Spoken Language Communications
Research Laboratories

Nobody is Perfect: ATR's Hybrid Approach to Spoken Language Translation

October 25, 2005

Michael Paul, Takao Doi, Youngsook Hwang,
Kenji Imamura, Hideo Okuma, Eiichiro Sumita

ATR Spoken Language Communication Research Laboratories
Kyoto, Japan

Spoken Language Communications Research Laboratories

Translation Output Examples
(Japanese-to-English)

input	engine	output	eval
かわいい気に入ったわ試着してもいいですか [cute] [i like it][can i try it on]	1	how cute i like it may i try this on	OK
背中のマッサージはいかがですか [how about a back massage]	1	how about a pedicure	NG
一緒にいきましょう [let's go together]	1	let's go to the same	NG
朝食込みですか [is breakfast included]	1	is this all inclusive	NG

Spoken Language Communications Research Laboratories

Translation Output Examples

(Japanese-to-English)

input	engine	output	eval
かわいい気に入ったわ試着してもいいですか [cute] [i like it][can i try it on]	2	may i try it on cute i like it	??
背中のマッサージはいかがですか [how about a back massage]	2	would you like a facial massage on my back	NG
一緒にいきましょう [let's go together]	2	let's go together	OK
朝食込みですか [is breakfast included]	2	breakfast that's included	??

Translation Output Examples

(Japanese-to-English)

input	engine	output	eval
かわいい気に入ったわ試着してもいいですか [cute] [i like it][can i try it on]	3	it is lovely the which it likes it is possible to try on is	NG
背中のマッサージはいかがですか [how about a back massage]	3	how about a back massage	OK
一緒にいきましょう [let's go together]	3	it will go together	??
朝食込みですか [is breakfast included]	3	does this include breakfast	OK

ATR Spoken Language Communications Research Laboratories

Select the Best Translation

(Japanese-to-English)

input	engine	output	eval
かわいい気に入ったわ試着してもいいですか [cute] [i like it][can i try it on]	1	how cute i like it may i try this on	OK
背中のマッサージはいかがですか [how about a back massage]	3	how about a back massage	OK
一緒にいきましょう [let's go together]	2	let's go together	OK
朝食込みですか [is breakfast included]	3	does this include breakfast	OK

Spoken Language Communications Research Laboratories

Talk Outline

- **ATR's hybrid approach to speech translation**
  - C^3 (Corpus Centered Computation) project
  - MT engines
  - method to select best translation
- **application to IWSLT05 translation task**
  - goals
  - track participation
  - discussion of evaluation results
- **conclusion**

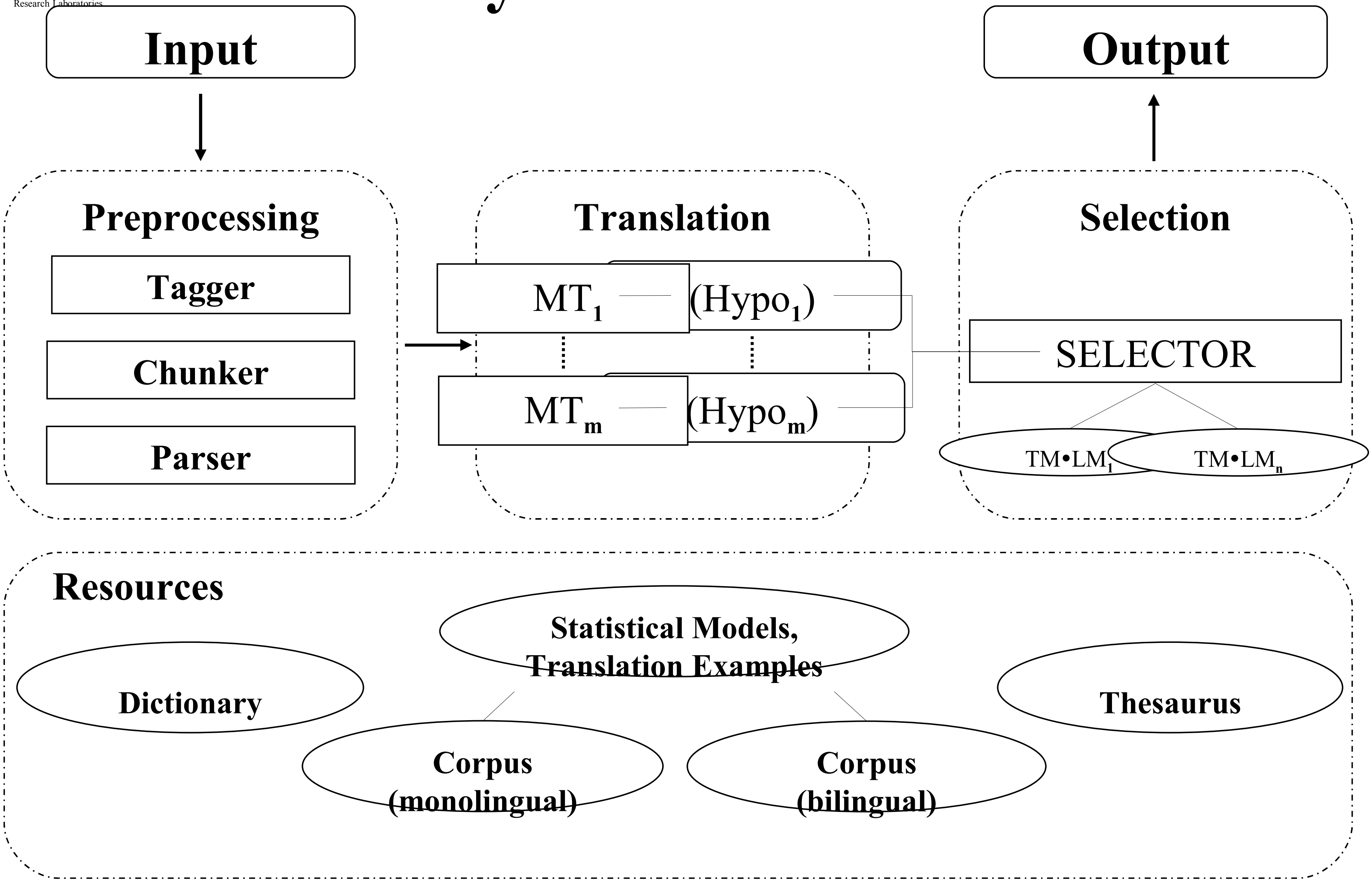
C^3 = C-cube
(Corpus Centered Computation)

- **C^3 places corpora at the center of translation technology**
  - **translation knowledge** is extracted from corpora
  - **translation quality** is improved by referring to corpora
  - **selection of best translation** is based on corpora

- **Example-based Machine Translation (EBMT)**
  - **uses corpora directly**
  - **retrieves translation examples** that matches the input closely
  - **adjust examples** to obtain translation

- **Statistical Machine Translation (SMT)**
  - **learns statistical models** for language and translation from corpora and dictionaries
  - **searches for best translation at run-time** according to its models

System Outline



ATR
Spoken Language Communications Research Laboratories

Element MT Engines

Type	MT engine	Description
SMT	SAT	example-based greedy decoder using IBM-4 models
	PBHMTM	word-graph-based decoder using phrase-based HMM translation models
	MSEP	phrase-based SMT engine using morpho-syntactic (part-of-speech, chunk) information
	HPATR2	SMT engine based on syntactic transfer
EBMT	HPAT	syntactic-transfer-based EBMT based on hierarchical phrase alignments
	HPATR	syntactic-transfer-based EBMT incorporating word-level SMT methods
	D3	DP-match-driven EBMT engine
	EM	translation memory
SELECTOR		multi-engine output selection method using multiple statistical models

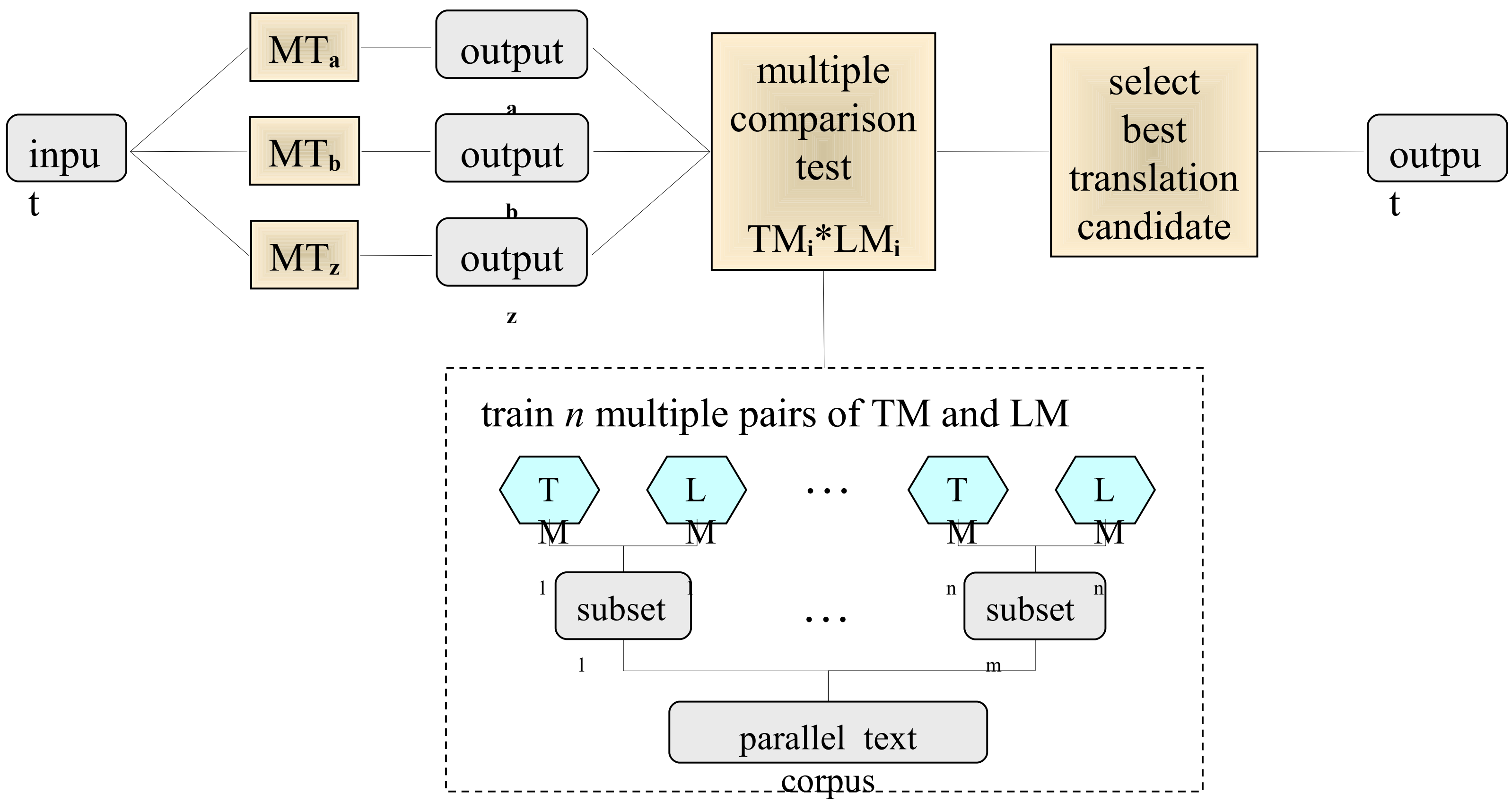
Features of Element MT Engines

	SMT				EBMT			
	SAT	PBHMTM	MSEP	HPATR2	HPATR	HPAT	D3	EM
Resource	corpus	corpus	corpus, chunker	corpus, parser	corpus, parser	corpus, parser thesaurus	corpus, thesaurus, bilingual dictionary	corpus
Unit	sent. &word	phrase	phrase	phrase	phrase	phrase	sent.	sent.
Coverage	wide	wide	wide	wide	wide	wide	narrow	narrow
Quality	very good	good	good	good	good	good	very good	very good
Speed	modest	slow	slow	modest	modest	fast	fast	fast

Spoken Language Communications
Research Laboratories

Selection of Best Translation

- calculate scores based on language and translation models
- apply multiple comparison test
- check significance of score differences



ATR
Spoken Language Communications Research Laboratories

Selection of Best Translation

determine priority order of element MT engines
→ translate development set and evaluate MT outputs (WER)

calculate and assign multiple statistical scores (TM_i ・ LM_i $1 \leq i \leq n$) to each translation hypothesis of the given test sentence

apply pair-wise comparison test (→ Kruskal –Wallis test) in order to check whether the MT output score of first engine is better than MT output of second MT engine

if a significantly better MT output can be found, use this one for the comparison with remaining MT outputs. Otherwise, **select the best MT output according to the priority order**

continue significance test for remaining MT engines and output selected translation

Spoken Language Communications Research Laboratories

Talk Outline

- **ATR's hybrid approach to speech translation**
  - C^3 (Corpus Centered Computation)
  - MT engines
  - method to select best translation
- **application to IWSLT05 translation task**
  - goals
  - track participation
  - discussion of evaluation results
- **conclusion**

ATR Spoken Language Communications Research Laboratories

Our Goals for IWSLT 2005

Effects of Training Data Size	
variable amounts of training data	20K → 170K → 540K

Effects of NLP Tools	
preprocessing of training data	tagger, (chunker, parser)

Effects of Multi-Engine Approach	
combining mult. MT engines	➢ SELECTOR vs. best element MT engine ➢ upper boundary (ORACLE experiment)

Track Participation

Translation Direction: **(JE)** Japanese-to-English
(CE) Chinese-to-English

Data Track: **(C)** C-STAR Track
(T) Supplied+Tool Data Track
(S) Supplied Data Track

MT engine	JE			CE		
	C (→ 5)	**T** (→ 7)	**S** (→ 3)	**C** (→ 7)	**T** (→ 7)	**S** (→ 3)
SAT	○	○	○	○	○	○
PBHMTM	○	○	○	○	○	○
MSEP	×	○	N/A	○	○	N/A
HPATR2	○	○	N/A	○	○	N/A
HPAT	×	○	N/A	N/A	N/A	N/A
HPATR	×	×	N/A	○	○	N/A
D3	○	○	N/A	○	○	N/A
EM	○	○	○	○	○	○

Spoken Language Communications
Research Laboratories

Priority Order of Element MT Engines

language	data track	priority order
JE	C	EM≫D3>HPATR2>PBHMTM
	T	EM≫D3>HPAT>HPATR2>PBHMTM>SAT>MSEP
	S	EM≫PBHMTM>SAT
CE	C	EM≫D3>HPATR2>HPATR>MSEP>PBHMTM>SAT
	T	EM≫MSEP>D3>HPATR>PBHMTM>HPATR2>SAT
	S	EM≫PBHMTM>SAT

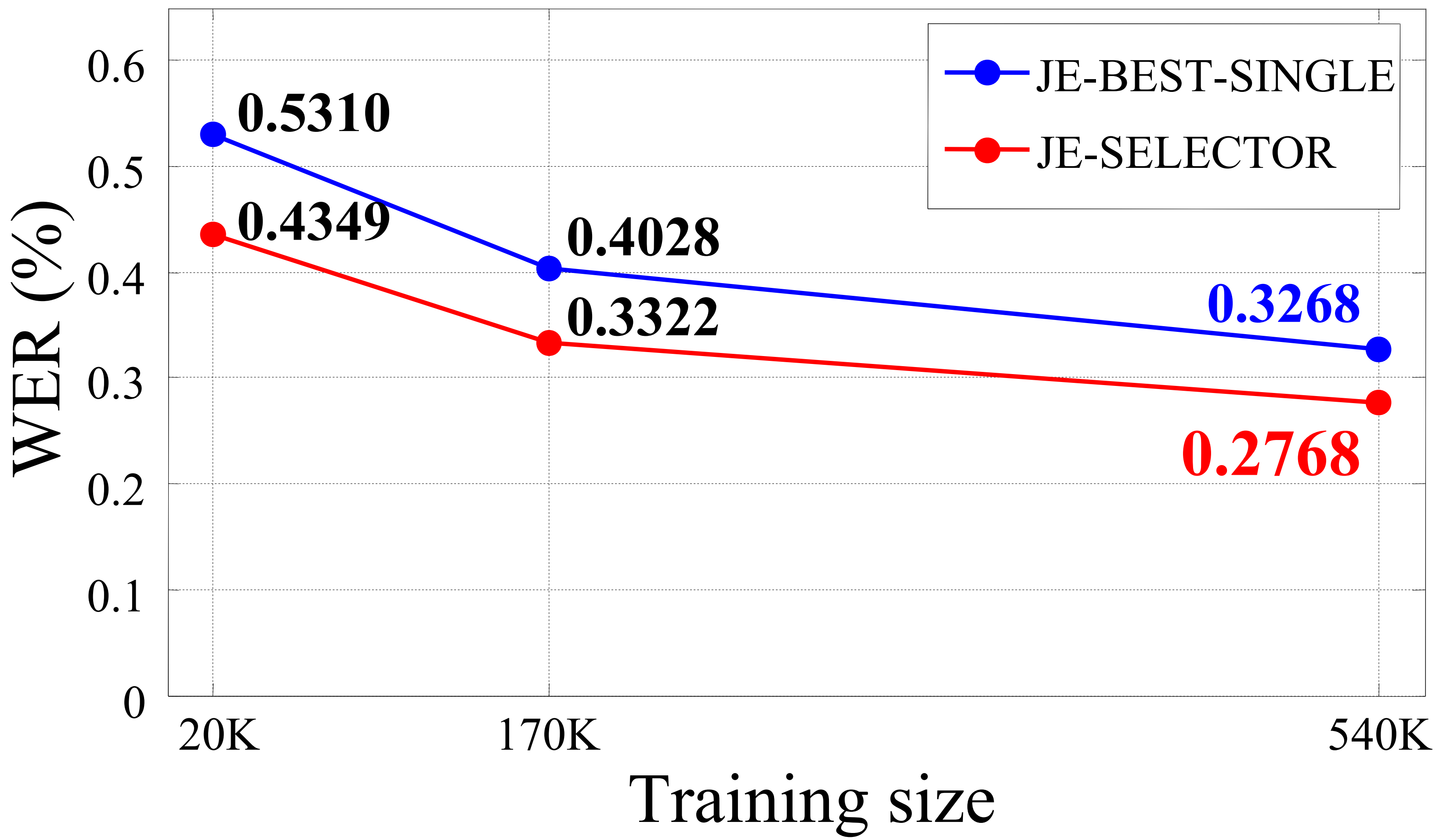
- **large differences** between languages and data tracks
- selection of **optimal combination** difficult
- highest priority to EM, rest MT order optimized on develop set

Evaluation Results
(official run submissions)

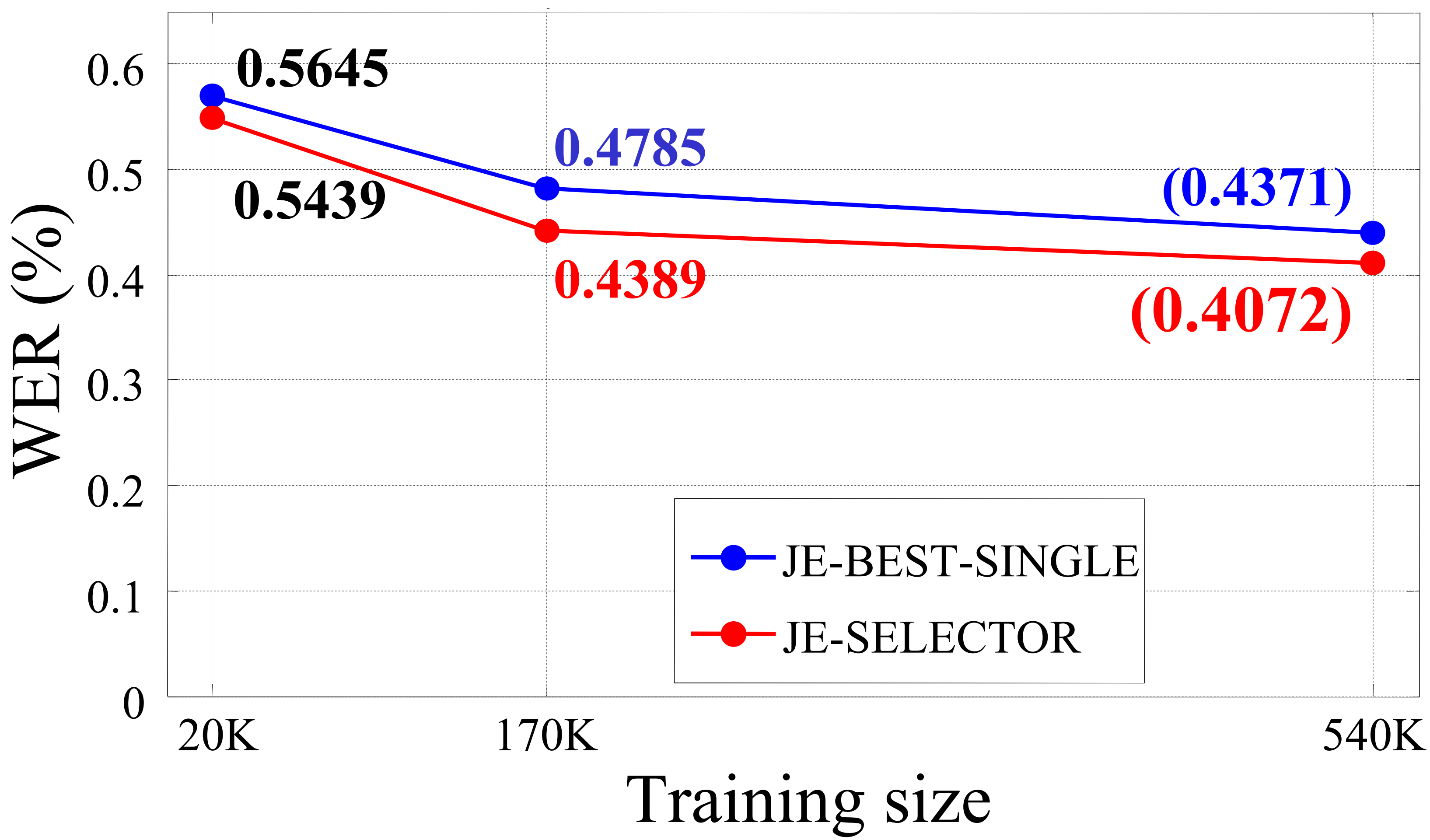
language	data track	automatic evaluation				
		BLEU	NIST	METEOR	WER	GTM
JE	C	**0.6873**	**10.7375**	**0.8102**	**0.2768**	**0.6934**
	T	0.4774	8.1720	0.6658	0.4349	0.5520
	S	0.3744	7.7368	0.6008	0.5568	0.4822
CE	C	**0.5031**	**8.6875**	**0.6845**	**0.4389**	**0.5898**
	T	0.3804	6.7540	0.5819	0.5439	0.4950
	S	0.3938	8.0004	0.6291	0.5235	0.5533

- better performance for JE data tracks compared to CE data tracks
- large gain for JE-T (vs. JE-S) due to word normalization
- side-effects of NLP tools for CE-T

Effects of Training Data Size
(Japanese-to-English)



Effects of Training Data Size
(Chinese-to-English)



Effects on NLP Tools

MT engine	JE			CE		
	C (→ 5)	**T (→ 7)**	**S (→ 3)**	**C (→ 7)**	**T (→ 7)**	**S (→ 3)**
SAT	○	○	○	○	○	○
PBHMTM	○	○	○	○	○	○
MSEP	×	○	N/A	○	○	N/A
HPATR2	○	○	N/A	○	○	N/A
HPAT	×	○	N/A	N/A	N/A	N/A
HPATR	×	×	N/A	○	○	N/A
D3	○	○	N/A	○	○	N/A
EM	○	○	○	○	○	○

- 3MT = SAT, PBHMTM, EM

Effects on NLP Tools

language	data track	WER
JE	T (3MT)	**0.5221**
	S	0.5568
CE	T (3MT)	**0.5913**
	S	0.5235

- comparison of JE-S vs. JE-T and CE-S vs. CE-T using the three element MT engines of the Supplied Track (SAT,PBHMTM,EM)
- medium improvement of 3.5% in WER for JE
- degradation in performance for CE due to word segmentation differences and lower coverage of our in-house tagging tool

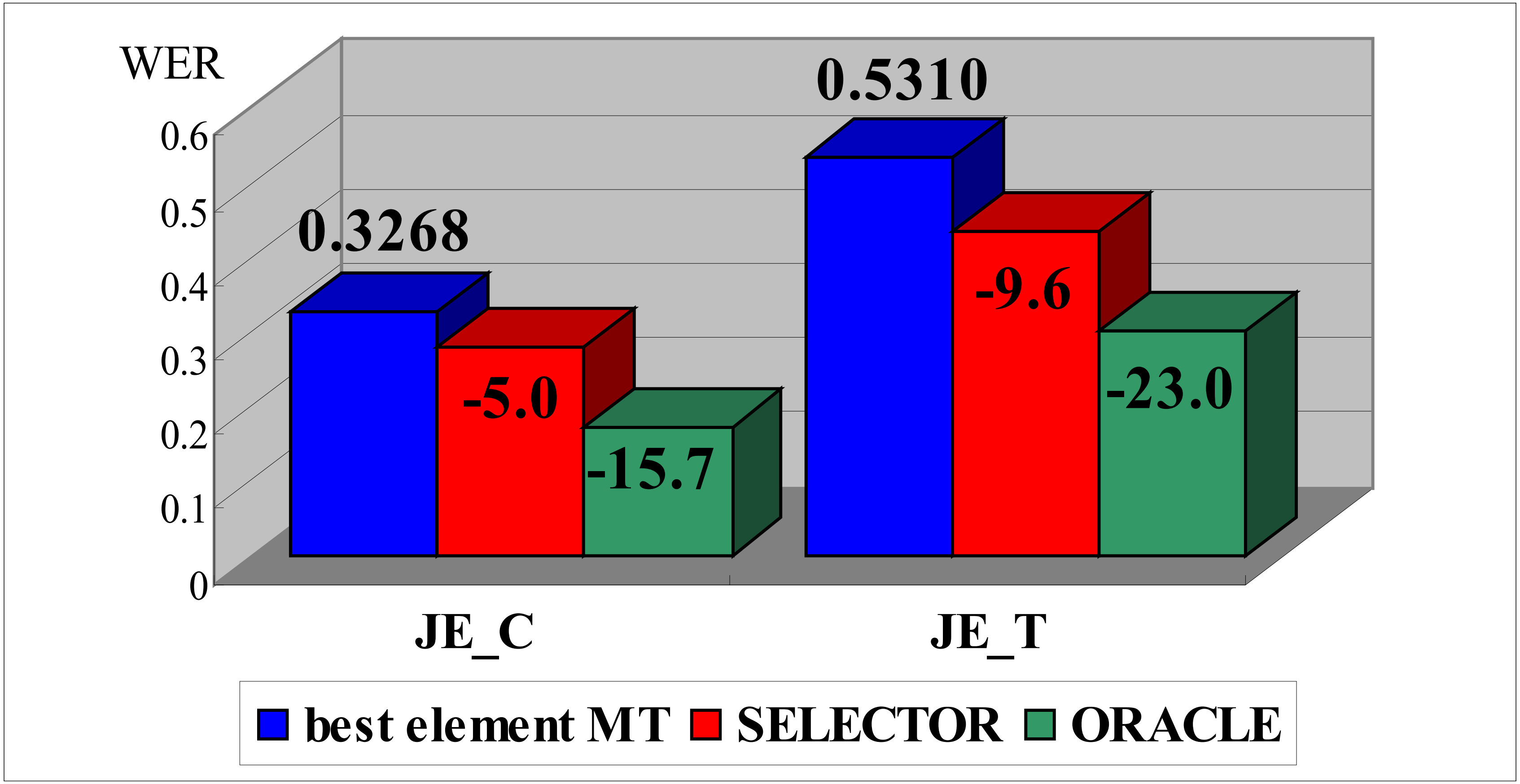
ATR Spoken Language Communications Research Laboratories

Effects on Multi-Engine Approach

MT engine	WER of JE systems		
	C	T	S
SAT	0.3404	0.5541	0.5664
PBHMTM	**0.3268**	**0.5310**	**0.5589**
MSEP	0.3956	0.5384	
HPATR2	0.3457	0.5478	
HPAT	0.4526	0.5427	
HPATR	0.4137	0.5507	
D3	0.3971	0.5650	
EM	0.5995	0.8949	0.9426

- SMT engines outperformed EBMT engines
- best performing systems for JE is PBHMTM

Effects of Multi-Engine Approach



- SELECTOR **outperforms all element MT engines**
- **5% gain** for JE-C and even up to 10% for JE-T
- SELECTOR **does not tap the full potential of element MT engines**

Distribution of Selected JE Hypotheses

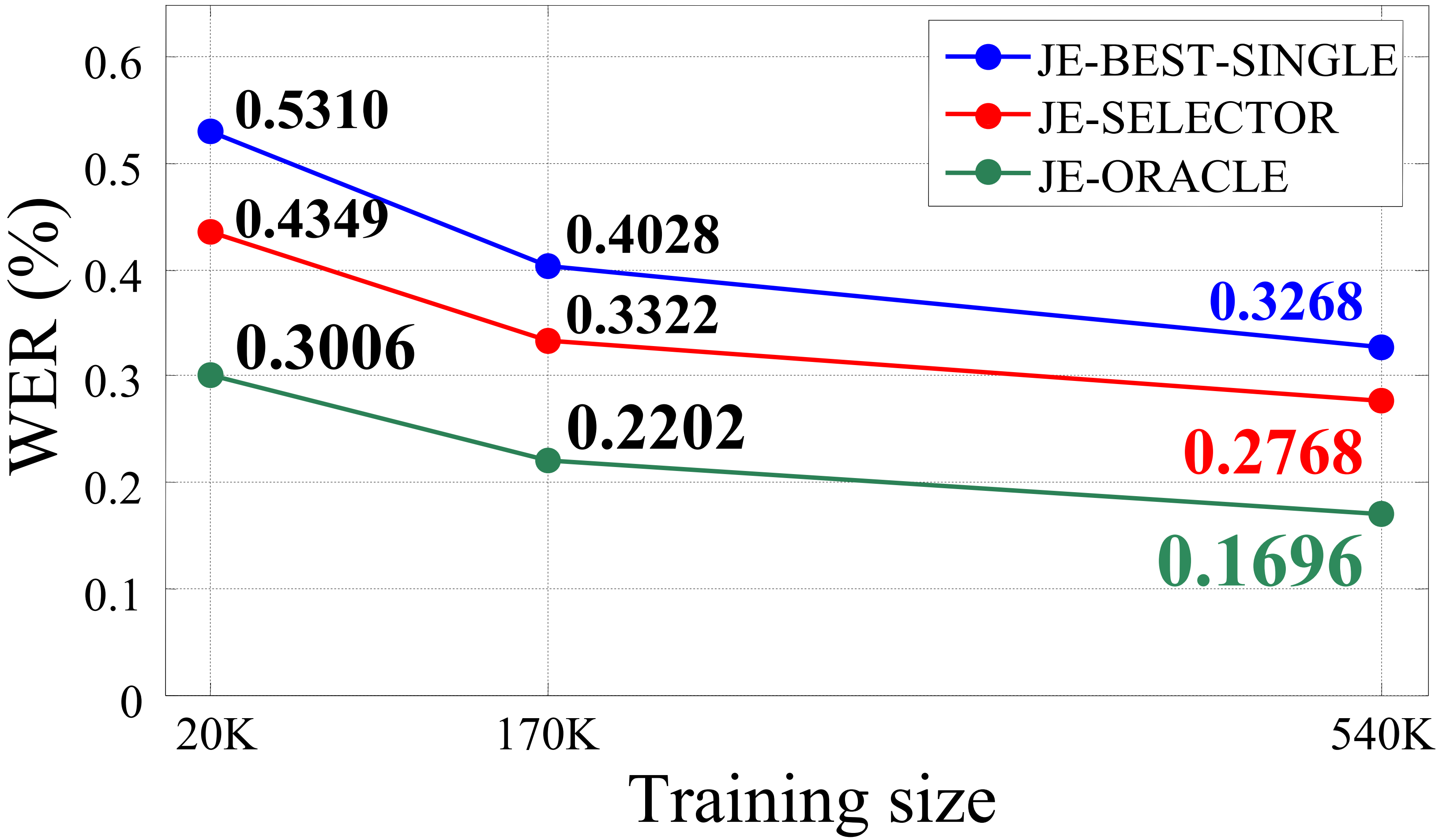
MT engine	SELECTOR (%)		
	C	T	S
SAT	9.9	3.0	5.9
PBHMTM	16.4	**23.3**	**84.4**
MSEP	×	10.9	
HPATR2	17.2	12.0	
HPAT	×	×	
HPATR	×	17.4	
D3	12.1	19.4	
EM	**44.4**	14.0	9.7

MT engine	ORACLE (%)		
	C	T	S
SAT	8.5	11.6	**46.8**
PBHMTM	7.7	9.1	44.6
MSEP	9.9	9.3	
HPATR2	6.7	17.6	
HPAT	**32.8**	11.9	
HPATR	9.3	**21.5**	
D3	8.5	10.1	
EM	16.6	8.9	8.5

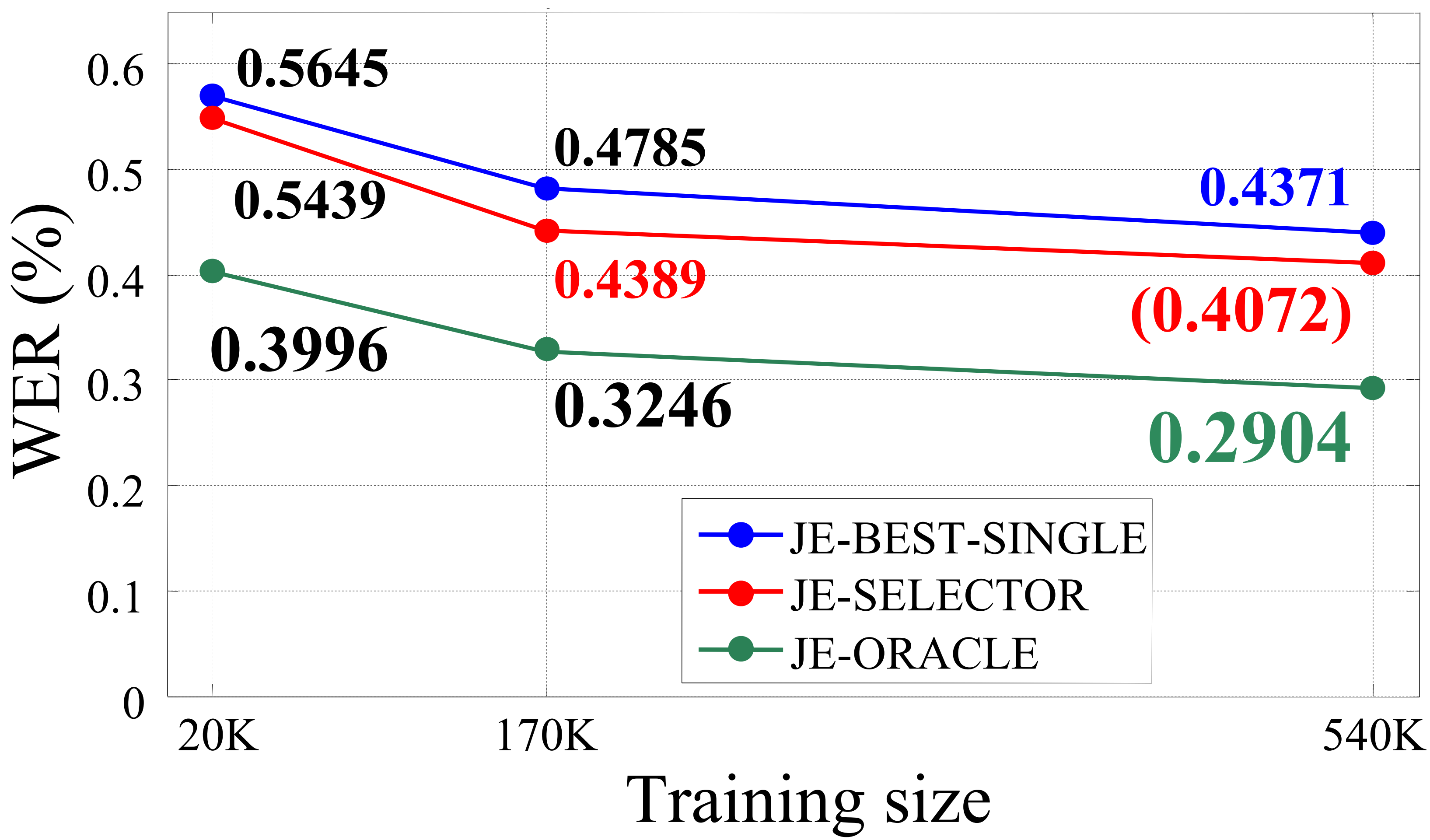
- **SELECTOR biased toward SMT** engines
- **features beyond statistical TM・ LM score required** to improve system performance

ATR
Spoken Language Communications Research Laboratories

Upper Boundary
(Japanese-to-English)



Upper Boundary
(Chinese-to-English)



Spoken Language Communications Research Laboratories

Lessons learned from IWSLT 2005

Effects of Training Data Size	
variable amounts of training data	**increase in training data led to improved results**

Effects of NLP Tools	
preprocessing of training data	**preprocessing of the training data** was **important** to achieve high trans.quality

Effects of Multi-Engine Approach	
combining mult. MT engines	**significant gain obtained**, but still plenty of room for improvement

Spoken Language Communications Research Laboratories

Conclusion

- the proposed **hybrid approach was successful** on the IWSLT05 translation task
- **the proposed selection method outperformed all element MT engines** gaining 4-5% in WER towards the best MT engine
- **SMT-based selection of multiple MT outputs underachieved its task**

Future Work

- **additional features** besides the utilized statistical model scores have to be incorporated into the selection process in order **to tap the full potential of the element MT engines**
- **improve system performance of element MT engines** in order to rise the upper boundary