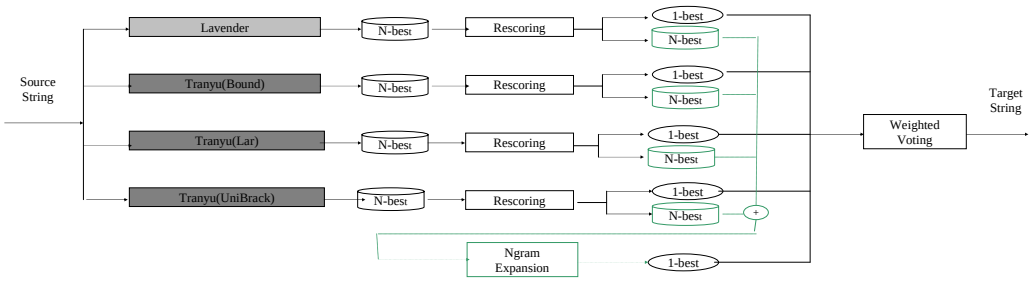
I²R's Machine Translation System for IWSLT 2009

Xiangyu Duan, Deyi Xiong, Hui Zhang, Min Zhang and Haizhou Li
Institute for Infocomm Research, Singapore
{xduan, dyxiong, mzhang, hli}@i2r.a-star.edu.sg, zhangh1982@gmail.com

1. Three Layer Machine Translation Framework: individual systems + rescoring + system combination



Individual Systems

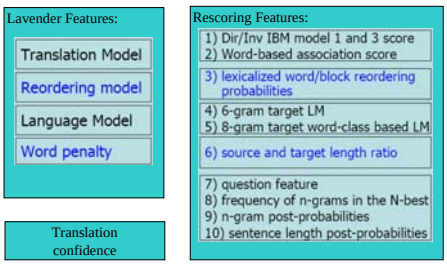
Lavender: Phrasal Translation System

Tranyu: Max-entropy-based ITG Translation System

- **Bound:** boundary words based reordering model
- **Lar:** boundary words based reordering + linguistically annotated reordering mode
- **UniBrack:** boundary words based reordering + syntax-driven bracketing model

Rescoring

Rich additional feature functions



System Combination

- ***n*-gram expansion:** generation of new hypotheses
  - Collect all the n-grams from the original n-best
  - Continuously expand the partial hypothesis through the n-grams
- **Weighted voting**: sentence level system combination
  - 1-bests of four individual systems: Lavender, Tranyu(Bound), Tranyu(Lar) and Tranyu(UniBrack)
  - 1-best of n-gram expansion

2. Experimental Results

2.1 Baseline Performance

Alignmentor		*Baseline*	*+dev*
Berkeley	*Dev*	0.4630	-
	Tst	0.4526	**0.4629**

Lavender as our baseline system
Alignmentor: alignment by agreement (Berkeley)
+dev: add development set (CSTAR 2003) into training set
Tst: test set in lab (IWSLT 2008 test set)

2.2 Performance of Word Alignments Combination

Performance of combination of Berkeley alignments

	grow	***grow-diag***	***Grow+grow-diag***
Dev	0.4630	0.4609	0.4635
Tst	0.4526	0.4472	0.4598
Tst (+dev)	0.4629	0.4523	**0.4732**

Performance of combination between Giza++ and Berkeley's alignments

	Giza++_grow-diag-final-and + Berkeley_grow+grow-diag
Dev	0.4760
Tst	0.4636
Tst (+dev)	0.4710

2.3 Performance of Rescoring

		Lavender	***Tranyu: Bound***	***Tranyu: UniBrack***	***Tranyu: LAR***
Dev	***before***	0.4635	0.4719	0.4478	0.4597
	after	0.4858	0.4951	0.4882	0.5008
Tst	***before***	0.4598	0.4521	0.4471	0.4594
	after	0.4618	0.4715	0.4743	0.4760
Tst (+dev)	***before***	0.4732	0.4604	0.4572	0.4589
	after	0.4799	0.4755	**0.4816**	0.4790

- rescoring on Lavender's outputs improves performance marginally
- rescoring on syntax-based systems improves performance dramatically. UniBrack achieves the best performance after rescoring.

2.4 Performance of System Combination

	Dev	***Tst***	***Tst (+dev)***
UniBrack	0.4882	0.4743	0.4816
n-gram expan	0.5106	0.4841	0.4880
weighted vote	0.5185	0.4897	**0.4944**

3. Official Result of BTEC: Chinese-To-English

Primary	***Contrastive1***	***Contrastive2***	***Contrastive3***
0.4595	0.4599	0.4441	0.4527

Primary: weighted vote on *train+d+t*
•*Contrastive1*: weighted vote on *train+d*
•*Contrastive2*: n-gram expansion on *train+d+t*
•*Contrastive3*: UniBrack on *train+d+t*

4. Conclusion

•System combination framework incorporates mainly two kinds of our in-house SMT systems
  - phrase-based system: Lavender
  - ITG-based system: Tranyu

•Following the workflow of word alignment combination, rescoring and system combination, we improve performance significantly over baseline.