音声翻訳技術

Graham Neubig
奈良先端科学技術大学院大学（NAIST）
2015/5/11
http://phontron.com/slides/neubig15jsai-slides.pdf

共著者：
中村哲、戸田智基、 Sakriani Sakti 、
叶高朋、大串正矢、藤田朋希、
清水宏晃、小田悠介、三重野隆史、 Do Quoc Truong

音声翻訳

引用元：情報通信研究機構　YouTube

音声翻訳システム

音声認識

こんにちは、駅はどこですか？

機械翻訳

Hello, where is the station?

音声合成

次世代の音声翻訳に向けた２つの課題

- 音声翻訳結果を早く聞き手に届けられるか？
  →同時音声翻訳
- 音声の表現力を維持して翻訳できるか？
  →音響特徴の翻訳
- 音声認識誤りに頑健に翻訳できるか？
  →複数の候補の考慮と同時最適化

今回の範囲外：
→ 句読点の挿入 [Peitz+11]
→ 言いよどみへの対処 [Wang+10]

同時音声翻訳

音声翻訳システム

- ある言語の音声から違う言語の音声へ翻訳

音声認識

こんにちは、駅はどこですか？

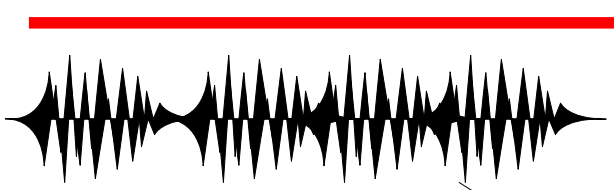
機械翻訳

Hello, where is the station?

音声合成

遅延の問題

- 従来のシステムは１文の入力が終わるまで翻訳しない！



遅延

音声認識

こんにちは、駅はどこですか？

機械翻訳

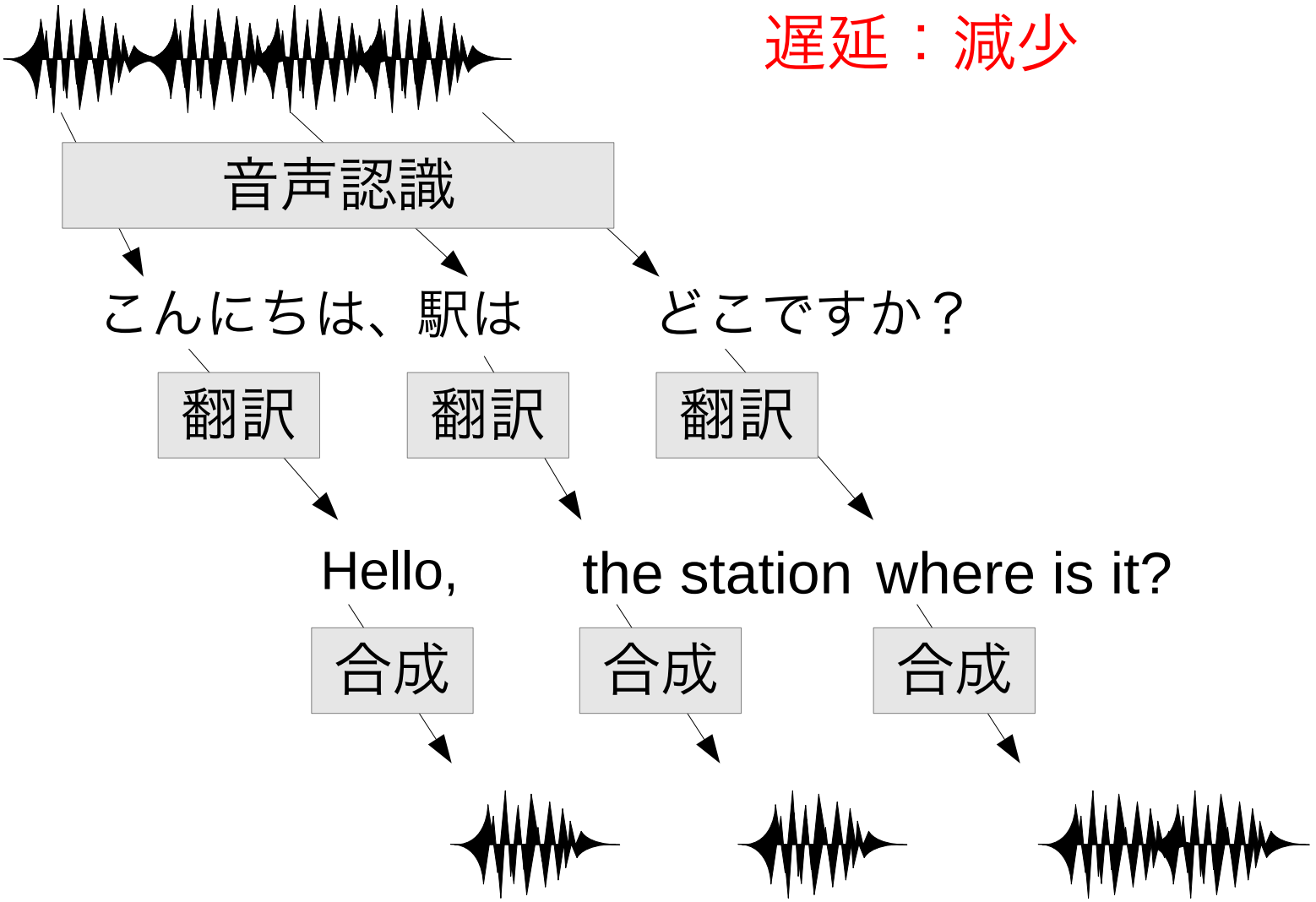
Hello, where is the station?

音声合成

目標：遅延の低減

- １文が完全に終わる前に適切なタイミングで翻訳開始

遅延：減少



遅延の削減に関わる研究

- [Fugen+ 2007]
  - 言語モデルや音響情報に基づくタイミング決定
- [Bangalore+ 2012]
  - 音声認識の無音区間（pausing）を用いて文を分割
- [Rangarajan+ 2013]
  - 予測された句読点の挿入位置（コンマ、ピリオド、その他）を使用
  - 数種類の手法を比較検討　…　句読点による手法が最高性能
- [Fujita+ 2013]
  - 翻訳モデルにおける並べ替えの位置を推定
  - 並べ替えが起きなさそうな場所で翻訳

訳出タイミングの最適化
[Oda+ ACL14]

訳出タイミング決定法の問題

すべてヒューリスティクスに基づく手法

音韻的情報、言語的情報　…

- 分割位置が翻訳精度に与える影響を考慮せず
- 翻訳器に対して分割位置が最適化されていない

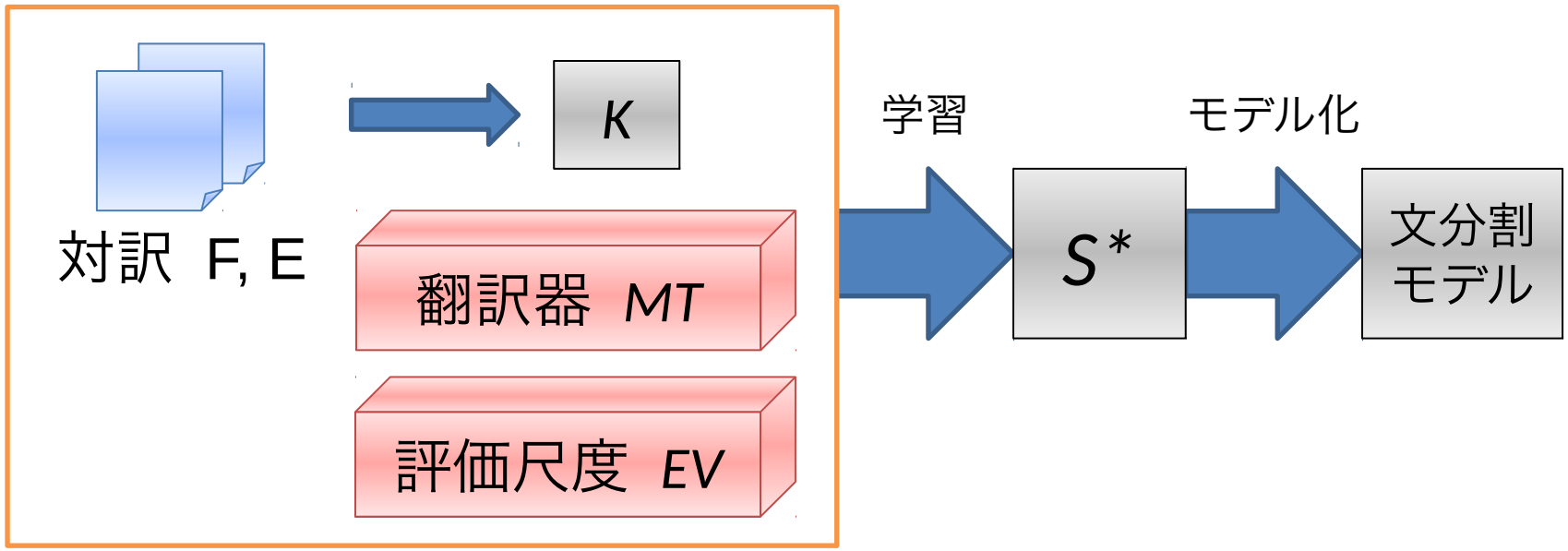
タイミングを最適化したい！

定式化

1)学習データ (対訳コーパス) 全体で分割する数 K を決定

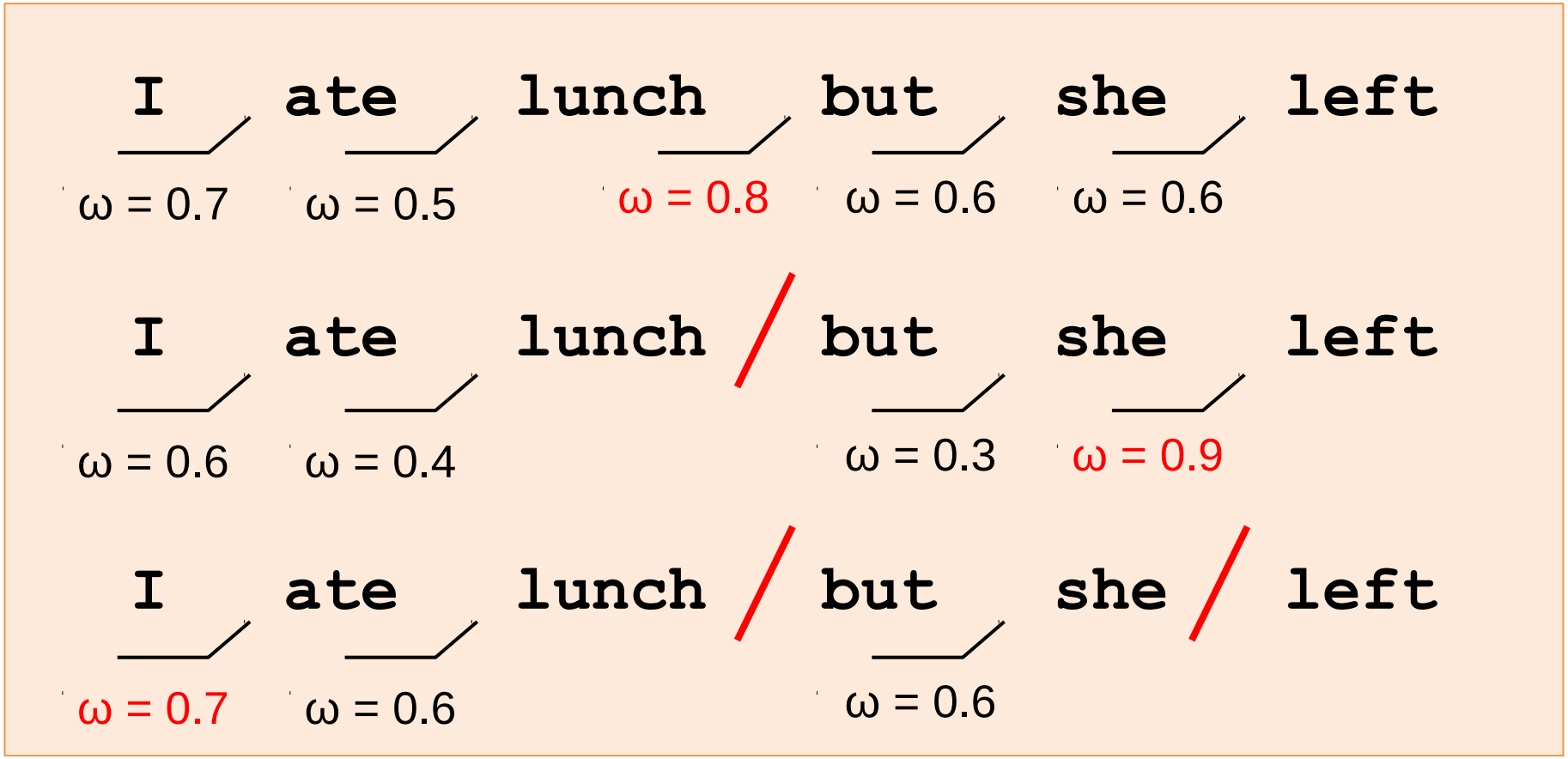
2)K 個の分割位置を学習データから選択

3)分割位置の素性をモデル化



手法 1: 貪欲法による分割

- 次の分割位置を決めるとき、今までに選んだ分割位置を保持
  (＝貪欲法 : greedy search)



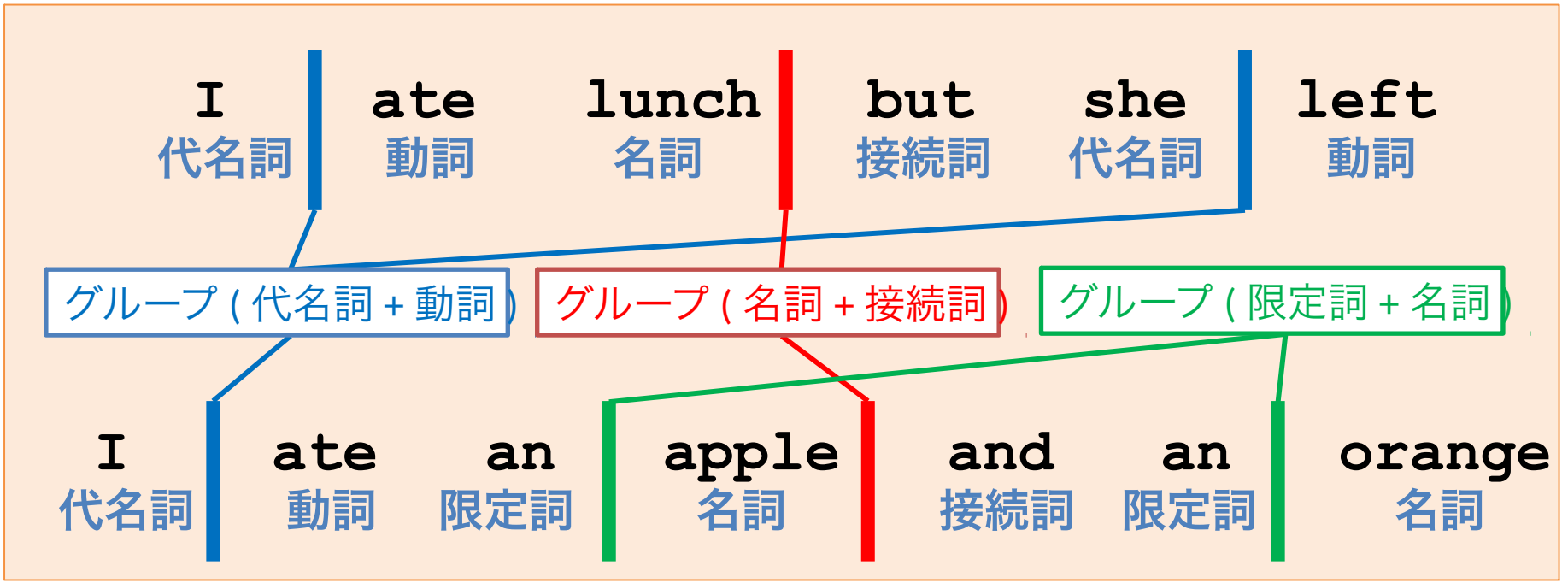
→ 選ばれた分割位置の素性を SVM で学習

手法２：素性のグループ化

- ω は複雑な関数、ノイズが多い
  - 貪欲法では偶然 ω が良くなる分割位置で過学習

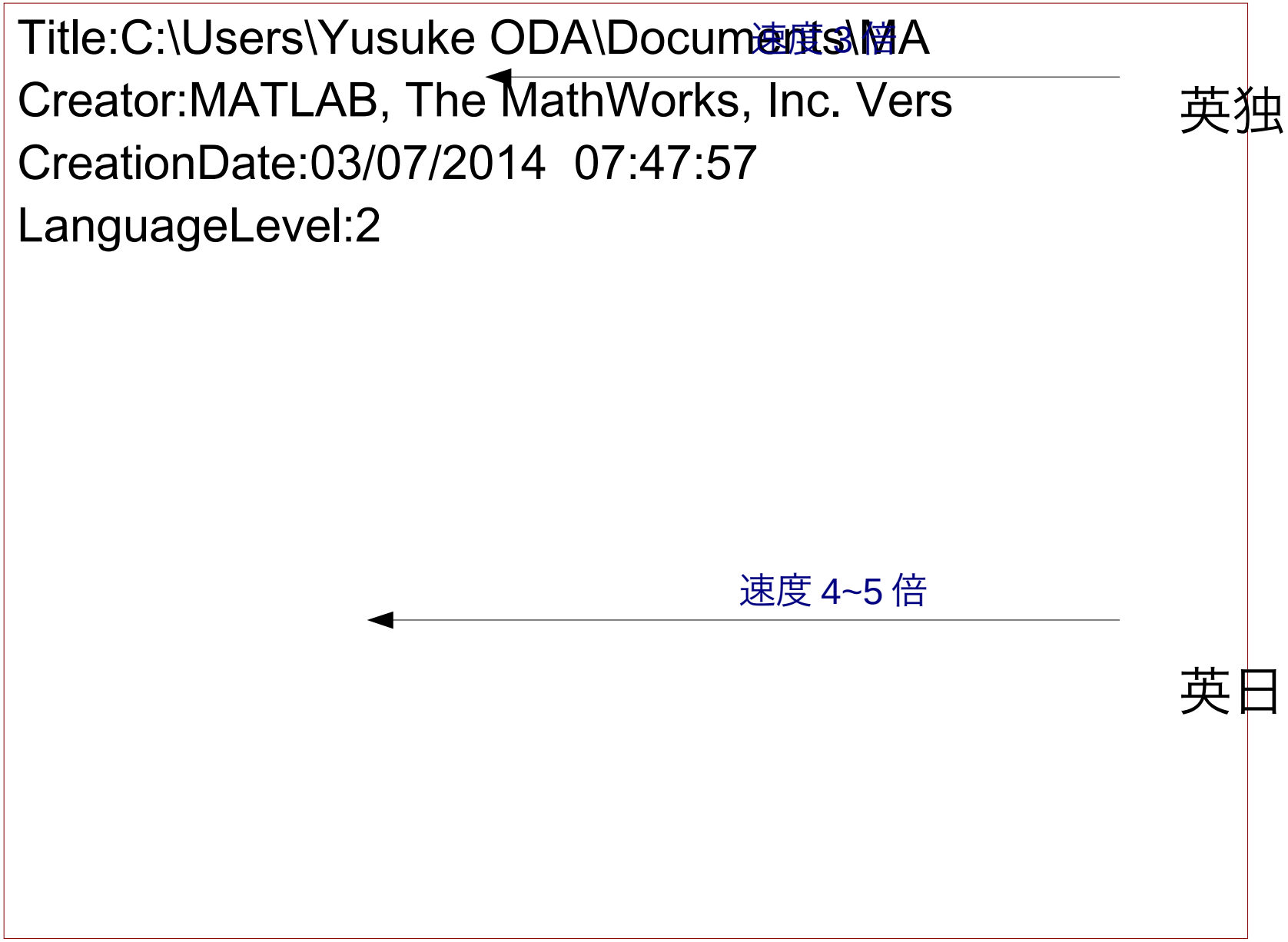
解決策　…　同じ素性を持つ分割位置をグループ化、同時に分割

例：前後の品詞



動的計画法 (DP) で探索、
探索で素性が得られるので　モデル化は不要
正則化の導入も可能

実験結果（BLEU）

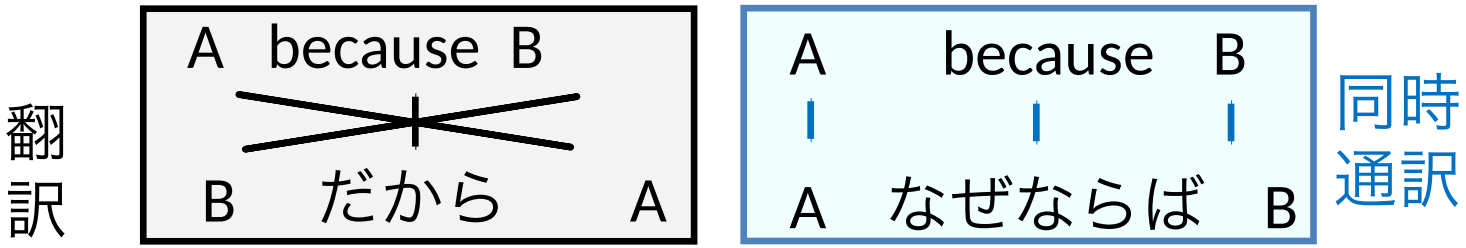

Title:C:\Users\Yusuke ODA\Documents\MA
Creator:MATLAB, The MathWorks, Inc. Vers
CreationDate:03/07/2014 07:47:57
LanguageLevel:2
速度3倍
英独
速度 4~5 倍
英日

システムデモ
([Fujita+13] の手法)

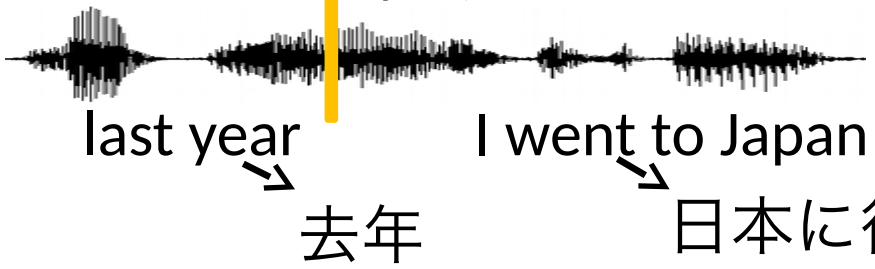
同時通訳データを用いた同時音声翻訳 [Shimizu+ 13, Shimizu+ 14]

同時通訳者の技術

- 語彙の選択
  - 文法構造が異なる言語対において並び替えを減少させる狙い



- サラミテクニック [Jones 02]
  - １つの長い文を複数のチャンクに分割



遅延時間を短縮するために
同時通訳者は様々な技術を駆使

同時通訳データ

- 収録材料
  - TED 講演（英語 → 日本語）

利点：翻訳データ（字幕）と同時通訳データを比較

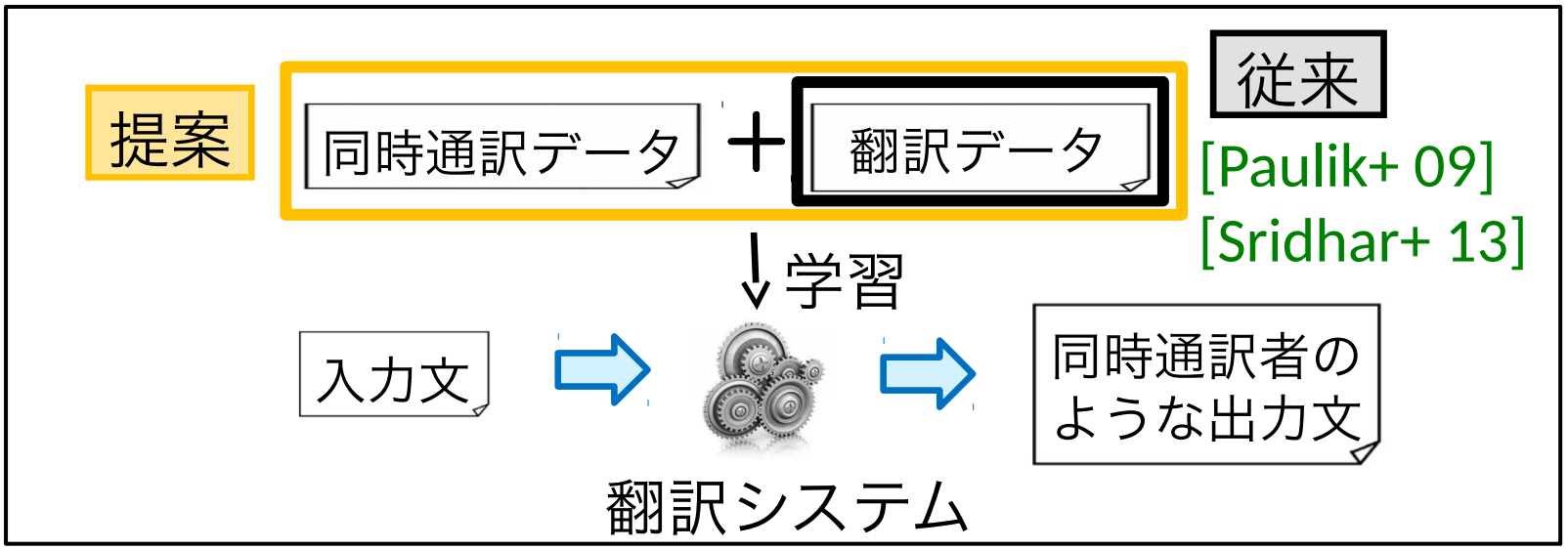
- 同時通訳者
  - 通訳経験年数が異なる三人
  - 経験が長い順に S, A, B ランク

Experience	Rank
15 years	S rank
4 years	A rank
1 year	B rank

利点：同時通訳の訳出の違いを分析

同時通訳データの適用

- アプローチ
  - 同時通訳者のように訳出する同時音声翻訳の構築

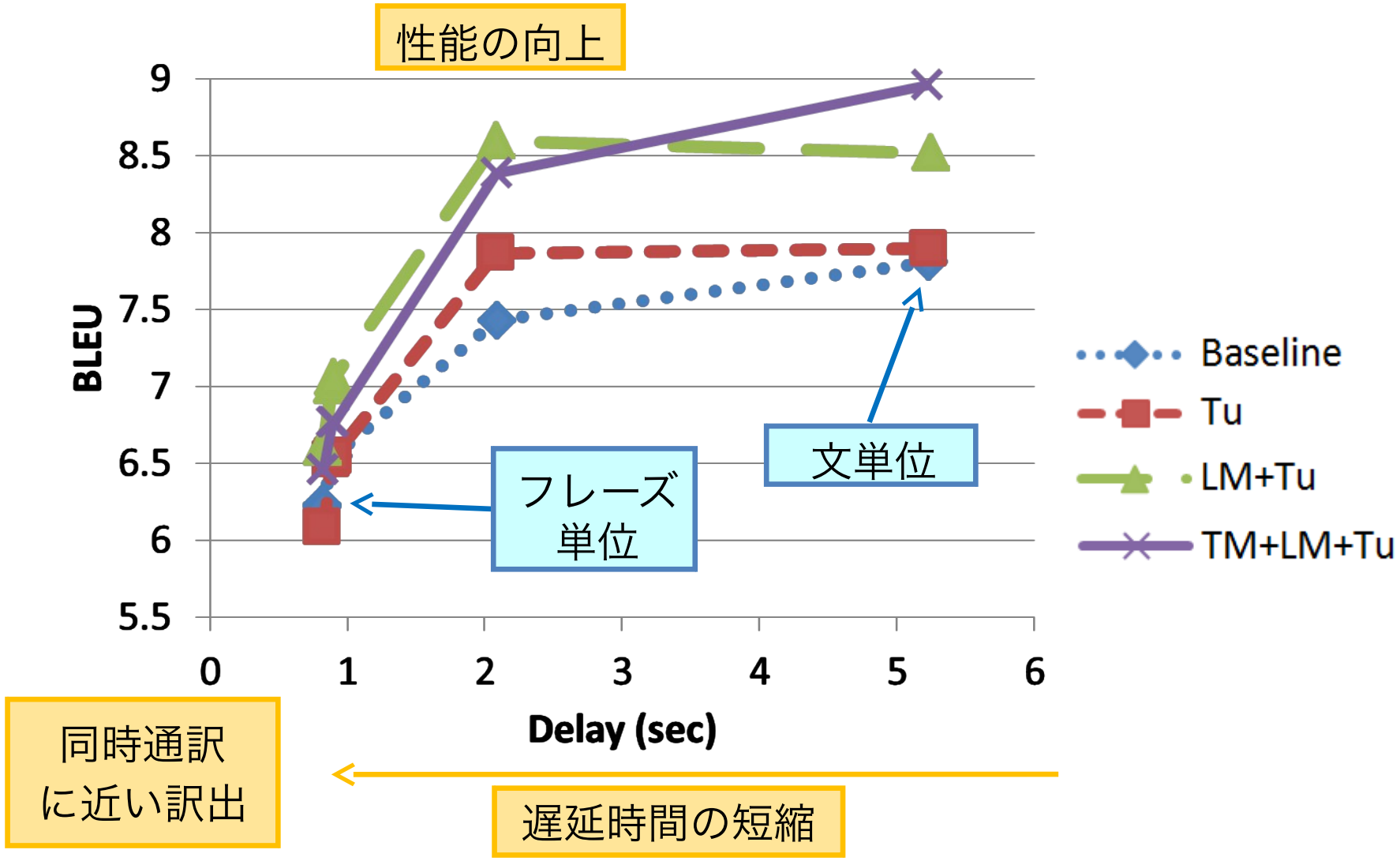


分野適応の技術を用いた通訳者の再現

- チューニング (Tu)
  - 同時通訳者の訳出結果に近づくようにパラメータが調節されることを期待
- 言語モデル (LM) ：線形補間
  - 同時通訳者のような語順や語彙選択を期待
- 翻訳モデル (TM) ： fill-up 法 [Bisazza+ 11]
  - LM と同様，同時通訳者のような語彙選択を期待

機械翻訳システムの学習における３つの過程に
同時通訳データを利用

通訳データを用いた評価実験



訳出の例

	例文
入力	If you look at in the context of the history you can see what this is doing
正解	過去から流れを見てみますと災害はこのように増えています
従来	見てみると歴史の中で見ることができますこれがやっていること
提案	では歴史の中で見ることができますこれがやっていること

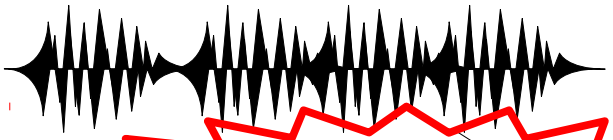
- 単語数の減少
  - チューニングによって短いフレーズを好むパラメータに調整
- 翻訳結果の 25% の文が「で」から開始
  - 同時通訳者が次の文を待つまでの沈黙を回避

音響特徴の翻訳

声の特徴は多く語る

問題！

- 音声認識の時点で声の特徴が失われる…



音声認識

こんにちは、駅はどこですか？

機械翻訳

Hello, where is the station?

音声合成

声の特徴を翻訳する音声翻訳
[Truong+ 15]

実験題材：強調の翻訳 [Kano+12, Kano+13]

Hello, I'm Mike. My membership number is 581.

No !, my ID is 581.

Five **Eight** One

（強調）

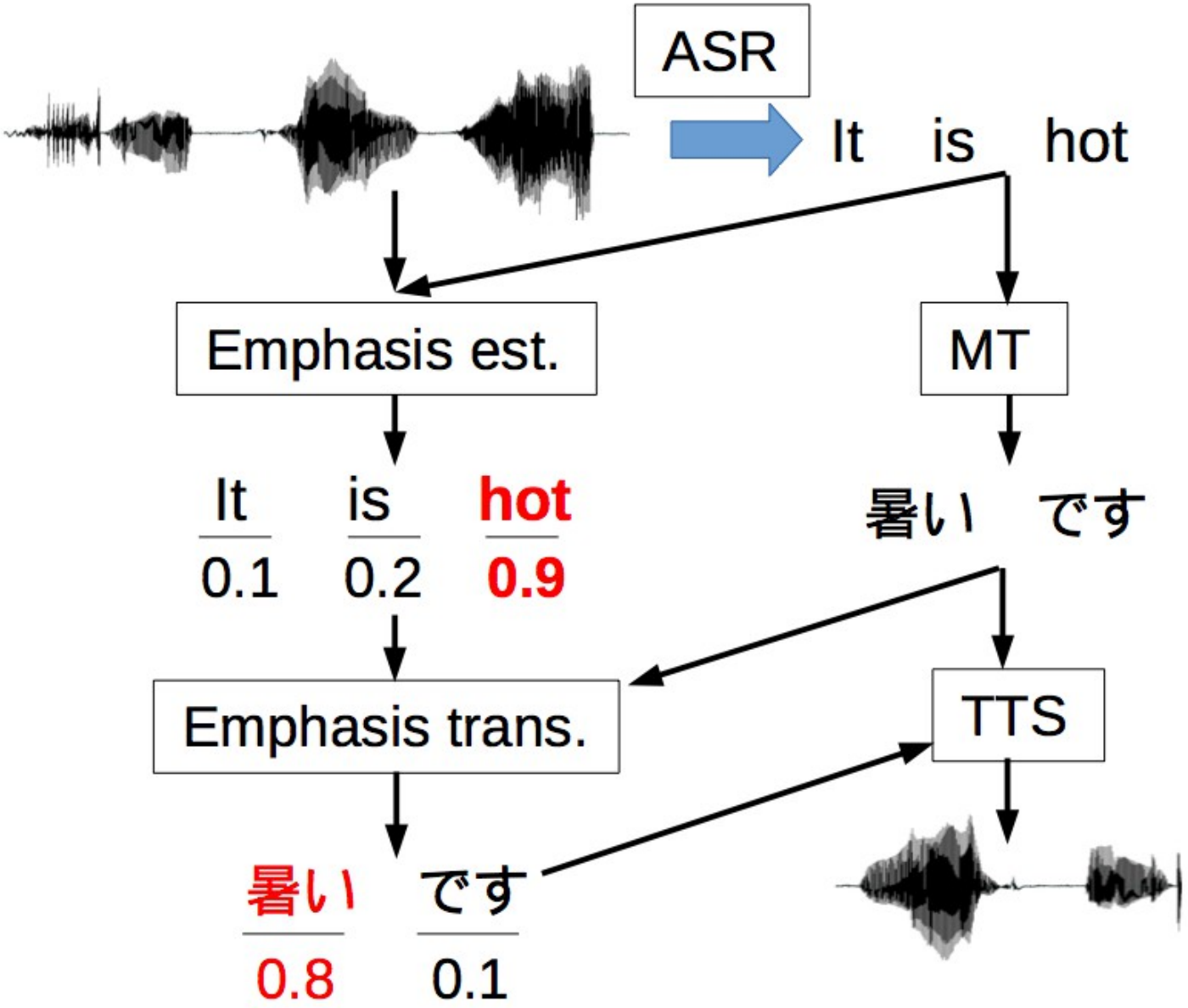
会員番号 *511* のマイク様ですね？

失礼しました . 会員番号 *581* のマイク様ですね？

ご **はち** ご

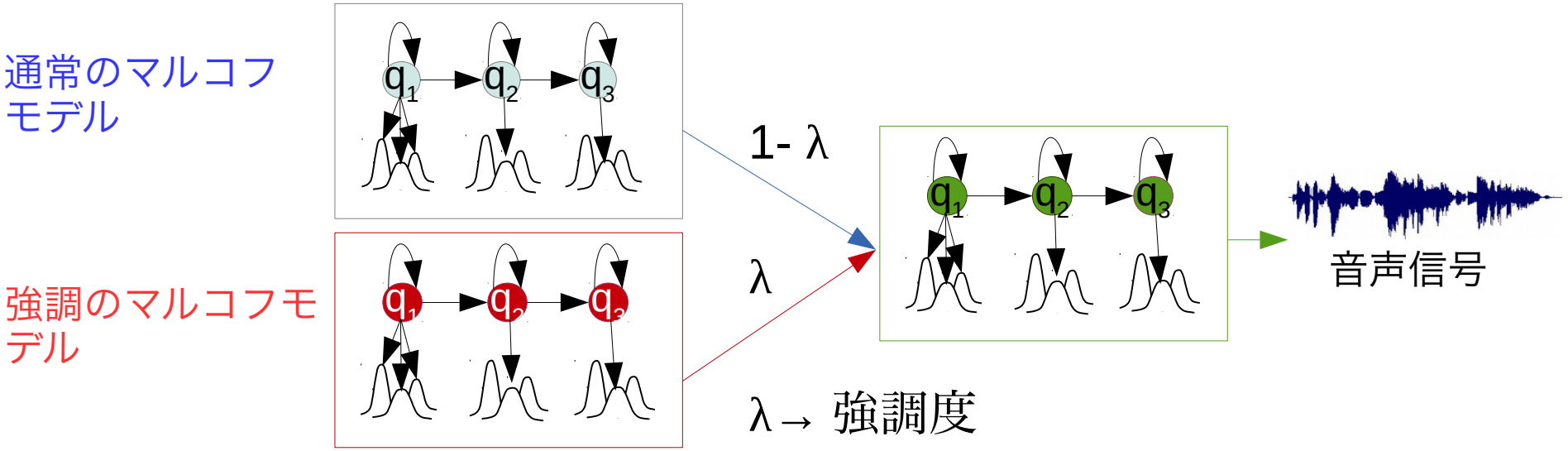
（強調）

音声の特徴の翻訳


ASR
It is hot
Emphasis est.
MT
It is hot
0.1 0.2 0.9
暑い です
Emphasis trans.
TTS
暑い です
0.8 0.1

音響翻訳の枠組み

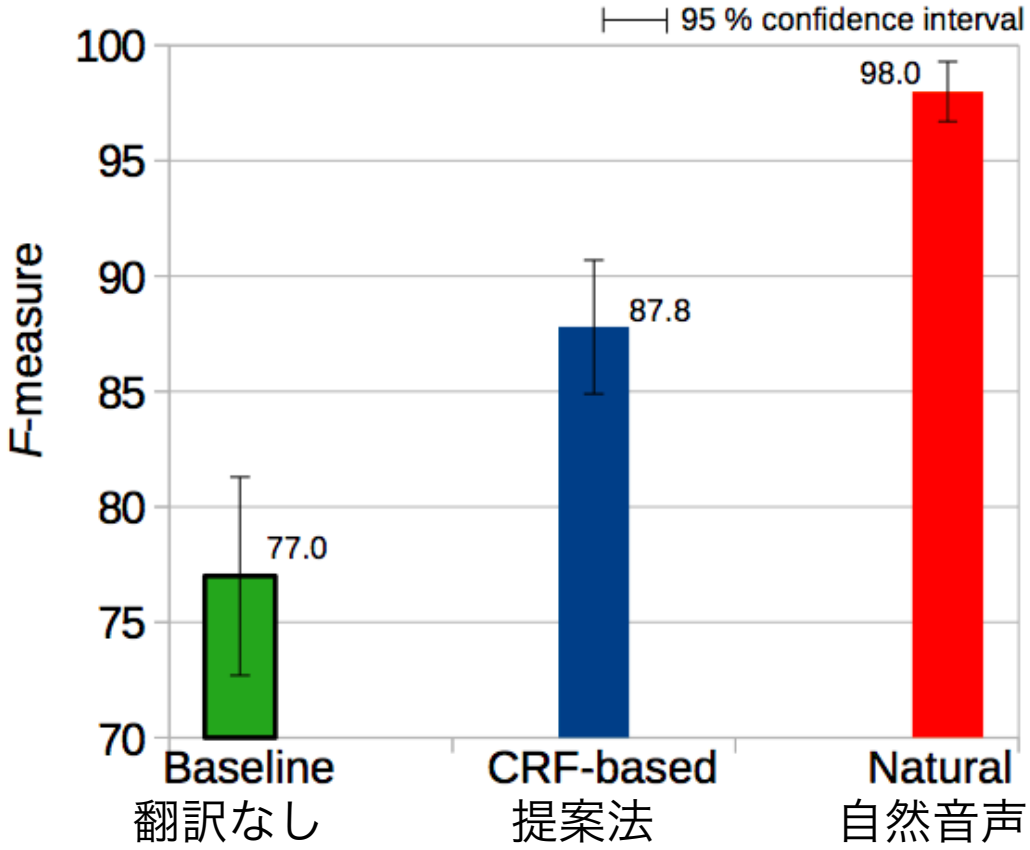
- 強調あり音声の認識・合成：
  重回帰隠れセミマルコフモデル (MR-HSMM)



- 音響特徴の翻訳：条件付き確率場 (CRF)

実験結果

- 人間に聞かせ「どの単語が強調？」と聞いた時の正解率



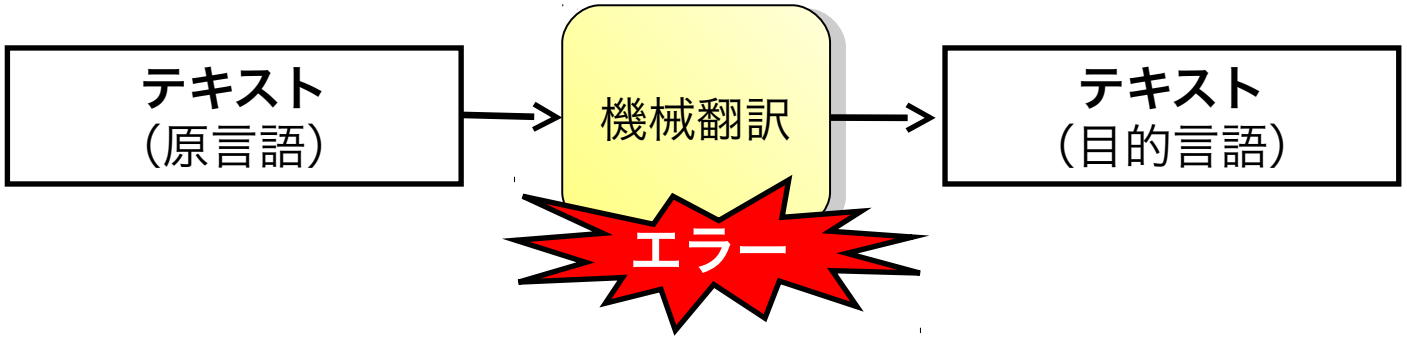
例１：

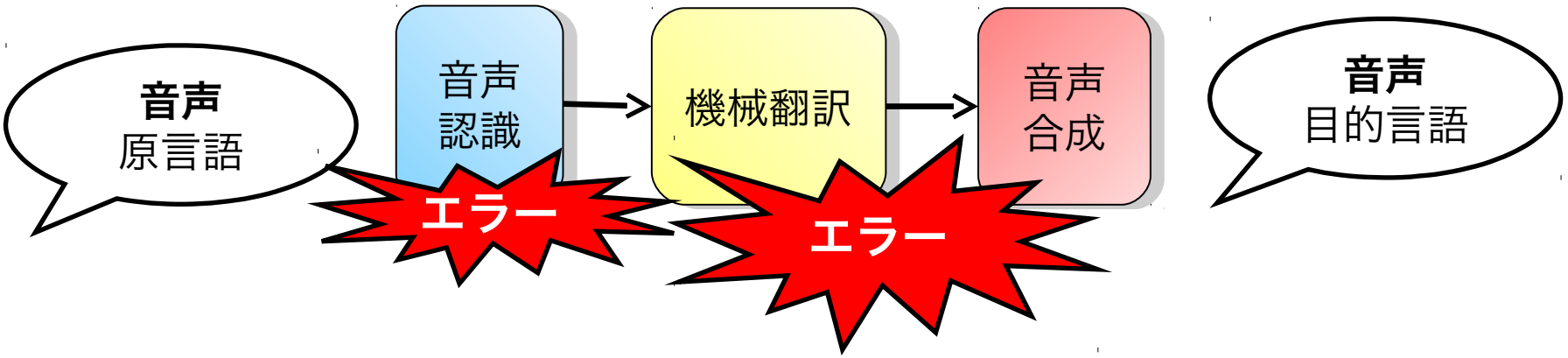
例２：

- 人間は翻訳された強調が認識可！

音声認識誤りに頑健な音声翻訳

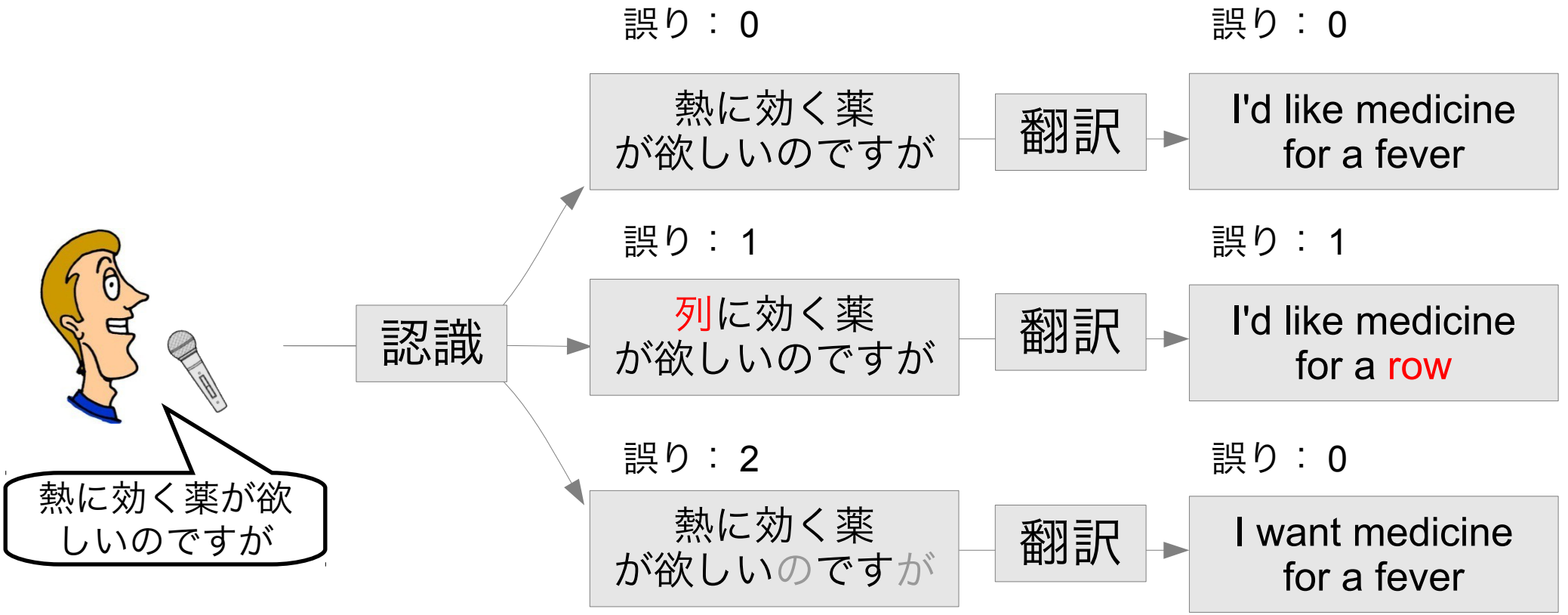
音声翻訳独特の問題点





音声認識の誤りが機械翻訳に影響を及ぼす 33

音声認識誤りと翻訳



- 認識誤りへの対応
  - 複数の候補の中からどの認識候補を利用するかを選択
  - 誤りの中でもマシなものになるように学習

認識・翻訳のパラメータ最適化

- 各モデルのスコアを組み合わせた解のスコア

	LM	TM	RM	
○ Taro visited Hanako	-4	-3	-1	-8
X the Taro visited the Hanako	-5	-4	-1	-10
X Hanako visited Taro	-2	-3	-2	-7 ← 最大 X

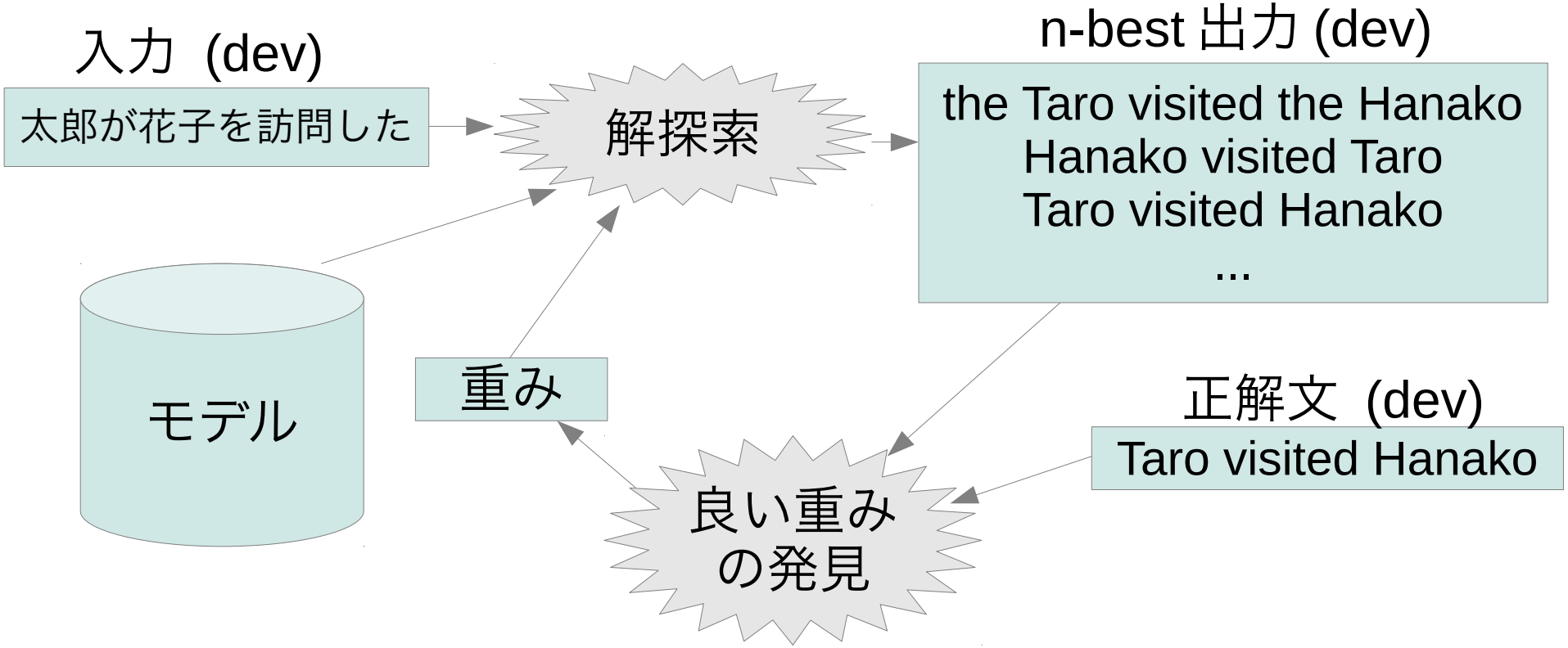
- スコアを重み付けると良い結果が得られる

	LM	TM	RM	
○ Taro visited Hanako	0.2*-4	0.3*-3	0.5*-1	-2.2 ← 最大 ○
X the Taro visited the Hanako	0.2*-5	0.3*-4	0.5*-1	-2.7
X Hanako visited Taro	0.2*-2	0.3*-3	0.5*-2	-2.3

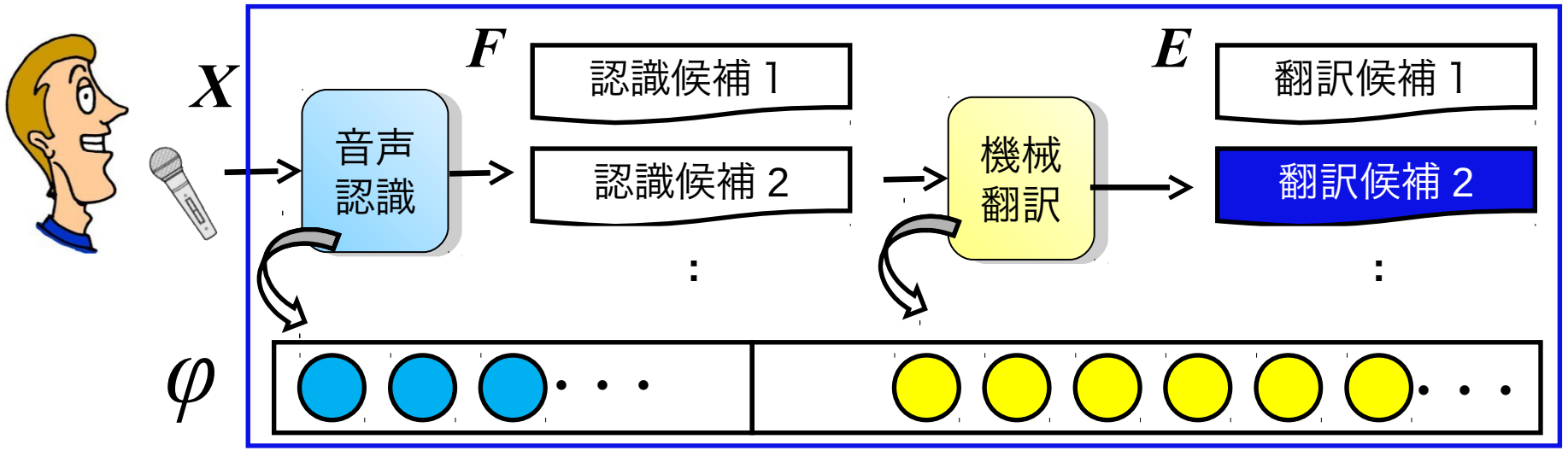
- チューニングは重みを発見 : $w_{LM}=0.2$ $w_{TM}=0.3$ $w_{RM}=0.5$

誤り率最小化 (MERT)

- 翻訳精度を重み調整の反復で最適化



複数の候補を考慮した認識・翻訳



$$\hat{E} = \arg\max_{E}(P(E, F \mid X, \lambda_{ASR}, \lambda_{MT}))$$

$\hat{E}$:選択される翻訳候補 F :認識仮説 E :翻訳仮説

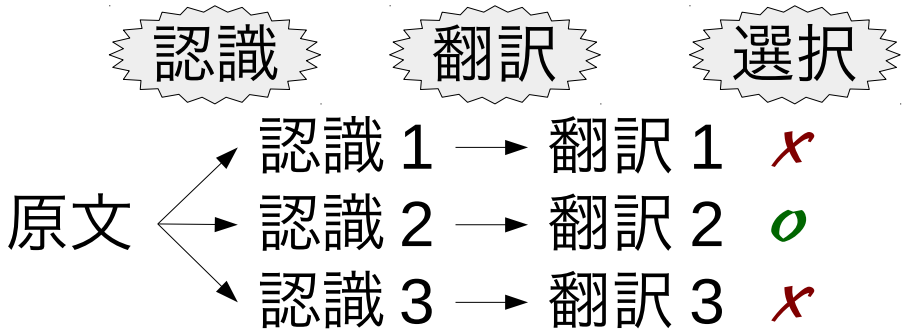
$P(E, F \mid X, \lambda_{MT,ASR})$:認識、翻訳の同時確率

$\lambda_{ASR,MT}$:

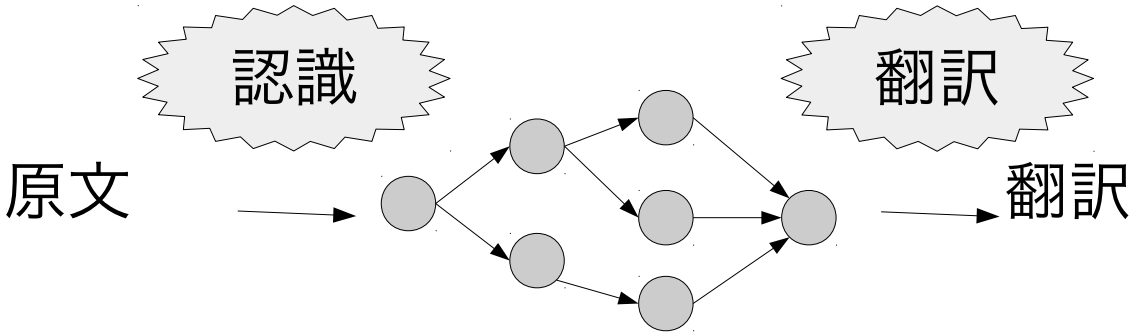
37

候補の表現法

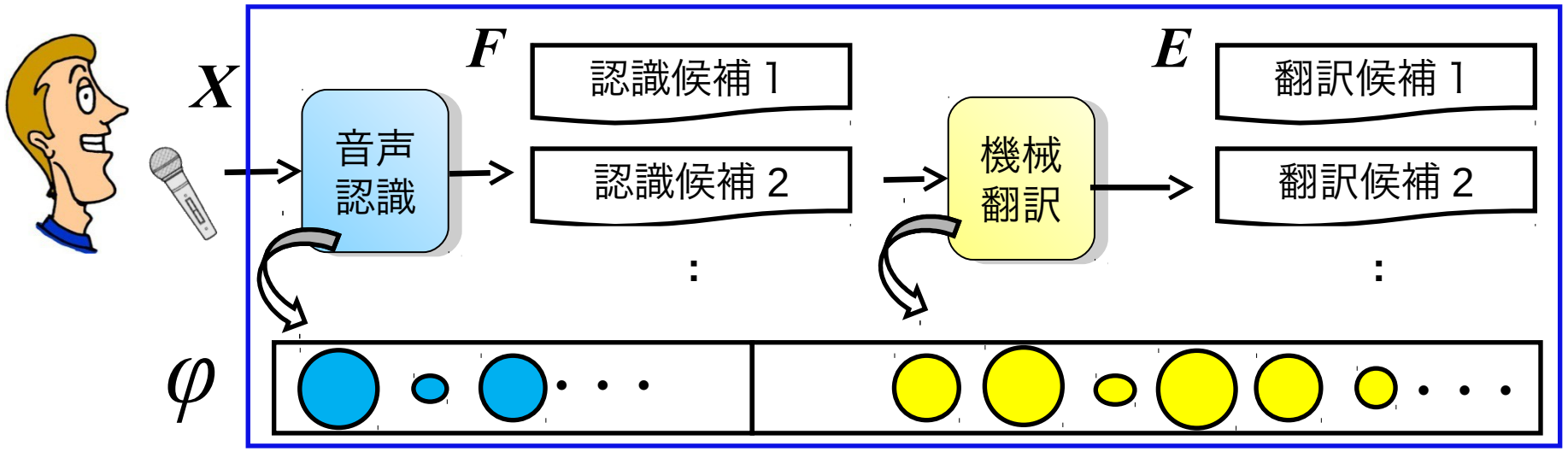
n-best リスト



ラティス



同時最適化 [Zhang+04]



$$\boldsymbol{\lambda}_{MT}, \boldsymbol{\lambda}_{ASR} = \arg\max_{\boldsymbol{\lambda}_{MT}, \boldsymbol{\lambda}_{ASR}} \left(BLEU(\boldsymbol{E}_{ref}, \hat{\boldsymbol{E}}) \right)$$

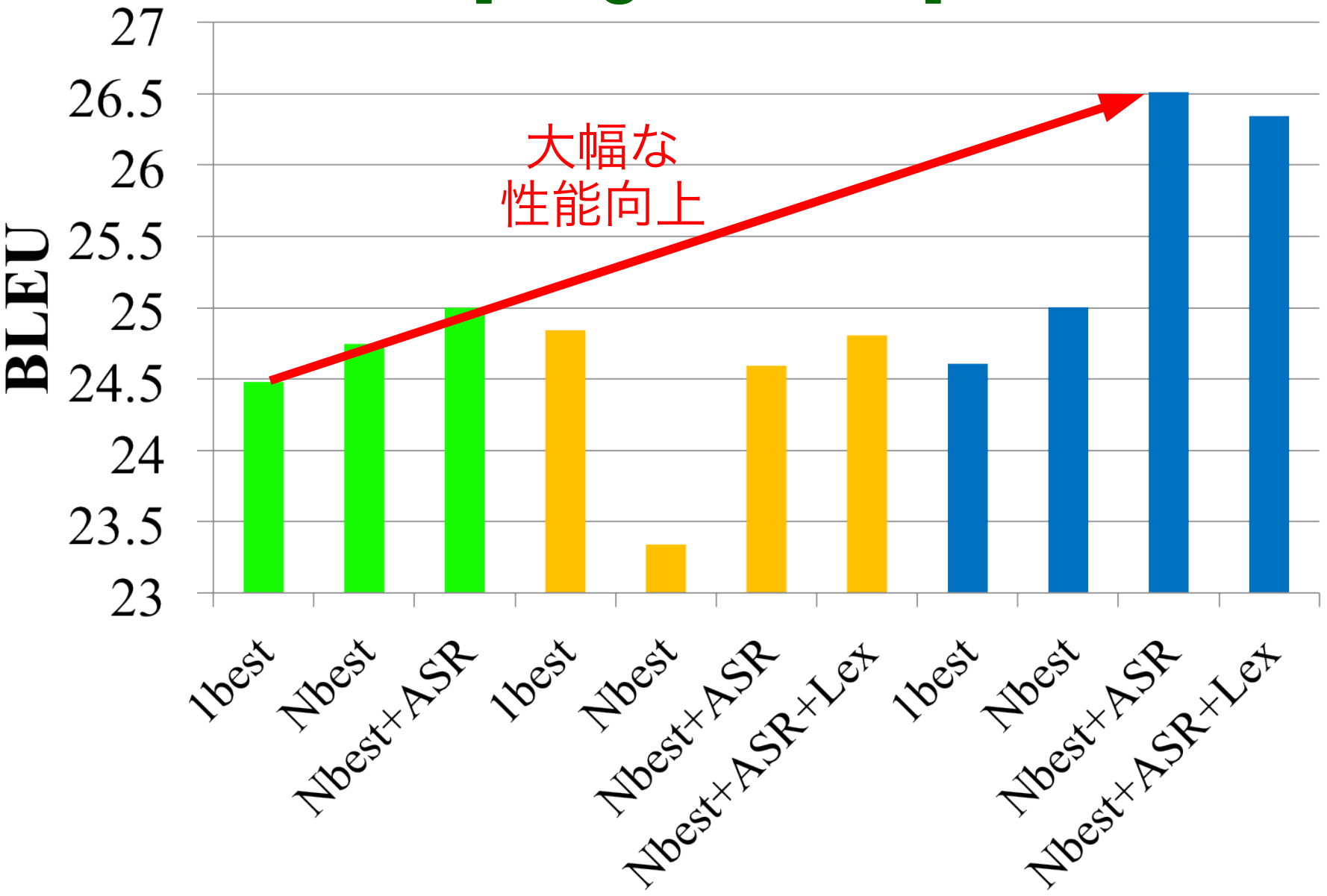
$\boldsymbol{E}_{ref}$:英語参照文　　$\hat{\boldsymbol{E}}$: 翻訳文

$BLEU$: 翻訳の評価尺度 [Papineni et al.,2002.]

$\boldsymbol{\lambda}_{ASR}, \boldsymbol{\lambda}_{MT}$:最適化した重みベクトル

39

実験結果
[Ohgushi+13]



まとめ

まとめ

- より早く、より豊かに音声翻訳結果を聞き手に届ける取り組み
- 今後の課題：
  - 同時音声翻訳における未来の情報の予測 [Grissom+14,Oda+15]
  - より豊かな音響情報の翻訳
  - end-to-end システムの開発

参考文献

参考文献

- S. Bangalore, V. K. R. Sridhar, P. K. L. Golipour, and A. Jimenez. Real-time incremental speech-to-speech translation of dialogs. In Proc. NAACL, 2012.
- C. Fugen, A. Waibel, and M. Kolss. Simultaneous translation of lectures and speeches. Machine Translation, 21(4):209–252, 2007.
- T. Fujita, G. Neubig, S. Sakti, T. Toda, and S. Nakamura. Simple, lexicalized choice of translation timing for simultaneous speech translation. In Proc. InterSpeech, pages 3487–3491, 2013.
- A. Grissom II, H. He, J. Boyd-Graber, J. Morgan, and H. Daume III. Don't until the final verb wait: Reinforcement learning for simultaneous machine translation. In Proc. EMNLP, pages 1342–1352, 2014.
- R. Jones. Conference interpreting explained, volume 6. 2002.
- T. Kano, S. Sakti, S. Takamichi, G. Neubig, T. Toda, and S. Nakamura. A method for translation of paralinguistic information. In Proc. IWSLT, pages 158–163, 2012.
- T. Kano, S. Takamichi, S. Sakti, G. Neubig, T. Toda, and S. Nakamura. Generalizing continuous-space translation of paralinguistic information. In Proc. InterSpeech, 2013.
- H. Ney. Speech translation: Coupling of recognition and translation. In Proc. ICASSP, pages 517–520, 1999.
- Y. Oda, G. Neubig, S. Sakti, T. Toda, and S. Nakamura. Optimizing segmentation strategies for simultaneous speech translation. In Proc. ACL, pages 551–556, 2014.
- Y. Oda, G. Neubig, S. Sakti, T. Toda, and S. Nakamura. Syntax-based simultaneous translation through prediction of unseen syntactic constituents. In Proc. ACL, 2015.
- M. Ohgushi, G. Neubig, S. Sakti, T. Toda, and S. Nakamura. An empirical comparison of joint optimization techniques for speech translation. In Proc. InterSpeech, pages 2619–2622, 2013.
- M. Paulik and A. Waibel. Extracting clues from human interpreter speech for spoken language translation. In Proc. ICASSP, pages 5097– 5100. IEEE, 2008.
- S. Peitz, M. Freitag, A. Mauser, and H. Ney. Modeling punctuation prediction as machine translation. In Proc. IWSLT, pages 238–245, 2011.
- V. K. Rangarajan Sridhar, J. Chen, S. Bangalore, A. Ljolje, and R. Chengalvarayan. Segmentation strategies for streaming speech translation. In Proc. NAACL, pages 230–238, 2013.
- H. Shimizu, G. Neubig, S. Sakti, T. Toda, and S. Nakamura. Constructing a speech translation system using simultaneous interpretation data. In Proc. IWSLT, pages 212–218, 2013.
- H. Shimizu, G. Neubig, S. Sakti, T. Toda, and S. Nakamura. Collection of a simultaneous translation corpus for comparative analysis. In Proc. LREC, 2014.
- W. Wang, G. Tur, J. Zheng, and N. F. Ayan. Automatic disfluency removal for improving spoken language translation. In Proc. ICASSP, pages 5214–5217, 2010.

追加資料

同時音声翻訳の評価 [Mieno+ 15]

- 人手評価で、様々な遅延と翻訳精度を持った文を評価者に表示し、評価関数を計算