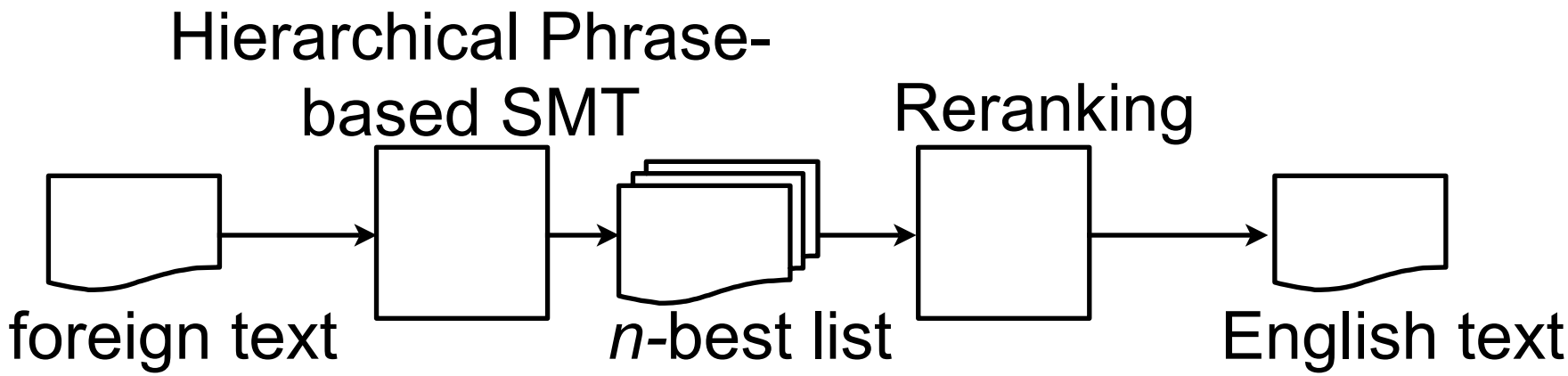
NTT Statistical Machine Translation for IWSLT 2006

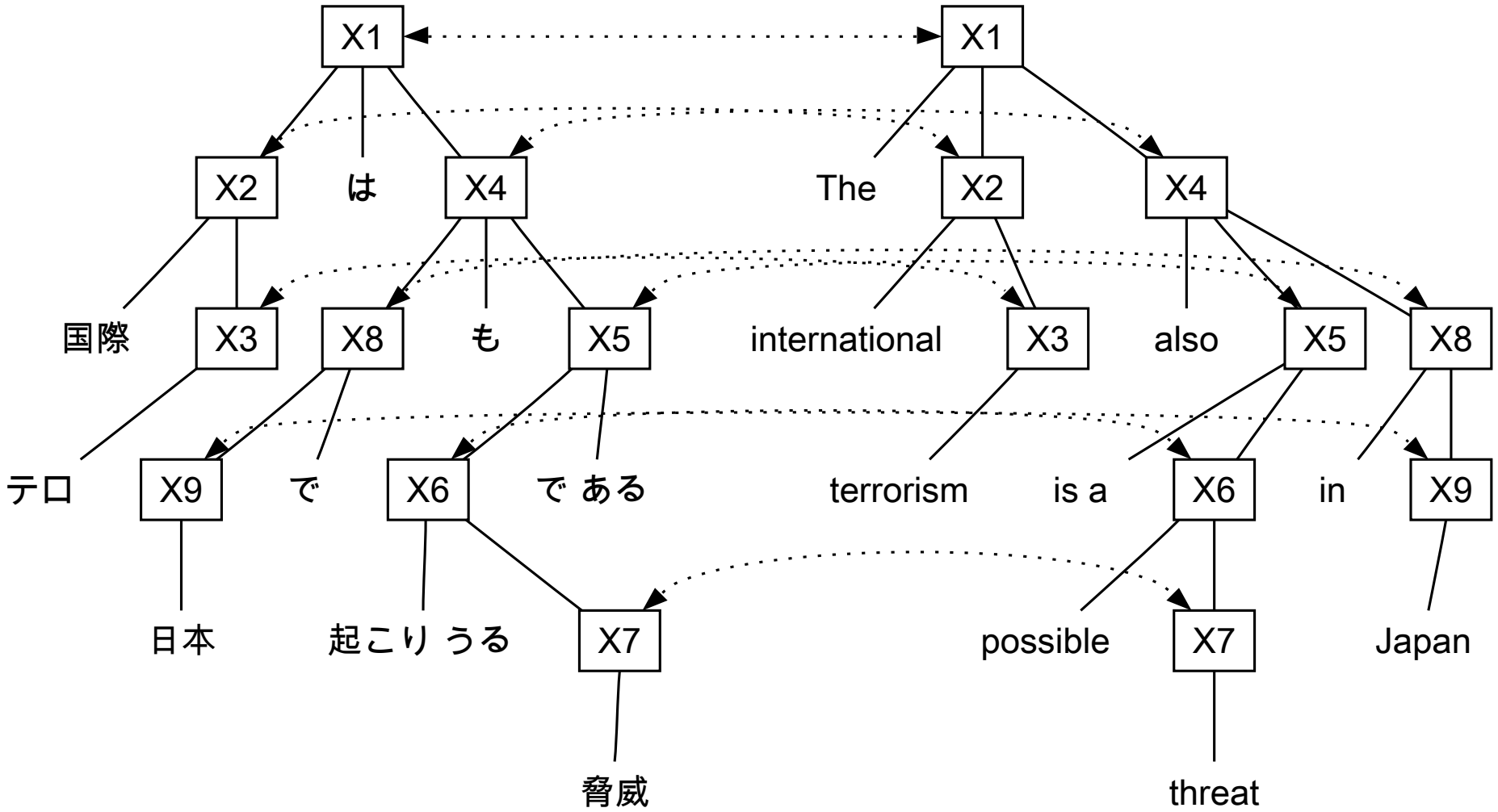
Taro Watanabe, Jun Suzuki, Hajime Tsukada
and Hideki Isozaki
NTT Communication Science Labs.
{taro,jun,tsukada,isozaki}@cslab.kecl.ntt.co.jp

Overview



- Hierarchical Phrase-based SMT achieved by:
  - Target normalized form
  - Earley-style top-down parsing
- Reranking via voted perceptron

Hierarchical Phrase-based SMT

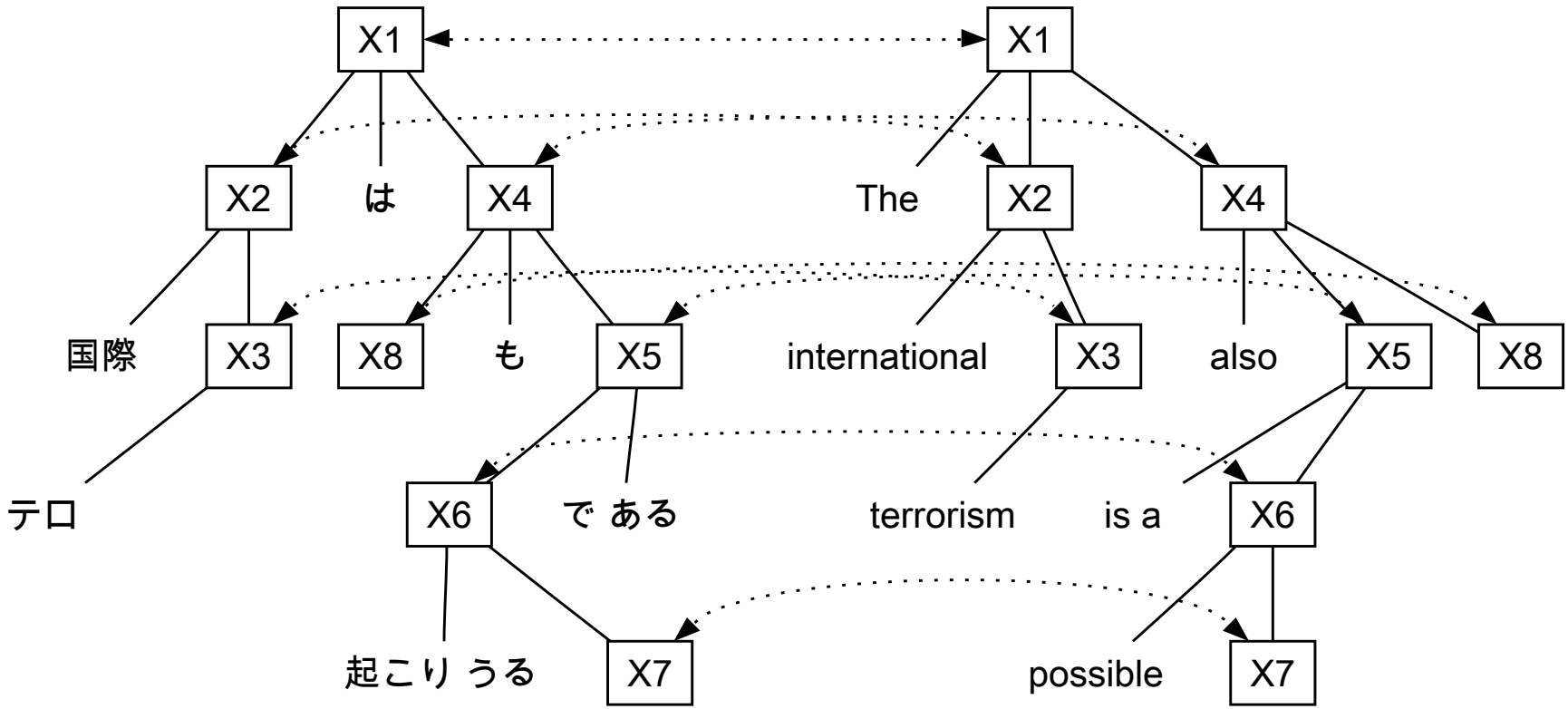


Simplified Grammar

$X_{\boxed{1}}$ 份子	|||	members of $X_{\boxed{1}}$
定 将 $X_{\boxed{1}}$	|||	decided to $X_{\boxed{1}}$
$X_{\boxed{1}}$ 是 $X_{\boxed{2}}$	|||	is $X_{\boxed{1}}$ $X_{\boxed{2}}$
$X_{\boxed{1}}$ 不再 $X_{\boxed{2}}$	|||	no longer $X_{\boxed{1}}$ $X_{\boxed{2}}$
$X_{\boxed{2}}$ 一个 最 $X_{\boxed{1}}$	|||	One of the most $X_{\boxed{1}}$ $X_{\boxed{2}}$
$X_{\boxed{2}}$ 的 $X_{\boxed{1}}$ 表示 哀悼	|||	condolences to the $X_{\boxed{1}}$ $X_{\boxed{2}}$

- Target normalized form
  - Phrase-prefixed structure for target-side (GNF-like structure)
  - Arbitrary structure for source-side
  - Constrained rule-extraction

Decoding by Top-down parsing



- Earley-style parsing on source-side
  - Straight-forward intersection with ngram
- Similar to a phrase-based decoding algorithm

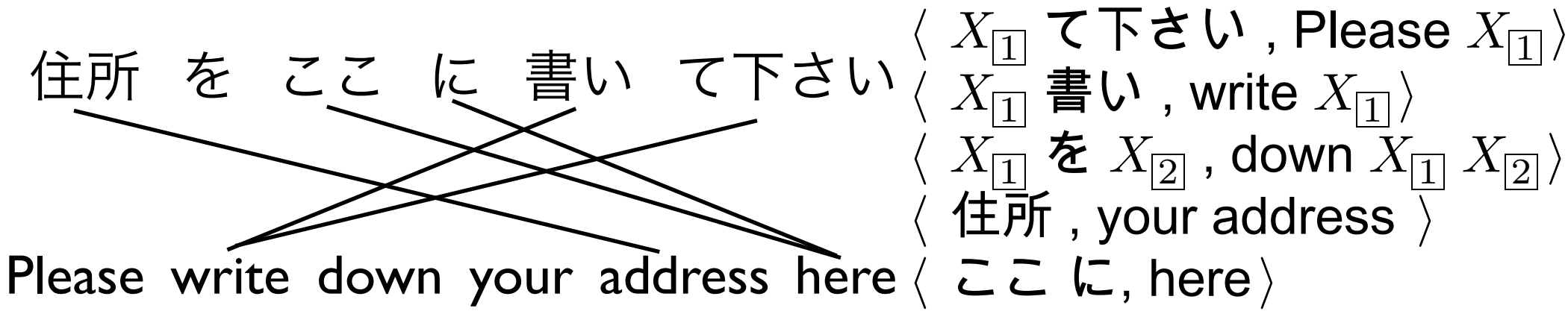
Log-linear Approach

$$\hat{e}_1^I = \underset{e_1^I}{\operatorname{argmax}} \frac{\exp\left(\sum_{m=1}^{M} \lambda_m h_m(e_1^I, f_1^J)\right)}{\sum_{e'_1{}^{I'}} \exp\left(\sum_{m=1}^{M} \lambda_m h_m(e'_1{}^{I'}, f_1^J)\right)}$$

- Mixed-case 5-gram
- Rule translation probabilities
- Lexical weights
- Insertion/deletion penalties
- Backtrack penalties
- # of words/# of rules

Reranking by Voted Perceptron

- Ranking Voted Perceptron with BLEU-based updates
- Features
  - SC: Scores from the baseline decoder
  - AL: Word-pairs from IBM Model Viterbi alignment
  - RU: Rules & Rule pattern



Reranking Algorithm

$D = \{D^1, ..., D^M\}$: Development set
$C^m = \{c_1^m, ..., c_N^m\}$: The original N-best list of D^m
$X^m = \{x_1^m, ..., x_N^m\}$: (reordered) N-best list of D^m
$Ranking(W, C^m)$: returns N-best list of C^m reordered based on the score, $s_n^m = \langle W, \phi(c_n^m) \rangle$

for $t = 1, ..., T$ **do**
  for $m = 1, ..., M$ **do**
    $X^m \leftarrow Ranking(W, C^m)$
    for $i = 1, ..., |X^m|$ **do**
      for $j = i + 1, ..., |X^m|$ **do**
        if $BLEU(x_j^m) > BLEU(x_i^m) \& WER(x_j^m) \leq WER(x_i^m)$ **then**
          $W = W + (BLEU(x_j^m) - BLEU(x_i^m)) \times (\phi(x_j^m) - \phi(x_i^m))$
        end if
      end for
    end for
    $V_t = W$
  end for
end for

Update all incorrect ranking pair through pair-wise comparison

Approximated BLEU

- Very frequent updates required:
  - Computation of doc-set BLEU is impossible
- Sentence-wise BLEU?
  - Bad objective: 27.78 to 25.95 in MTEval 2006
- Approximated BLEU:
  - doc-set BLEU of 1-best
  - Compute difference for each segment

Tasks

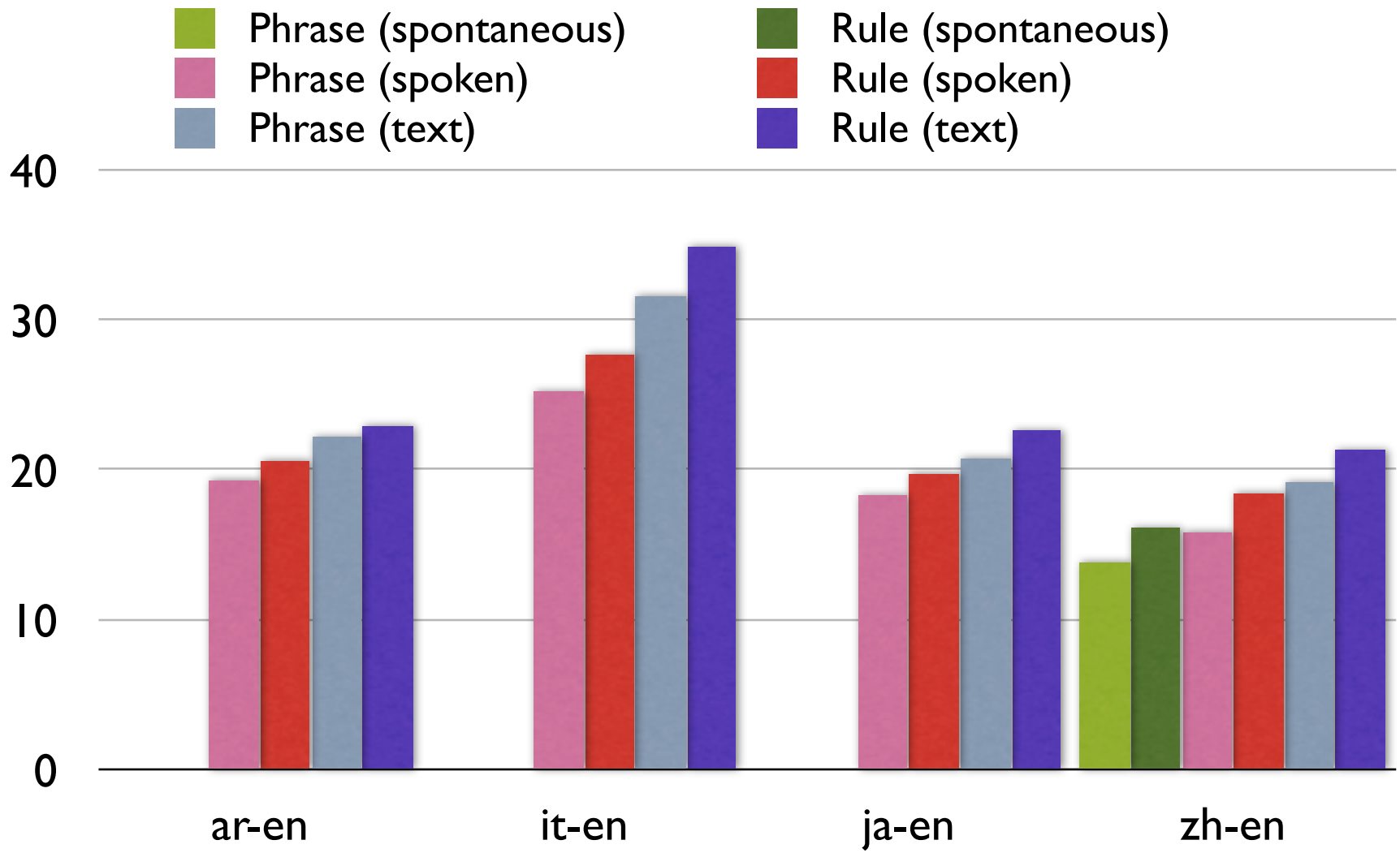
- ASR's 1-best translation
- Case-restoration/punctuation-insertion required
- Preprocessing:
  - Case/punctuation-preserved English-side + lower-cased/punctuation-removed source-side
  - Induce multiple alignments from differently preprocessed corpora (punct-removed, etc.)
  - Aggregate rules from differently aligned corpora

Official Results

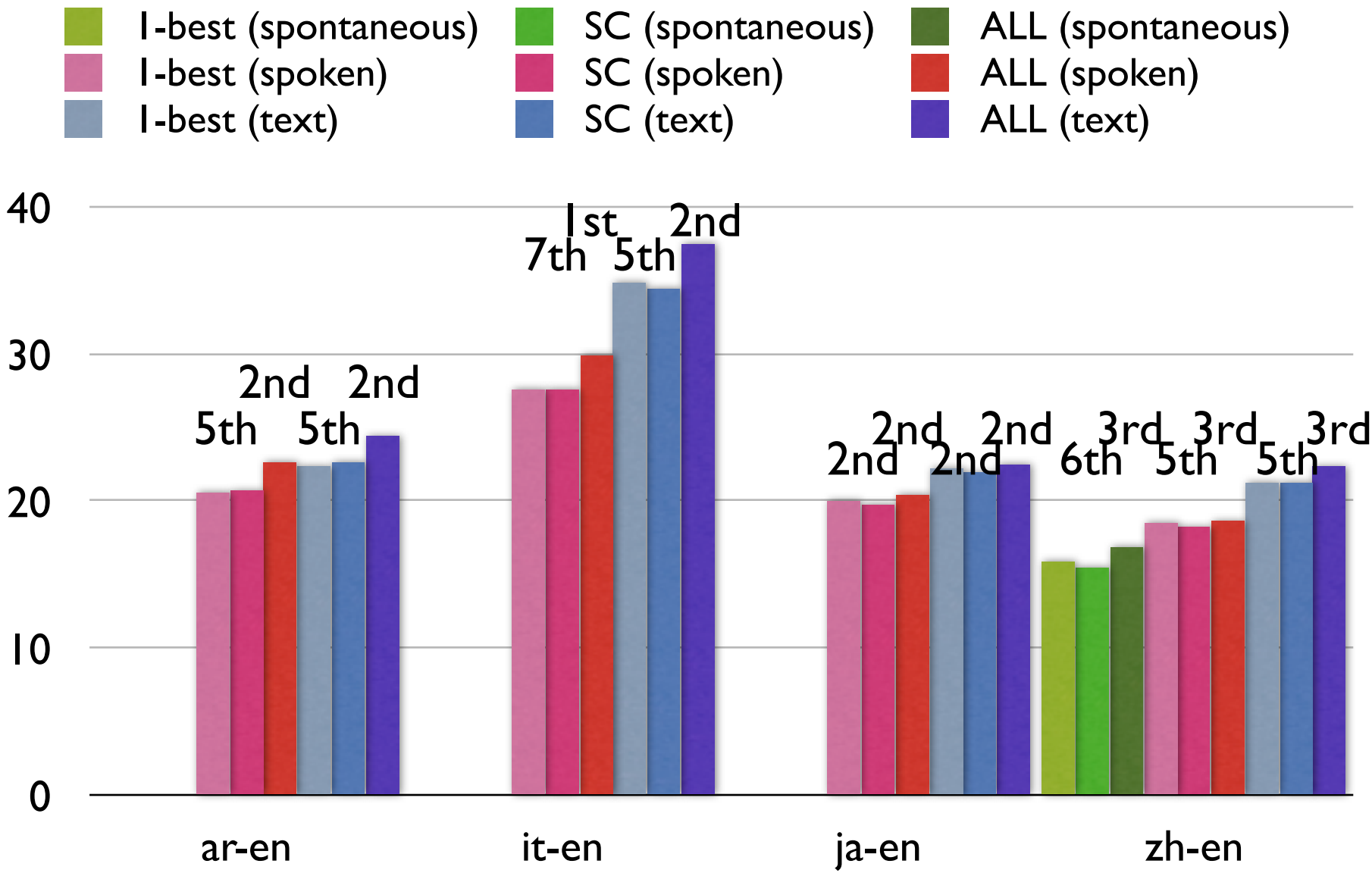
		BLEU	NIST	METEOR	mPER	mWER
ar-en	spoken	20.71 (5th)	4.84	43.97	64.67	56.65
	text	22.65 (5th)	5.33	47.76	62.79	54.15
it-en	spoken	27.69 (7th)	6.70	56.07	57.00	48.13
	text	34.49 (5th)	7.83	64.31	50.79	41.57
ja-en	spoken	19.84 (2nd)	5.48	45.00	71.08	55.12
	text	22.03 (2nd)	5.91	48.77	69.02	52.17
zh-en	spontaneous	15.59 (6th)	4.18	39.46	70.20	59.72
	spoken	18.34 (5th)	4.53	42.15	68.44	57.71
	text	21.35 (5th)	5.13	47.43	65.47	53.70

Remarks: reranked with only SC features

Results on Hierarchical Phrase-based SMT



Results on Reranking



Conclusion

- Better than *non*-hierarchical translation
- Benefit from sparse features (RU, AL) in reranking
- Hierarchical Phrase-based SMT as a baseline
  - Traget normalized form + top-down parsing
- Reranking by Voted Perceptron
  - BLEU-based updates + Approximated-BLEU