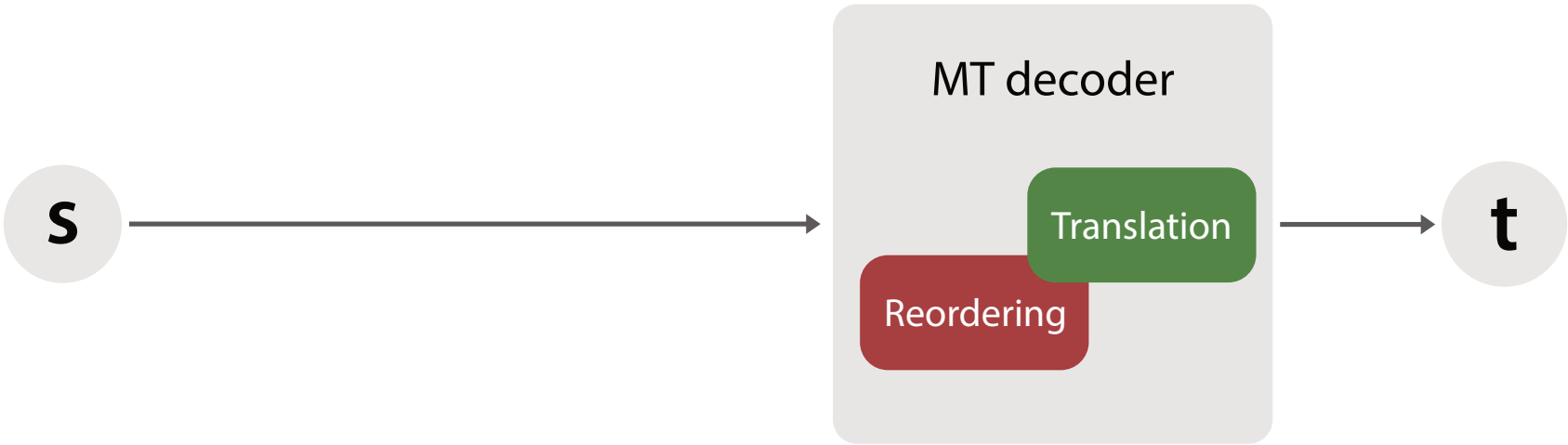
Preordering and Word Order Freedom in Machine Translation

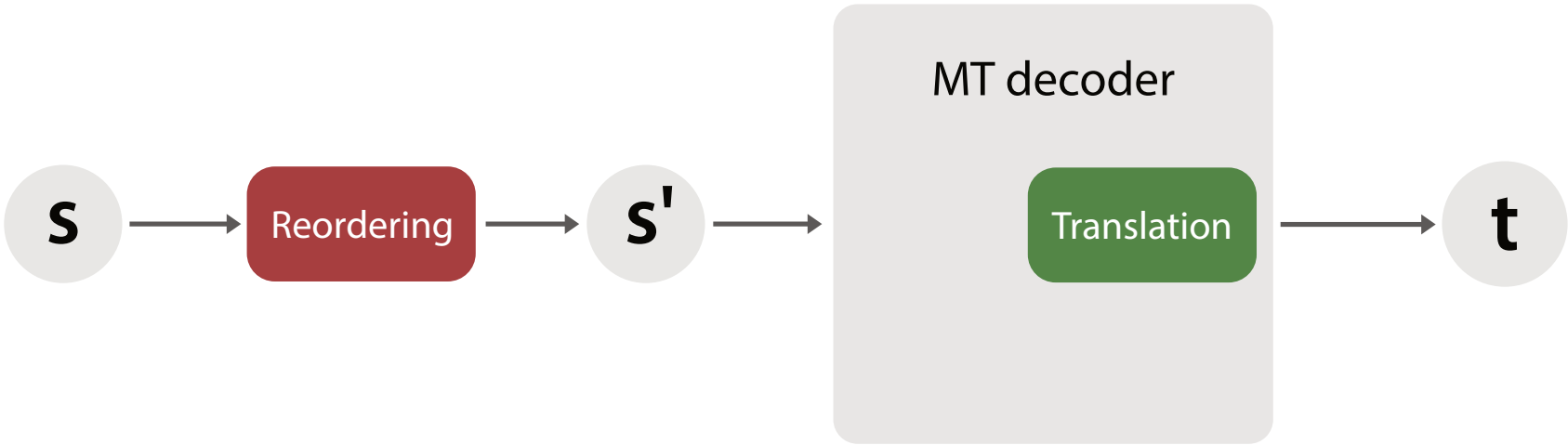
Joachim Daiber, Miloš Stanojević, Wilker Aziz and Khalil Sima'an

Institute for Logic, Language and Computation
University of Amsterdam

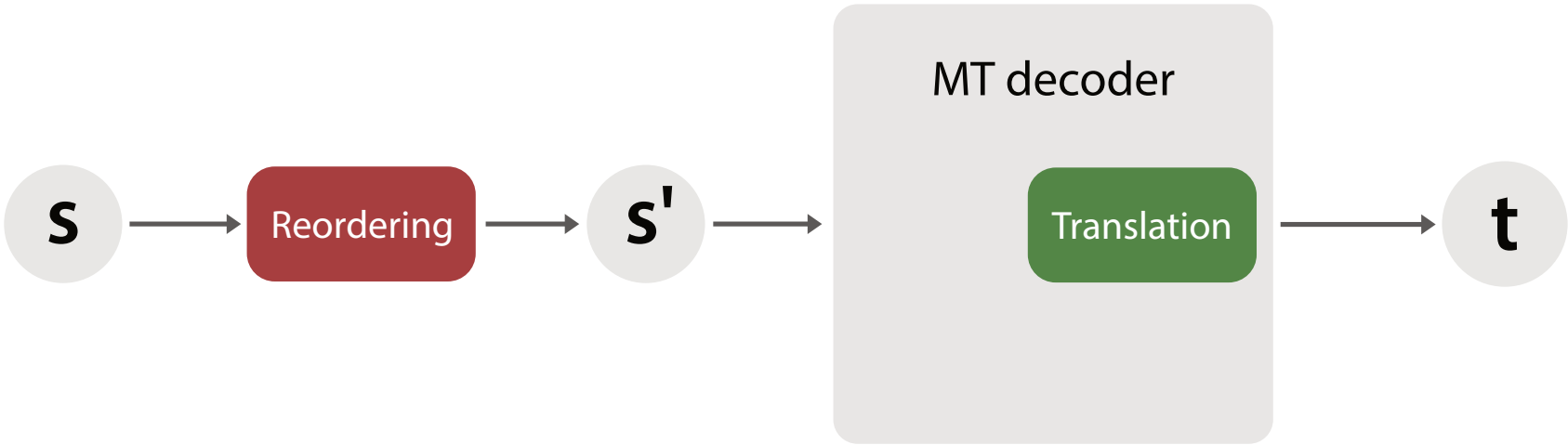
Reordering in PBMT



Reordering in PBMT



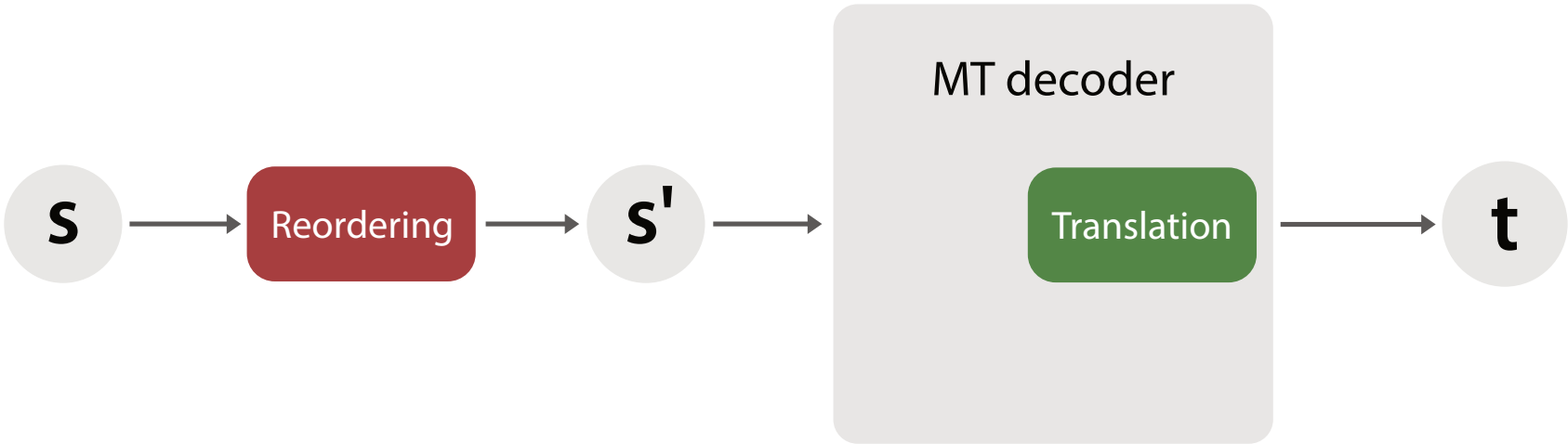
Reordering in PBMT



Preordering

\+ Works well for English–Japanese, English–Korean, ...

Reordering in PBMT



Preordering

+ Works well for English–Japanese, English–Korean, ...
– Not that well for English–German, English–Czech, ...

Reordering in PBMT

Why the partial success?

- **Free word order** languages are difficult: many potential target word orders

Reordering in PBMT

Why the partial success?

- **Free word order** languages are difficult: many potential target word orders
- **Source side often not enough** to predict a unique target word order

Reordering in PBMT

Why the partial success?

- **Free word order** languages are difficult: many potential target word orders
- **Source side often not enough** to predict a unique target word order

Potential solution:

- Narrow down space of target word orders and let decoder decide

Reordering in PBMT

Why the partial success?

- **Free word order** languages are difficult: many potential target word orders
- **Source side often not enough** to predict a unique target word order

Potential solution:

- Narrow down space of target word orders and let decoder decide
- Prepare MT system for this kind of input

Outline

1. Word Order Freedom
   - What exactly do we mean by free word order?
   - Can we measure it?

Outline

1. Word Order Freedom
   - What exactly do we mean by free word order?
   - Can we measure it?
2. Translation with Permutation Lattices
   - Permutation Lattices
   - Lattice Silver Training

Outline

1. Word Order Freedom
   - What exactly do we mean by free word order?
   - Can we measure it?
2. Translation with Permutation Lattices
   - Permutation Lattices
   - Lattice Silver Training
3. Experiments with two language pairs
   - English–Japanese
     Strict word order language
   - English–German
     Free word order language

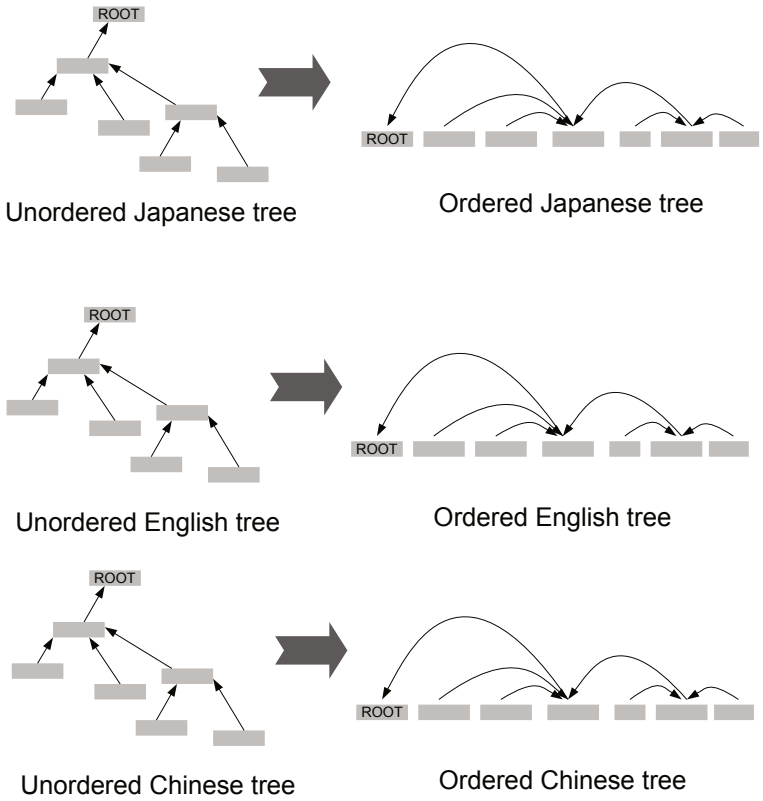
What Does Free Word Order Mean?

- Multiple word orders expressing the same meaning

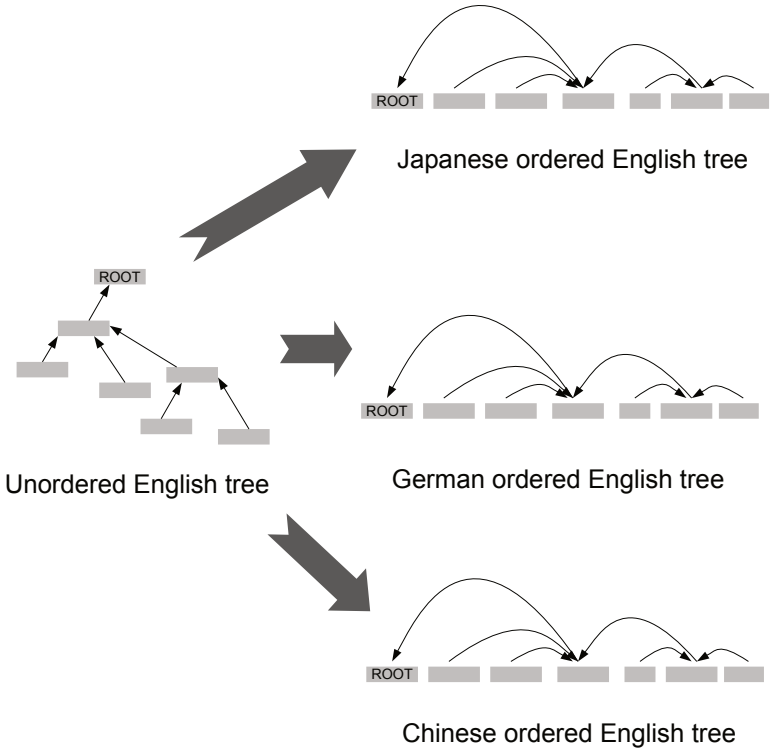
What Does Free Word Order Mean?

- Multiple word orders expressing the same meaning
  $\rightarrow$ Fuzzy definition!
- Can we quantify and compare languages' word order freedom?

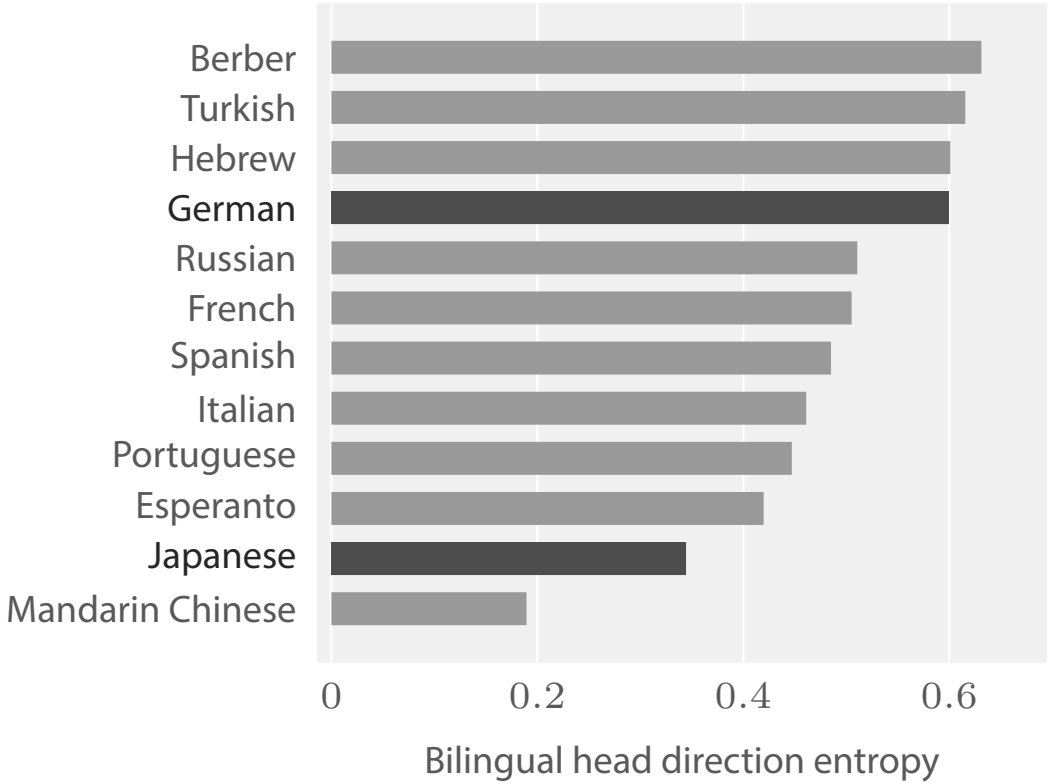
Estimating Word Order Freedom: Futrell et al. (2015)



Estimating Word Order Freedom: Our Bilingual Approach



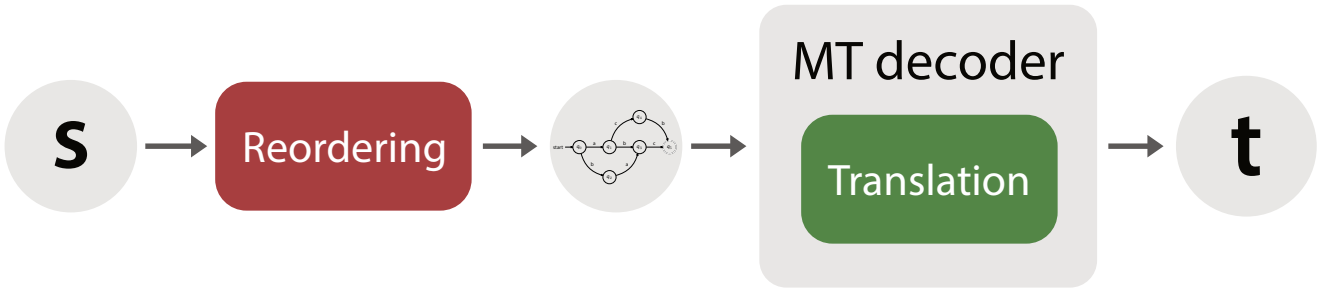
Word Order Freedom



Free Word Order

What to do about Free Word Order?

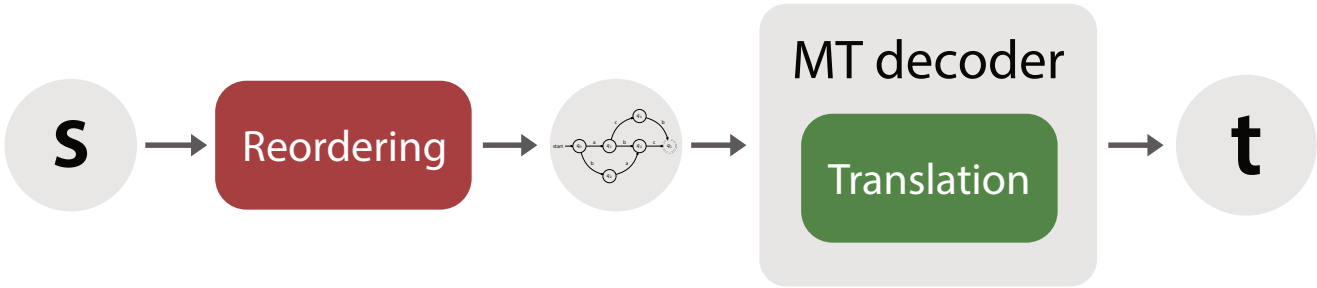
- Instead of choosing 1-best word order, work with *n*-best



Free Word Order

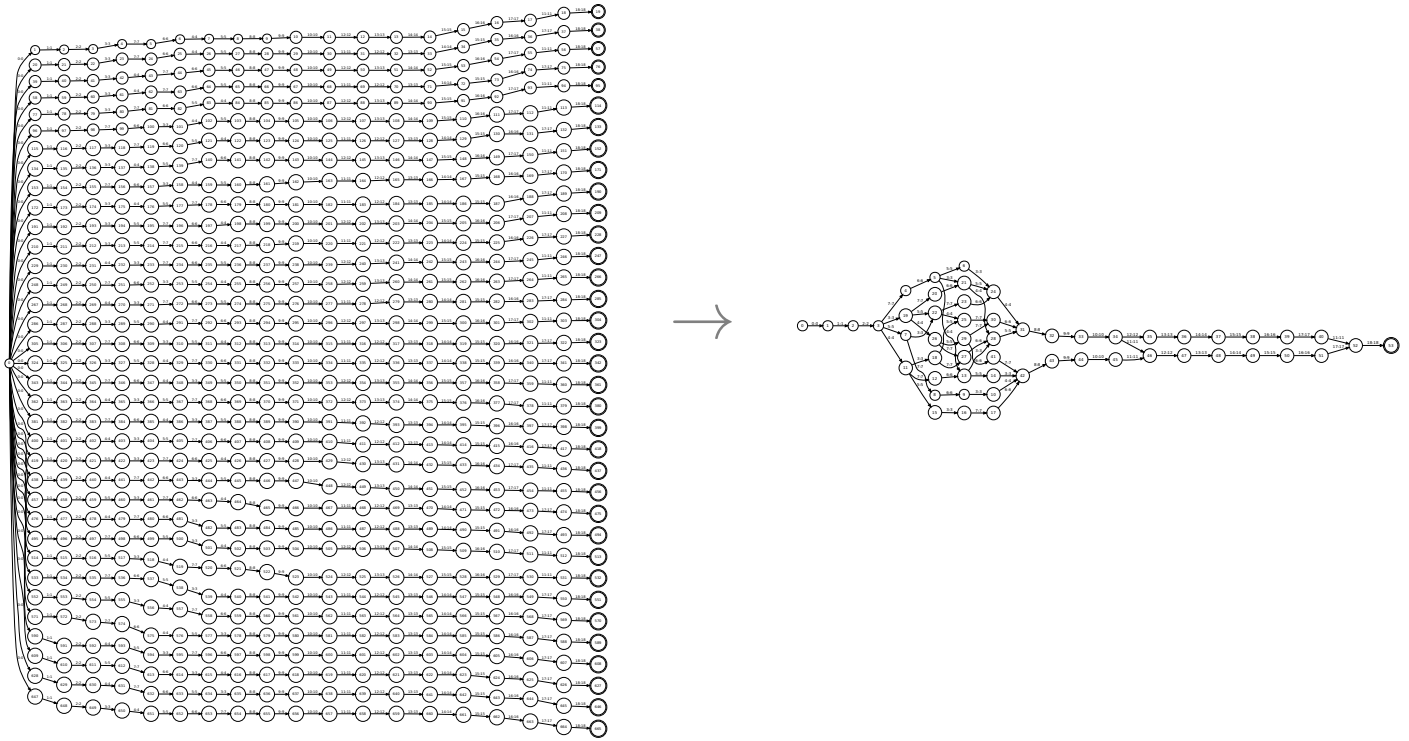
What to do about Free Word Order?

- Instead of choosing 1-best word order, work with *n*-best



- Decoding with *n*-best permutations:
  - **1) Inference:** Compactly encode *n*-best WOs into lattice
  - **2) Training:** Estimate parameters for translating lattices

Permutation Lattices



Lattice Training

Q: How to extract the phrase table?

Gold Training

(target-order $\mathbf{s}'$, $\mathbf{t}$)

+ compact
+ less noisy
– may not cover preordered input

Lattice Training

Q: How to extract the phrase table?

Gold Training	**Predicted Training**
(target-order $\mathbf{s}'$, $\mathbf{t}$)	(predicted $\mathbf{s}'$, $\mathbf{t}$)
+ compact	? less compact
+ less noisy	– more noisy
– may not cover preordered input	+ covers preordered input

Lattice Training

Q: How to extract the phrase table?

Gold Training

(target-order $\mathbf{s}'$, $\mathbf{t}$)

+ compact
+ less noisy
– may not cover preordered input

Predicted Training

(predicted $\mathbf{s}'$, $\mathbf{t}$)

? less compact
– more noisy
+ covers preordered input

Lattice Silver Training

($\mathbf{s}'$ from *n*-best most similar to gold, $\mathbf{t}$)

+ compact
+ less noisy
+ covers preordered input

Preordering Models

Reordering Grammar

Stanojević and Sima'an (2015)

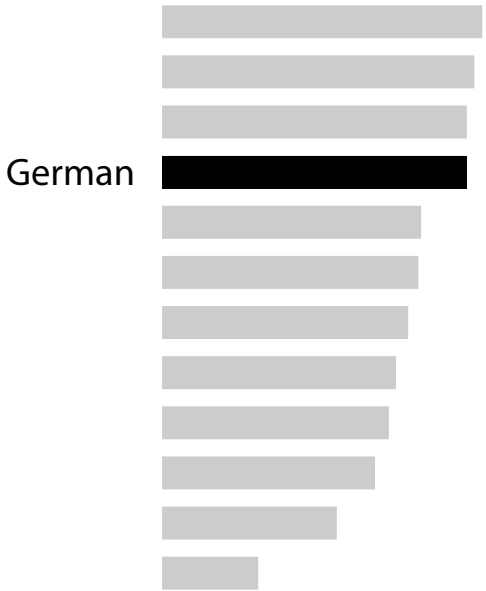
- Hierarchical, probabilistic model
- Unsupervised (no syntax)
- Beyond ITG (based on PETs)

Neural Preordering

de Gispert et al. (2015); Jehl et al. (2014)

- Preorders dep. tree recursively
- NN to decide pairwise ordering of subtrees
- We extend search to *n*-best branch-and-bound

Translation Experiment: English→German

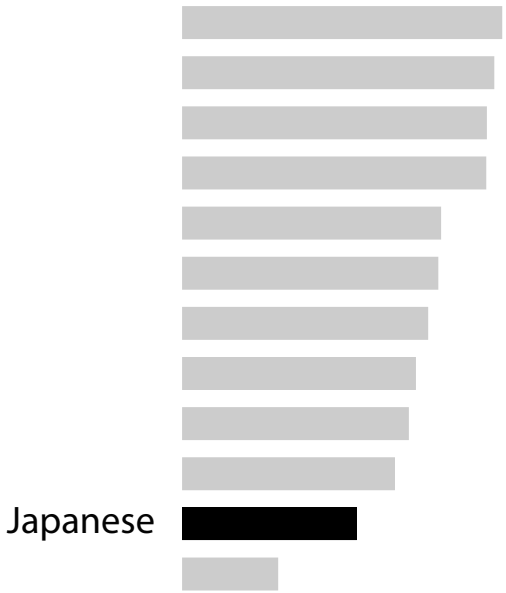


WMT newstest 2015.

- Big oracle WO gain:
  +4.92 over baseline (21.76 BLEU)
- 1-best with DL 6 performs poorly
- Lattices with Silver Training and DL 0
  on par with baseline +0.12

Software: Moses (with extended lattice support). Optimizer: Batch MIRA (15 iterations, BLEU averaged over 3 runs).
Training data: 4.5m sentence pairs (WMT16). LM (5-gram): 189m sentences (WMT16). Dev/Test: newstest2014/newstest2015.

Translation Experiment: English→Japanese



NTCIR Patent translation task.

- Big oracle WO gain:
  +4.57 over baseline (29.65 BLEU)
- 1-best with DL 6 performs well: +2.49
- Lattices with DL 0 enables additional improvement: +2.85

Software: Moses (with extended lattice support). Optimizer: Batch MIRA (15 iterations, BLEU averaged over 3 runs).
Training data: NTCIR-8 Patent Translation Task. Target tokenization: KyTea 5 Neubig et al. (2011). LM (5-gram): NTCIR-8. Dev/Test: NTCIR-7/NTCIR-9.

Summary of our Contributions

Bilingual Head Direction Entropy

- Measure of word order freedom for various languages

Permutation Lattices produced from preordering models:

- Valuable tool to address WO uncertainty
- Additional improvements over 1-best for strict WO languages
- Enable translation of freer WO languages with auto. preordering
  $\rightarrow$ **Lattice Silver Training** for model estimation

Thank You!

Any questions?

de Gispert, A., Iglesias, G., and Byrne, W. (2015). Fast and accurate preordering for smt using neural networks. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2015)*.

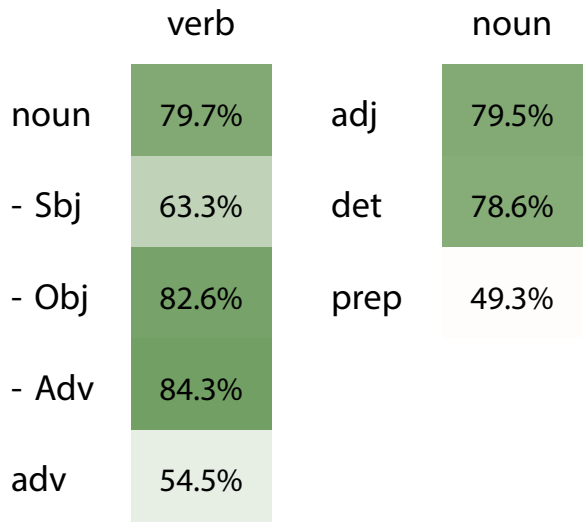
Futrell, R., Mahowald, K., and Gibson, E. (2015). Quantifying word order freedom in dependency corpora. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pages 91–100, Uppsala, Sweden. Uppsala University, Uppsala, Sweden.

Jehl, L., de Gispert, A., Hopkins, M., and Byrne, B. (2014). Source-side preordering for translation using logistic regression and depth-first branch-and-bound search. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 239–248 (12 pages), Gothenburg, Sweden. Association for Computational Linguistics.

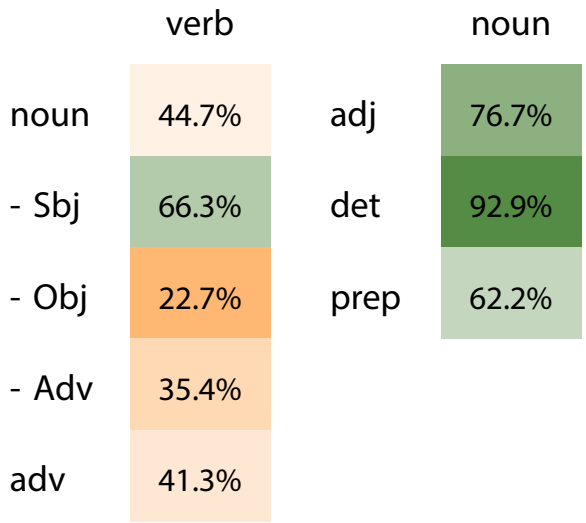
Neubig, G., Nakata, Y., and Mori, S. (2011). Pointwise prediction for robust, adaptable japanese morphological analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 529–533, Portland, Oregon, USA.

Stanojević, M. and Sima'an, K. (2015). Reordering Grammar Induction. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 44–54, Lisbon, Portugal. Association for Computational Linguistics.

Word Order Freedom: Micro Level



(a) English–Japanese



(b) English–German

Number of Permutations vs. Lattice Size

		Lattice	
Permutations	Kendall τ	States	Transitions
Monotone	83.78	23	22
5	84.69	24	52
10	85.23	33	69
100	86.20	72	138
1000	86.75	123	233

Table: Permutations and lattice size (En–De).

Experiments: Neural Reordering on English-German

		Translation	Word order
	DL	BLEU	Kendall τ
Baseline	6	21.76	54.75
Oracle order	6	26.68	58.05
	0	26.41	57.92
First-best	6	21.21[A]	53.44
Lattice (silver)	0	21.88[B]	54.51

Software: Moses (with extended lattice support).
Optimizer: Batch MIRA (15 iterations, BLEU averaged over 3 runs).
Training data: 4.5m sentence pairs (WMT16). LM (5-gram): 189m sentences (WMT16).
Dev/Test: newstest2014/newstest2015.

[A]Stat. significant against baseline. [B]Stat. significant against first-best.

Experiments: Reordering Grammar on English-Japanese

		Translation	Word order
	DL	BLEU	Kendall τ
Baseline	6	29.65	44.87
Oracle order	6	34.22	56.23
	0	30.55	53.98
First-best	6	32.14[A]	49.68
Lattice	0	32.50[AB]	50.79

Software: Moses (with extended lattice support). Optimiser: Batch MIRA (15 iterations).
Training data: NTCIR-8 Patent Translation Task. Target tokenization: KyTea 5 Neubig et al. (2011). LM (5-gram): NTCIR-8. Dev/Test: NTCIR-7/NTCIR-9.

[A]Stat. significant against baseline. [B]Stat. significant against first-best.