Speech Translation: from Single-best to N-Best to Lattice Translation

Ruiqiang ZHANG *Genichiro KIKUI*

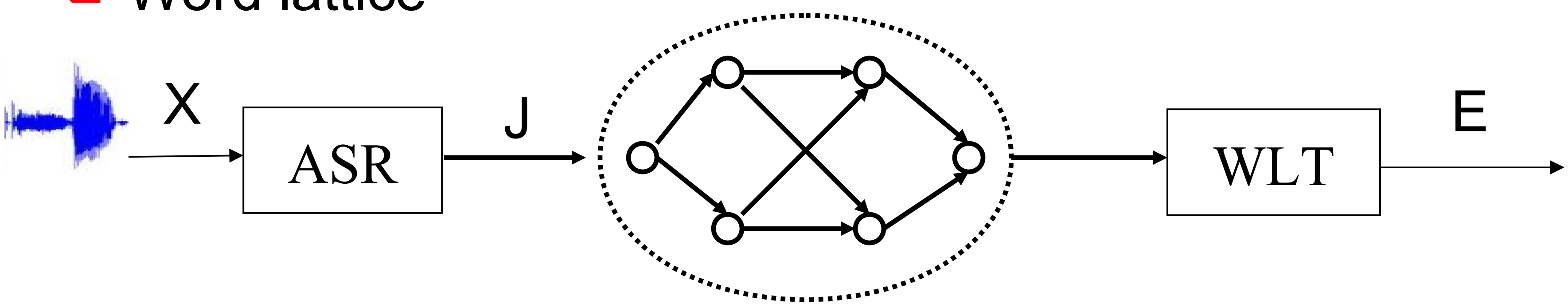
ATR Spoken Language Communication Laboratories

Speech Translation Structure

- Single-best only

- N-best hypothesis translation

- Word lattice



References

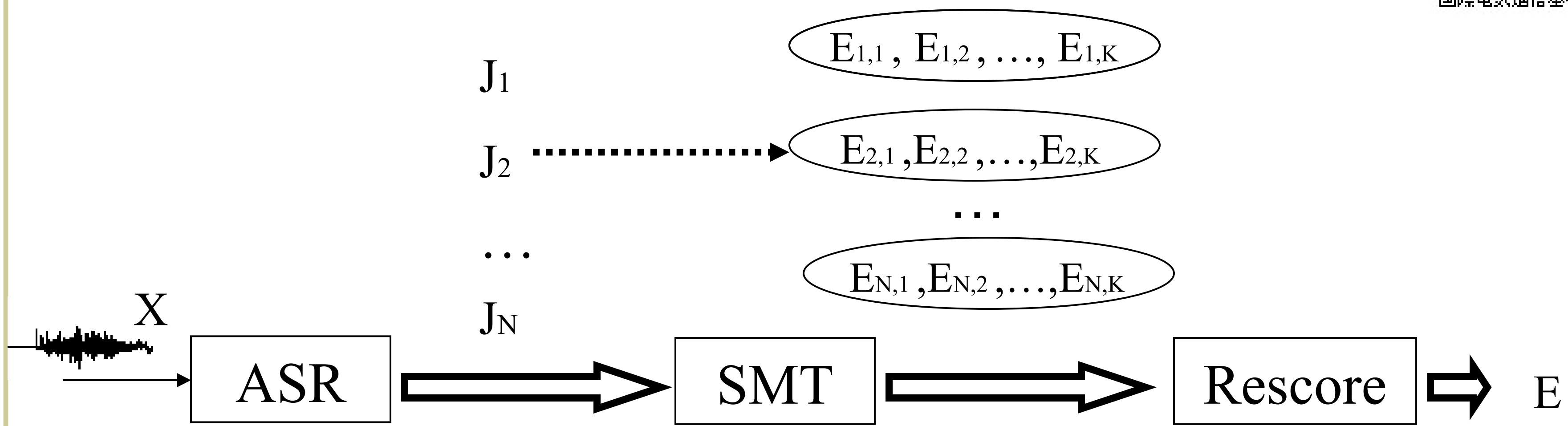
ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社
国際電気通信基礎技術研究所

- Ney (ICASSP 1999) . Speech translation: Coupling of recognition and translation.
- Casacuberta (2002). Architectures for speech-to-speech translation using finite-state transducer
- Zhang(Coling 2004). A unified approach in speech-to-speech translation
- Saleem(ICSLP 2004). Using word lattice information for a tight coupling in speech translation systems
- Matusov(Eurospeech 2005). On the Integration of Speech Recognition and SMT
- (Bozarov, Zhang)(Eurospeech 2005). Speech Translation by Confidence Measure

Outline

- N-best translation
- Word lattice translation
- IWSLT 2005 evaluation
- Conclusions

N-best Hypothesis Translation



- J_1, J_2 and J_N : N-best speech recognition hypotheses
- $E_{2,1}$, $E_{2,2}$ … $E_{2,K}$: K-best translation hypotheses produced from J_2
- Rescore: to rescore all NxK translations

Rescore: Integration of ASR and SMT

ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社国際電気通信基礎技術研究所

- Statistical theory

$$\begin{aligned}\hat{E} &= \arg\max_{E} P(E|X) \\ &= \arg\max_{E} P(E)P(X|E) \\ &= \arg\max_{E} \left\{ P(E) \sum_{J} P(X,J|E) \right\} \\ &= \arg\max_{E} \left\{ P(E) \sum_{J} P(X|J)P(J|E) \right\}\end{aligned}$$

- Make approximations

$$\left\langle \hat{E}, \hat{J} \right\rangle = \arg\max_{E,J} \{ P(E)P(X \mid J)P(J \mid E) \}$$

Rescore: Log-linear models

$$\hat{E} = \arg\max_{E} \sum_{m=1}^{M} \lambda_m \log P_m(X, E)$$

$$P(E \mid X) = \frac{\exp(\sum_{m=1}^{M} \lambda_m f_m(X, E))}{\sum_{E'} \exp(\sum_{m=1}^{M} \lambda_m f_m(X, E))}$$

E : all possible translation hypotheses

$P_m(X, E)$: m-th feature in log value

λ : weight of each feature

Parameter optimization

ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社 国際電気通信基礎技術研究所

Objective function :

$$\lambda_1^M = optimize \; D(R_s, \widehat{E}_s)$$

$\widehat{E}_s$ translation output after log-linear model rescoring

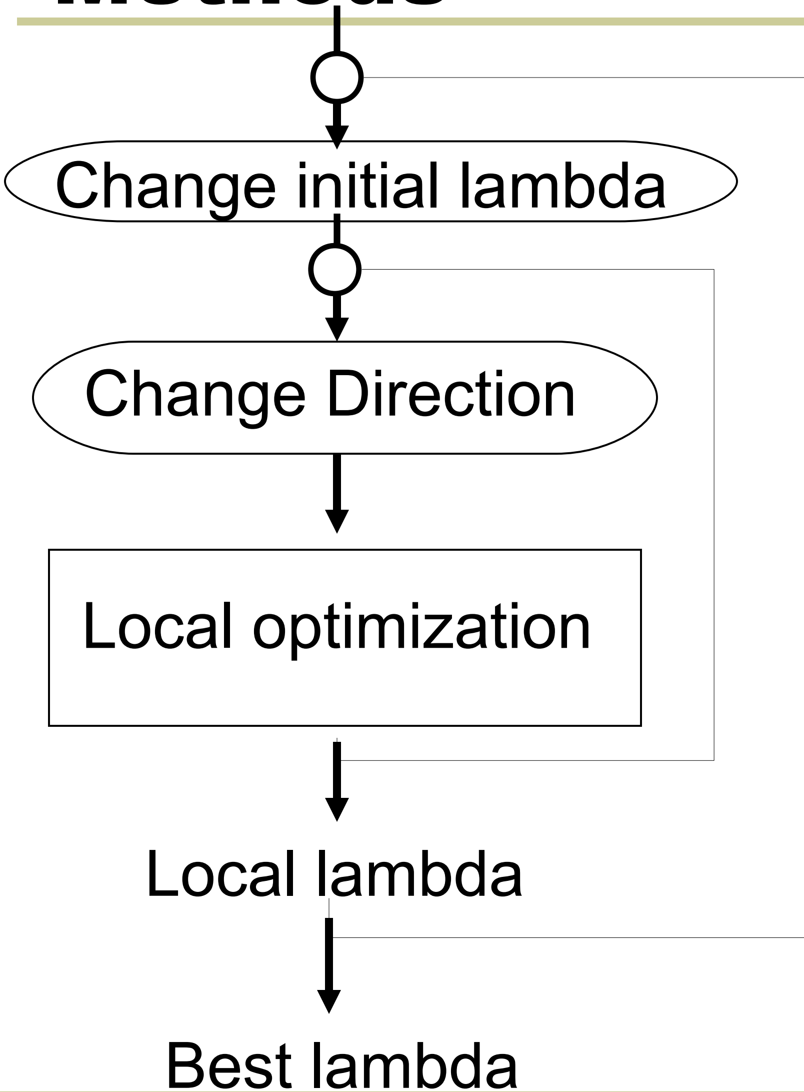
R_s References of English sentences. 16 reference sentences for each English sentence

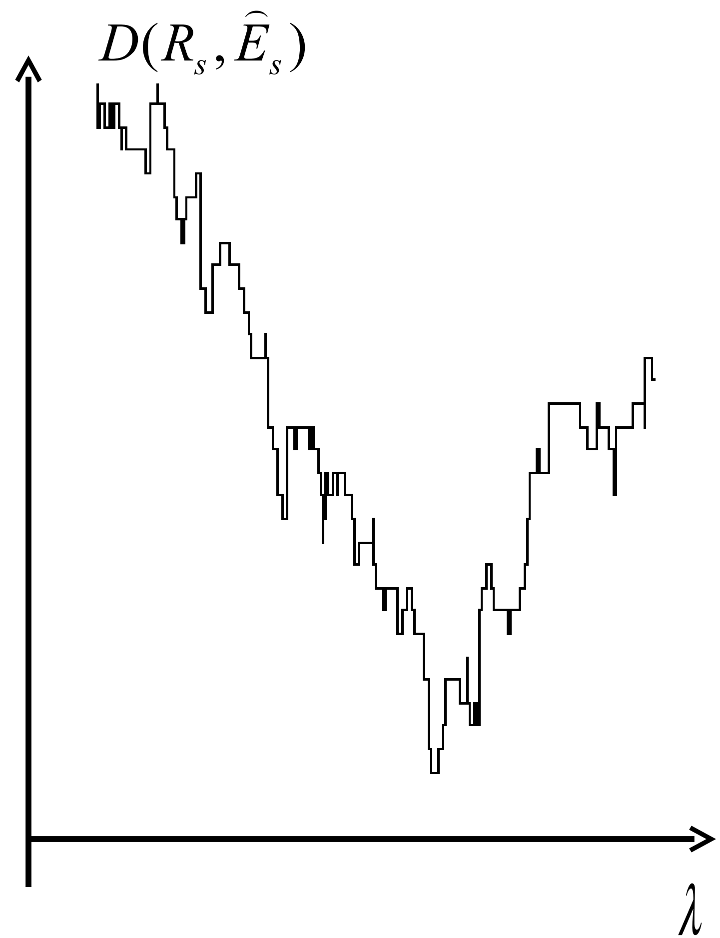
$D(R_s, \widehat{E}_s)$ Automatic translation quality metrics. BLEU, NIST, mWER and mPER

Translation Assessment $D(R_s, \hat{E}_s)$

- N-gram methods
  - BLEU: A weighted geometric mean of the n-gram matches between test and reference sentences plus a short sentence penalty
  - NIST: An arithmetic mean of the n-gram matches between test and reference sentences
- Word error rate
  - mWER: multiple reference word error rate.
  - mPER: multiple reference position independent word error rate

Optimization:Direction Set Methods





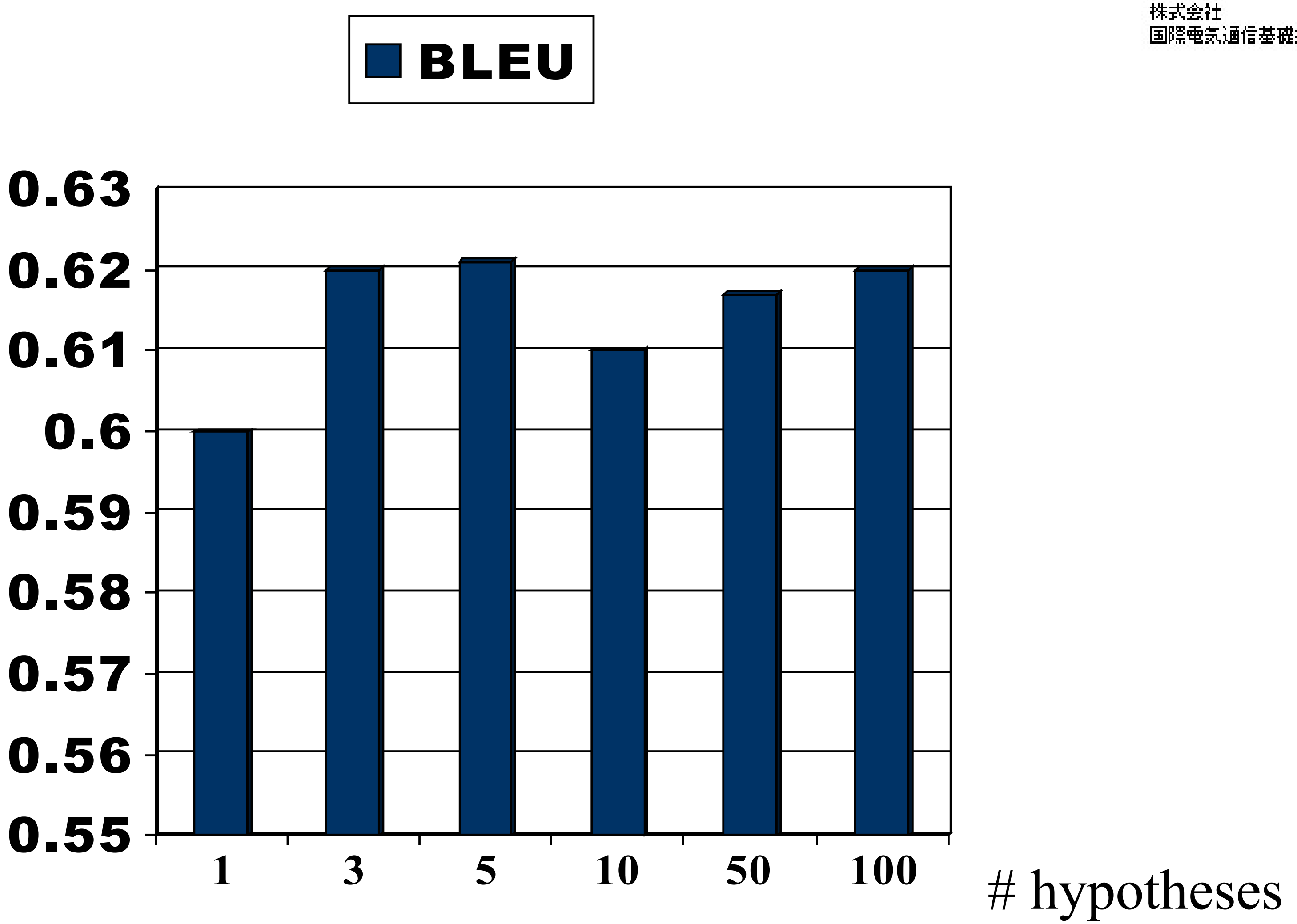
Features from ASR

- Acoustic Model (AM) scores
  - Gaussian mixture output probability density function(pdf)
- Language Model(LM) scores
  - N-gram language model

Features from Phrase-based SMT

- **Target language model** (trigram)
- **Target class language model: SRILM cluster (5-gram)**
- **Target phrase language model**:
- **Phrase translation model:**
- **Distortion model** :
- **Length model**:
- **NULL word translation model:**
- **Jump model**:
- **Long distance target LM**: (9-gram) for rescore
- **Long distance class LM**: (11-gram)

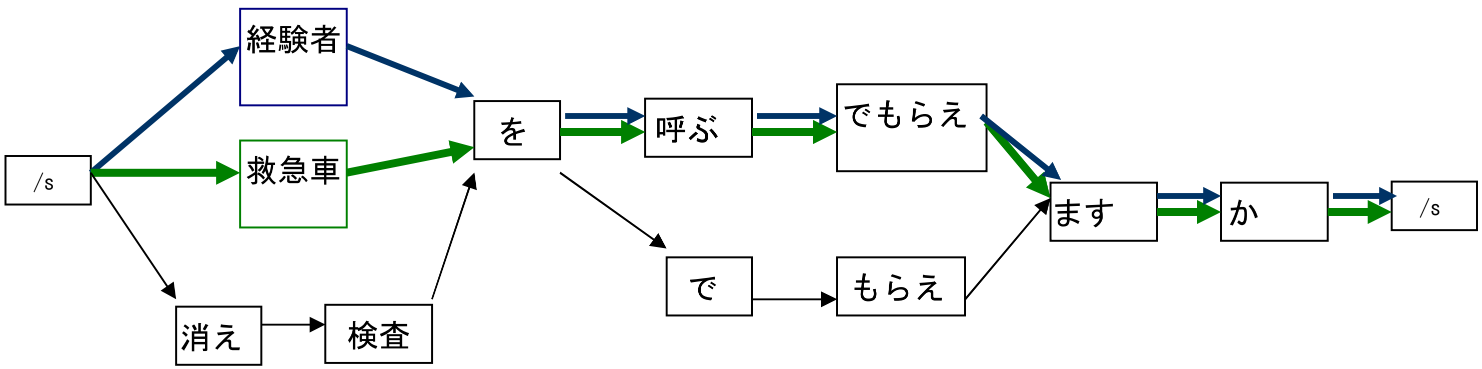
An Experimental Results of N-best Translation



Word Lattice Translation

Recognition Word Lattice

ATR
ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社
国際電気通信基礎技術研究所



ASR First-best: 経験者を呼ぶでもらえますか

First-best translation: **Could I get a job**

ASR correct recognition: 救急車を呼ぶでもらえますか

Word Lattice translation: **Could you call an ambulance**

Machine Translation for Text

ATR
ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社
国際電気通信基礎技術研究所

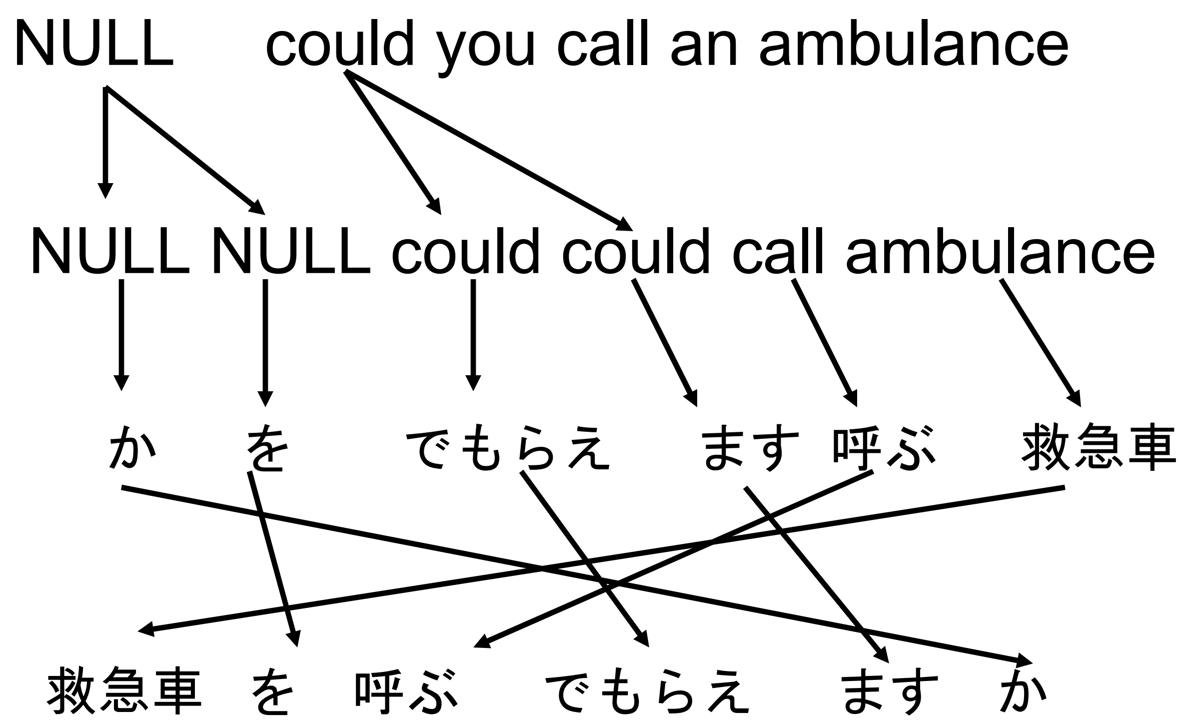
could you call an ambulance

NULL model

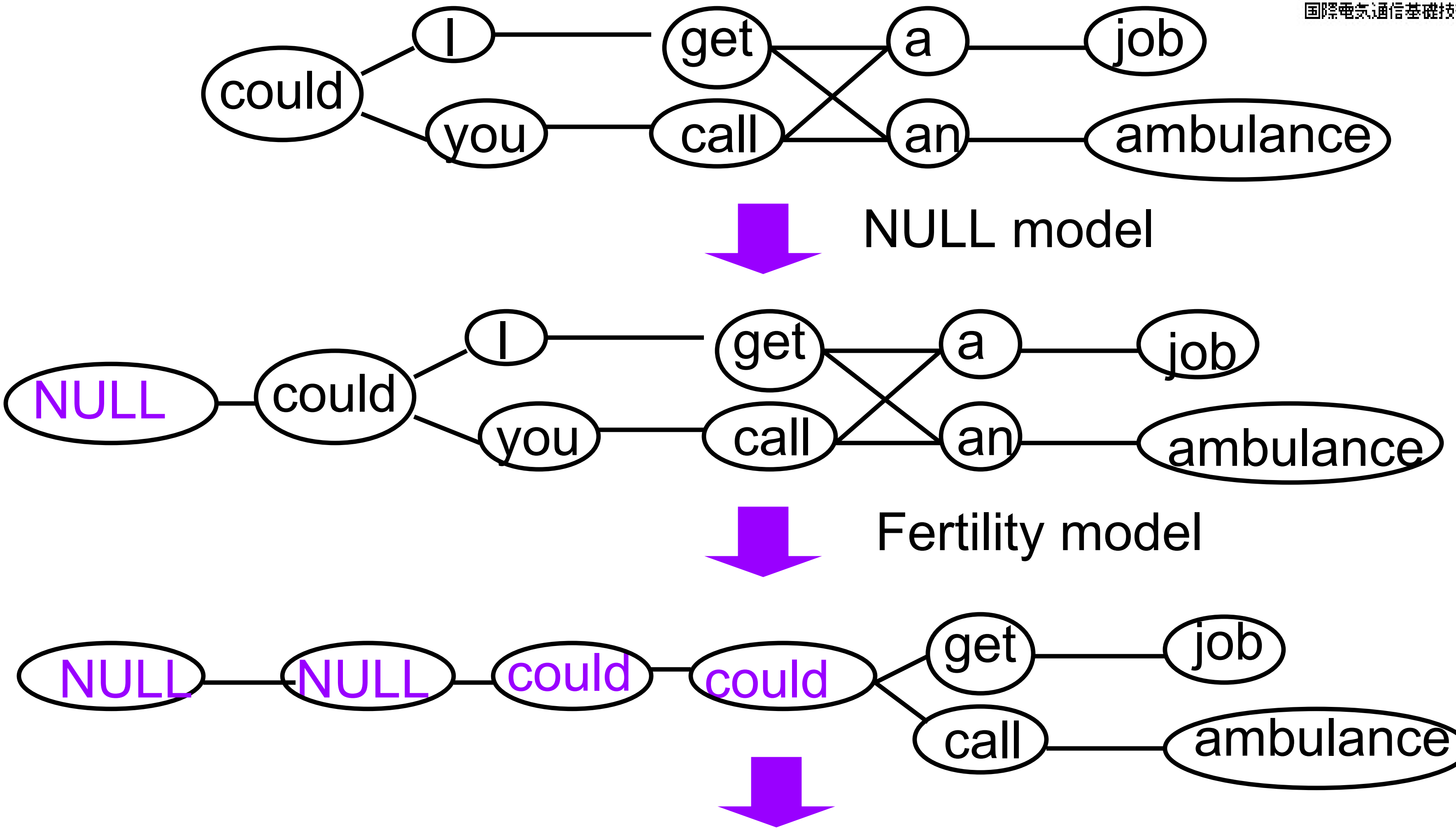
Fertility model

Lexical model

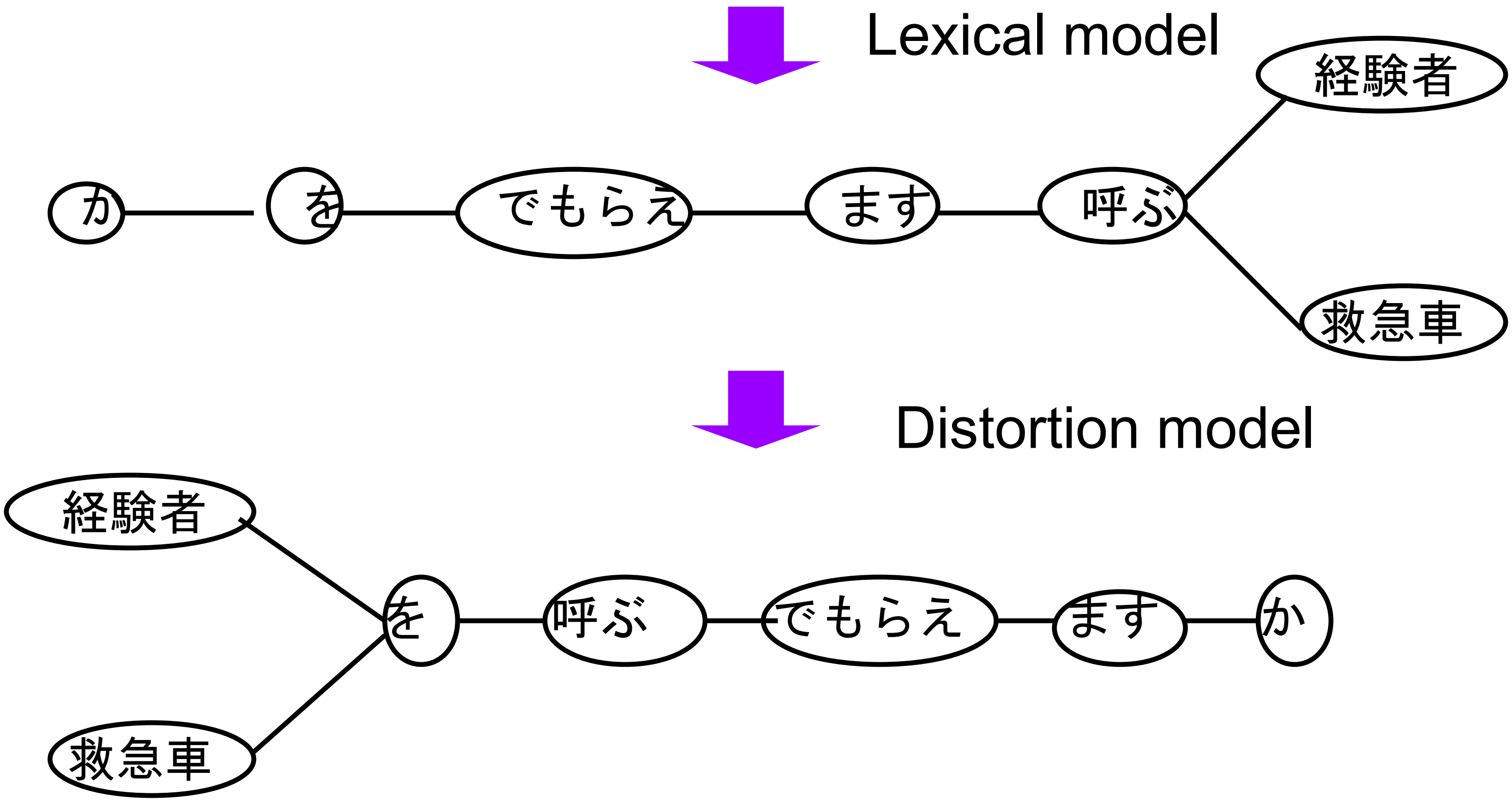
Distortion model



Machine Translation for Lattice



Machine Translation for Lattice



How We Translate Word Lattice

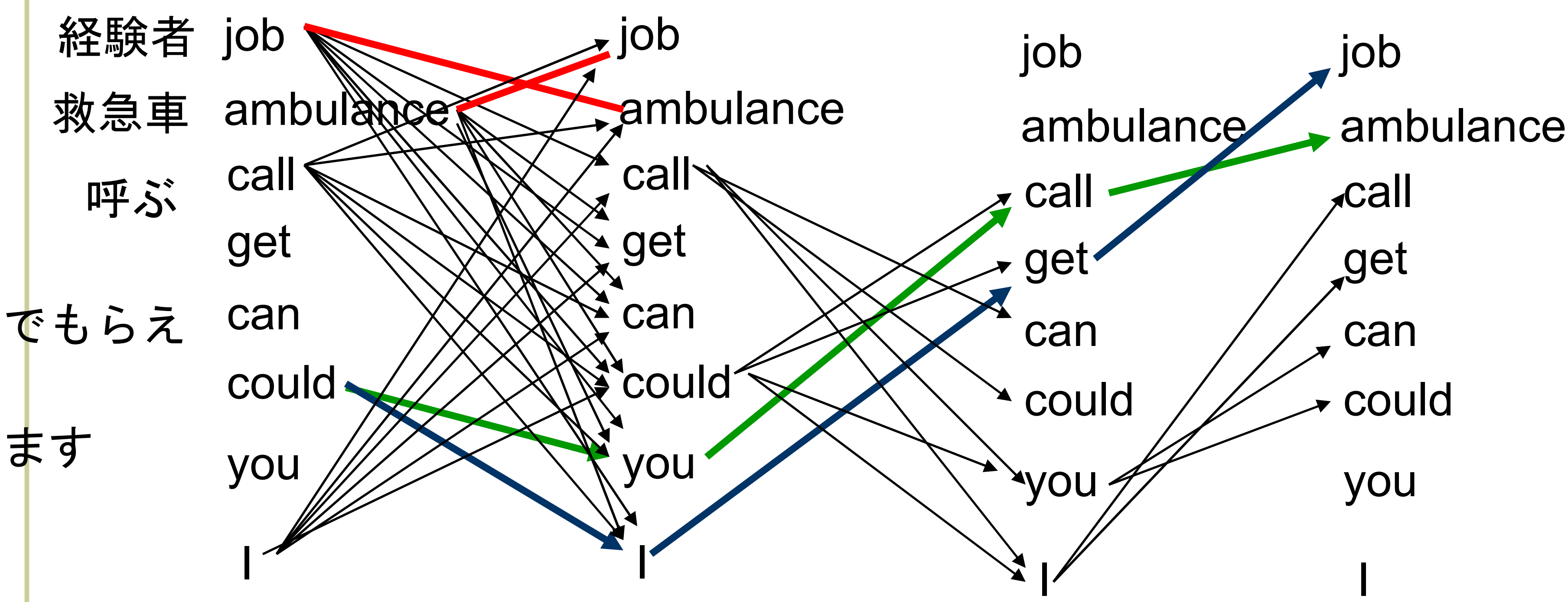
ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社
国際電気通信基礎技術研究所

- Two-step decoding: beam-search + A* search
- beam search: construct translation word graph (TWG)
  - An edge in the word lattice is mapped to an edge in the TWG
  - A path in the TWG corresponds to a path in the word lattice
  - Lower-scored edges are pruned.
  - Simple translation models are used.
- A* search:
  - Search the TWG with a higher-grade translation models(IBM model4)

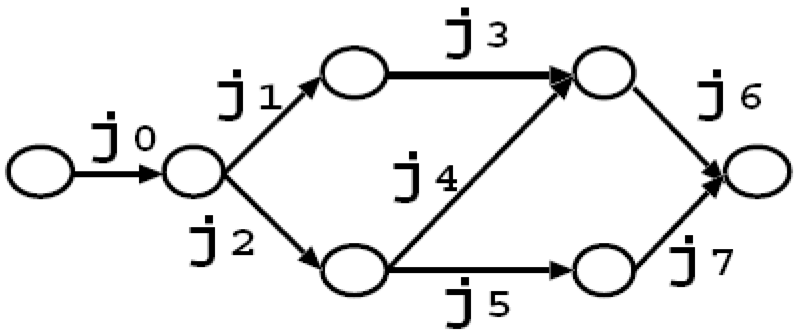
Illustration of Constructing TWG (Translation Word Graph)

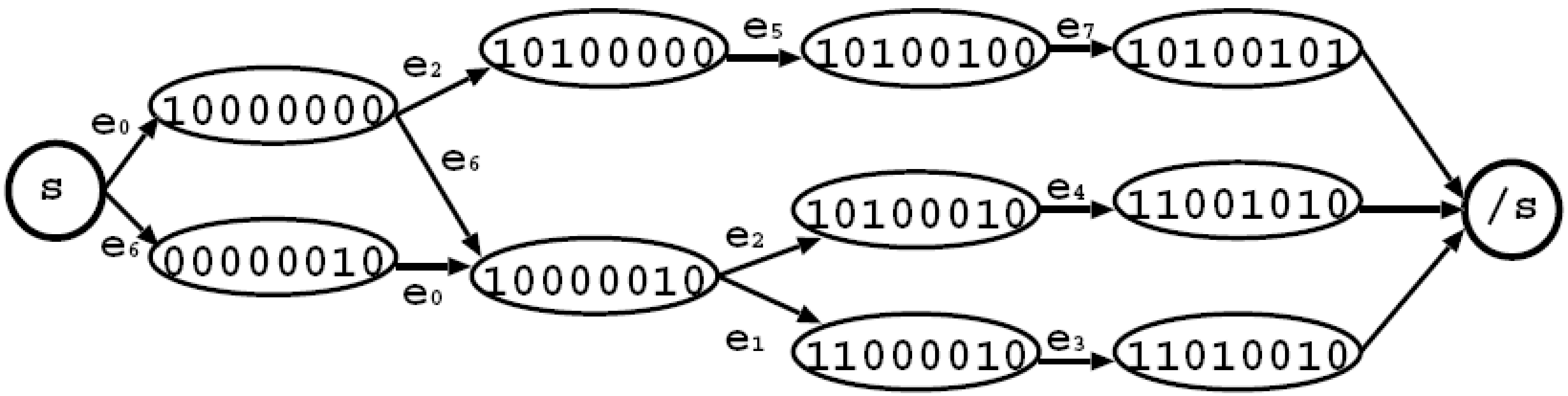
ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社 国際電気通信基礎技術研究所

Beam-search: threshold pruning



Translation Word Graph (example)





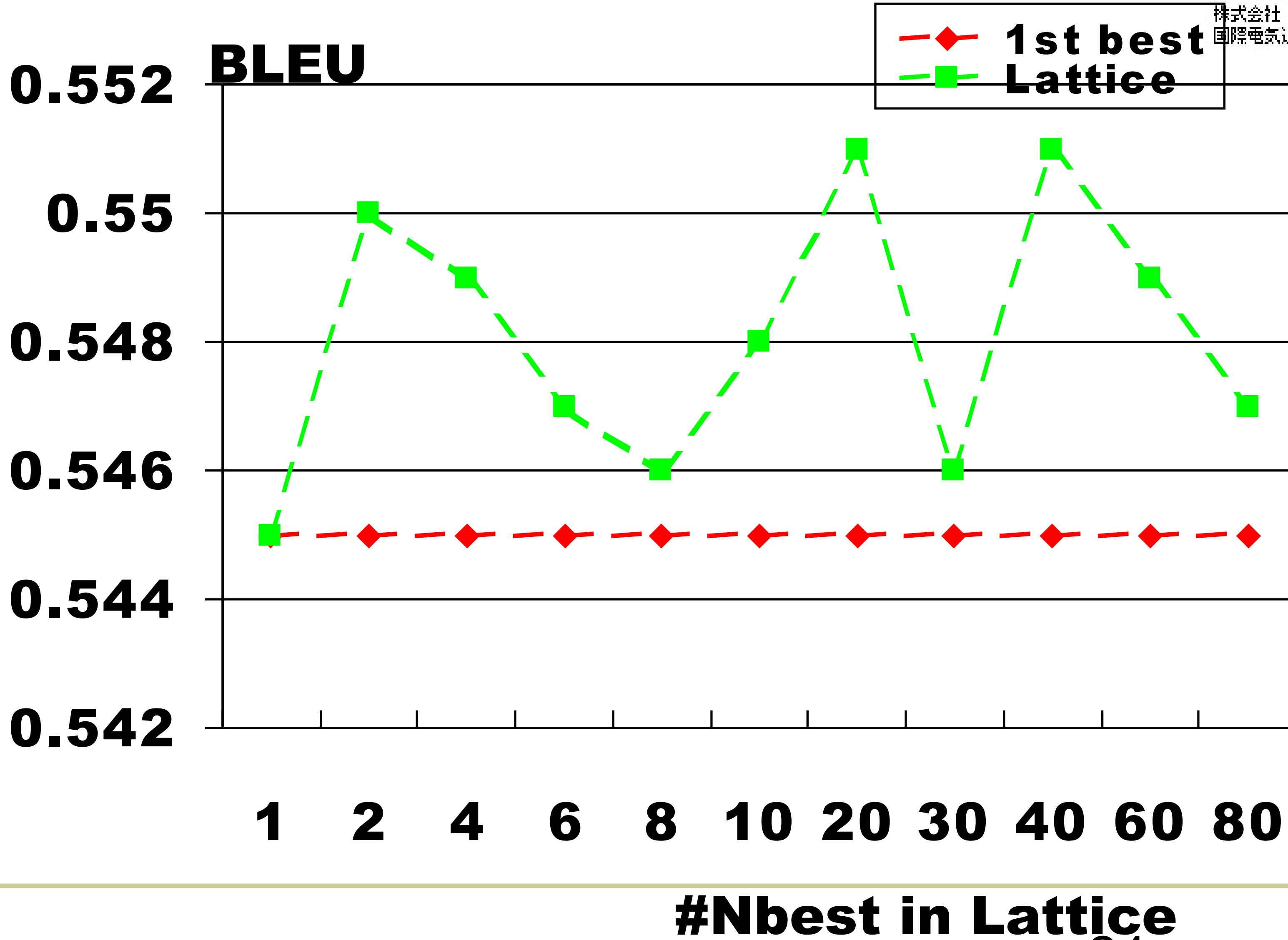
A* Search

- A* search
  - Forward score: Accumulated from the start node to current node, using IBM Model4 model
  - Heuristic score: Accumulated from the current node to the end node
- Approximations are made on the models dependent on the length of source sentence:
  - distribution model
  - NULL word

Features in Speech Translation Models

$$
\begin{aligned}
\hat{E} = \arg\max_{E} \{ & \lambda_0 \log P_{pp} + \lambda_1 \log P_{lm}(E) \\
& + \lambda_2 \log P_{lm}(POS(E)) + \lambda_3 \log P(\phi_0) \\
& + \lambda_4 \log N(\Phi \mid E) + \lambda_5 \log \mathrm{T}(J \mid E) \\
& + \lambda_6 \log D(E, J) \}
\end{aligned}
$$

Effect of Word Lattice Translation



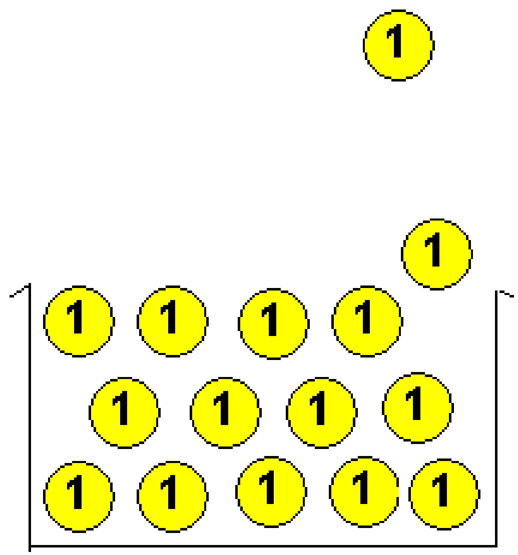
Beam-size effect in WLT

ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社
国際電気通信基礎技術研究所

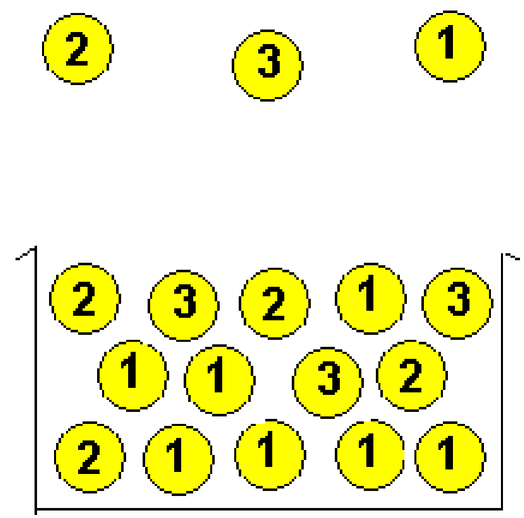
1 A translation of 1st best ASR hypotheses

2 A translation of 2nd best ASR hypotheses

3 A translation of 3rd best ASR hypotheses



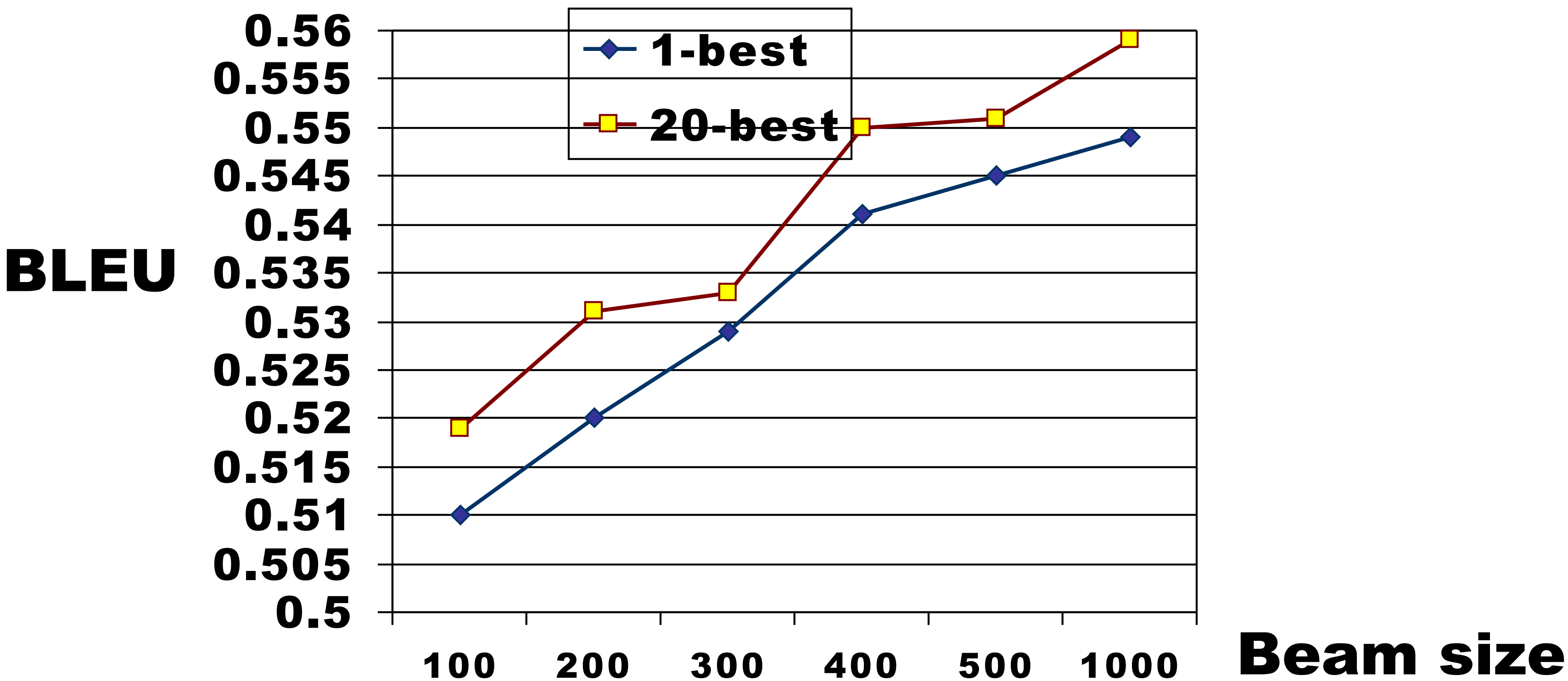
Single-best translation



Word lattice translation

Beam-size Effects in WLT (N-best=20)

- Promising hypotheses pruned in WLT but saved in single-best translation under the same beam size

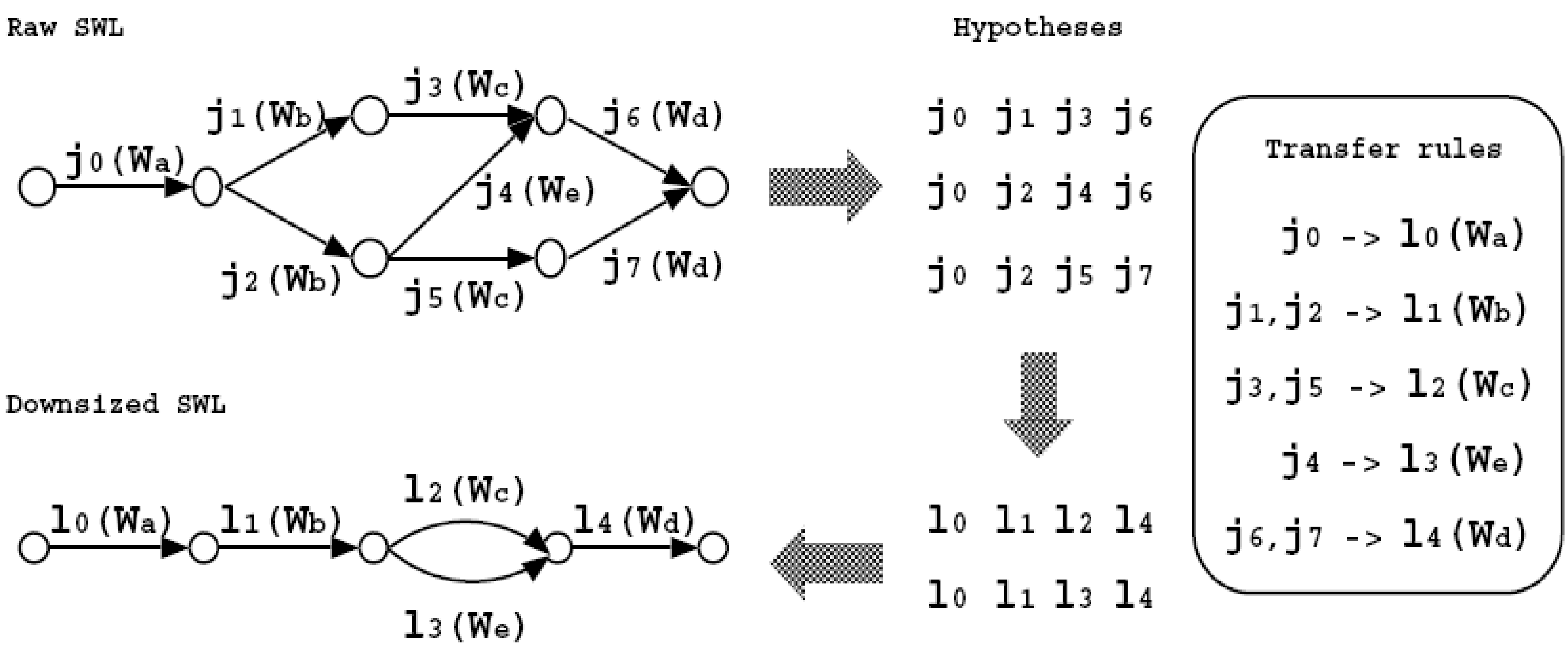


Why Word Lattice Minimization

ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社
国際電気通信基礎技術研究所

- Raw lattice is too huge
- A lot of duplicated word IDs in the lattice
- Significant are the top N-best hypotheses
- Minimization under the light of machine translation
- Minimization can make decoding fast
- Minimization can reduce translation error; reduce pruning error in decoding

Word Lattice Minimization



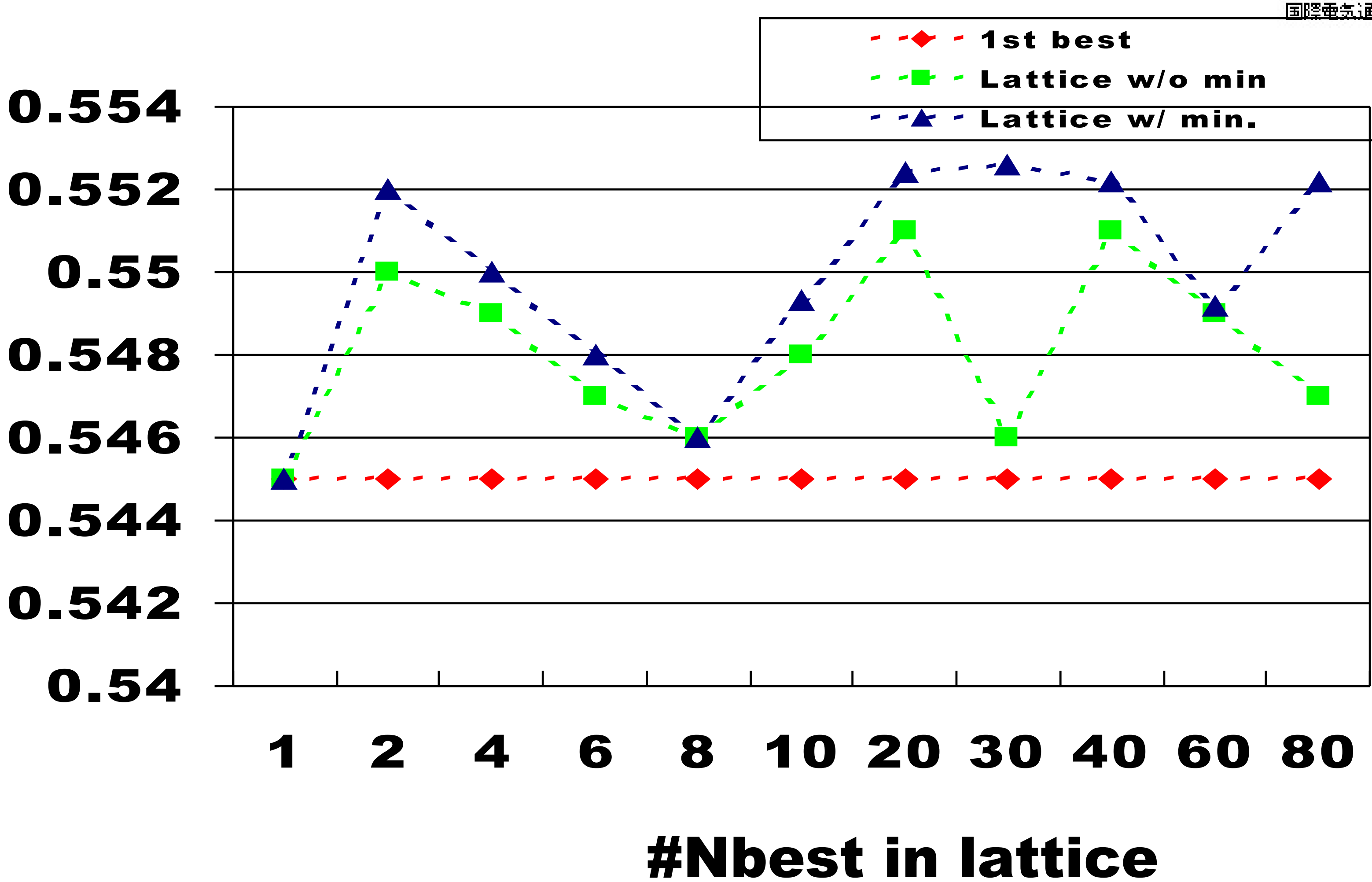
Word Lattice ?? N-best ??

ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社 国際電気通信基礎技術研究所

- After lattice minimization, the output is not a lattice again. Only N-best with new assigned edge ids.
- After lattice minimization, the ASR score lost in single edge. Instead, we use ASR path score to represent single edge's score.

Effect of Lattice Minimization

ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社
国際電気通信基礎技術研究所



Posterior Probability

ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社
国際電気通信基礎技術研究所

- Integrating acoustic model and language model probabilities
- Indicating relative accuracy of N-best hypotheses

$$p(J_j \mid X) = \frac{e^{\lambda \log score_j}}{\sum_{i=1}^{N} e^{\lambda \log score_i}}$$

$\log score_i$:log-scale ASR score (AM+LM)

Confidence Measure Filtering

ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社国際電気通信基礎技術研究所

- ASR hypotheses with very low posterior probability degrade translations
- A predefined confidence threshold, T, is applied to remove the most unlikely ASR hypotheses
- By comparing a hypothesis's posterior probability to the single-best hypothesis's posterior probability multiplied by T, $P_{first\text{-}best}$*T, remove the smaller.

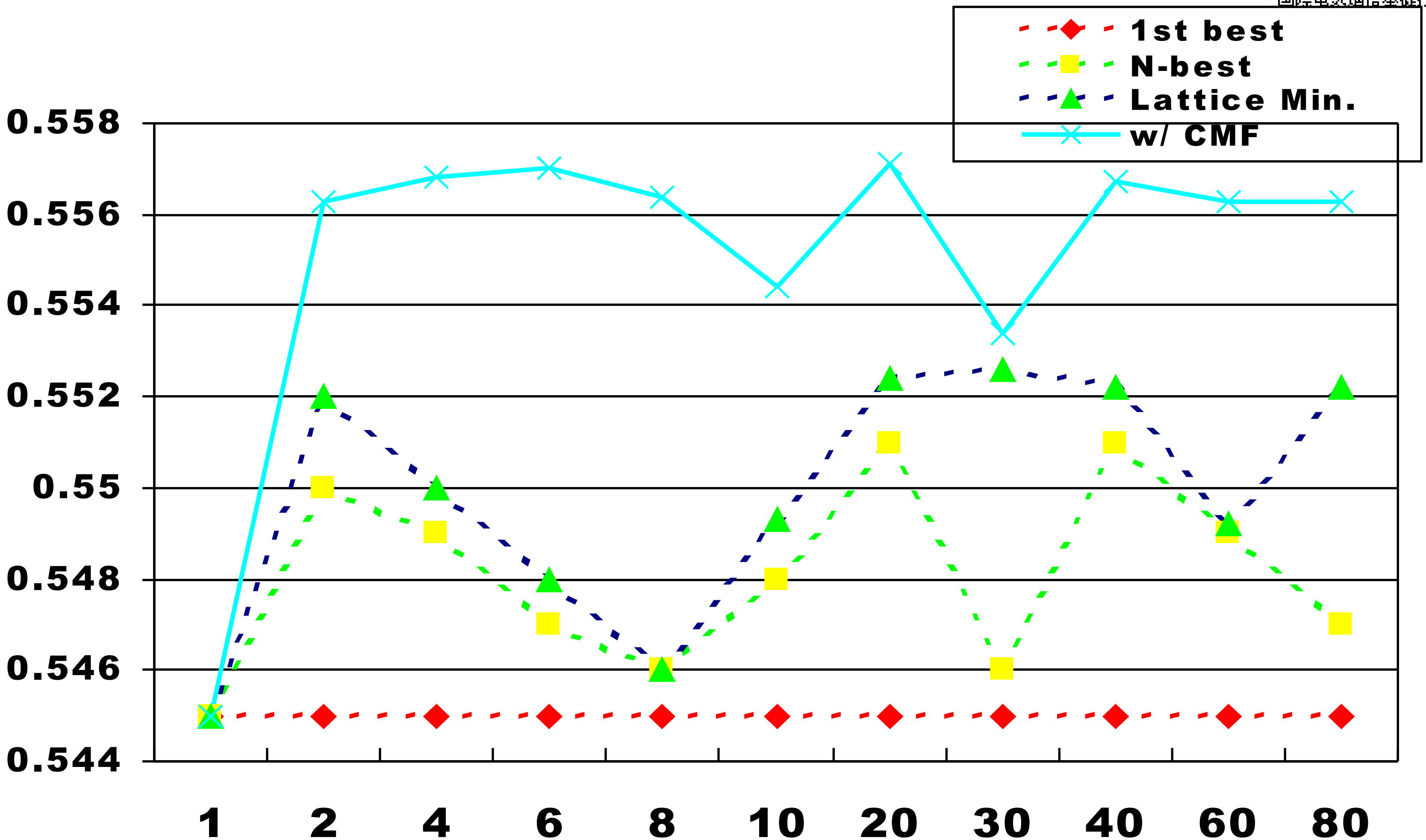
Confidence Measure Filtering

ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社 国際電気通信基礎技術研究所

ASR Output	ASR score	PP=Posterior probability	cmf= PP/PP1-st	Decision cmf>0.5 ?
1st cand.	0.55	0.215	1	PASS
2nd cand.	0.50	0.196	0.912	PASS
3rd cand.	0.45	0.176	0.818	PASS
4th cand.	0.40	0.157	0.730	PASS
5th cand.	0.30	0.118	0.549	PASS
6th cand.	0.20	0.078	0.363	FAIL
7th cand.	0.10	0.039	0.181	FAIL
8th cand.	0.05	0.020	0.09	FAIL

SUM=2.25

Effect of CM filtering



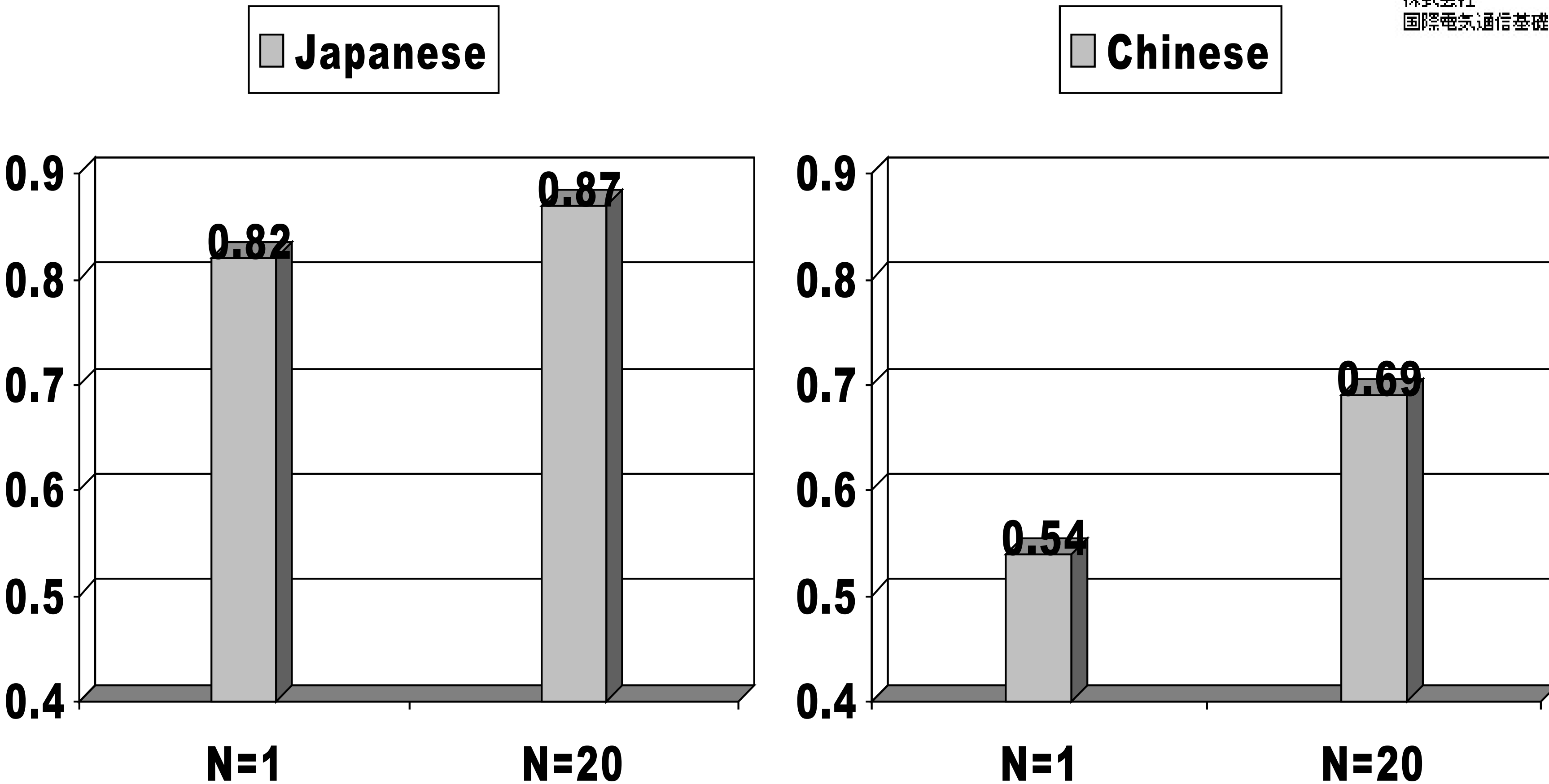
IWSLT 2005 Evaluation

IWSLT 2005 Evaluation(training data)

ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社
国際電気通信基礎技術研究所

Language pair	Data track	Data size	perplexity	
			Testset	Dev.data
C/E	Supplied +tagger	20K	65.4	53.8
	C-star	172K	69.3	52.2
J/E	Supplied +tagger	20K	54.9	53.7
	C-star	463K	22.5	31.6

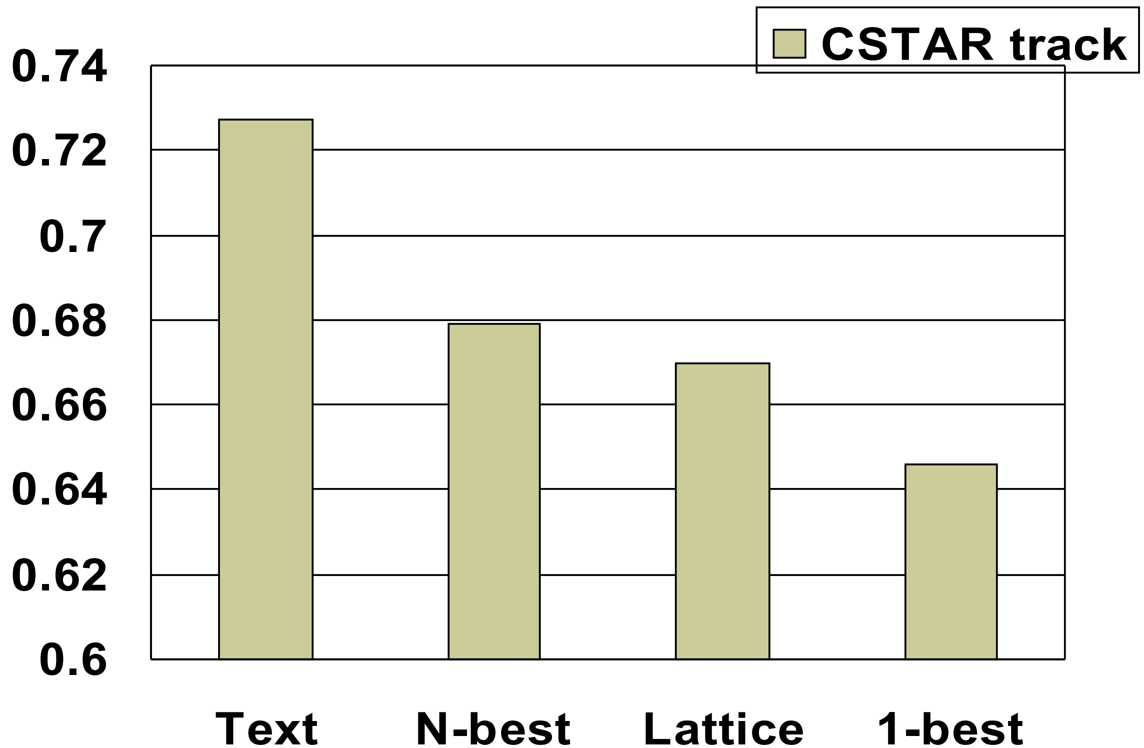
Test Data Analysis



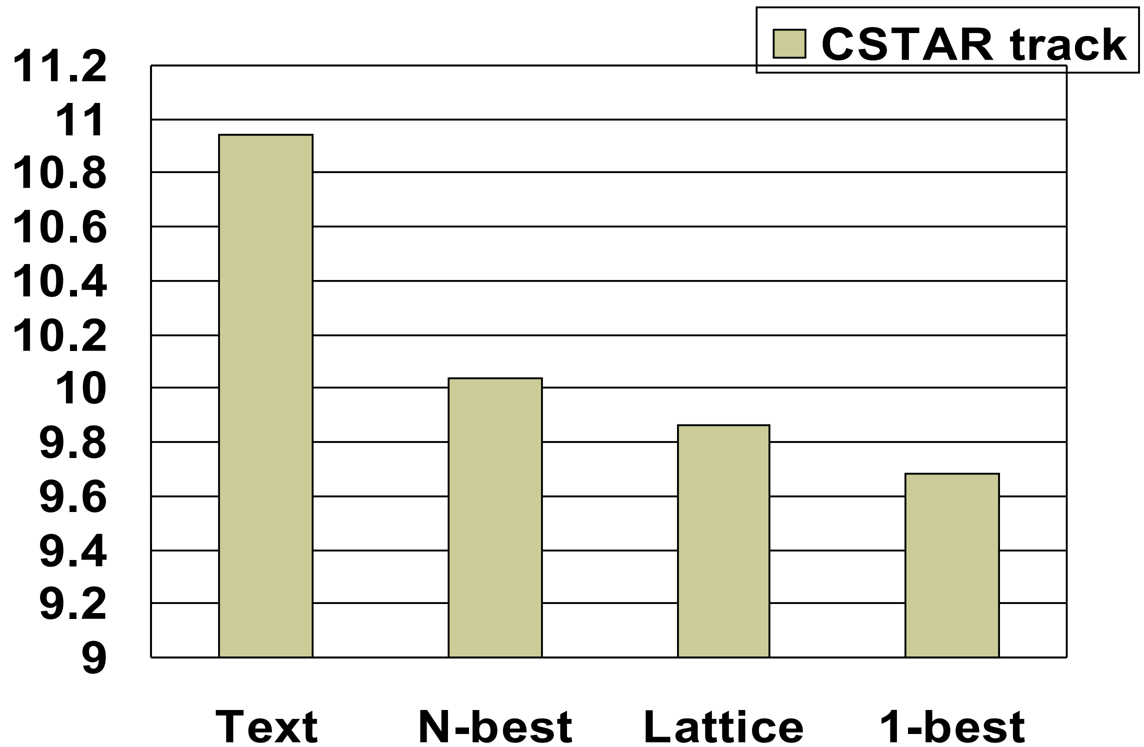
ASR Recognition Accuracy

Test Data Results (J/E BLEU C-star track)

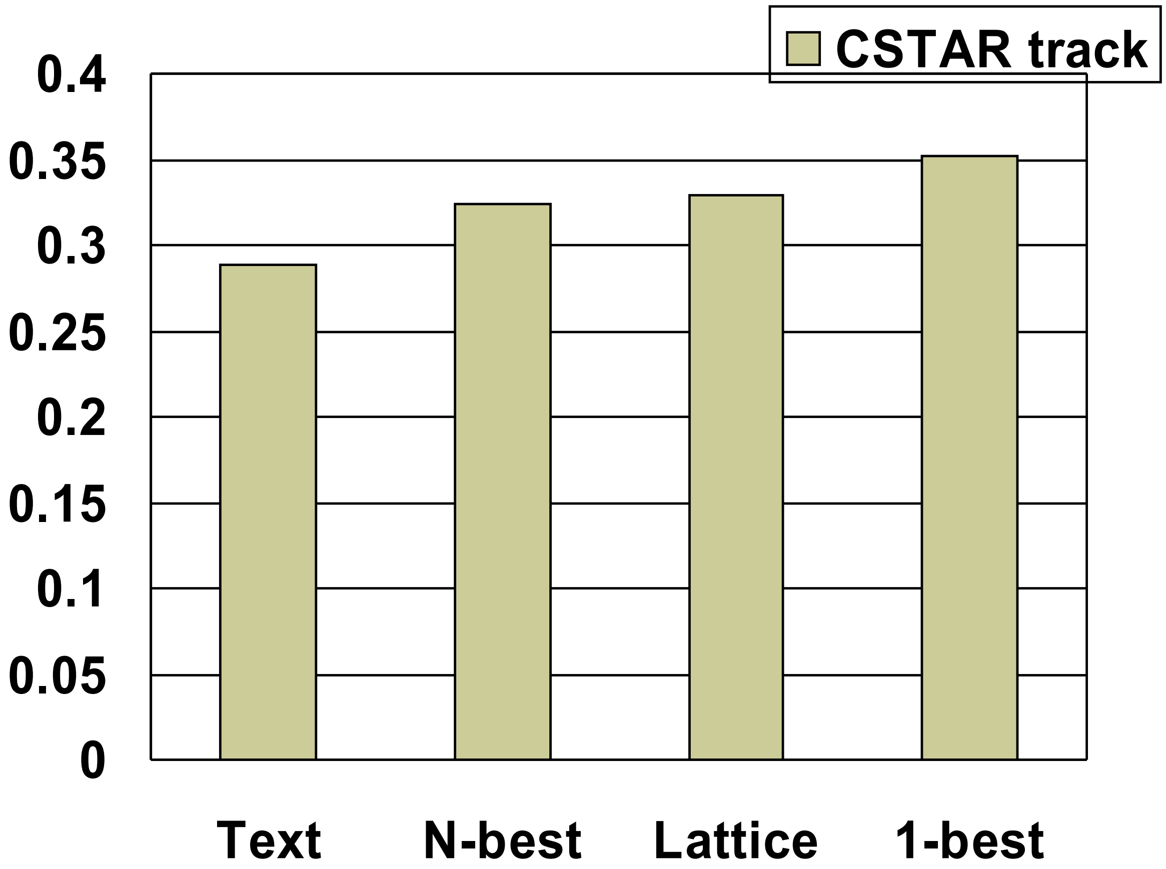
ATR
ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社
国際電気通信基礎技術研究所



Test Data Results (J/E NIST C-star track)



Test Data Results(J/E WER C-star track)



Evaluation Results (CE)

Data track	Input	BLEU	NIST	WER	PER	METEOR	GTM
Supplied+tools	Text	0.305	7.20	0.607	0.494	0.574	0.471
	Nbest	0.267	6.19	0.645	0.546	0.506	0.421
	Sbest	0.251	5.93	0.683	0.581	0.479	0.395
Cstar	Text	0.421	8.17	0.518	0.422	0.642	0.547
	Nbest	0.375	6.80	0.561	0.486	0.560	0.493
	Sbest	0.340	6.76	0.619	0.525	0.531	0.461

Evaluation Results(JE)

Data track	Input	BLEU	NIST	WER	PER	METEOR	GTM
Supplied+ tagger	Text	0.388	4.39	0.563	0.519	0.520	0.431
	Nbest	0.383	4.27	0.574	0.530	0.513	0.422
	Lattice	0.378	4.18	0.578	0.534	0.511	0.420
	Sbest	0.366	4.50	0.576	0.527	0.508	0.412
Cstar	Text	0.727	10.94	0.289	0.243	0.80	0.716
	Nbest	0.679	10.04	0.324	0.281	0.760	0.670
	Lattice	0.670	9.86	0.329	0.289	0.763	0.665
	Sbest	0.646	9.68	0.352	0.304	0.741	0.645

Remarks

- Text translation (0.727) > N-best translation (0.679)
- N-best translation (0.679) > lattice translation (0.67)
- Lattice translation (0.670) > single-best translation (0.646)

- Training data size influences speech translation

Analysis: Lattice Translation Worse than N-best Translation

- We used the same number of ASR hypotheses in N-best translation and lattice translation
- In beam search, N-best translation and lattice translation used the same beam size and threshold in pruning
- Model approximations and inaccuracy: distortion, null, acoustic model, language model.

Comparisons of the structures

- Single-best translation
  - ✓ Simple, direct
  - ✓ ASR and SMT isolated optimization
  - ✓ MT flexible, easy to upgrade, multiple translation engines
  - × Non-robust to ASR WER
- N-best hypothesis translation
  - ✓ Robust, resistant to ASR WER
  - ✓ MT flexible, multiple translation engines
  - × Slow, duplicate calculation
- Word lattice translation
  - ✓ Reduce computing cost, efficient
  - ✓ Speech translation system, ASR and SMT, overall optimized
  - × MT inflexible

Conclusions

ATR
ADVANCED TELECOMMUNICATIONS RESEARCH INSTITUTE INTERNATIONAL
株式会社
国際電気通信基礎技術研究所

- We applied two approaches to improve ASR single-best translation.
- By applying a log-linear model, N-best translation approach can improve single-best translation effectively.
- We observed improved speech translation performance in word lattice translation:
  - Confidence measure filtering
  - Word lattice reduction