Bilingual Markov Reordering Labels
for Hierarchical SMT

Gideon Maillette de Buy Wenniger and Khalil Sima'an

gemdbw AT gmail.com   k.simaan AT uva.nl
http://staff.science.uva.nl/~gemaille/ http://staff.science.uva.nl/~simaan/

Statistical Language Processing and Learning Lab
Institute for Logic Language and Computation
University of Amsterdam, the Netherlands

October 25th, 2014

The incoherence of translation reordering

Sentence type	Sentence contents
Source Sentence	der handlungsspielraum der beiden betroffenen regierung ist also durch das internationale recht begrenzt .
Reference	any action by the two governments concerned is therefore limited by this international law .
Hiero (Baseline)	the margin for manoeuvre of two government is concerned by the international community limited .

Hiero and Memento

Question: what do they have in common?
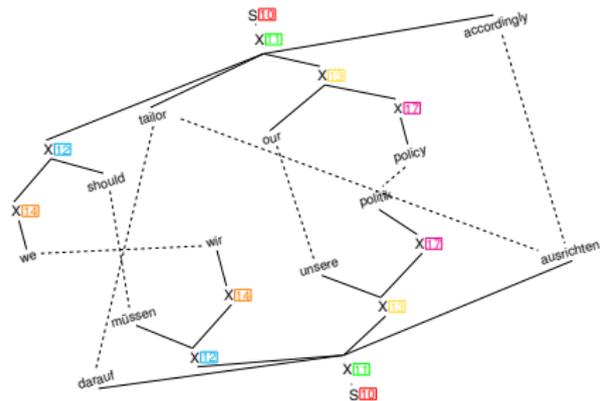
Lexicalization and Language model: the words are not enough

Coherence demands (reordering) context

Vision: Hierarchical Alignment Trees (HATs)
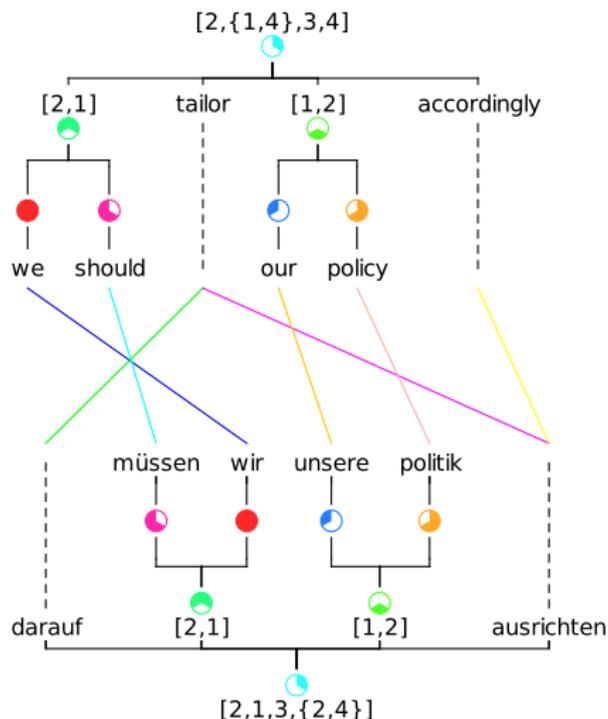
Outline

- Part 1:
  Bilingual Phrase Reordering Labels

- Part 2:
  Label Substitution Features

- Part 3:
  Experiments

- Conclusions

Part 1:
Bilingual Phrase Reordering Labels

NDT with Alignment structure
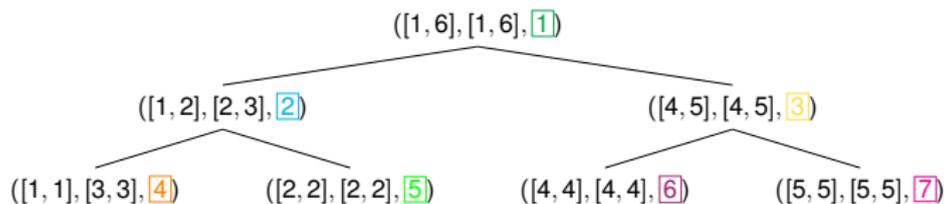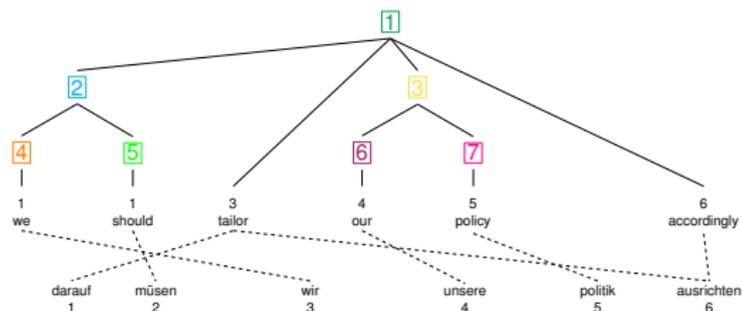
NDT with Alignment structure = HAT
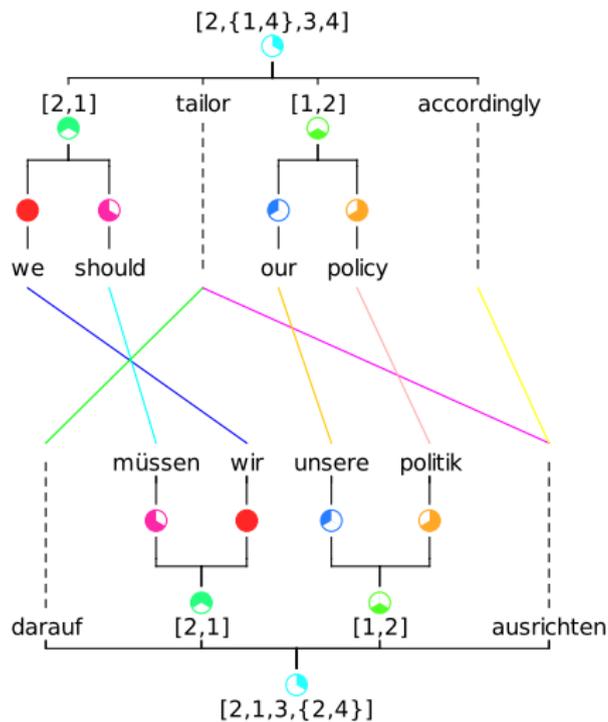
Reordering Labeled Grammar Extraction
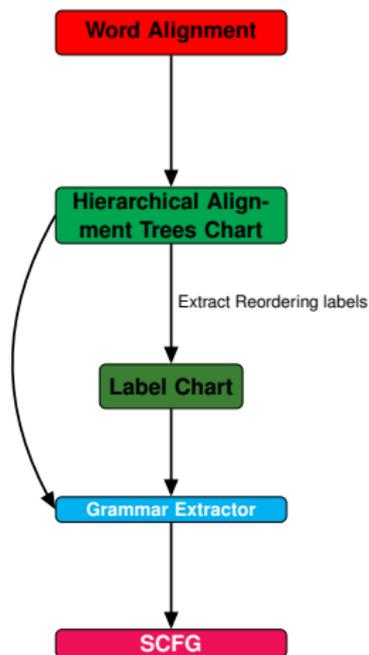
Bilingual Phrase Reordering label categories

- Phrase-Centric

- Parent-Relative

Phrase-centric reordering labels

- Complexity relation between base phrase and children in *HAT* determines label
- Five cases distinguished, ordered by increasing complexity
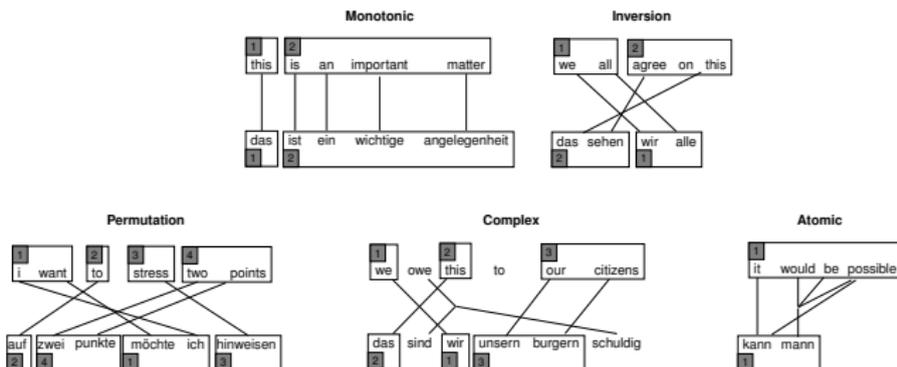
Known labels from ITG and Phrase pair Theory

Monotonic

- *Monotonic*: If the alignment can be split into two monotonically ordered parts.
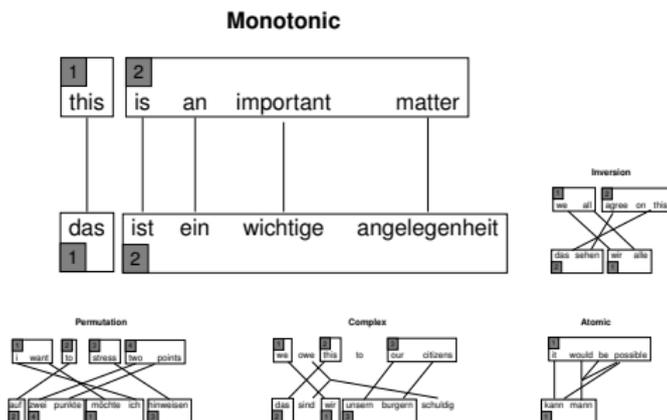
Inverted

- *Inverted*: If the alignment can be split into two inverted parts.
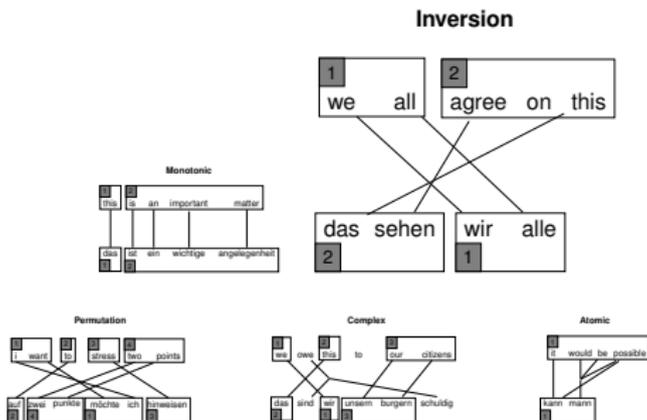
Atomic

- *Atomic*: If the alignment does not allow the existence of smaller (child) phrase pairs.
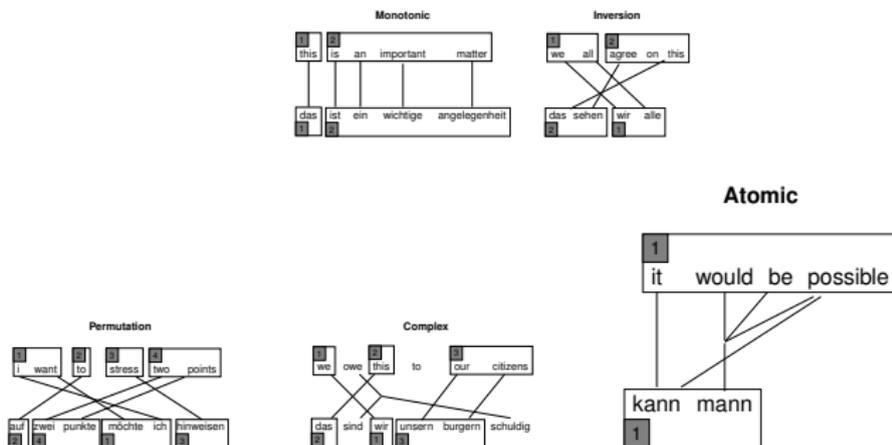
New labels based on HATs

Permutation

- *Permutation*: If the alignment can be factored as a permutation of more than 3 parts.
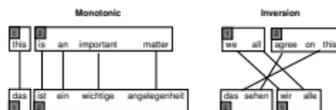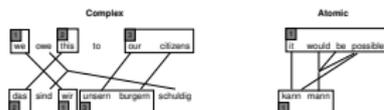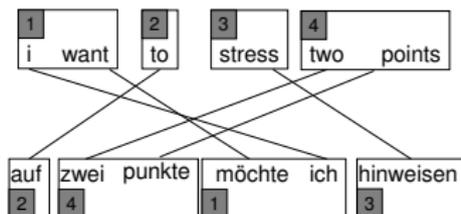
Complex

- *Complex*: No alignment factorization as a permutation of parts, but smaller phrase pair is contained (i.e., it is composite).
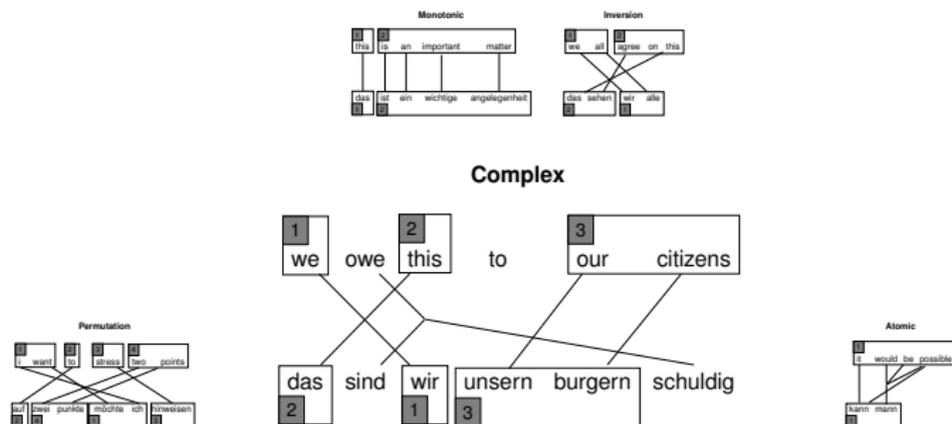
Phrase-Centric labeled derivation

S $_{10}$

S $_{10}$

Phrase-Centric labeled derivation

Phrase-Centric labeled derivation

Phrase-Centric labeled derivation
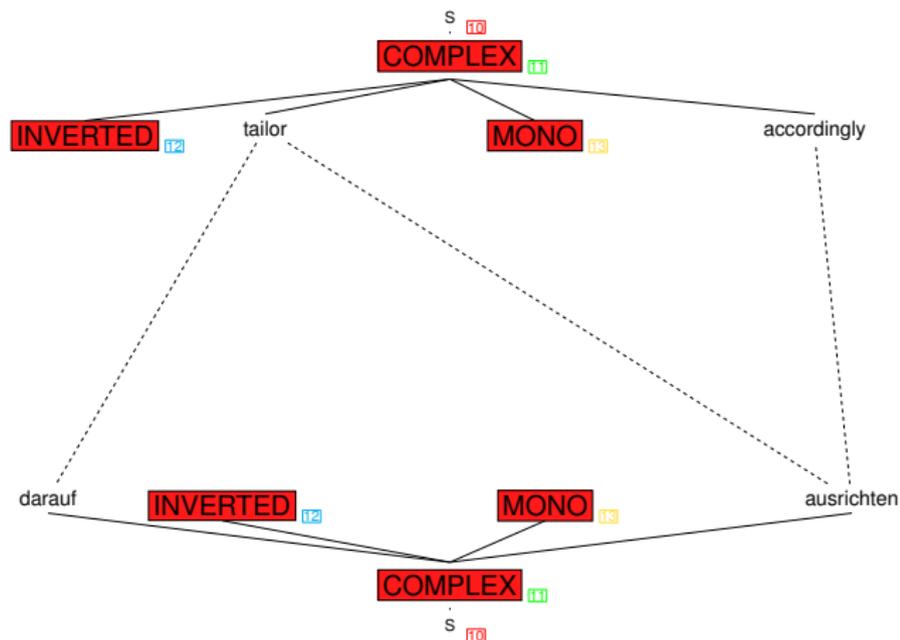
Phrase-Centric labeled derivation
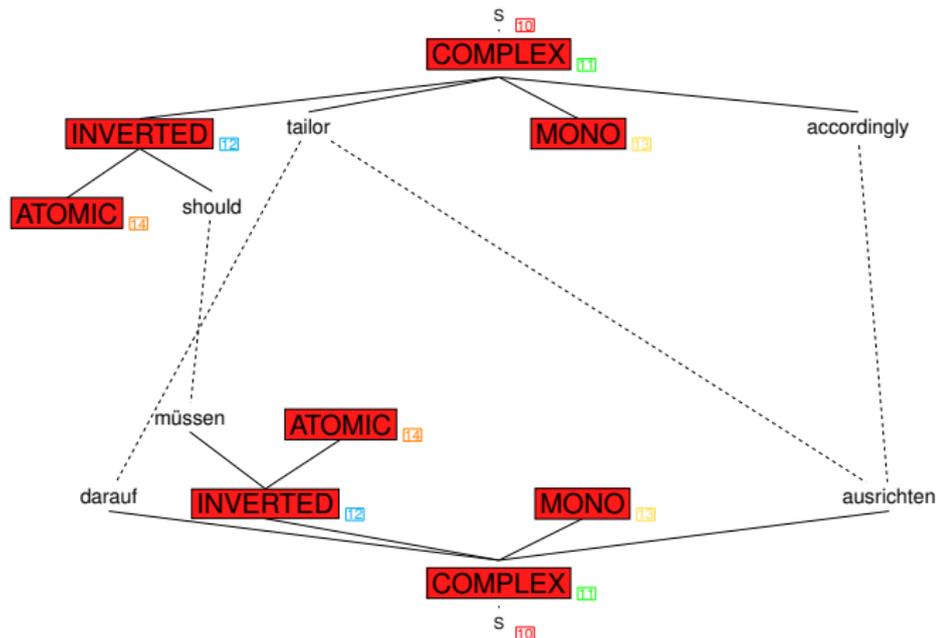
Phrase-Centric labeled derivation
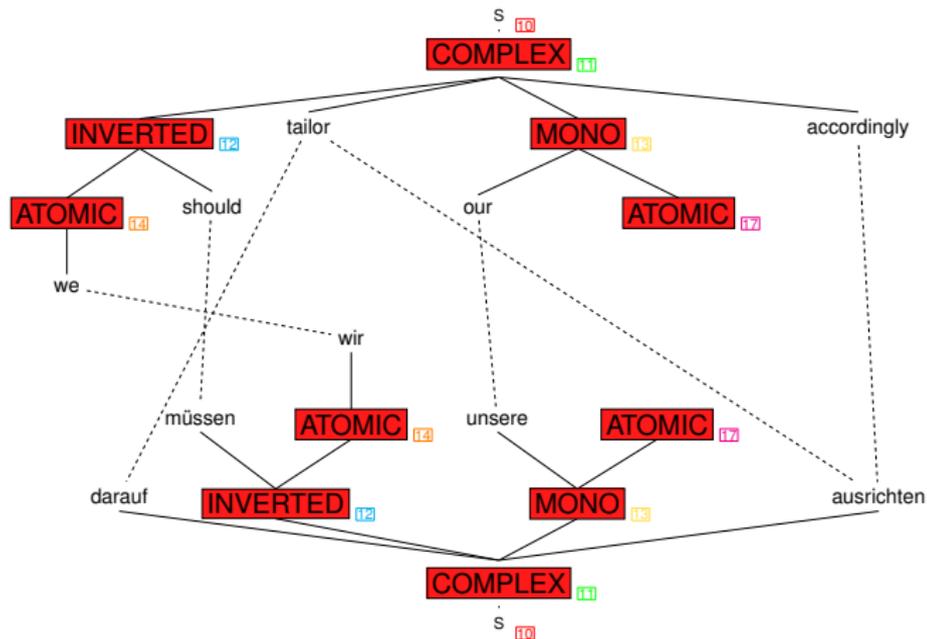
Phrase-Centric labeled derivation
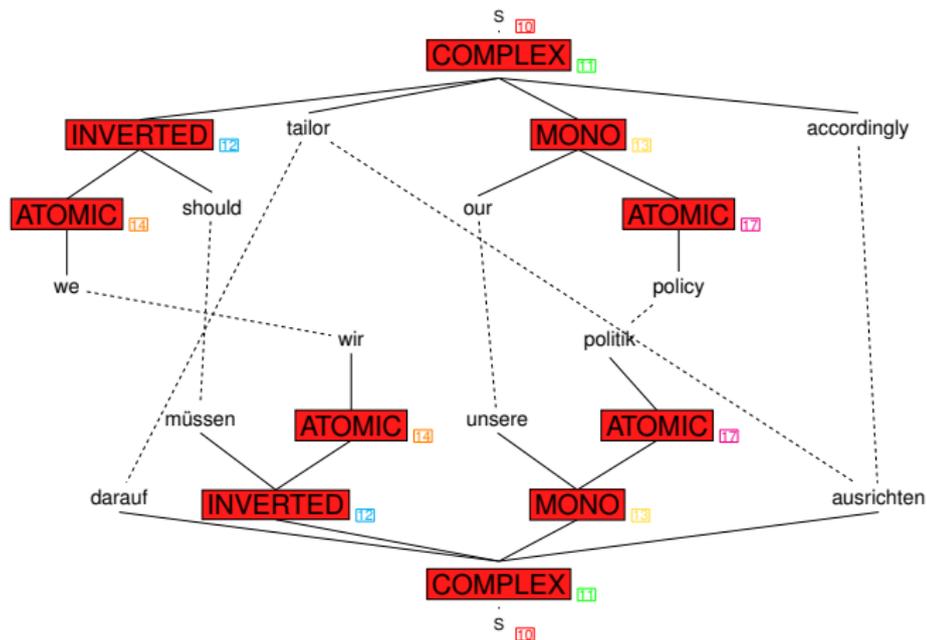
Parent-Relative reordering labels

- Describe type of reordering relative to embedding "parent" phrase

- First-order view on reordering

- (Details ommitted due to time constraints)

Part 2:
Label Substitution Features

Label substitution features

- Unique feature for every label pair $\langle L_\alpha, L_\beta \rangle$

- Marks specific LHS substitutes specific gap

- Two more coarse features:
  - ▸ Match
  - ▸ Nomatch
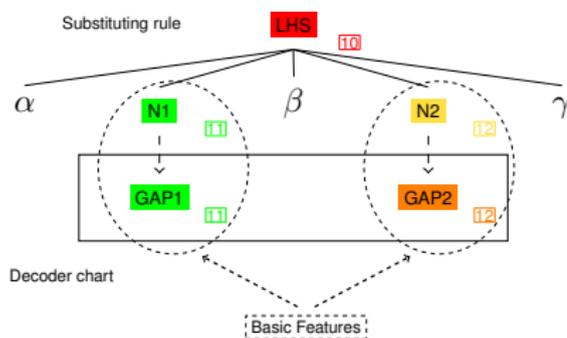
Part 3: Experiments

Motivating Example - After

Sentence type	Sentence contents
Source Sentence	der handlungsspielraum der beiden betroffenen regierung ist also durch das internationale recht begrenzt .
Reference	any action by the two governments concerned is therefore limited by this international law .
Hiero (Baseline)	the margin for manoeuvre of two government is concerned by the international community limited .
Our System	**the scope of the two governments concerned is therefore limited by international law .**

Experimental Setup

- German-English and Chinese-English language pairs
- Data properties

Language pair	dataset type	size	data origin
German-English	train	1M	Europarl
	dev	2K	WMT-07 - dev
	test	2K	WMT-07 - test
Chinese-English	train	7.34M	*MultiUn + Hong Kong Parallel Text*
	dev	2K	*Multiple Translation Chinese*
	test	2K	*Multiple Translation Chinese*

  ▸ Max sentence length 40
- Language model
  ▸ 4-gram language model
  ▸ Kneser-Ney discounting

Experimental Setup - Evaluation

- Evaluation Metrics
  - BLEU
  - METEOR
  - Translation Error Rate (TER)
  - KENDALL-Reordering Score (KRS)

- 3 runs all experiments

- Significance Tests
  - Re-sampling test from MultEval
  - Sign test, used for KRS

Baselines

- Comparison against Hiero and SAMT baselines

- Experiments with Joshua

- Default decoding settings used

Bilingual Phrase reordering labels

Two alternative labeling schemes:

- Hiero-0th
  - Phrase-centric bilingual reordering labels

- Hiero-1st
  - Parent-relative bilingual reordering labels

Two constraint types:

- Strict constraints

- Soft constraints

System Name	DEV				TEST			
	BLEU ↑	METEOR ↑	TER ↓	KRS ↑	BLEU ↑	METEOR ↑	TER ↓	KRS ↑
	German-English							
Hiero	**27.90**	32.69	58.22	66.37	**28.39**	32.94	58.01	67.44
SAMT	27.76	32.67	58.05	**66.84▲**	28.32	32.88	**57.70▲▲**	**67.63**
Hiero-0$^{th}_{ITG+}$	27.85	32.70	58.04▲▲	66.27	28.36	32.90▼	57.83▲▲	67.30
Hiero-0th	27.82	**32.75**	**57.92▲▲**	66.66	**28.39**	33.03▲▲	57.75▲▲	67.55
Hiero-1$^{st}_{Coarse}$	27.86	32.66	58.23	66.37	28.22▼	32.90	57.93	67.47
Hiero-1st	27.74▼	32.60▼▼	58.11	66.44	28.27	32.80▼▼	57.95	67.39
	Chinese-English							
Hiero	31.70	30.72	**61.21**	58.28	31.63	30.56	**59.28**	58.03
SAMT	**31.98▲**	30.81▲	61.83▼▼	**60.71▲▲**	31.87	30.61	59.97▼▼	**59.94▲▲**
Hiero-0$^{th}_{ITG+}$	31.54	**30.97▲▲**	62.79▼▼	59.54▲▲	**31.94▲▲**	**30.84▲▲**	60.76▼▼	59.45▲▲
Hiero-0th	31.66	30.95▲▲	62.20▼▼	60.00▲▲	31.90▲▲	30.79▲▲	60.11▼▼	59.68▲▲
Hiero-1$^{st}_{Coarse}$	31.64	30.75	61.37	59.48▲▲	31.57	30.57	59.58▼▼	59.13▲▲
Hiero-1st	**31.74**	30.79	61.94▼▼	**60.22▲▲**	31.77	30.62	60.13▼▼	59.89▲▲

Main Results Soft Constraints

	DEV				TEST			
System Name	BLEU ↑	METEOR ↑	TER ↓	KRS ↑	BLEU ↑	METEOR ↑	TER ↓	KRS ↑
	German-English							
Hiero	27.90	32.69	58.22	66.37	28.39	32.94	58.01	67.44
SAMT	27.76	32.67	58.05	**66.84**▲	28.32	32.88	**57.70**▲▲	**67.63**
Hiero-0$^{th}_{ITG+}$-Sft	28.00▲	32.76▲▲	**57.90**▲▲	66.17	**28.48**	32.98	57.79▲▲	67.32
Hiero-0th-Sft	28.01▲	32.71	57.95▲▲	66.24	28.45	32.98	57.73▲▲	67.51
Hiero-1$^{st}_{Coarse}$-Sft	27.94	32.69	57.91▲▲	66.26	28.45▲	32.94	57.75▲▲	67.36
Hiero-1st-Sft	**28.13**▲▲	**32.80**▲▲	57.92▲▲	66.32	28.45	**33.00**▲	57.79▲▲	67.45
	Chinese-English							
Hiero	31.70	30.72	61.21	58.28	31.63	30.56	59.28	58.03
SAMT	**31.98**▲	30.81▲	61.83▼▼	**60.71**▲▲	31.87	30.61	59.97▼▼	**59.94**▲▲
Hiero-0$^{th}_{ITG+}$-Sft	31.88▲	30.46▼▼	**60.64**▲▲	57.82▼	31.93▲▲	30.37▼▼	**58.86**▲▲	57.60▼
Hiero-0th-Sft	32.04▲▲	30.90▲▲	61.47▼▼	59.36▲▲	32.20▲▲	30.74▲▲	59.45▼	58.92▲▲
Hiero-1$^{st}_{Coarse}$-Sft	32.39▲▲	31.02▲▲	61.56▼▼	59.51▲▲	32.55▲▲	30.86▲▲	59.57▼▼	59.03▲▲
Hiero-1st-Sft	**32.63**▲▲	**31.22**▲▲	62.00▼▼	**60.43**▲▲	**32.61**▲▲	**30.98**▲▲	60.19▼▼	**59.84**▲▲

Label substitution feature weights analysis

Substituted label	Substituted to label							
	X	E.F.D.	E.F.I.	E.F.M.	L.B.I.	L.B.M.	R.B.I.	R.B.M.
	German-English							
E.F.D.	-3.09E-2	-5.67E-3	-2.86E-2	-3.95E-2	-2.00E-2	-3.12E-2	-3.19E-2	-2.84E-2
E.F.I.	-2.77E-2	-2.37E-2	-8.76E-3	-3.53E-2	-2.72E-2	-3.48E-2	-2.72E-2	-2.84E-2
E.F.M.	-9.00E-3	-2.56E-2	-3.31E-2	-2.08E-2	-5.29E-2	-2.98E-2	-3.10E-2	-4.22E-2
L.B.I.	-4.82E-2	-2.51E-2	-2.00E-2	-4.57E-2	**2.96E-2**	-2.50E-2	-2.01E-2	-2.95E-2
L.B.M.	-3.82E-2	-3.15E-2	-2.26E-2	-6.18E-2	-2.55E-2	-1.30E-2	-2.85E-2	-3.03E-2
R.B.I.	-3.06E-2	-2.61E-2	-2.66E-2	-6.99E-3	-2.51E-2	-2.84E-2	**8.22E-3**	-2.37E-2
R.B.M.	-8.69E-3	-1.35E-2	-3.18E-2	-2.24E-2	-1.93E-2	-2.10E-2	**3.56E-3**	**1.13E-2**
TOP.	-2.33E-2	-2.76E-2	-2.74E-2	-2.63E-2	-2.75E-2	-2.91E-2	-2.76E-2	-2.73E-2
	Chinese-English							
E.F.D.	-3.38E-2	**8.40E-3**	-1.99E-2	-1.94E-2	-2.79E-2	-2.98E-2	-2.53E-2	-2.81E-2
E.F.I.	-3.82E-2	-2.18E-2	**1.01E-2**	-5.19E-2	-2.27E-2	-3.24E-2	-2.57E-2	-3.05E-2
E.F.M.	-2.00E-2	-3.91E-2	-4.87E-2	**9.23E-4**	-4.10E-2	-1.76E-3	-5.52E-2	-4.43E-2
L.B.I.	-4.21E-2	-1.91E-2	-1.81E-2	-3.84E-2	**1.47E-2**	-2.02E-2	-2.97E-2	-3.12E-2
L.B.M.	-4.24E-2	-3.04E-2	-1.97E-2	-3.32E-2	-2.66E-2	**1.88E-3**	-2.67E-2	-4.15E-2
R.B.I.	-1.62E-2	-2.22E-2	-1.99E-3	-4.16E-2	-3.62E-2	-3.08E-2	**1.21E-2**	-3.53E-2
R.B.M.	**4.06E-3**	**8.44E-3**	-3.50E-2	-4.52E-2	-3.00E-2	-2.86E-2	-6.67E-3	**8.70E-4**
TOP.	-2.72E-2	-2.46E-2	-2.77E-2	-3.00E-2	-2.82E-2	-2.37E-2	-2.79E-2	-2.84E-2

Do we really need soft-matching?

- Best sytem strict matching (Chinese-English): 31.94 BLEU
- Best sytem soft-matching (Chinese-English): 32.61 BLEU
  - Improvement: 0.67 BLEU
- Labels are coarse (only 5 / 8 cases)
- Feature weights (Chinese-English) show strong preference matching

- Suggests soft-matching has strong merit, at least complementary (not entirely overlapping) to proper learning labels

Conclusions

- *Bilingual phrase reordering labels* improve reordering and lexical selection for Hierarchical SMT

- Using soft, not strict constraints is important to be successful

- Results also far superior to syntax-labeled translation (SAMT) for Chinese-English

- Major improvements for Chinese-English, up to $\pm$ 1 BLEU point

Questions?