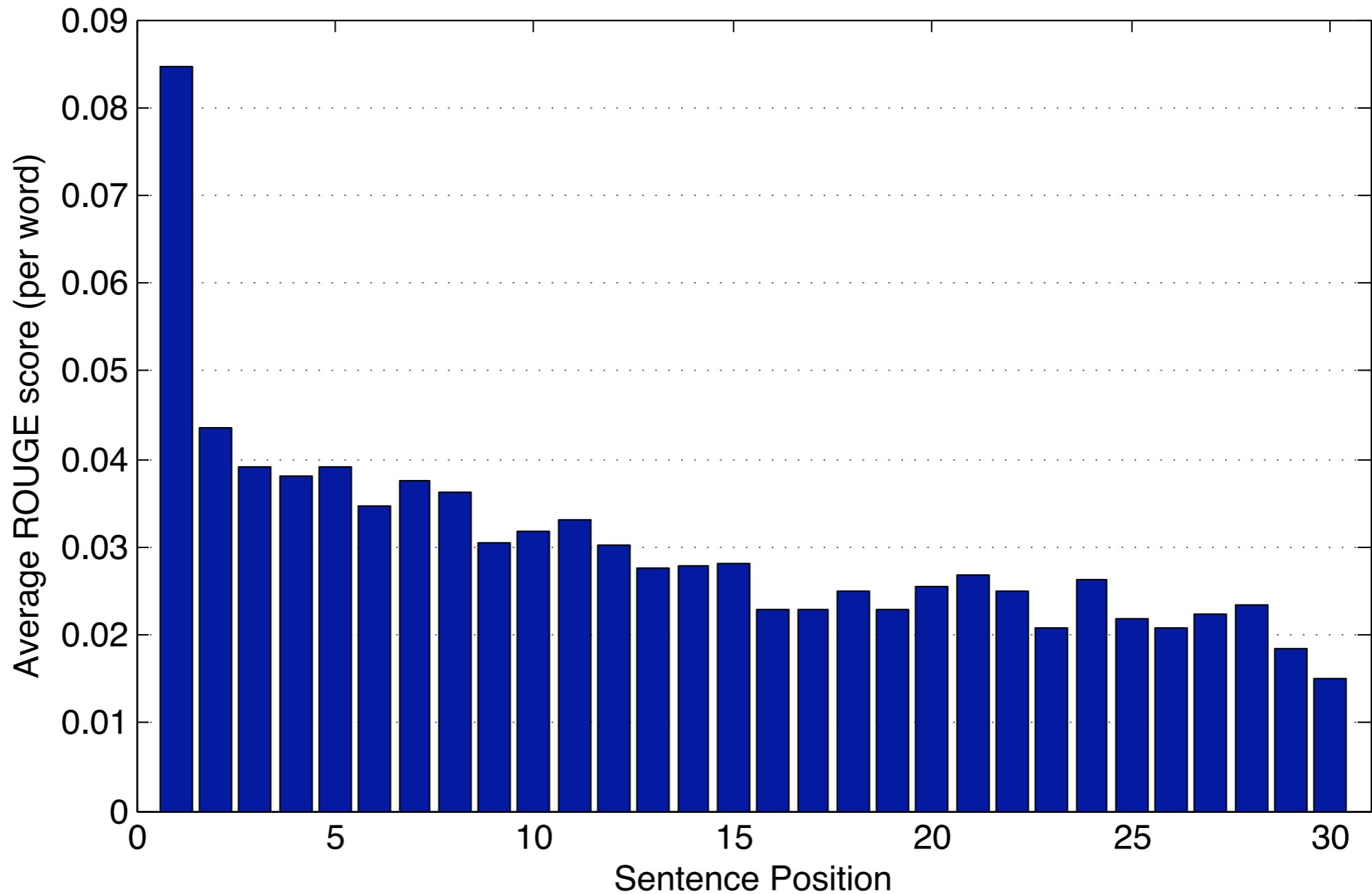
Fixing relative dates

Idea: targeted, high-precision sentence compression

The treaty was signed on Monday ← original



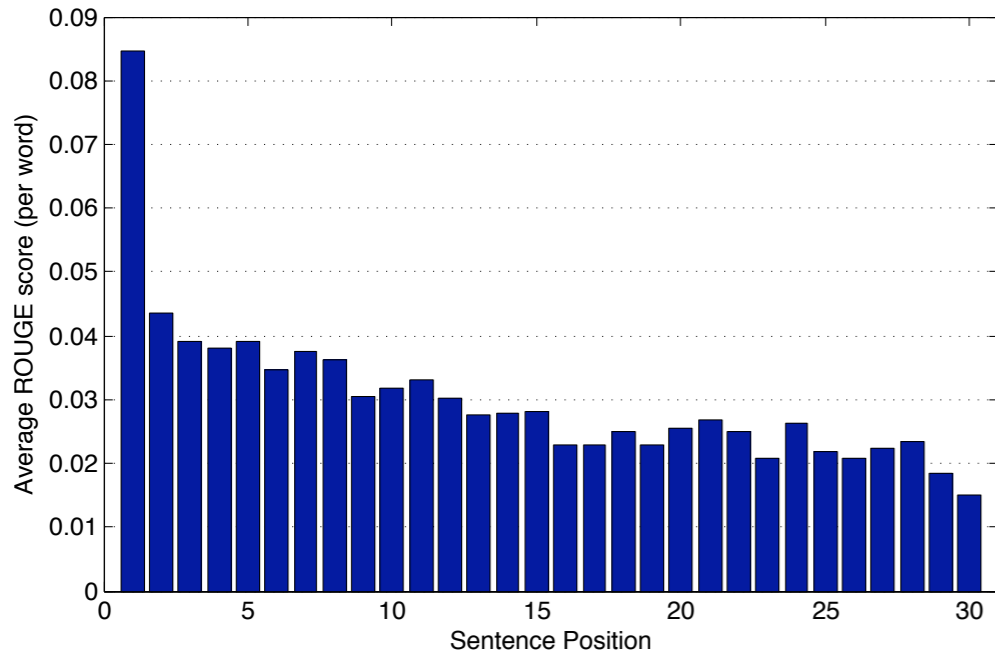
Leveraging sentence position



Early sentences are better, but by how much?

Leveraging sentence position

Set A

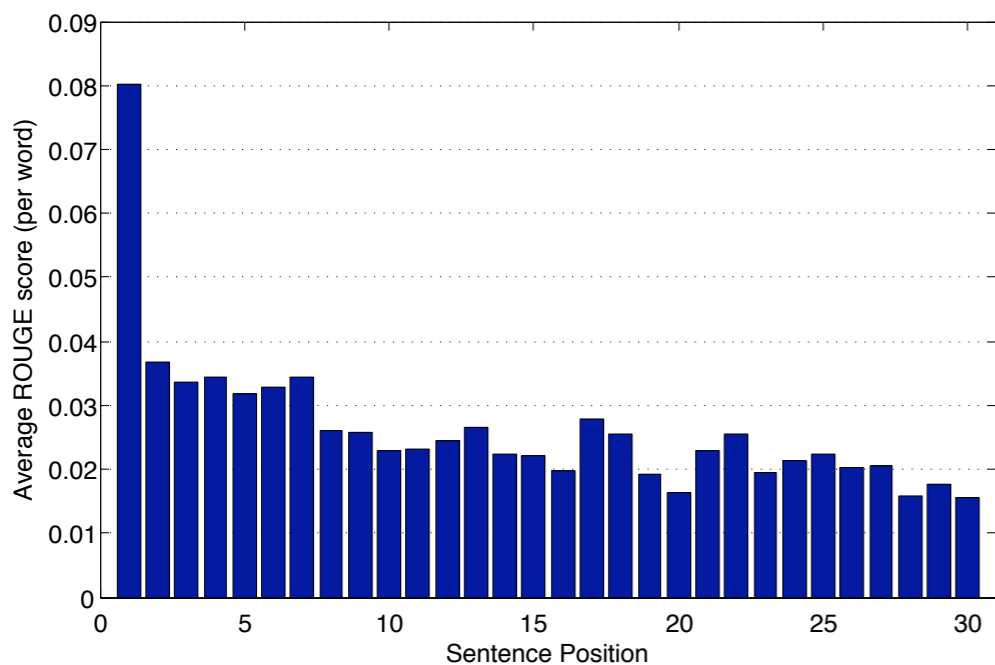


Upweight concepts in first sentences:

Set A: **x2**

Set B: **x3**

Set B



Data	Before	After	Change
08 Set A	0.1075	0.1169	+8.7%
09 Set A	0.1048	0.1220	+16.4%
08 Set B	0.0868	0.1137	+31.0%
09 Set B	0.0906	0.1059	+16.8%

ROUGE-2

Leveraging sentence position

Ratio in average ROUGE between
1st and 2nd sentences:

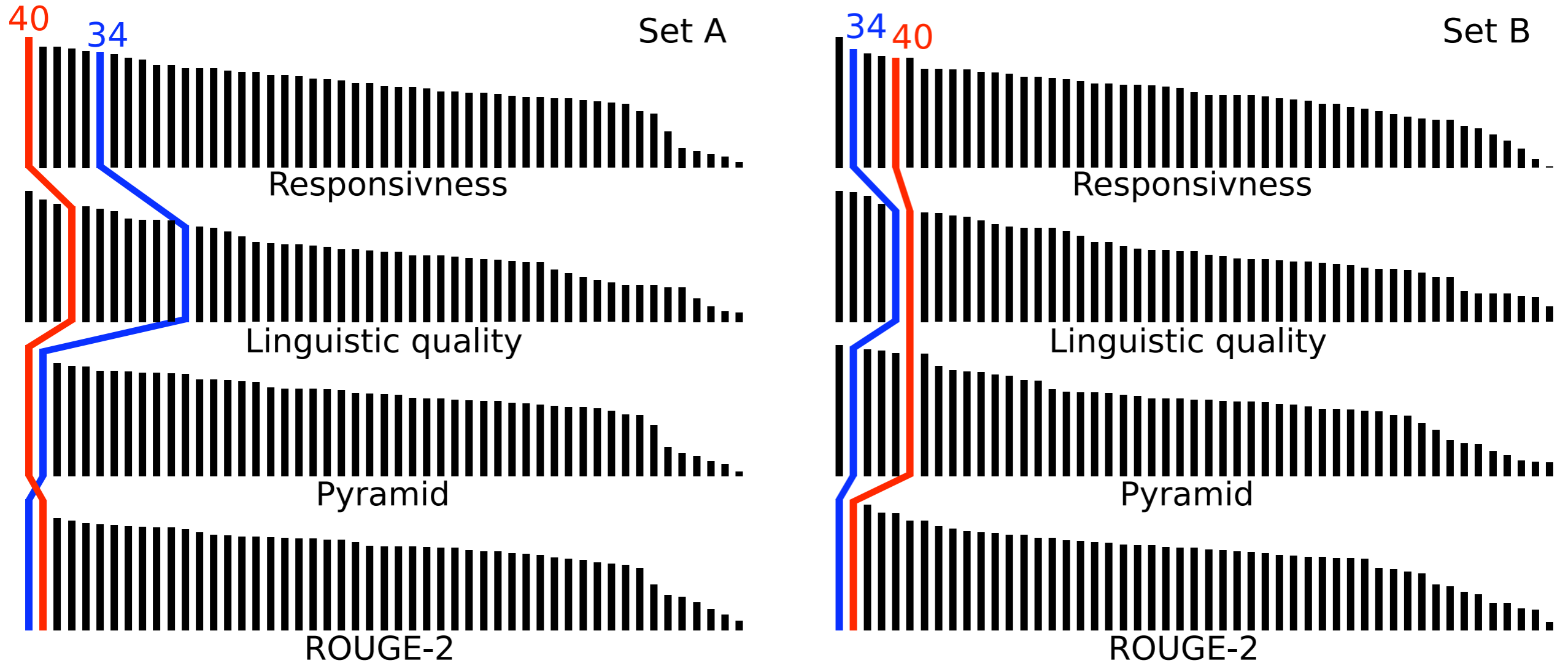
Associated Press: **2.0**

New York Times: **1.6**

Xinhua: **1.4**

More room to exploit position:
how to find the summary sentence?

Overall results



34: no compression; **40:** compression