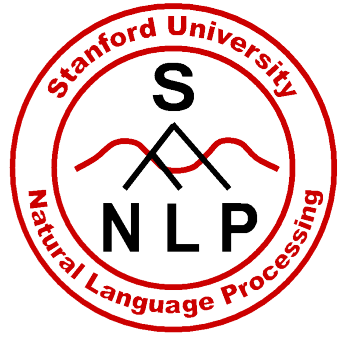
Improved Models of Distortion Cost for Statistical Machine Translation

Spence Green, Michel Galley, and Christopher D. Manning

Stanford University

June 4, 2010



Motivation

Why phrase-based MT?

- Fast, simple, and scalable
- Good performance for many language pairs
  (Zollmann et al., 2008; Lopez, 2008; etc.)

Reordering in (baseline) phrase-based decoders controlled by:

- A distortion cost model
- A distortion limit

Motivation

Why phrase-based MT?

- Fast, simple, and scalable
- Good performance for many language pairs
  (Zollmann et al., 2008; Lopez, 2008; etc.)

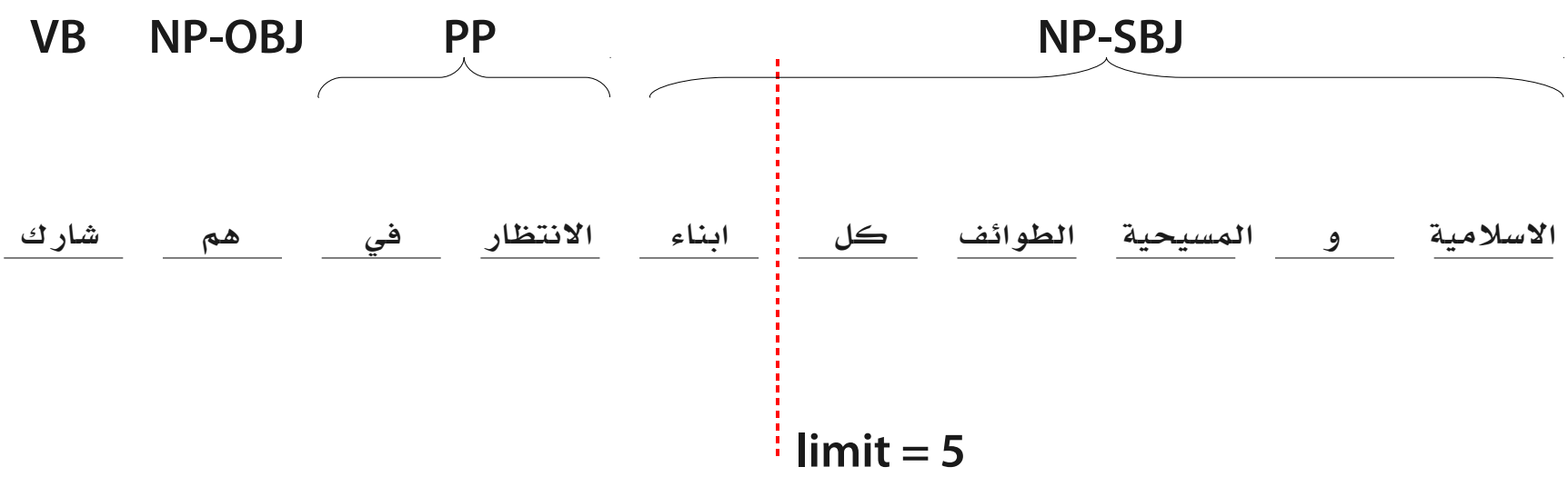
Reordering in (baseline) phrase-based decoders controlled by:

- A distortion cost model
- A distortion limit

Cost model is poor, so a low distortion limit is typically used

Motivating Example

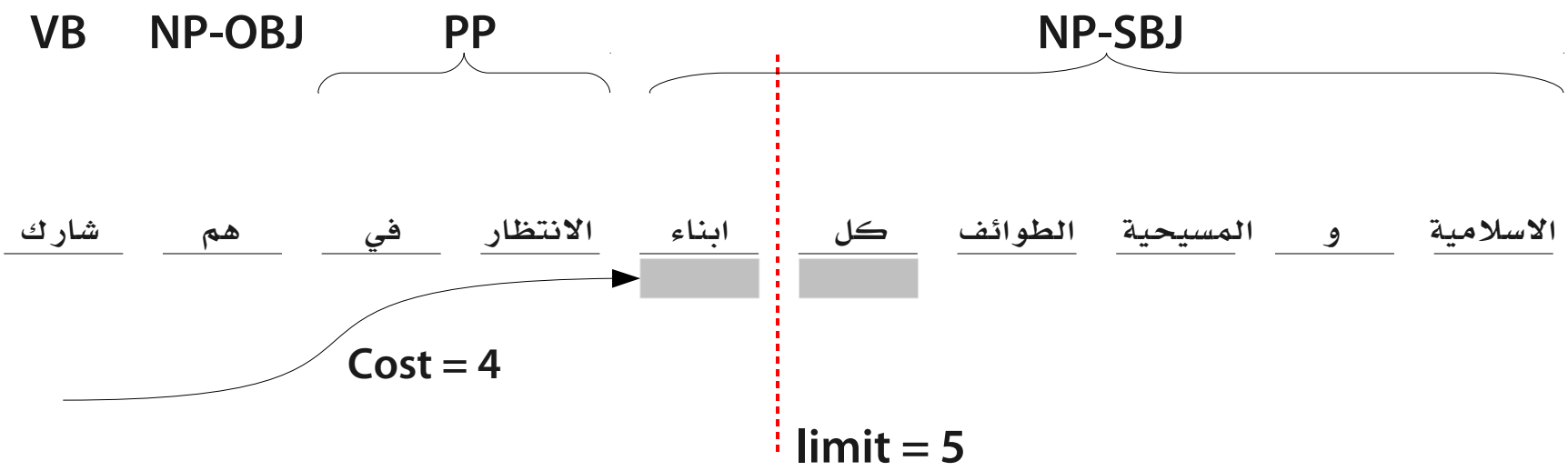
Gloss: **Followers of all of the Christian and Islamic sects engaged in waiting for them**



Hypothesis:

Motivating Example

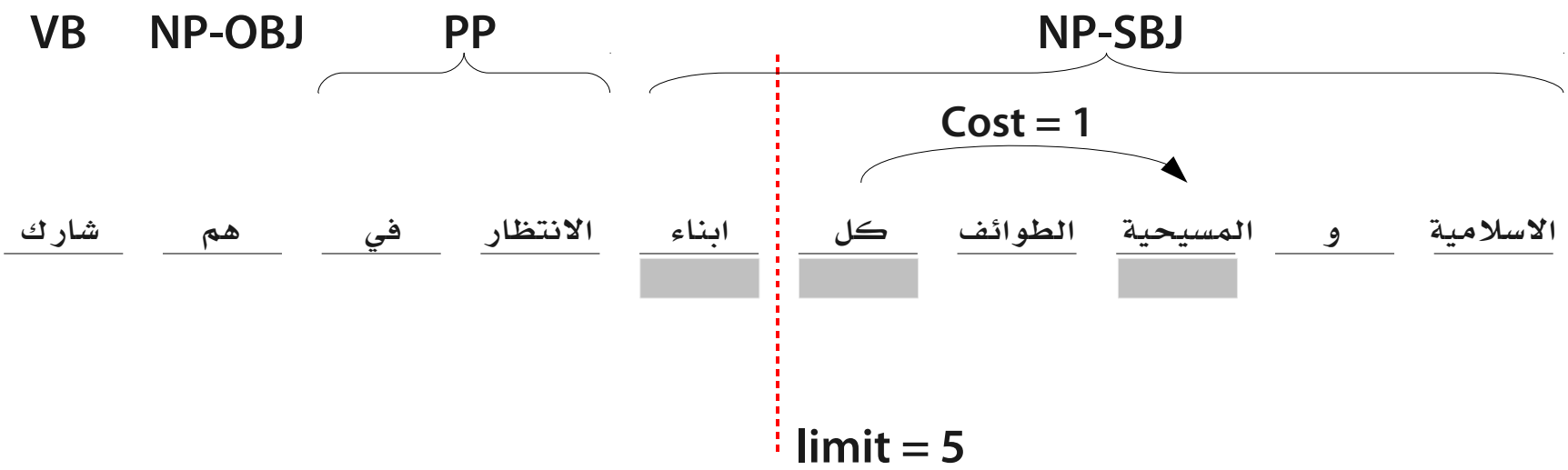
Gloss: **Followers of all of the Christian and Islamic sects engaged in waiting for them**



Hypothesis: Followers of all

Motivating Example

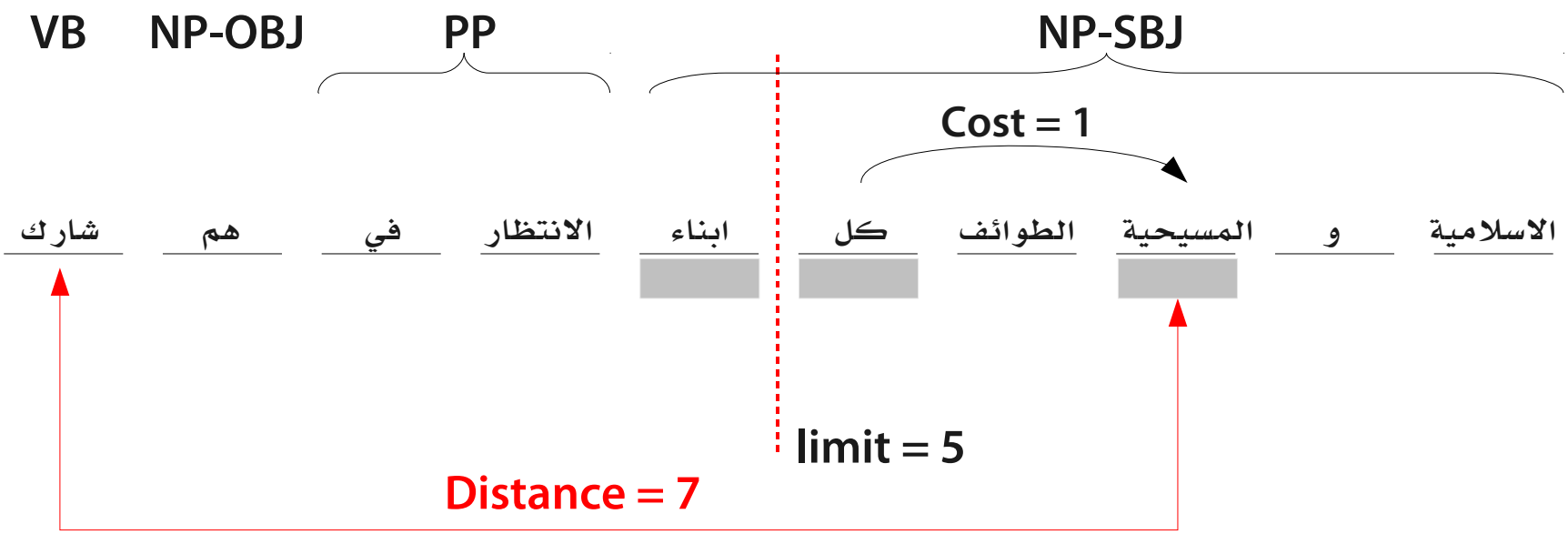
Gloss: **Followers of all of the Christian and Islamic sects engaged in waiting for them**



Hypothesis: Followers of all Christian

Motivating Example

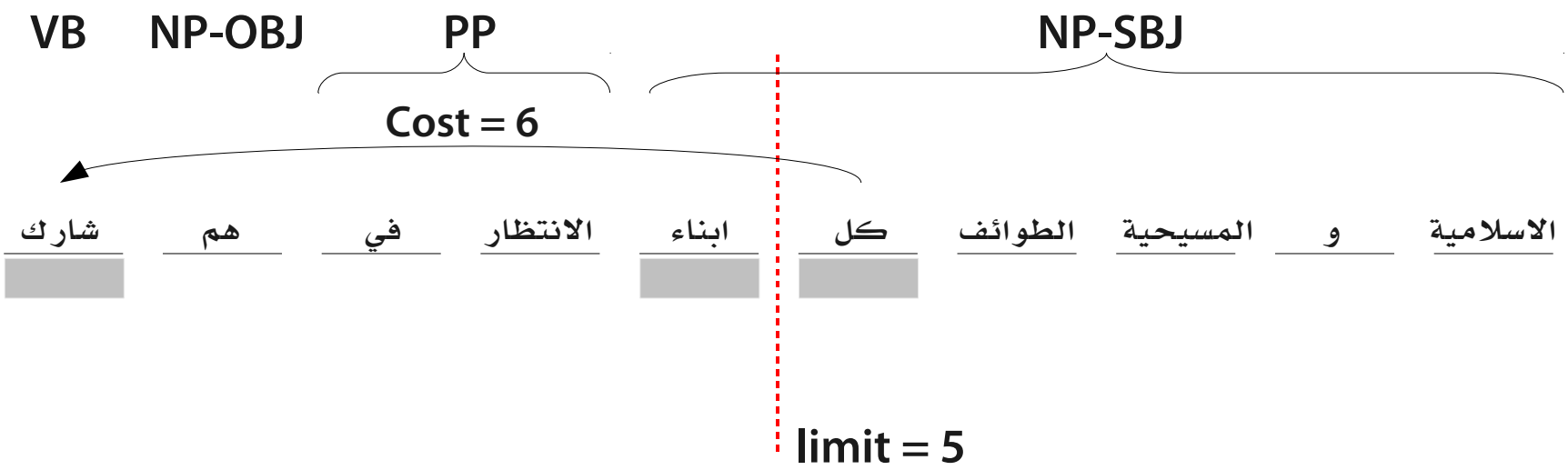
Gloss: **Followers of all of the Christian and Islamic sects engaged in waiting for them**



Hypothesis: Followers of all Christian

Motivating Example

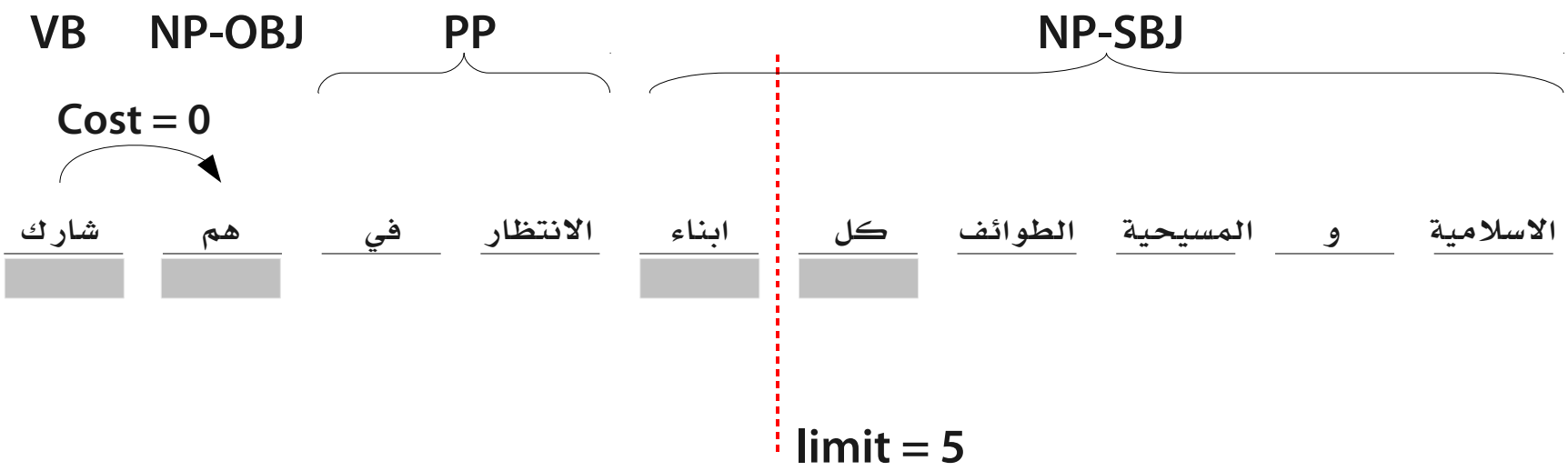
Gloss: **Followers of all of the Christian and Islamic sects engaged in waiting for them**



Hypothesis: Followers of all engaged

Motivating Example

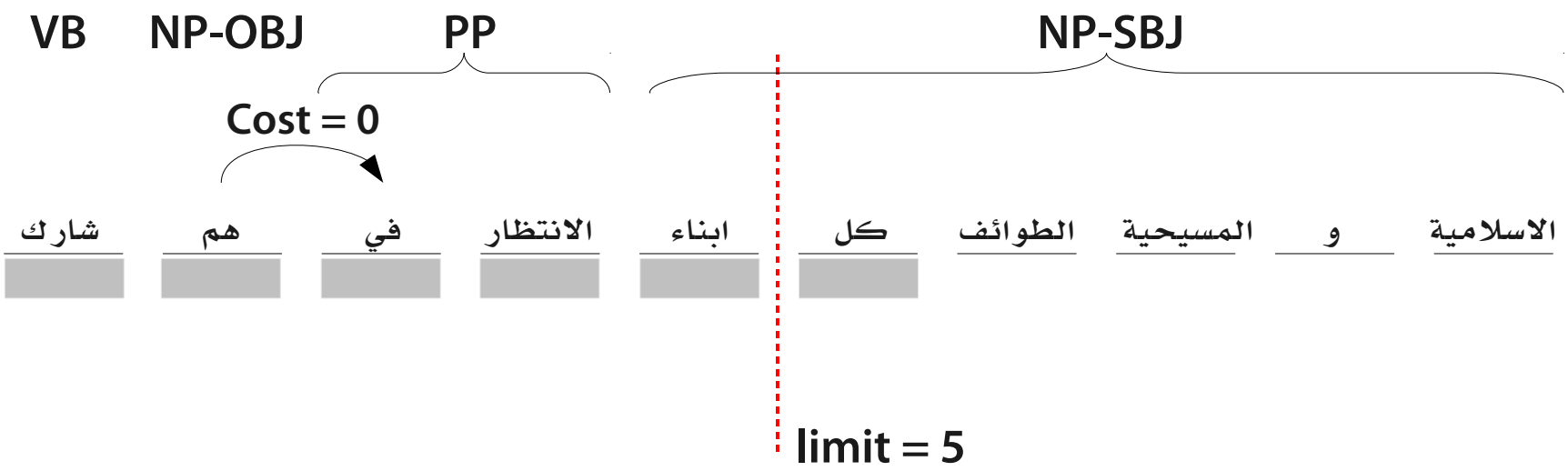
Gloss: **Followers of all of the Christian and Islamic sects engaged in waiting for them**



Hypothesis: Followers of all engaged them

Motivating Example

Gloss: **Followers of all of the Christian and Islamic sects engaged in waiting for them**



Hypothesis: Followers of all engaged them in waiting ...

Distortion Limit v. Distortion Cost

Cost is a *soft* constraint

- Does **not** prune the search space
- Feature in the log-linear decoder framework

Limit is a *hard* constraint

- Prunes translations from the search space

Distortion Limit v. Distortion Cost

Cost is a *soft* constraint

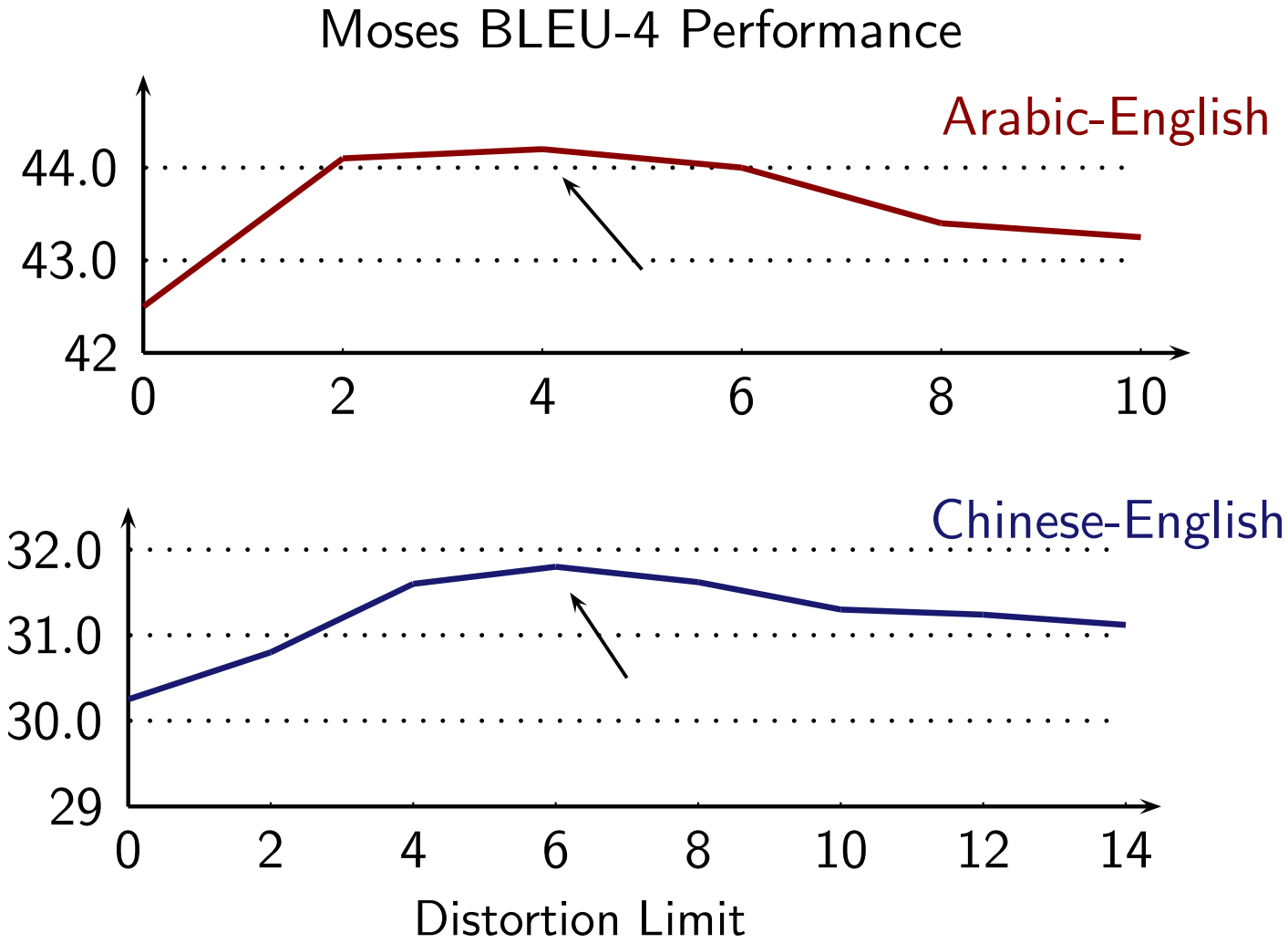
- Does **not** prune the search space
- Feature in the log-linear decoder framework

Limit is a *hard* constraint

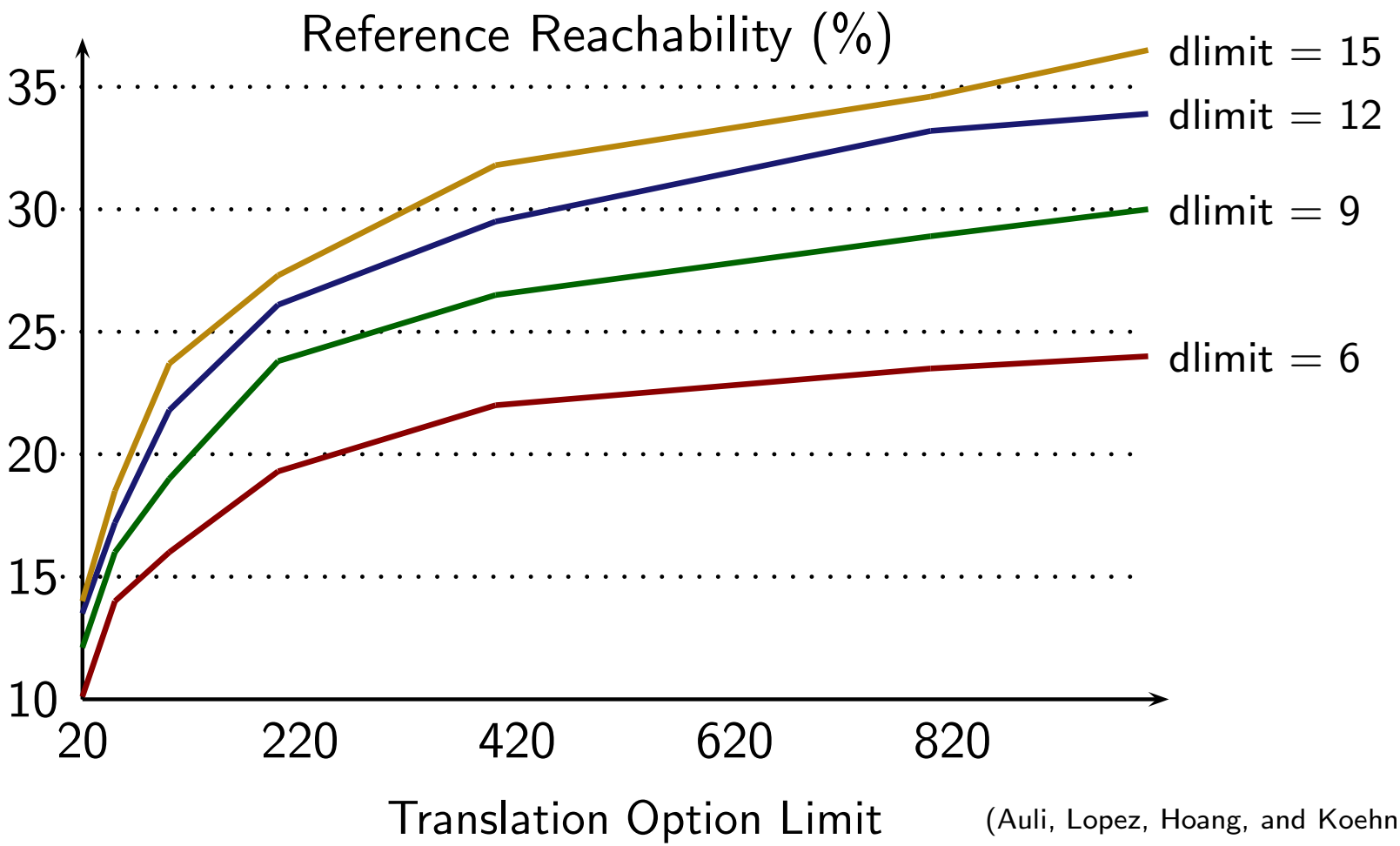
- Prunes translations from the search space

For Moses, low(er) distortion limit improves translation quality!

Translation Quality Decreases at High Distortion Limits



Hard Constraints Reduce Reference Reachability



(Auli, Lopez, Hoang, and Koehn, 2009)

A New Distortion Cost Model

Guide search *without* hard constraints

- Maintain baseline performance at high distortion limits
- Solution: Improve heuristic search with future cost estimation (Moore and Quirk, 2007)

Encourage linguistically-approriate reorderings

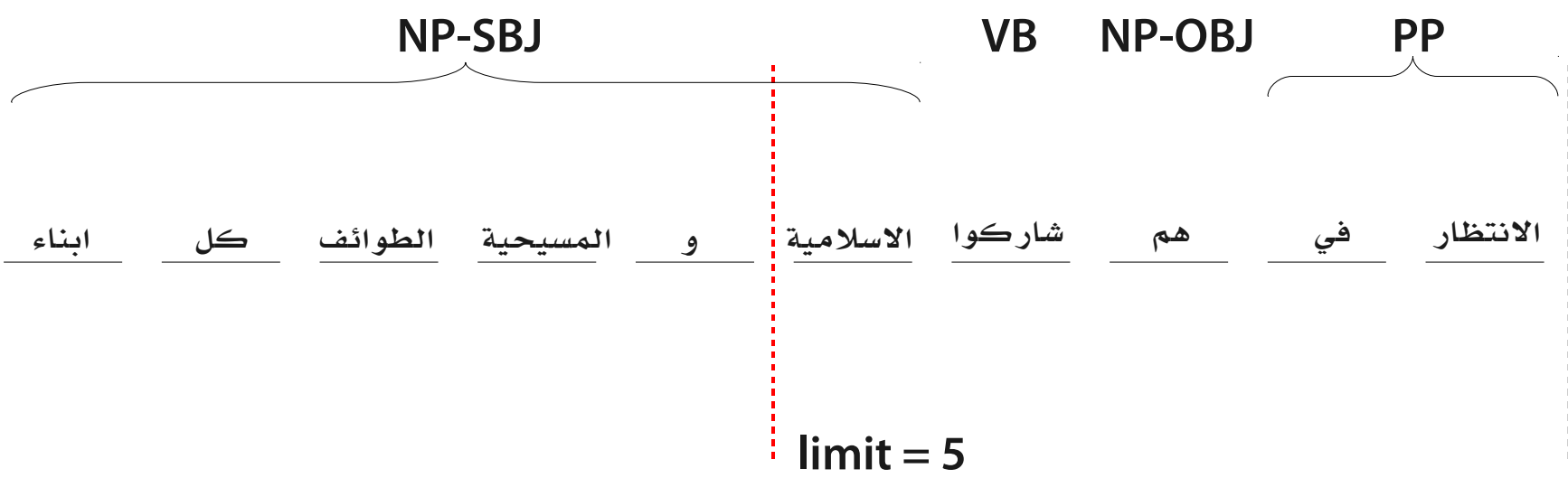
- Solution: Transition-based discriminative distortion model

Worst-case $O(n)$ cost computation

- Maintain linear running time of decoding!

Search Errors at High Distortion Limits

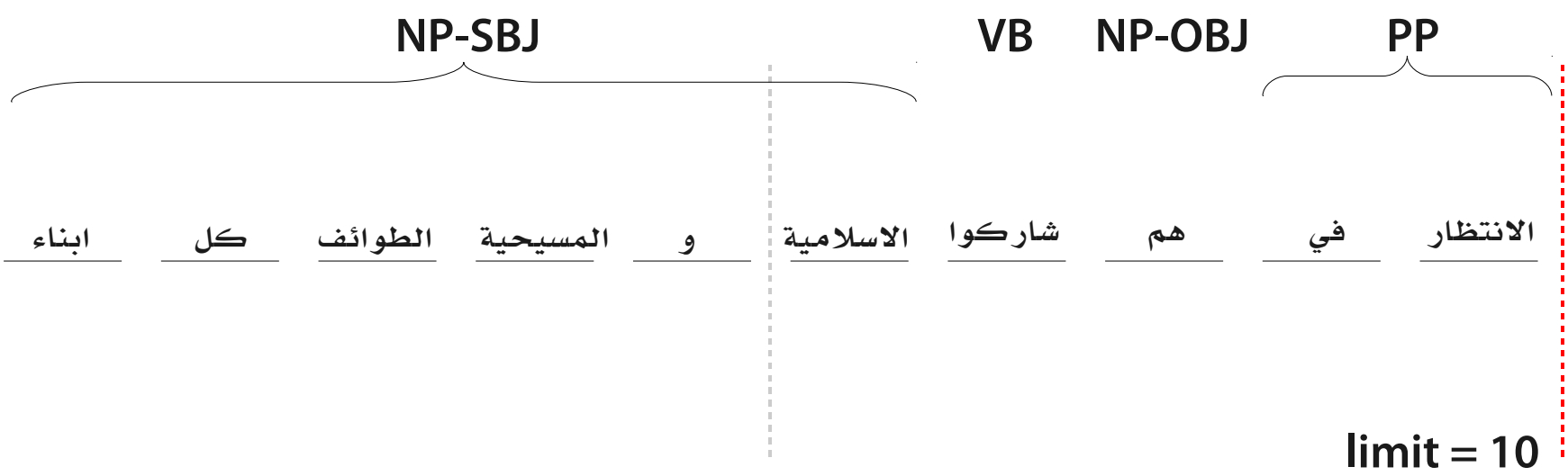
Gloss: **Followers of all of the Christian and Islamic sects engaged in waiting for them**



Hypothesis:

Search Errors at High Distortion Limits

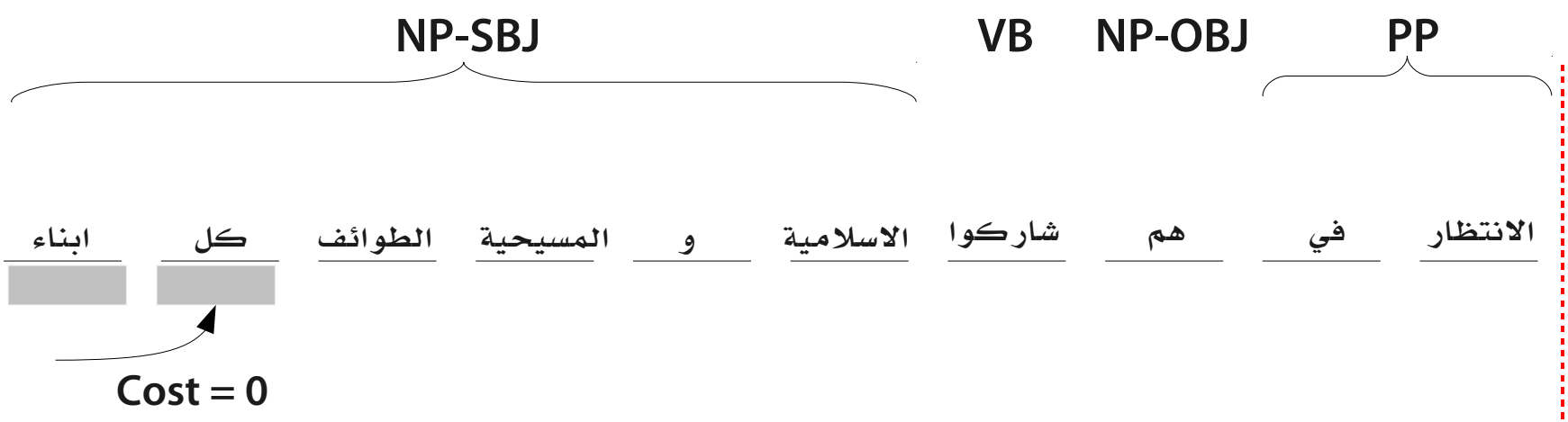
Gloss: **Followers of all of the Christian and Islamic sects engaged in waiting for them**



Hypothesis:

Search Errors at High Distortion Limits

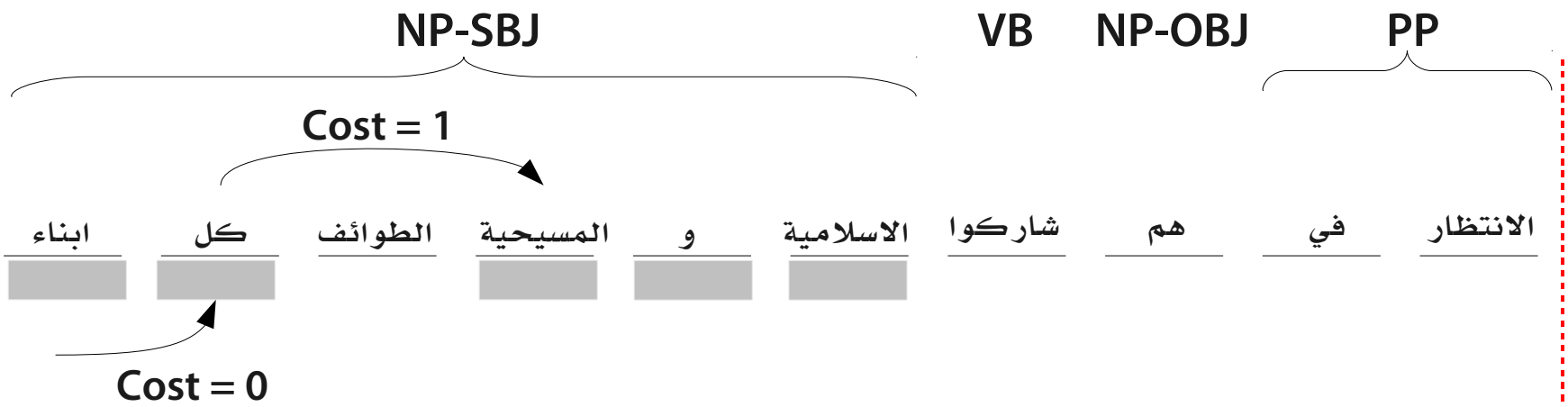
Gloss: **Followers of all of the Christian and Islamic sects engaged in waiting for them**



Hypothesis: Followers of all

Search Errors at High Distortion Limits

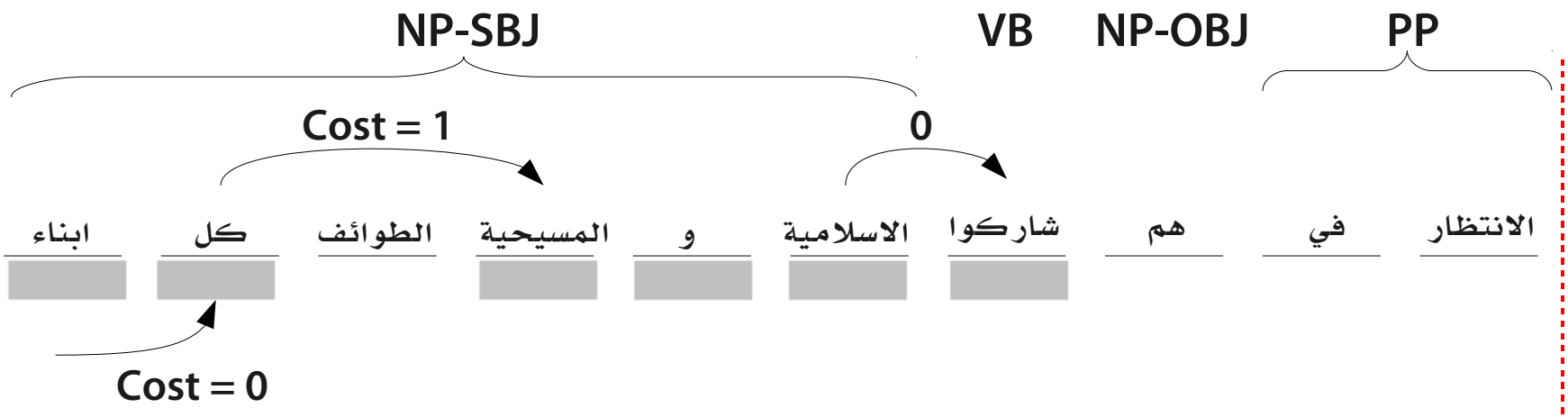
Gloss: Followers of all of the Christian and Islamic sects engaged in waiting for them



Hypothesis: Followers of all Christian and Islamic

Search Errors at High Distortion Limits

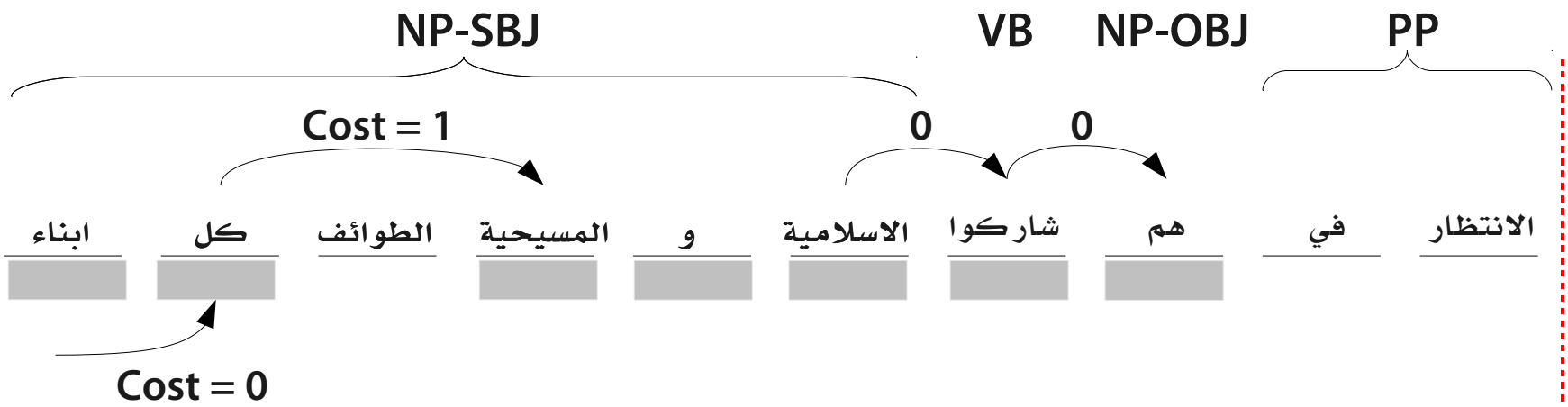
Gloss: Followers of all of the Christian and Islamic sects engaged in waiting for them



Hypothesis: Followers of all Christian and Islamic engaged

Search Errors at High Distortion Limits

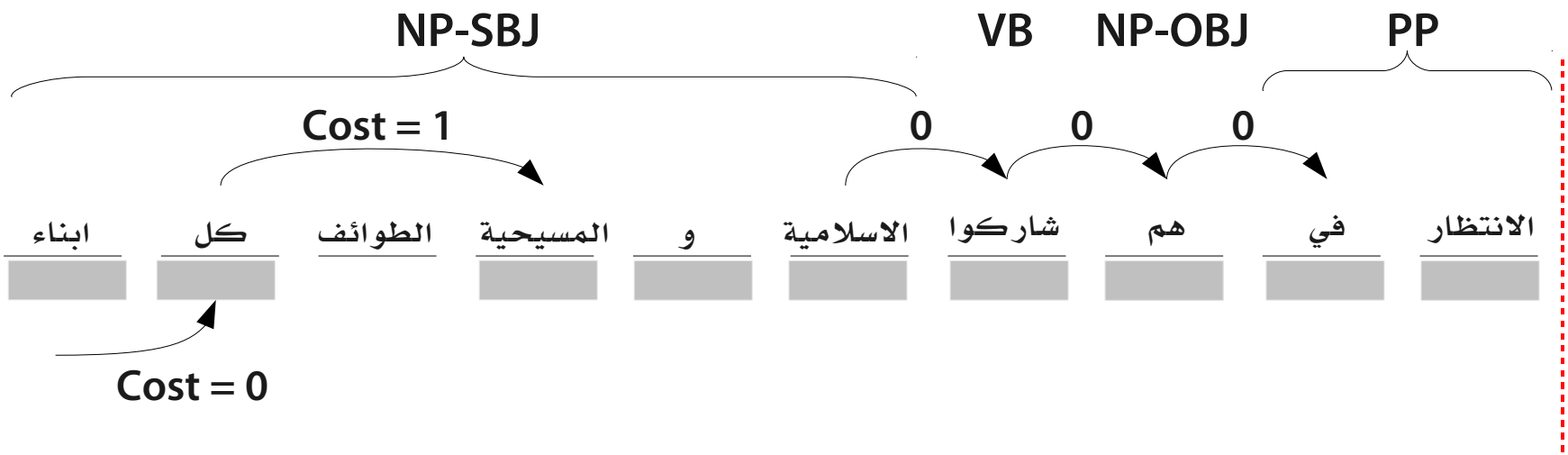
Gloss: Followers of all of the Christian and Islamic sects engaged in waiting for them



Hypothesis: Followers of all Christian and Islamic engaged them

Search Errors at High Distortion Limits

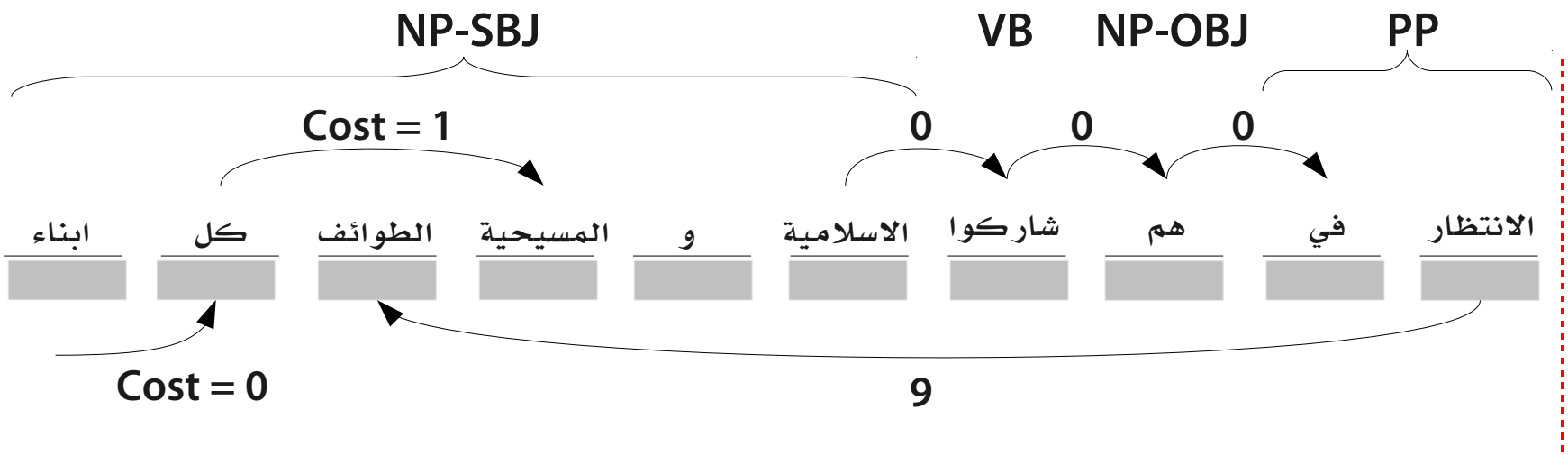
Gloss: Followers of all of the Christian and Islamic sects engaged in waiting for them



Hypothesis: Followers of all Christian and Islamic engaged them in waiting

Search Errors at High Distortion Limits

Gloss: **Followers of all of the Christian and Islamic sects engaged in waiting for them**



Hypothesis: **Followers of all Christian and Islamic engaged them in waiting sects**

An Admissible Future Cost Heuristic

$$
\begin{aligned}
s_j &\leftarrow \text{First } \textit{un}\text{covered source position} \\
s_{j'} &\leftarrow \text{First source position of phrase } p \\
\mathbf{C}_j &\leftarrow \text{Coverage set to the right of } s_j \\
D(s_{j'}, s_j) &\leftarrow \text{Linear distortion from } s_{j'} \text{ to } s_j
\end{aligned}
$$

When $j' > j$, the estimate is

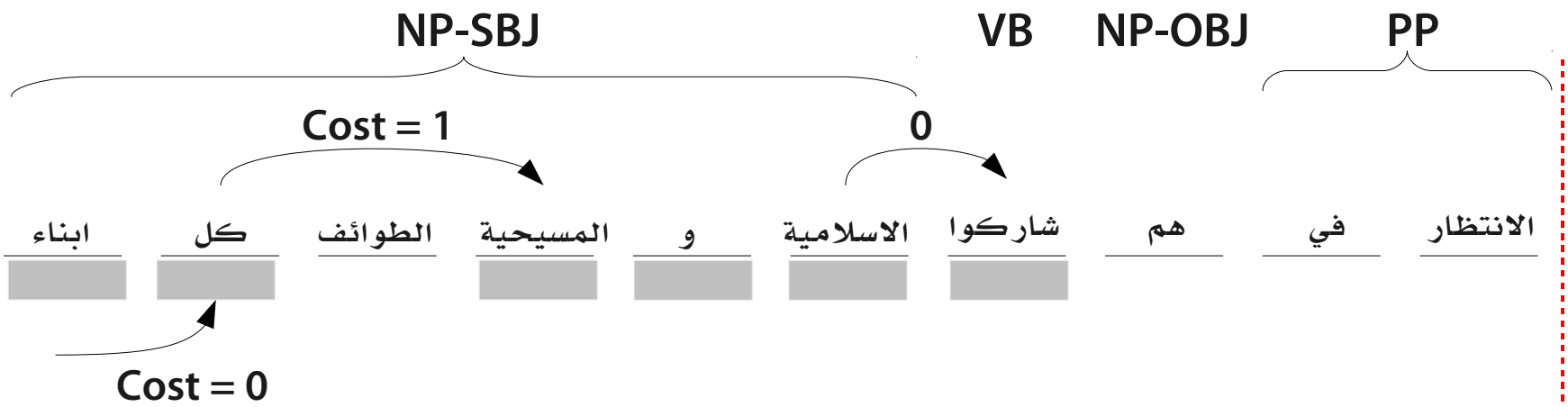
$$F = |\mathbf{C}_j| + D(s_{j'}, s_j)$$

Update the estimate at each translation step n

$$\Delta F = F_n - F_{n-1} \qquad n > 0$$

Linear Distortion with Future Cost

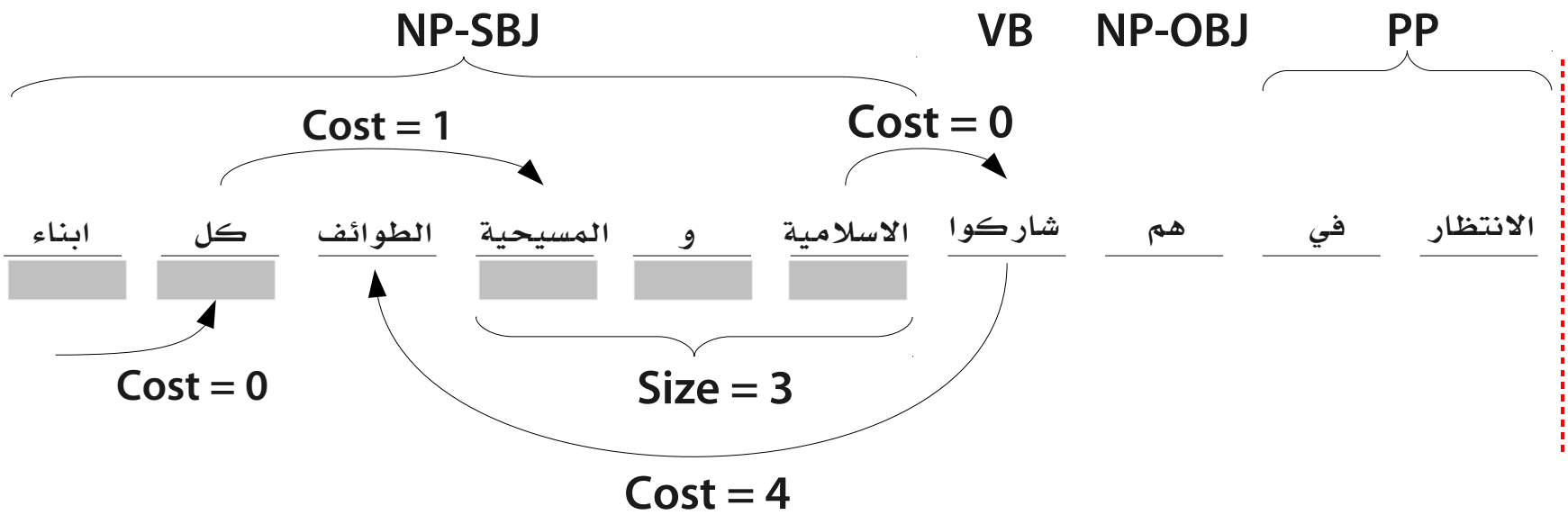
Gloss: Followers of all of the Christian and Islamic sects engaged in waiting for them



Hypothesis: Followers of all Christian and Islamic engaged

Linear Distortion with Future Cost

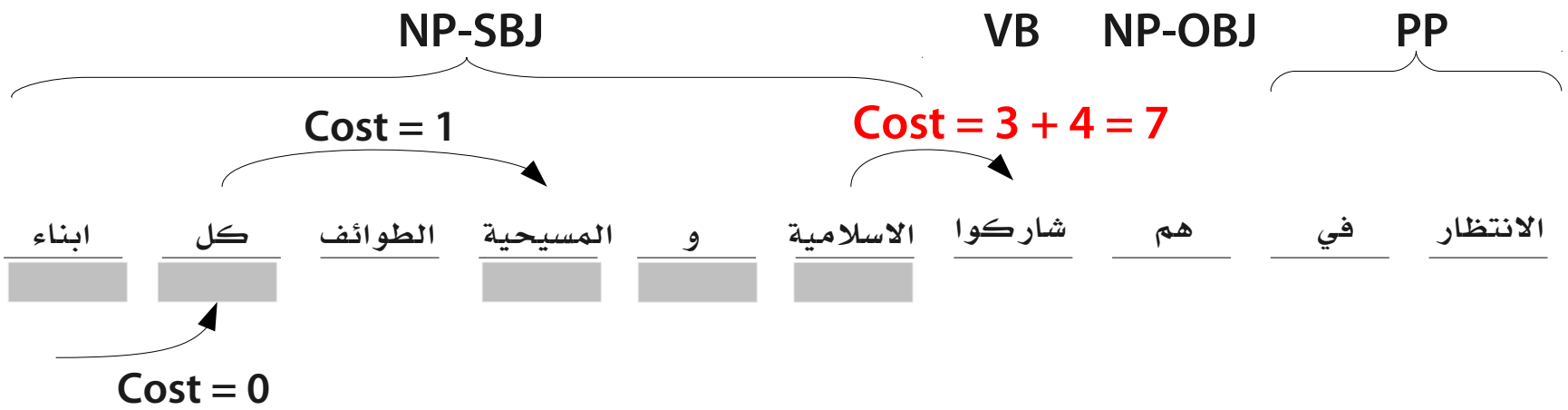
Gloss: Followers of all of the Christian and Islamic sects engaged in waiting for them



Hypothesis: Followers of all Christian and Islamic

Linear Distortion with Future Cost

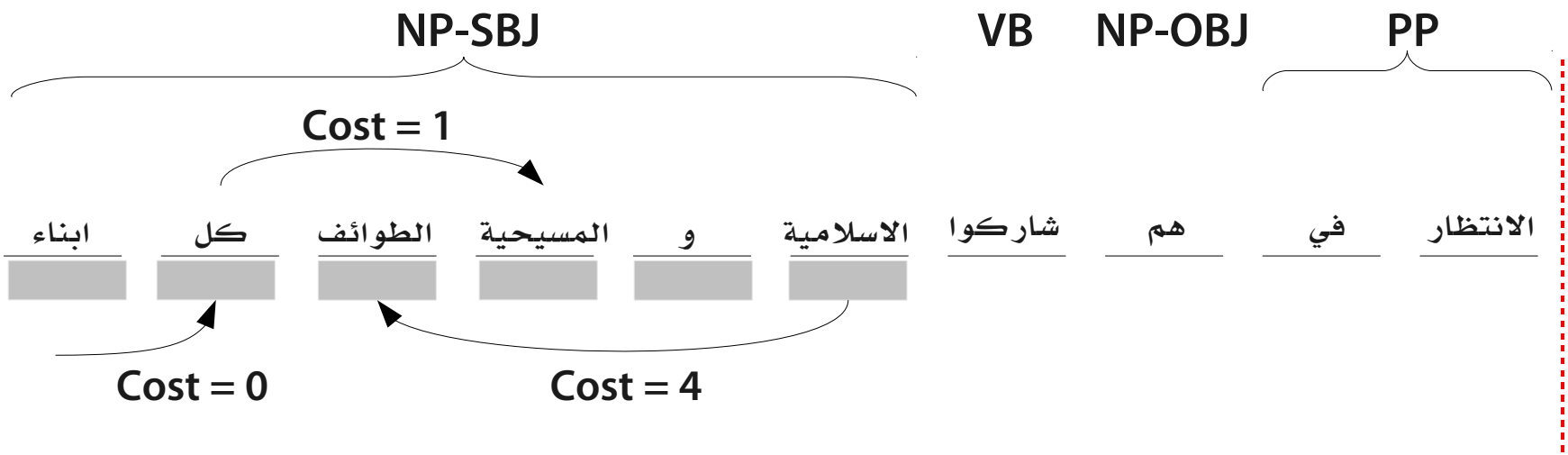
Gloss: Followers of all of the Christian and Islamic sects engaged in waiting for them



Hypothesis: Followers of all Christian and Islamic

Linear Distortion with Future Cost

Gloss: Followers of all of the Christian and Islamic sects engaged in waiting for them



Hypothesis: Followers of all Christian and Islamic sects

Transition-based Discriminative Distortion Cost

Problem: Cost model still penalizes **all** reorderings

- Consider a verb-final language like Japanese
- Skipping over the entire verb complement is *good*
- Model should prefer particular reorderings

A **transition-based** discriminative distortion model

- Idea: compute cost of word-to-word transitions
- Source side features
- Discretized transition classes

Procedure

1. Classify discretized transitions with a log-linear model

$$p_\lambda\left(D_{j,j'} \mid \bar{\mathbf{s}}, j, j'\right) \propto \exp\left[\,\bar{\lambda} \cdot \bar{\mathbf{h}}\left(\bar{\mathbf{s}}, j, j', D_{j,j'}\right)\,\right]$$

2. Train with sorted word-to-word alignments
   - e.g. Arabic $\Longrightarrow$ Arabic$'$ (English word order)
3. Query model for transitions at each translation step

Features

This evaluation

- Words and POS tags
- Relative source sentence position (discretized)
- Source sentence length (discretized)

Future work

- POS tag chains (bigram, trigram, etc.)
- Agreement morphology
- Subject has been translated? (binary, global)

Further Motivation

Incremental processing like shift-reduce parsing

- Process source items in decoding order
- Classes are discrete jumps instead of "operations"
- Beam search via the decoder

Implementation: Constant-time During Decoding

Nine discrete distortion classes

- Same number of training examples per class

Separation into inbound/outbound models (Al-Onaizan and Papineni, 2006)

- Simplifies caching during decoding
- Future work: Combine into a single model

Model has four decoder features

- Inbound and outbound scores
- **Alignment penalty**
- **Future cost estimate**

Evaluation

MT system is Phrasal (Cer et al., 2010)

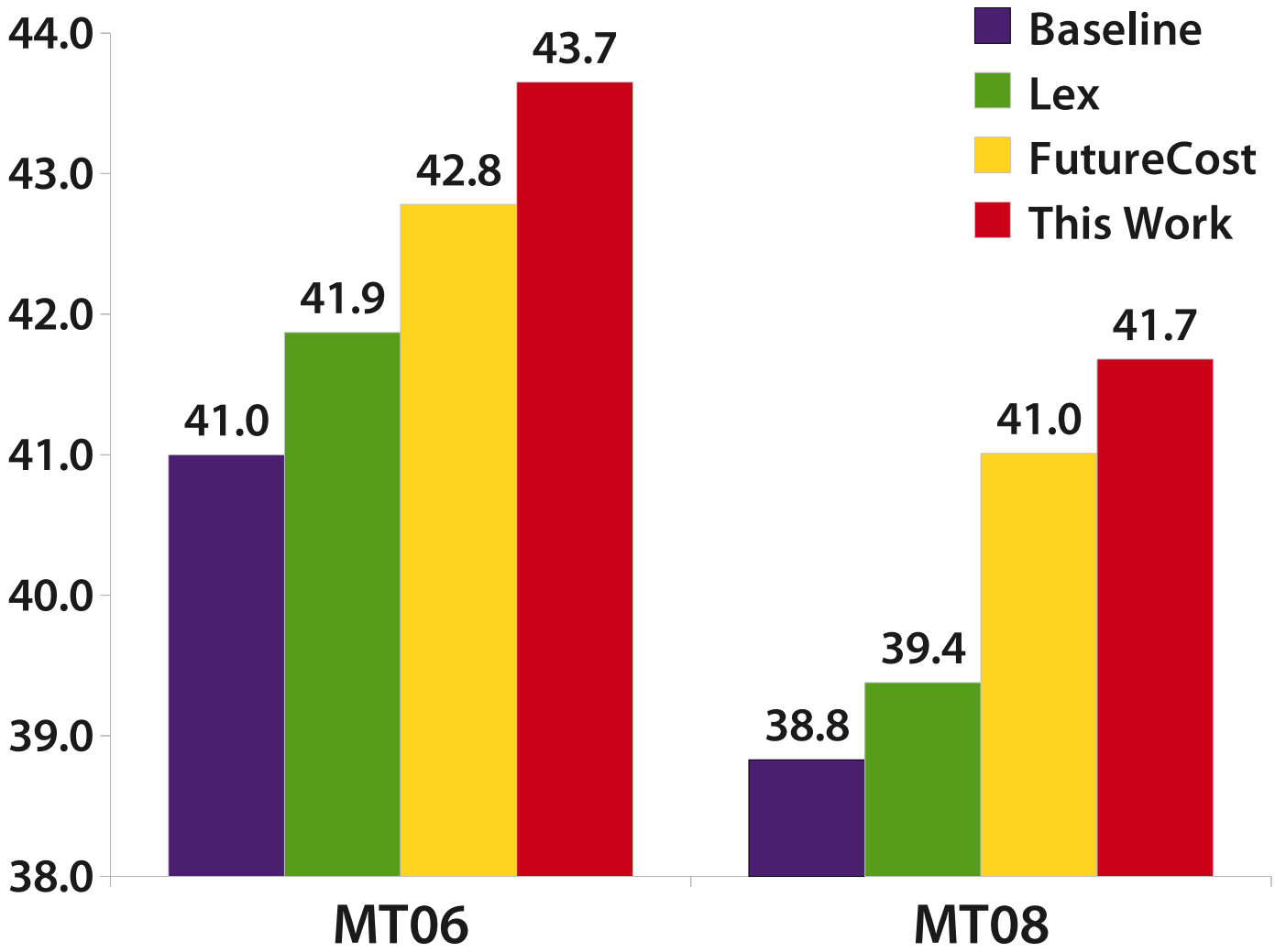
- Baseline: Moses feature set
- Lexicalized reordering model of Galley and Manning (2008)

NIST MT09 Ar-En constrained track training data

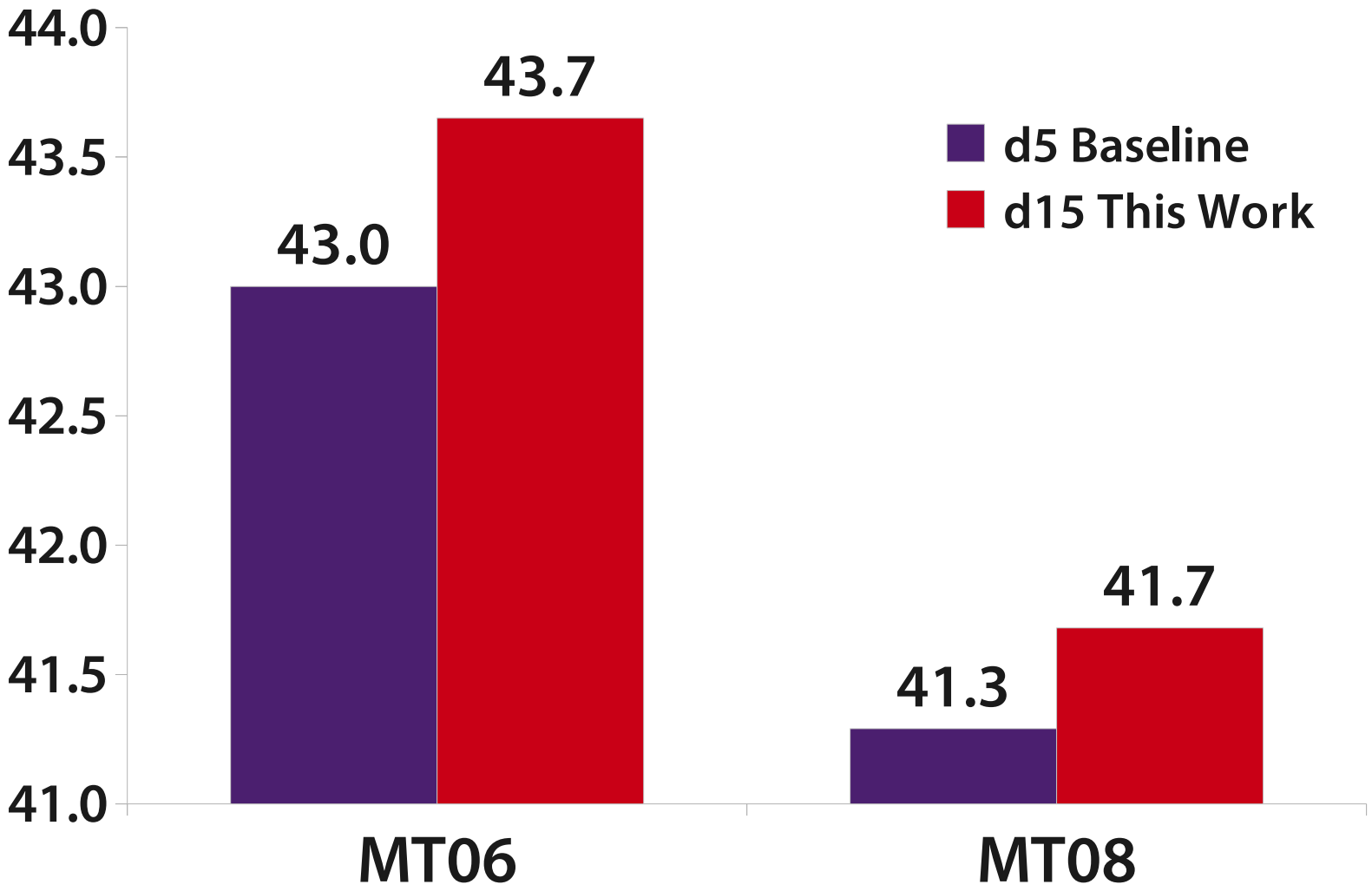
- Removed UN and comparable data
  - Same baseline, faster experiments
- 6.20M English and 5.73M Arabic tokens

Evaluated BLEU-4 on MT03/05/06/08 at dlimit = 15

High Distortion Limit (15)



Improvement Over the Low Distortion Limit Baseline



Conclusion

Contributions of this work

- Fixed search errors caused by linear distortion
- Added a distortion model with linguistic features
- Modest improvement over Moses at a high distortion limit

Software:

Phrasal	http://nlp.stanford.edu/phrasal/
Arabic NLP tools	http://nlp.stanford.edu/projects/arabic.shtml

Thank You!

Thanks to Daniel Cer and Claude Reichard.

Distortion Cost Curve for the adjective *American*