

I²R Multi-Pass Machine Translation System for IWSLT-2008

Boxing Chen, Deyi Xiong, Min
Zhang, Aiti Aw and Haizhou Li

Department of HLT
Institute for Infocomm Research
(I²R), Singapore

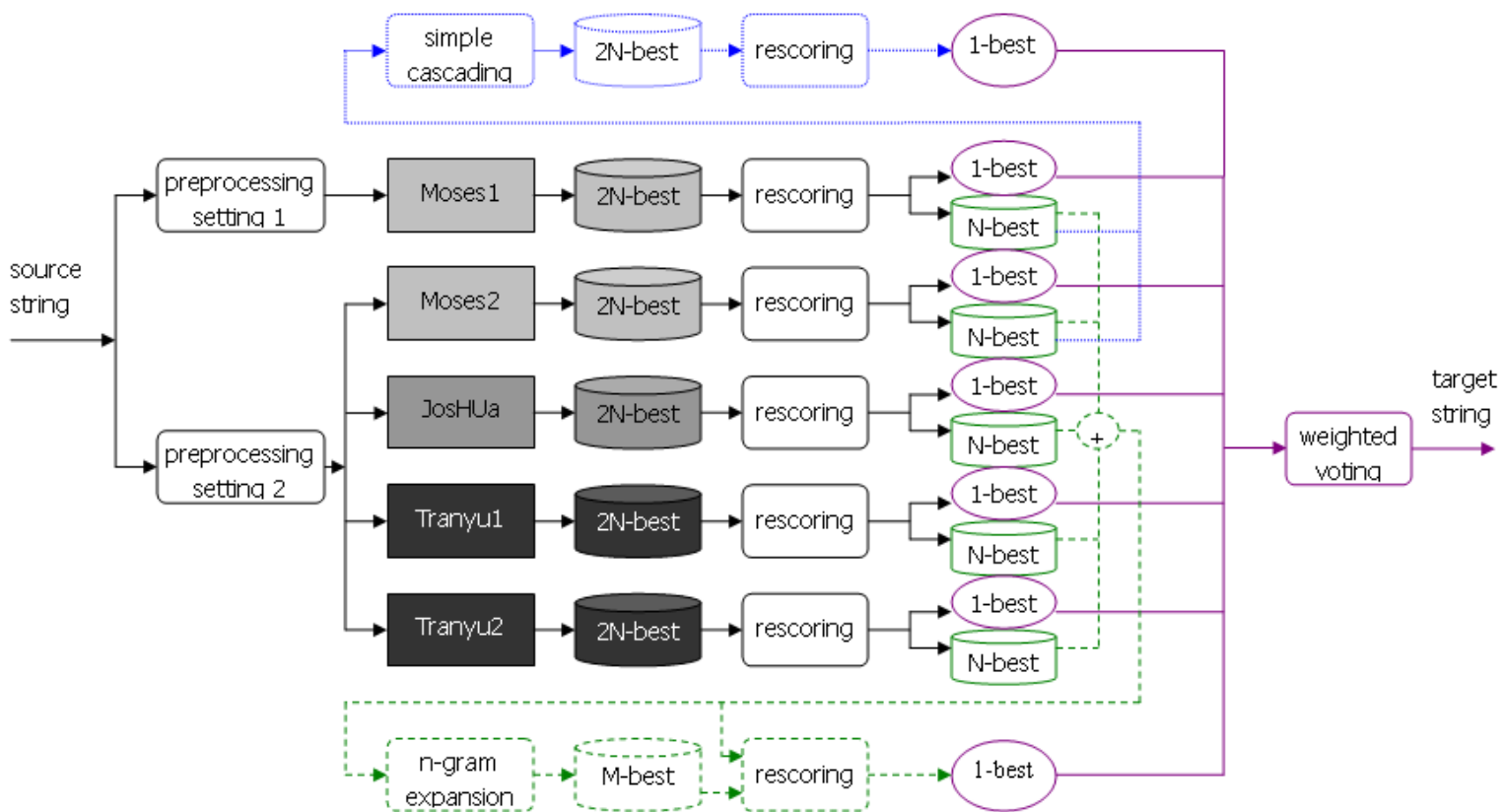
Tasks

- Chinese-to-English
 - BTEC task
 - Challenge task
- Chinese-to-English-to-Spanish PIVOT task
 - Joint effort with UPC-TALP
 - Will be reported by the co-worker of UPC

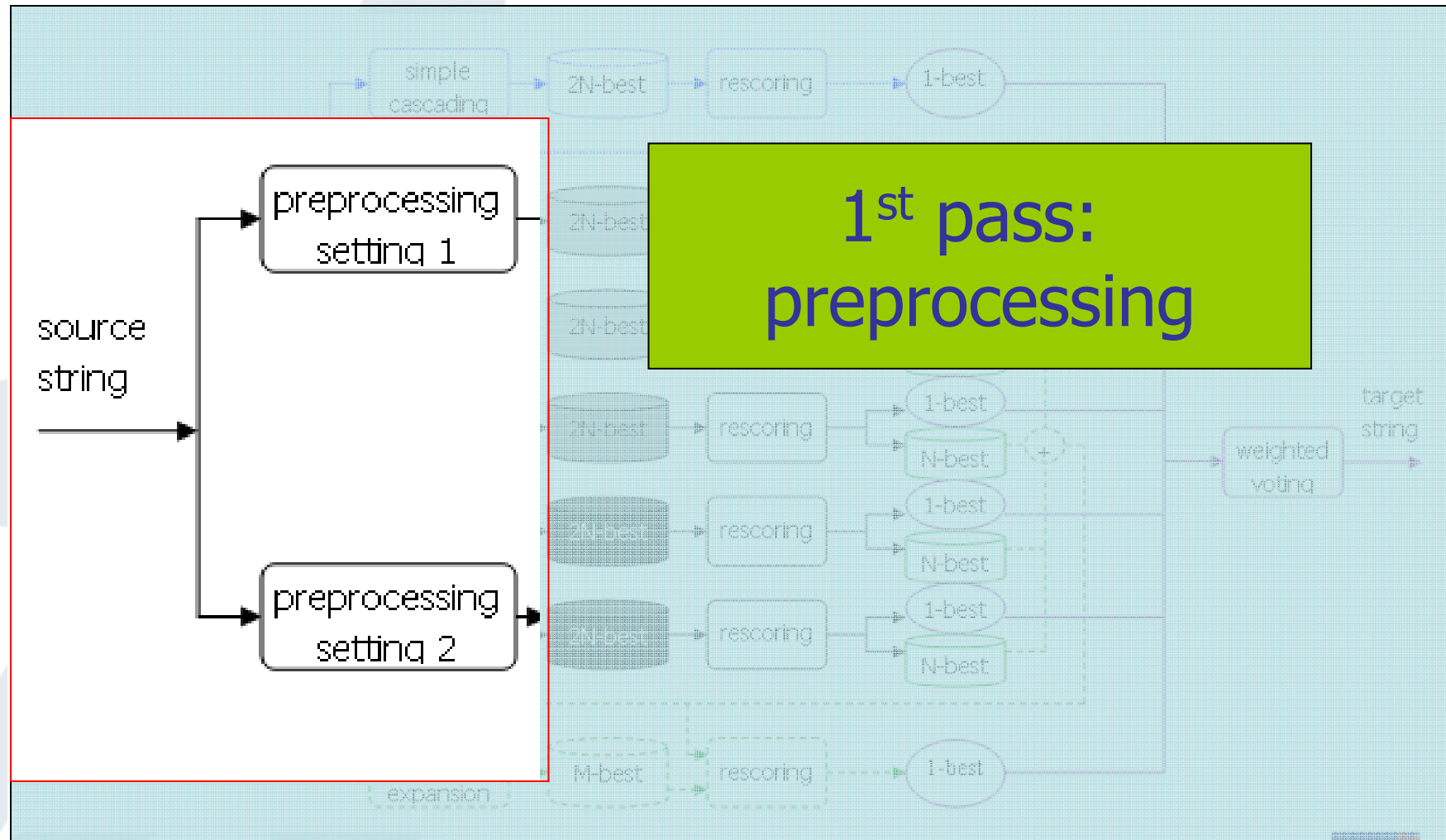
Outline

- Multi-pass MT System
 - System Architecture
 - 1st pass: preprocessing
 - 2nd pass: decoding
 - 3rd pass: rescoring
 - 4th pass: system combination
- Experiments and results
- Conclusion

System Architecture



Preprocessing

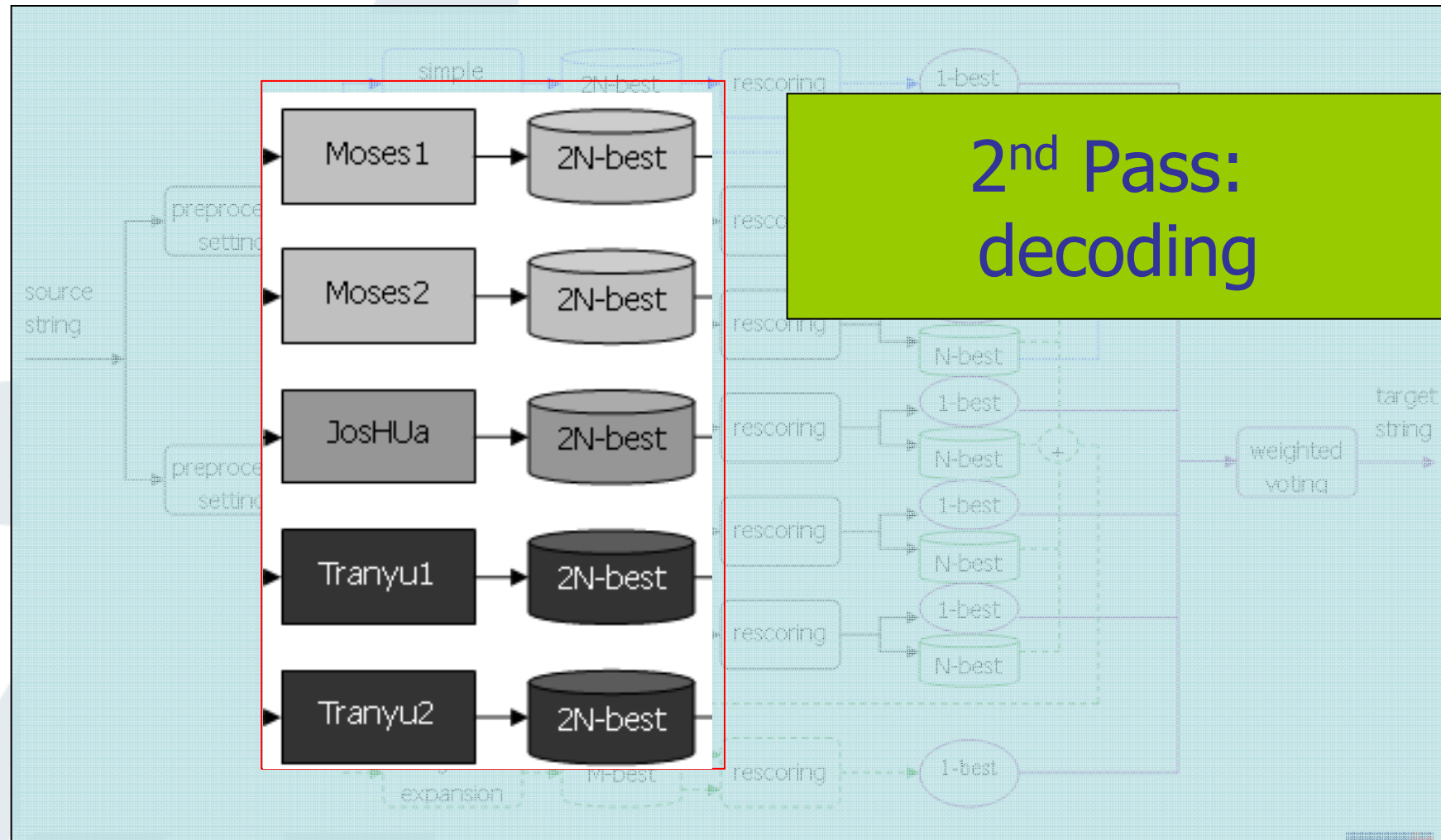


Preprocessing

- Preprocessing
 - Tools: LDC-SEG (L) , ICTCLAS (I)

	Preprocessing 1		Preprocessing 2	
	ch	en	ch	en
Tokenization	L	x	I	x
Txt-to-digit	x	x		
Lower-casing		x		x

Decoding



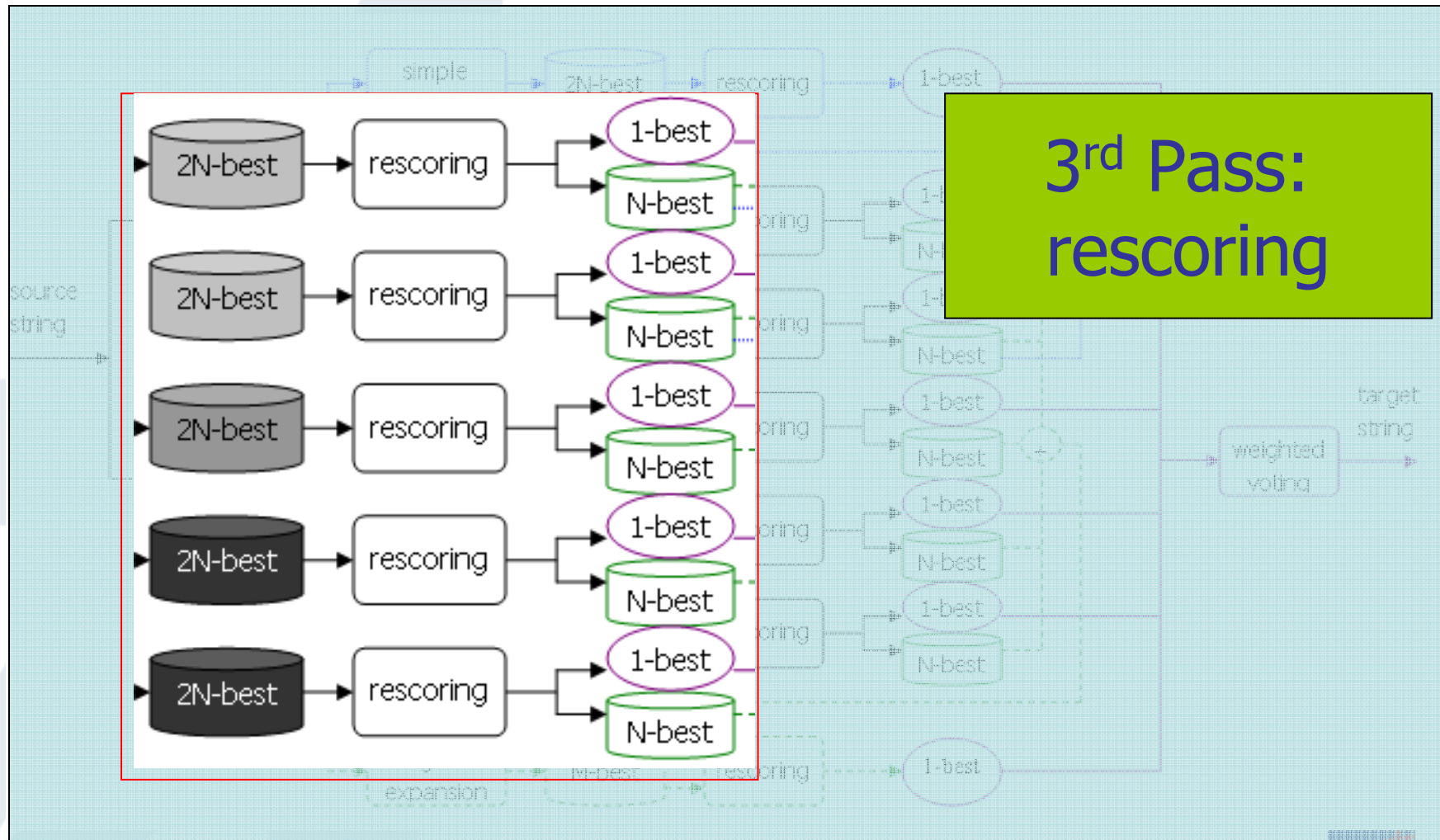
Decoding

- Preprocessing Setting
 - Moses1: preprocessing 1
 - All other 4 systems: preprocessing 2
- Moses: (open source)
 - Phrase-based system
- JosHUa: (open source)
 - Hierarchical phrase-based MT system
- Tranyu: (in-home)
 - BTG-based system

Tranyu

- Adapting BTG to phrasal translation
- CKY-style decoder
- Reordering models
 - MaxEnt-based
 - Features:
 - Boundary words (Xiong et al. 2006)
 - Linguistic annotations (Xiong et al. 2008)
- Systems
 - Tranyu 1: boundary words based reordering model
 - Tranyu 2: boundary words based reordering + linguistically annotated reordering model

Rescoring



Rescoring

- Rich additional feature functions

Moses Features:

Translation Model

Reordering model

Language Model

Word penalty

Translation
confidence

Rescoring Features:

- 1) Dir/Inv IBM model 1 and 3 score
- 2) Word-based association score

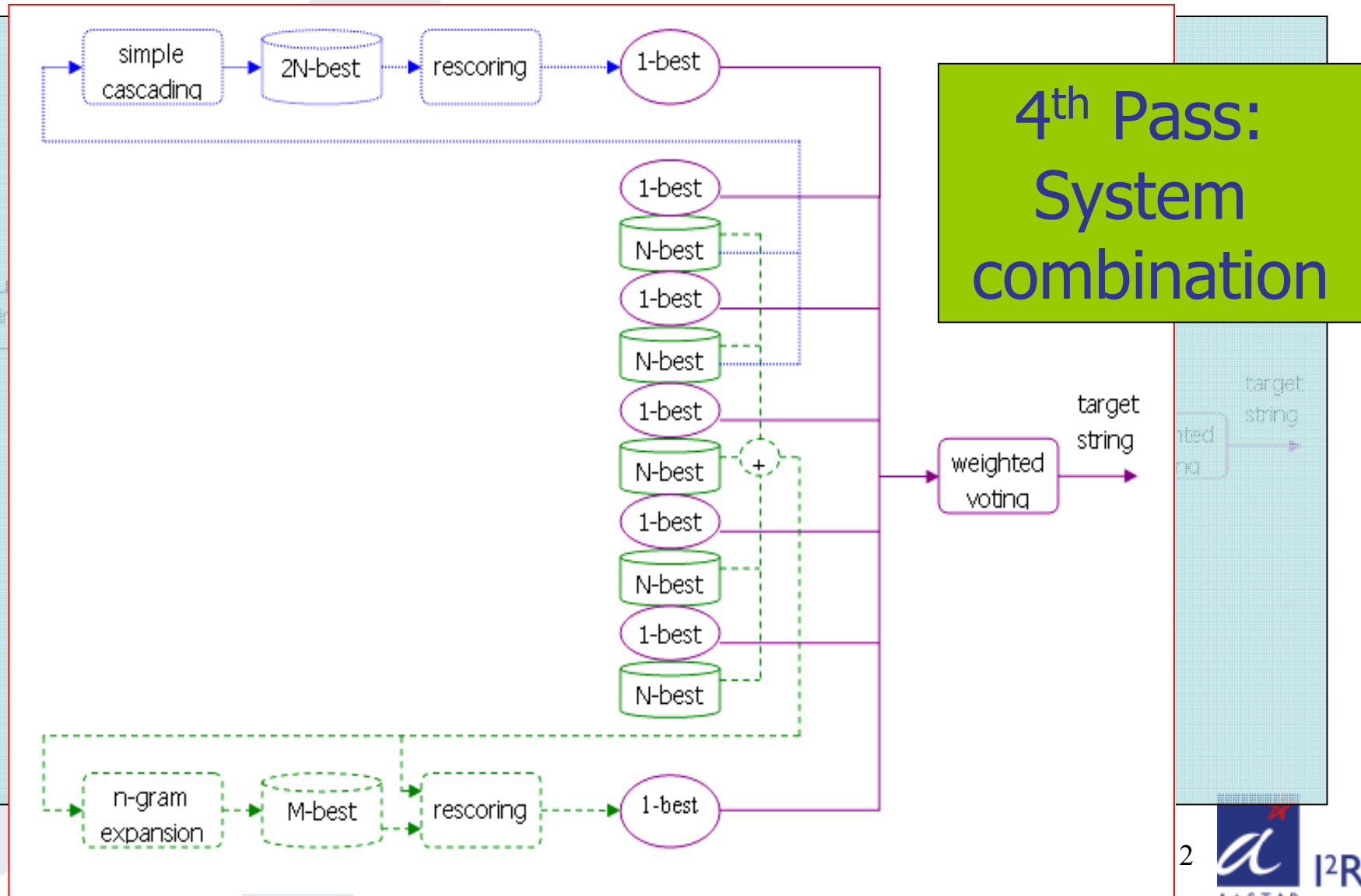
3) lexicalized word/block reordering probabilities

- 4) 6-gram target LM
- 5) 8-gram target word-class based LM

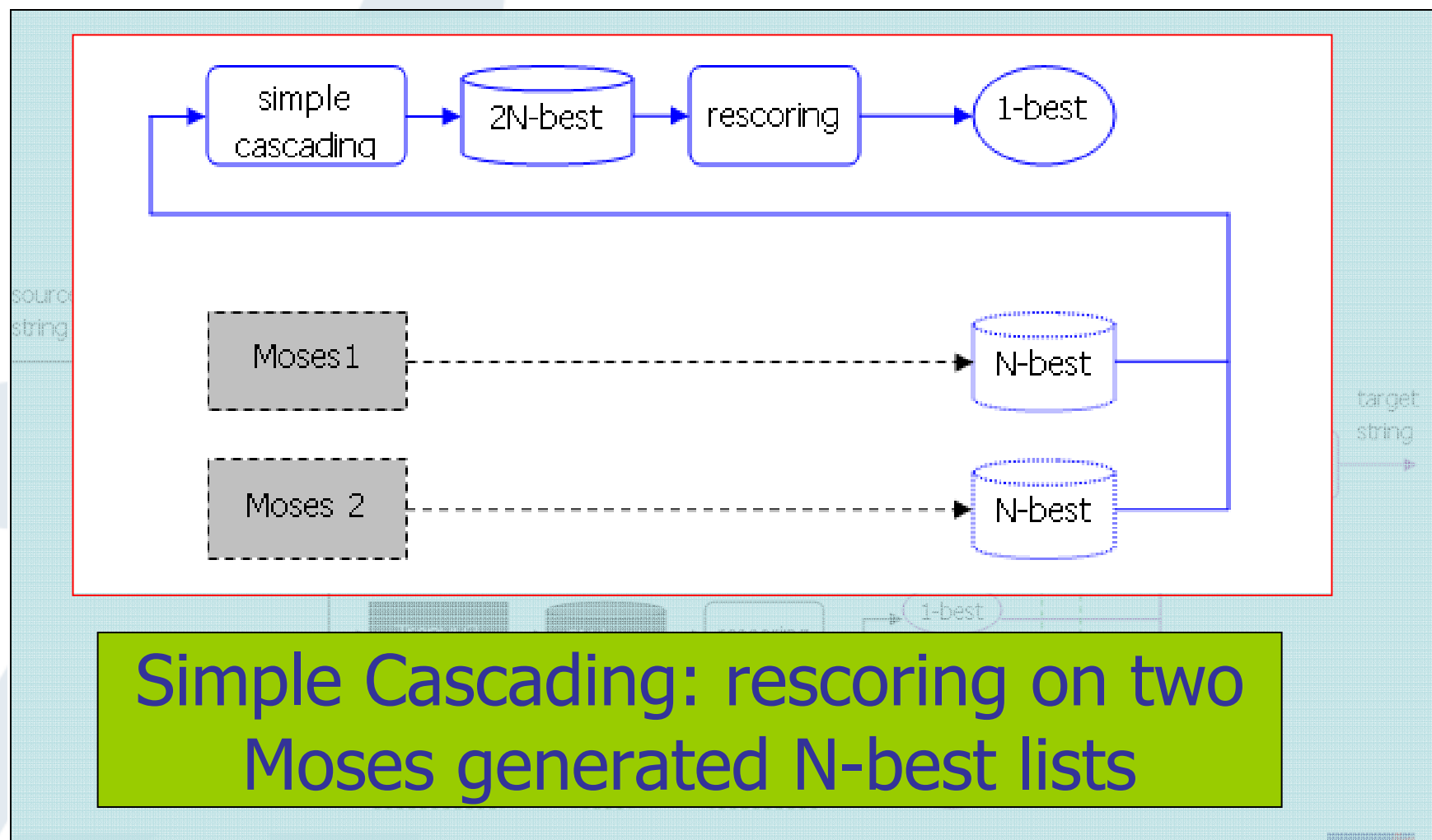
6) source and target length ratio

- 7) question feature
- 8) frequency of n-grams in the N-best
- 9) n-gram post-probabilities
- 10) sentence length post-probabilities

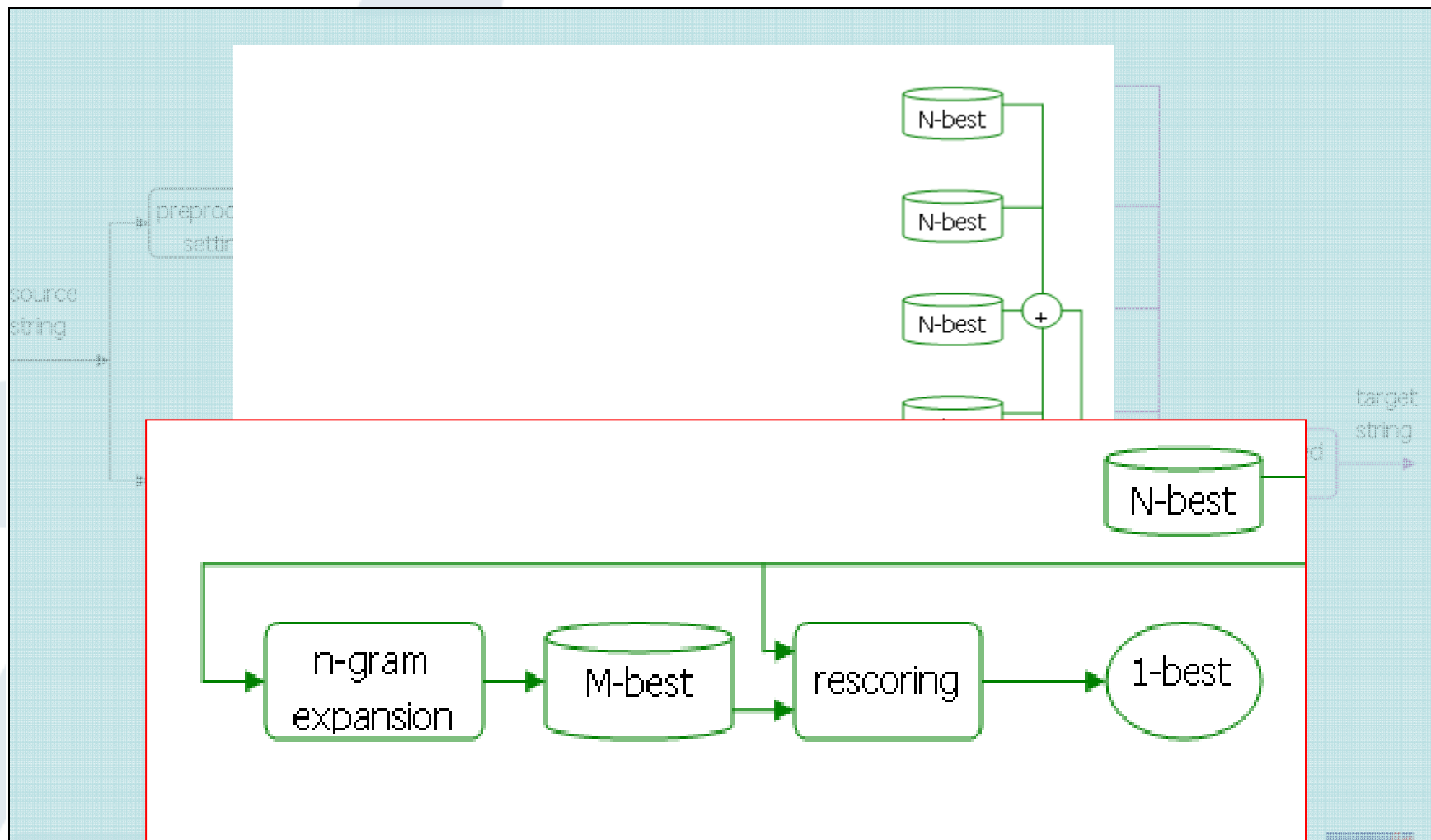
System Combination



System Combination: Simple Cascading



System Combination: N-gram expansion



System Combination: N-gram expansion

- n -gram expansion generates new hypotheses
 - Collect all the n -grams from the original N-best
 - Continuously expand the partial hypothesis through the n -grams.

Reference: my book is in the green basket .

Original entry: my book is in the green case .
my book is inside the green basket .

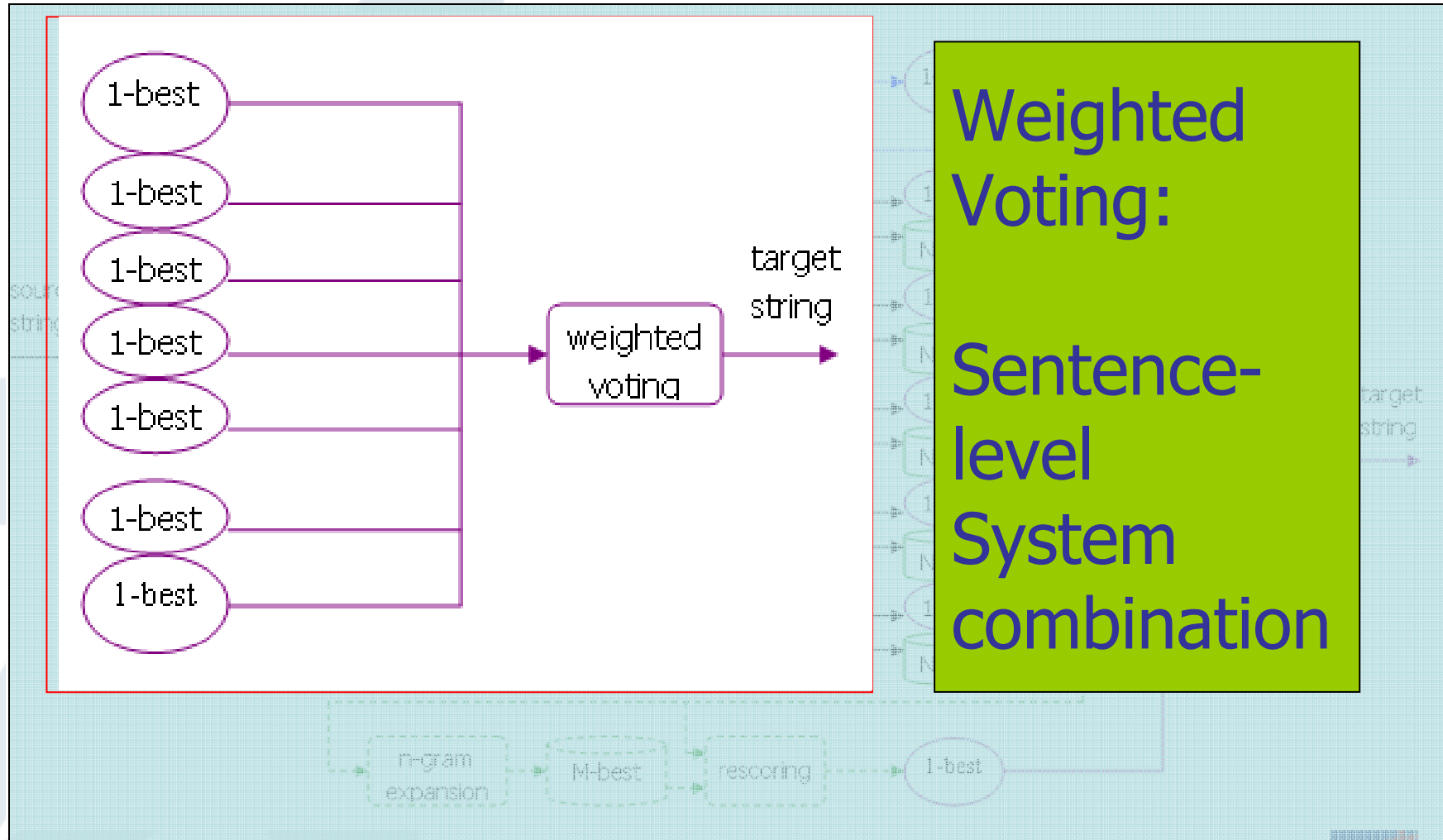
3-grams: my book is, book is in, is in the, in the green,
the green case, is inside the, the green basket ...

n -gram
expansion

Partial Hyp: my book is in
 n -gram: _____ is in the
New partial Hyp: my book is in the

New Hyp: my book is in the green basket .

System Combination: Weighted Voting



Experiments

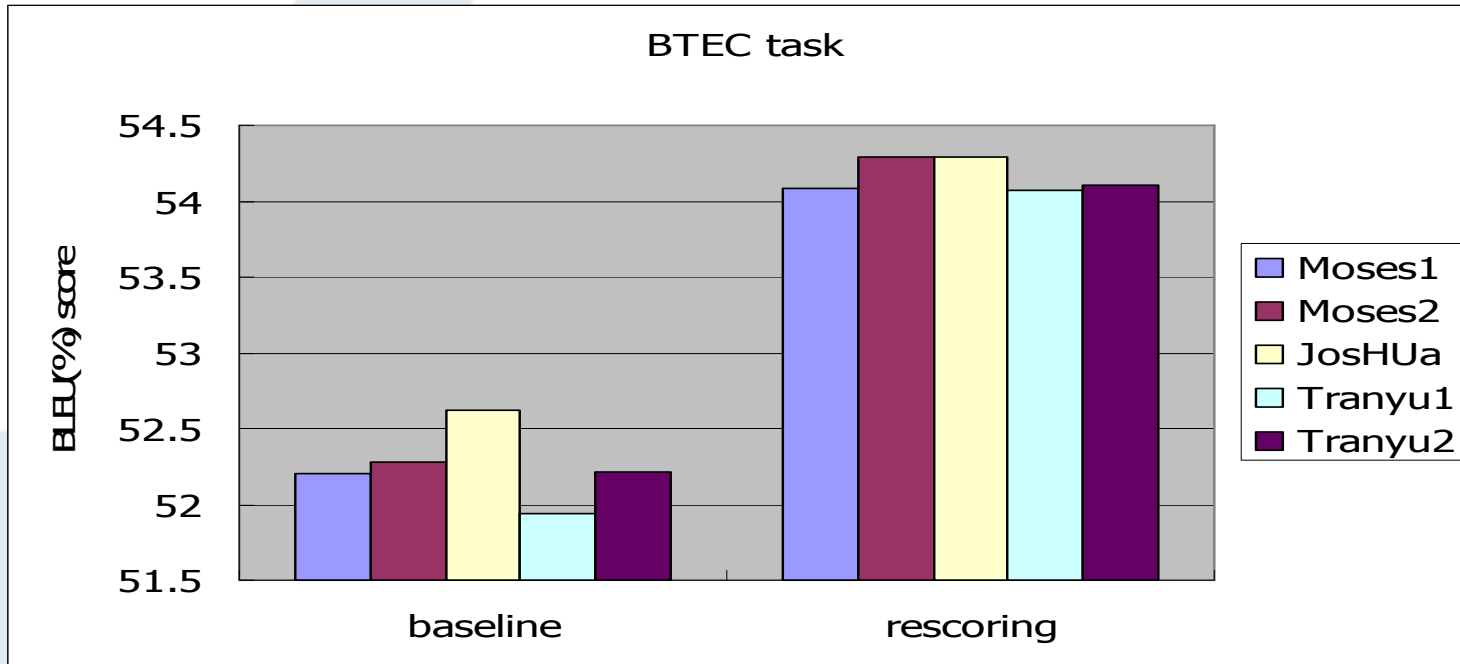
- Training data
 - Bilingual Training data:
 - BTEC supplied data (20K sentences)
 - HIT-corpus (132K sentences)
 - Olympic-corpus (54K sentences)
 - PKU-corpus (200K sentences)
 - Total: 399K sentence-pairs, 5.2M target words
 - Dev data: 6K sentences
 - Additional target data: Tanaka corpus
 - 155K sentence-pairs, 1.4M target running words

Effect of additional data

- BLEU% score on dev sets

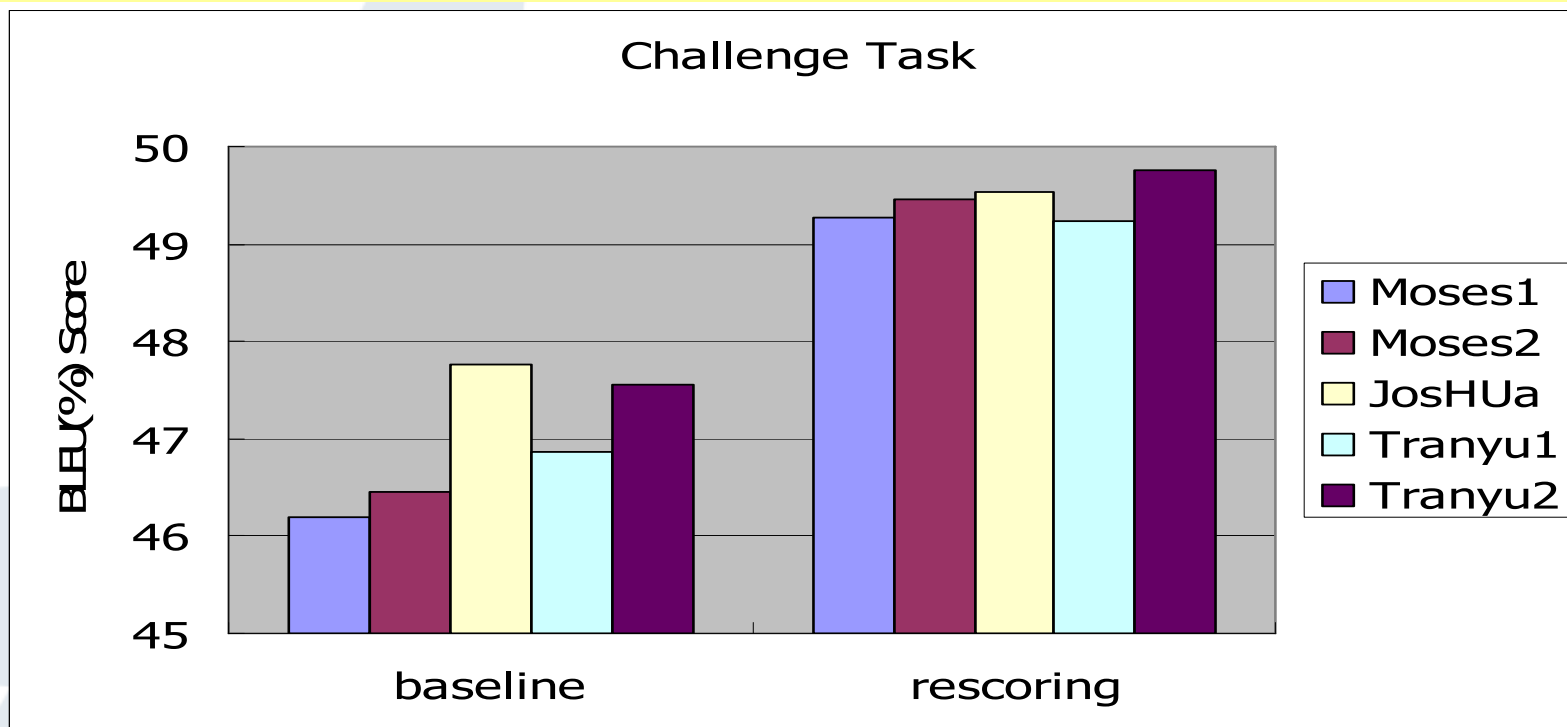
	CSTAR03 (BTEC)	DEV08 (Challenge)
Supplied data	40.96	36.12
+dev. data	45.76	42.29
+addi. Data	50.98	44.92
All data	52.28	46.45

Results (Base & Resc): BTEC task



- Performances of five systems are very similar:
- Baseline: JosHUa > Moses2 > Tranyu2 > Moses1 > Tranyu1
- Rescoring: Moses2 = JosHUa > Tranyu2 > Moses1 > Tranyu1
- Rescoring improved about **1.6-2 BLEU-score** for all systems

Results (Base & Resc): Challenge Task



- Performances of five systems are also similar:
- Baseline: JosHUa > Tranyu2 > Tranyu1 > Moses2 > Moses1
- Rescoring: Tranyu2 > JosHUa > Moses2 > Moses1 > Tranyu1
- Rescoring improved about **2-3 BLEU-score** for all systems

Results: 1st stage System Combination

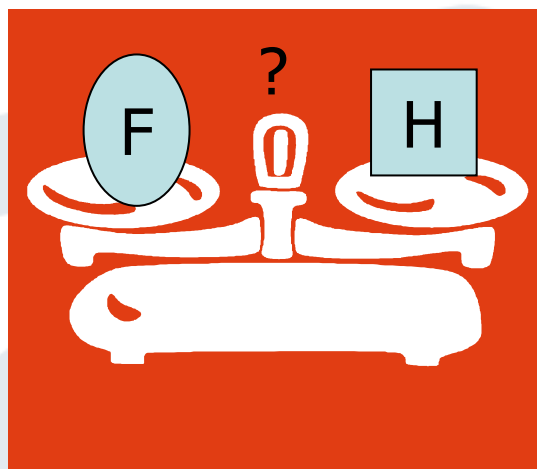
Simple Cascading

Advantages:

More features
(include local feat.
used in decoding)

Disadvantages:

Less distinct
hypotheses



N-gram expansion

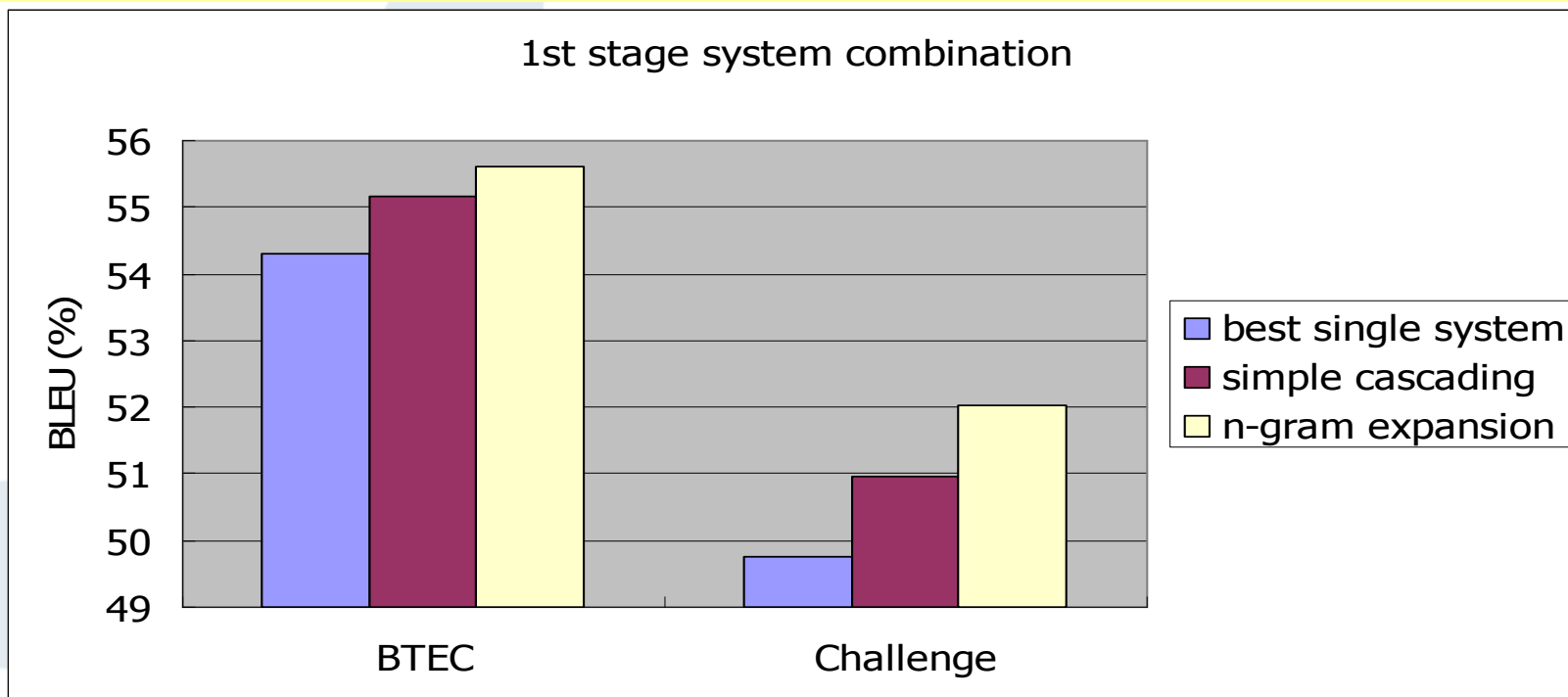
Advantages:

More distinct
hypotheses

Disadvantages:

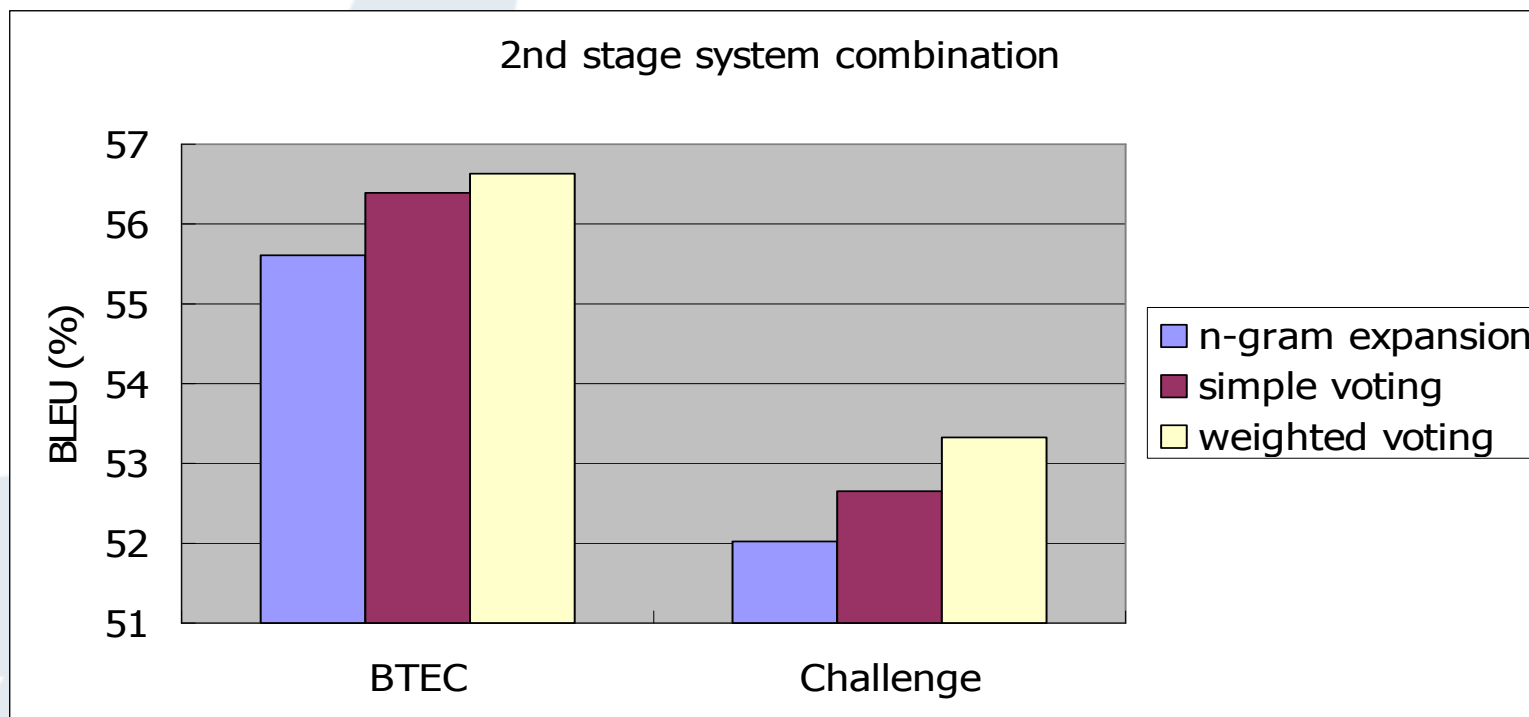
Less features (no
local features)

Results: SC vs NE



- N-gram expansion (NE) outperformed simple cascading (SC)
- Compared with the best single system:
 - Simple cascading obtained about **0.9/1.3 BLEU-score**
 - N-gram expansion obtained about **1.3/2.3 BLEU-score**

Results: weighted voting



- Simple voting: weights of all systems are set to 1.
- Compared with the results of n-gram expansion:
 - Simple voting obtained about **0.8/0.6 BLEU-score**
 - Weighted voting obtained about **1.0/1.3 BLEU-score**

Official score

- Chinese-to-English BTEC and Challenge tasks

		BLEU (%)	NIST	METEOR
BTEC task	ASR	43.57	6.87	0.6017
	CRR	49.26	7.65	0.6446
Challenge task	ASR	39.38	5.96	0.6142
	CRR	46.89	6.66	0.6560

Conclusion and Discussion

- Conclusion
 - Multi-decoder to produce N-best lists
 - Rich additional feature functions to do rescoring
 - n -gram expansion to generate new hypotheses
 - Two-stage system combination
- Comments
 - As one reviewer pointed out: *the 5 systems are all phrase-based system, so the N-best lists are quite similar. It could not provide enough space for system combination. This may be the primary reason that its evaluation result is not too outstanding.*

Thanks for your attention!
Any questions?