Example-based Decoding for Statistical Machine Translation

Taro Watanabe† and Eiichiro Sumita

taro.watanabe@atr.co.jp

ATR Spoken Language Translation Research Laboratories

Overview

- Statistical Machine Translation (SMT)
- Decoding Problem in SMT
- Example-based Decoder
- Experiments
- Summary

Statistical Machine Translation

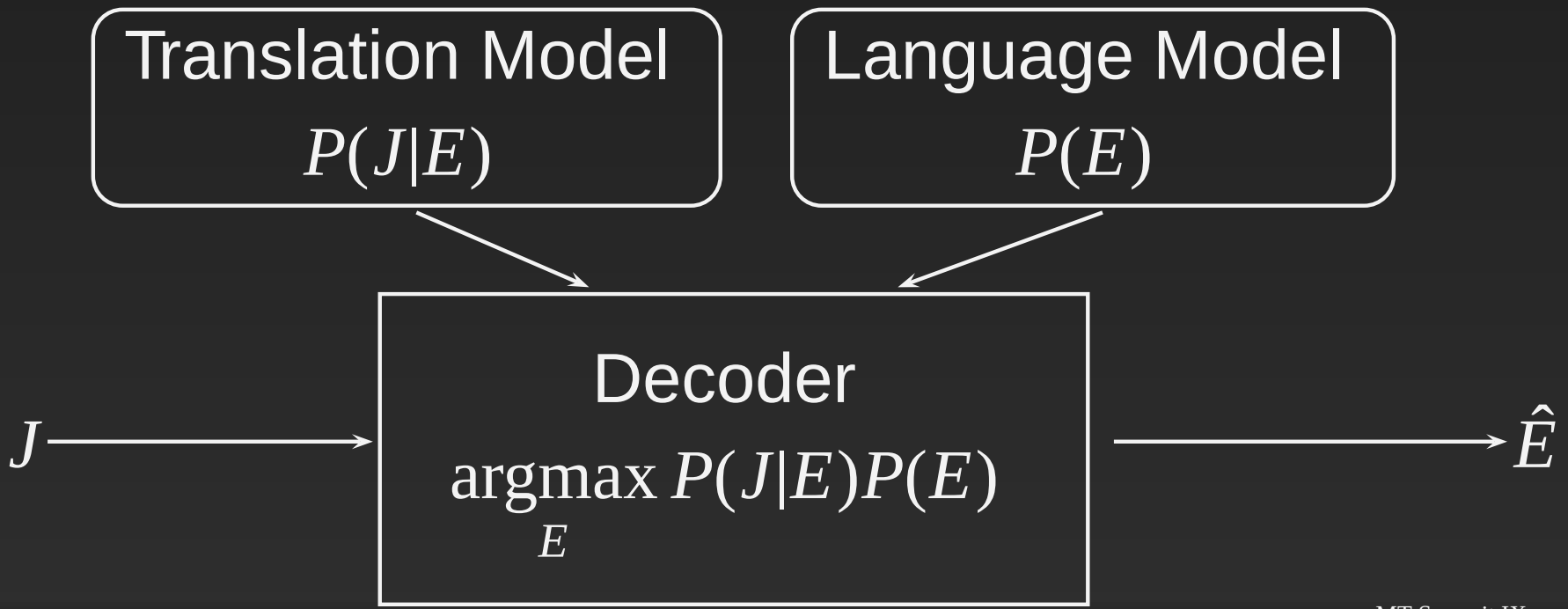
- Translation from J into E (Berger et al. 1993)

$$\begin{aligned}\hat{E} &= \operatorname*{argmax}_E P(E|J) \\ &= \operatorname*{argmax}_E P(E)P(J|E)\end{aligned}$$

Statistical Machine Translation

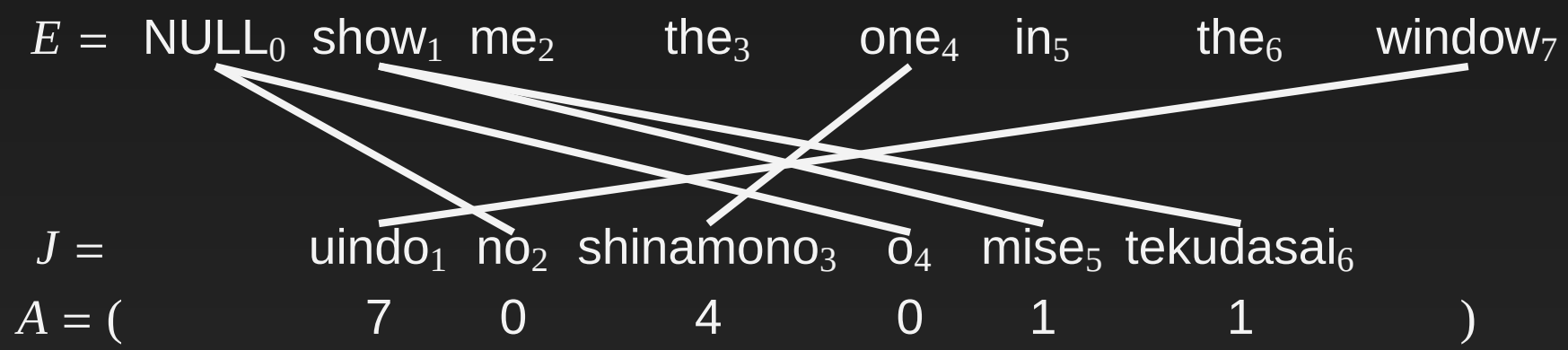
- Translation from J into E (Berger et al. 1993)

$$\begin{aligned}\hat{E} &= \operatorname*{argmax}_E P(E|J) \\ &= \operatorname*{argmax}_E P(E)P(J|E)\end{aligned}$$



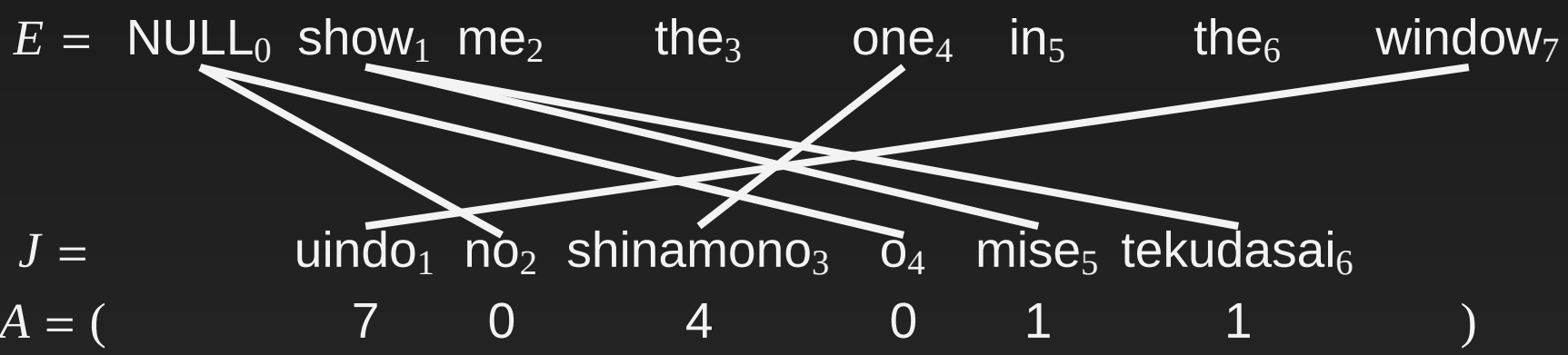
Word Alignment Based Statistical Translation

$$P(J|E) = \sum_{A} P(J, A|E)$$



Word Alignment Based Statistical Translation

$$P(J|E) = \sum_{A} P(J, A|E)$$



- Generative Process of $P(J, A|E)$
  IBM Model 1 — 5 (Berger, et al. 1993), HMM (Vogel, et al. 1996) etc.

Word-by-Word (or Phrase-by-Phrase) Decoding

- Prefix of partial translation with score by TM and LM
- Breadth first search to extend the prefix

Word-by-Word (or Phrase-by-Phrase) Decoding

- Prefix of partial translation with score by TM and LM
- Breadth first search to extend the prefix

input: 精算書 を 確認 さ せ てください

Word-by-Word (or Phrase-by-Phrase) Decoding

- Prefix of partial translation with score by TM and LM
- Breadth first search to extend the prefix

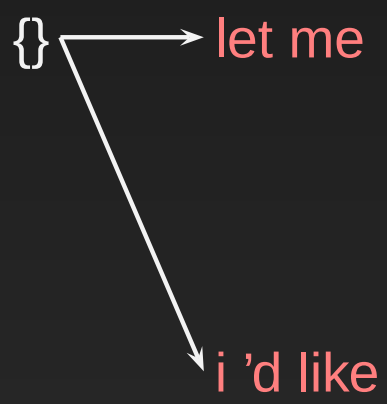
input: 精算書 を 確認 さ せ てください

{}

Word-by-Word (or Phrase-by-Phrase) Decoding

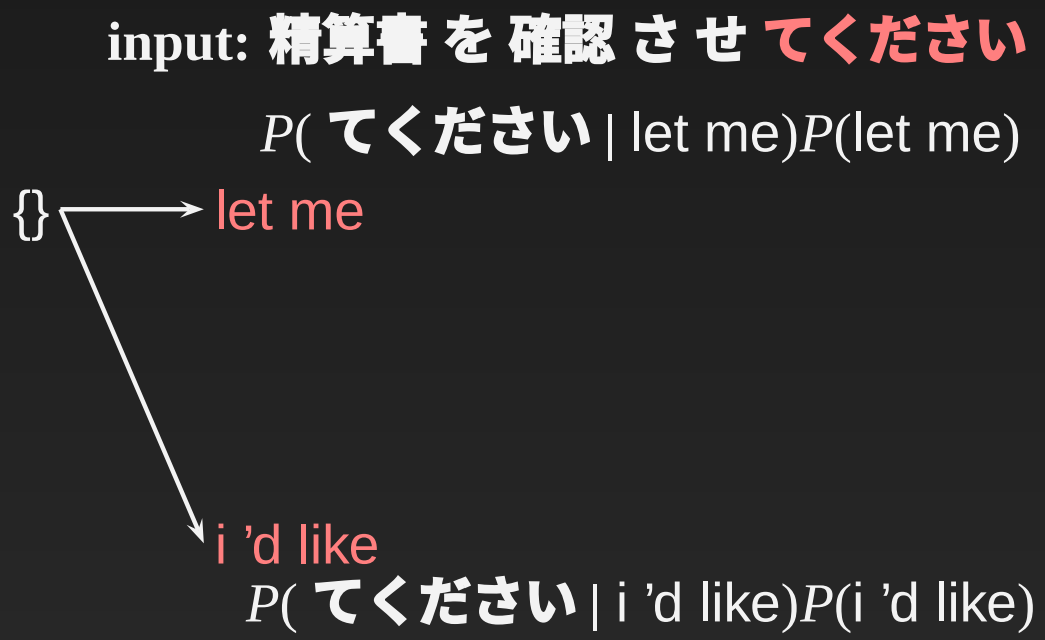
- Prefix of partial translation with score by TM and LM
- Breadth first search to extend the prefix

input: 精算書 を 確認 さ せ てください



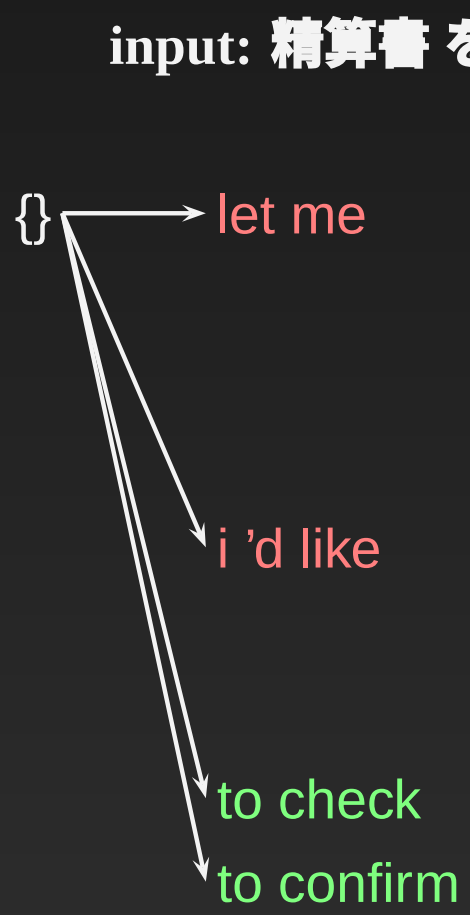
Word-by-Word (or Phrase-by-Phrase) Decoding

- Prefix of partial translation with score by TM and LM
- Breadth first search to extend the prefix



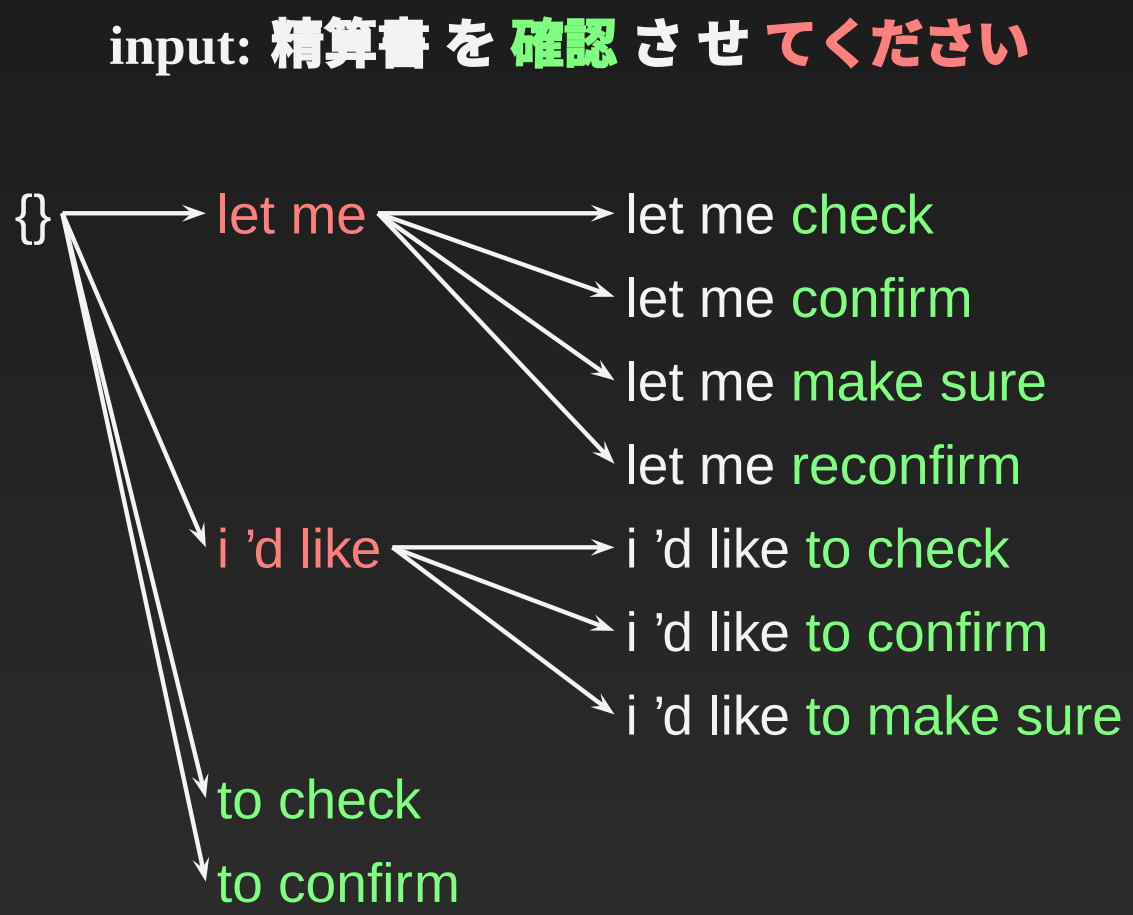
Word-by-Word (or Phrase-by-Phrase) Decoding

- Prefix of partial translation with score by TM and LM
- Breadth first search to extend the prefix



Word-by-Word (or Phrase-by-Phrase) Decoding

- Prefix of partial translation with score by TM and LM
- Breadth first search to extend the prefix



Decoding Problem in SMT

- Application to similar language pairs
  - Similar pairs : French–English, German–English, etc.
  - Distant pairs? : Japanese–English, Chinese-English, etc.

Decoding Problem in SMT

- Application to similar language pairs
  - Similar pairs : French–English, German–English, etc.
  - Distant pairs? : Japanese–English, Chinese-English, etc.
- Frequent insertion/deletion and intricated word alignments

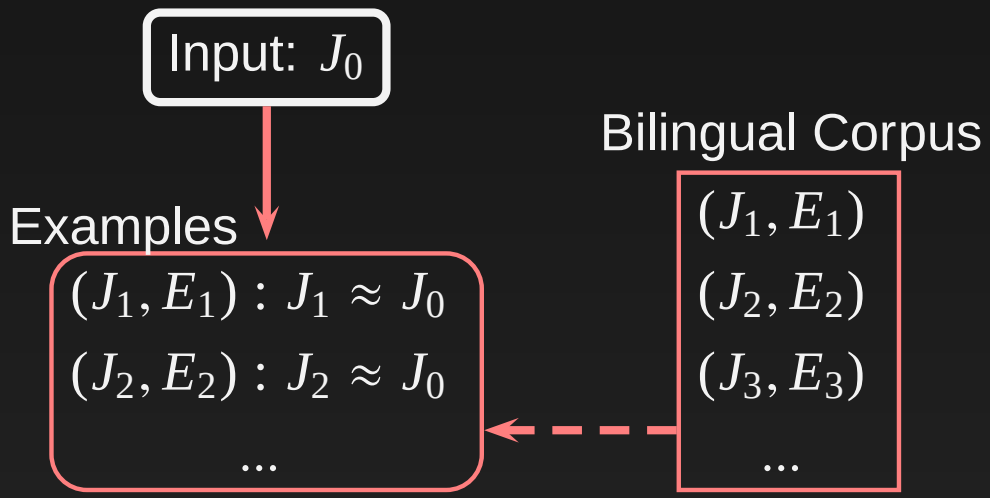
Decoding Problem in SMT

- Application to similar language pairs
  - Similar pairs : French–English, German–English, etc.
  - Distant pairs? : Japanese–English, Chinese-English, etc.
- Frequent insertion/deletion and intricated word alignments
- Word-by-word or phrase-by-phrase decoding
  - Frequent insertion/deletion — huge search space
  - Pruning is inevitable — search error

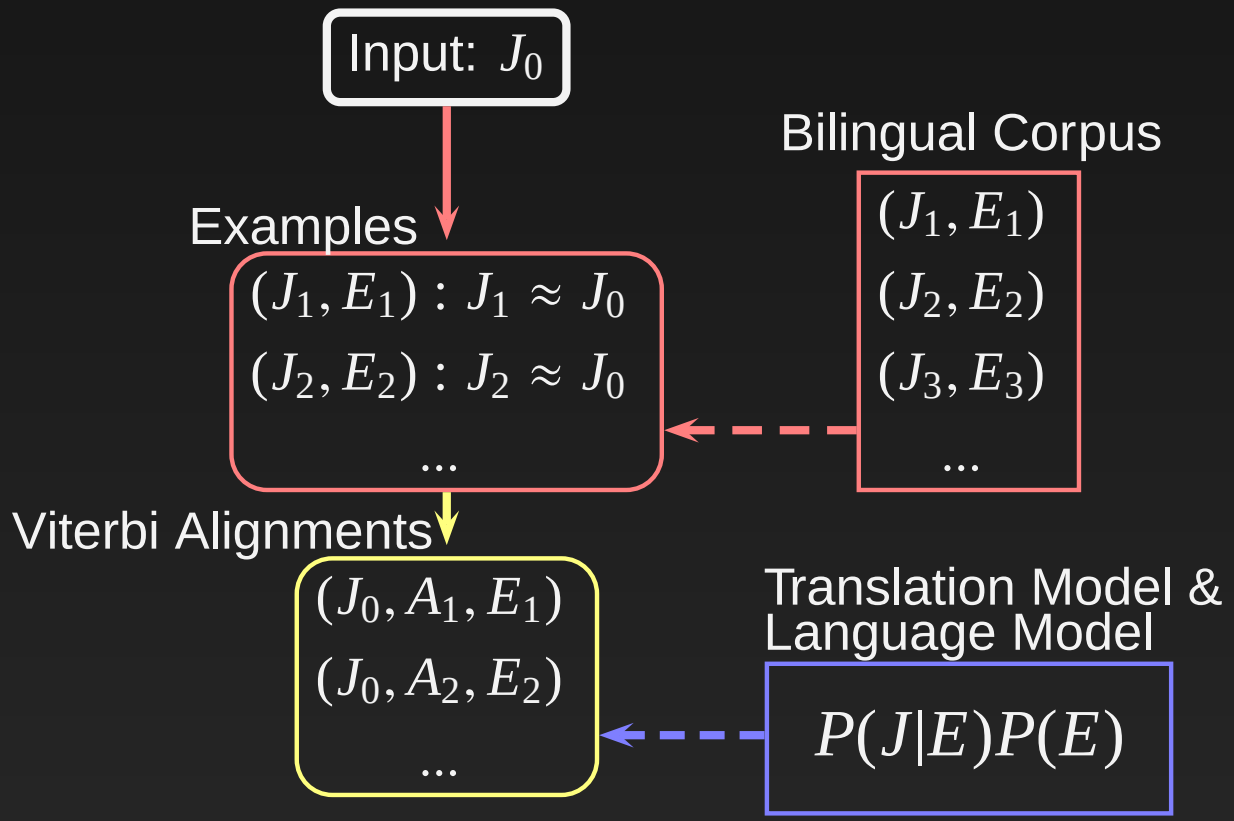
Example-based Decoder

Input: J_0

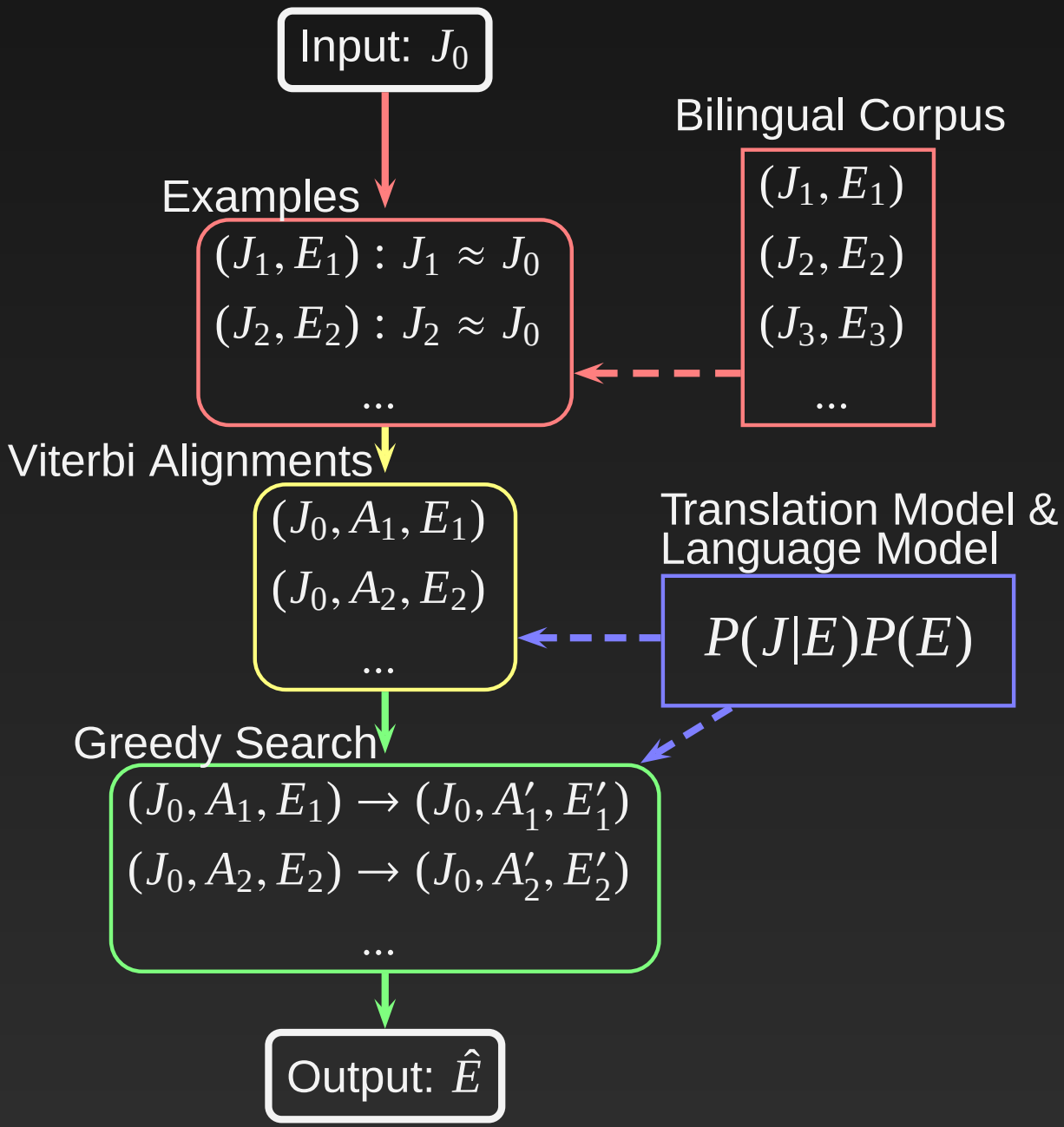
Example-based Decoder



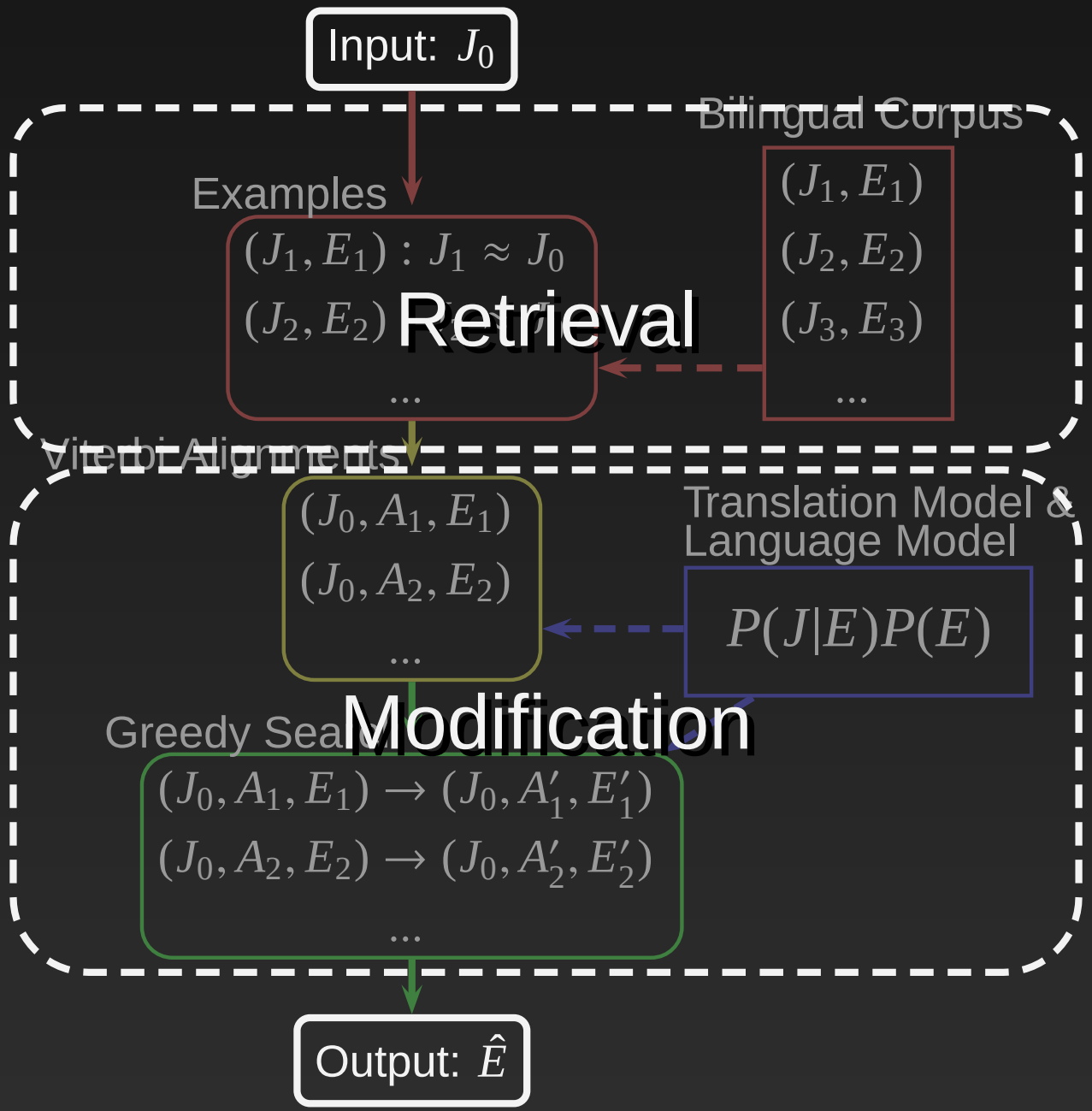
Example-based Decoder



Example-based Decoder



Example-based Decoder



Retrieval of Examples

- Similarity measure
  - Edit distance criteria

$$dis(J_k, J_0) = I(J_k, J_0) + D(J_k, J_0) + S(J_k, J_0)$$

$I(J_k, J_0), D(J_k, J_0), S(J_k, J_0)$ = # of ins/del/sub

Retrieval of Examples

- Similarity measure
  - Edit distance criteria

$$dis(J_k, J_0) = I(J_k, J_0) + D(J_k, J_0) + S(J_k, J_0)$$

$I(J_k, J_0), D(J_k, J_0), S(J_k, J_0)$ = # of ins/del/sub

  - *tf/idf* criteria: one translation pair = one document

$$P_{tf/idf}(J_k, J_0) = \sum_{i:J_{0,i} \in J_k} \frac{\log(N/df(J_{0,i}))/\log N}{|J_0|}$$

Retrieval of Examples

- Similarity measure
  - Edit distance criteria

$$dis(J_k, J_0) = I(J_k, J_0) + D(J_k, J_0) + S(J_k, J_0)$$

$I(J_k, J_0), D(J_k, J_0), S(J_k, J_0)$ = # of ins/del/sub

  - *tf/idf* criteria: one translation pair = one document

$$P_{tf/idf}(J_k, J_0) = \sum_{i:J_{0,i} \in J_k} \frac{\log(N/df(J_{0,i}))/\log N}{|J_0|}$$

$$score(J_k, J_0) = \begin{cases} (1.0 - \alpha)(1.0 - \frac{dis(J_k, J_0)}{|J_0|}) & dis(J_k, J_0) > 0 \\ \quad + \alpha P_{tf/idf}(J_k, J_0) & \\ 1.0 & \text{otherwise} \end{cases}$$

Modification of Examples

- For each translation example (J_k, E_k),
  1. Compute the viterbi alignment A_k for the pair (J_0, E_k)
  2. Perform greedy decoding algorithm thorough hill-climbing for (J_0, A_k, E_k) to obtain (J_0, A'_k, E'_k) by modifying A_k and E_k.

Modification of Examples

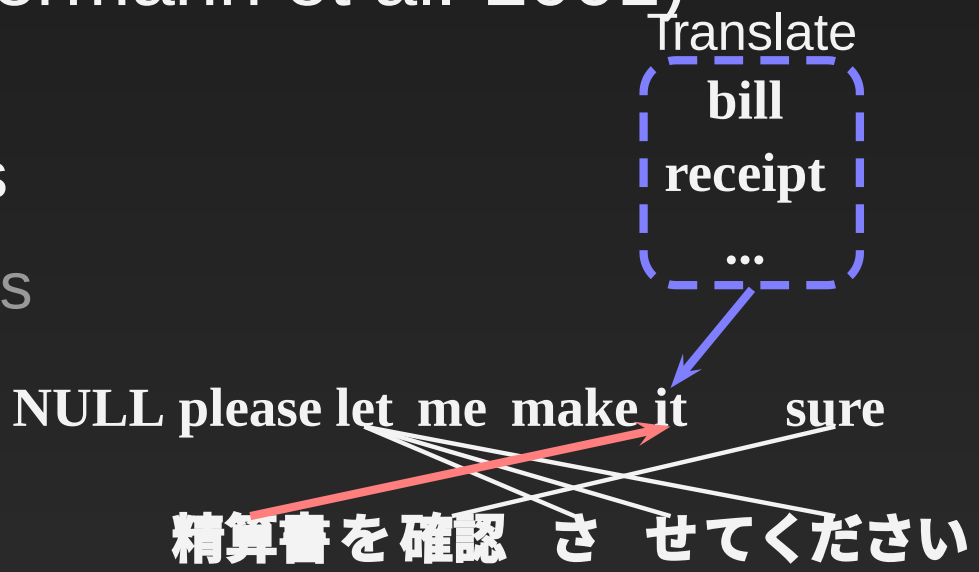
- For each translation example (J_k, E_k),
  1. Compute the viterbi alignment A_k for the pair (J_0, E_k)
  2. Perform greedy decoding algorithm thorough hill-climbing for (J_0, A_k, E_k) to obtain (J_0, A'_k, E'_k) by modifying A_k and E_k.
- Hill-climbing operators (Germann et al. 2001)
  - Translate words
  - Translate and align words
  - Translate and insert words
  - Move alignments
  - Swap segments
  - Remove words
  - Join words

Modification of Examples

- For each translation example (J_k, E_k),
  1. Compute the viterbi alignment A_k for the pair (J_0, E_k)
  2. Perform greedy decoding algorithm thorough hill-climbing for (J_0, A_k, E_k) to obtain (J_0, A'_k, E'_k) by modifying A_k and E_k.
- Hill-climbing operators (Germann et al. 2001)
  - Translate words
  - Translate and align words
  - Translate and insert words
  - Move alignments
  - Swap segments
  - Remove words
  - Join words

Modification of Examples

- For each translation example (J_k, E_k),
  1. Compute the viterbi alignment A_k for the pair (J_0, E_k)
  2. Perform greedy decoding algorithm thorough hill-climbing for (J_0, A_k, E_k) to obtain (J_0, A'_k, E'_k) by modifying A_k and E_k.
- Hill-climbing operators (Germann et al. 2001)
  - Translate words
  - Translate and align words
  - Translate and insert words
  - Move alignments
  - Swap segments
  - Remove words
  - Join words

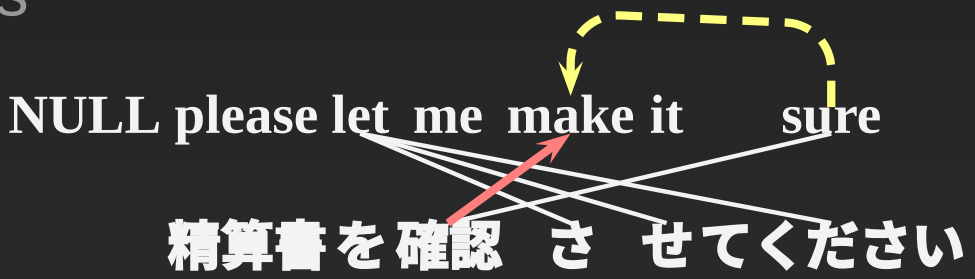


Modification of Examples

- For each translation example (J_k, E_k),
  1. Compute the viterbi alignment A_k for the pair (J_0, E_k)
  2. Perform greedy decoding algorithm thorough hill-climbing for (J_0, A_k, E_k) to obtain (J_0, A'_k, E'_k) by modifying A_k and E_k.
- Hill-climbing operators (Germann et al. 2001)
  - Translate words
  - Translate and align words
  - Translate and insert words
  - Move alignments
  - Swap segments
  - Remove words
  - Join words

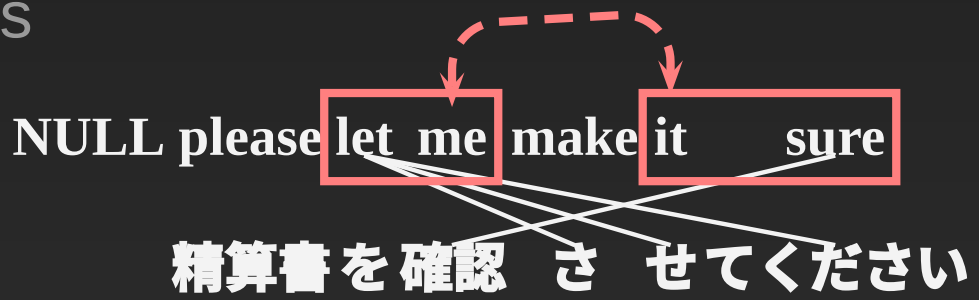
Modification of Examples

- For each translation example (J_k, E_k),
  1. Compute the viterbi alignment A_k for the pair (J_0, E_k)
  2. Perform greedy decoding algorithm thorough hill-climbing for (J_0, A_k, E_k) to obtain (J_0, A'_k, E'_k) by modifying A_k and E_k.
- Hill-climbing operators (Germann et al. 2001)
  - Translate words
  - Translate and align words
  - Translate and insert words
  - Move alignments
  - Swap segments
  - Remove words
  - Join words



Modification of Examples

- For each translation example (J_k, E_k),
  1. Compute the viterbi alignment A_k for the pair (J_0, E_k)
  2. Perform greedy decoding algorithm thorough hill-climbing for (J_0, A_k, E_k) to obtain (J_0, A'_k, E'_k) by modifying A_k and E_k.
- Hill-climbing operators (Germann et al. 2001)
  - Translate words
  - Translate and align words
  - Translate and insert words
  - Move alignments
  - Swap segments
  - Remove words
  - Join words



Modification of Examples

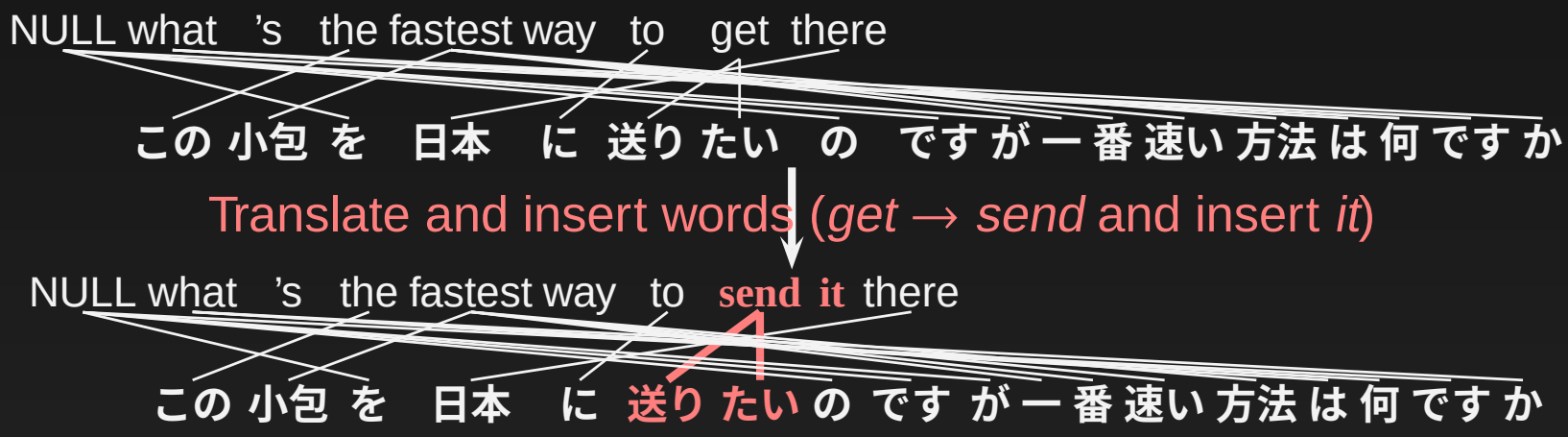
- For each translation example (J_k, E_k),
  1. Compute the viterbi alignment A_k for the pair (J_0, E_k)
  2. Perform greedy decoding algorithm thorough hill-climbing for (J_0, A_k, E_k) to obtain (J_0, A'_k, E'_k) by modifying A_k and E_k.
- Hill-climbing operators (Germann et al. 2001)
  - Translate words
  - Translate and align words
  - Translate and insert words
  - Move alignments
  - Swap segments
  - Remove words
  - Join words

Modification of Examples

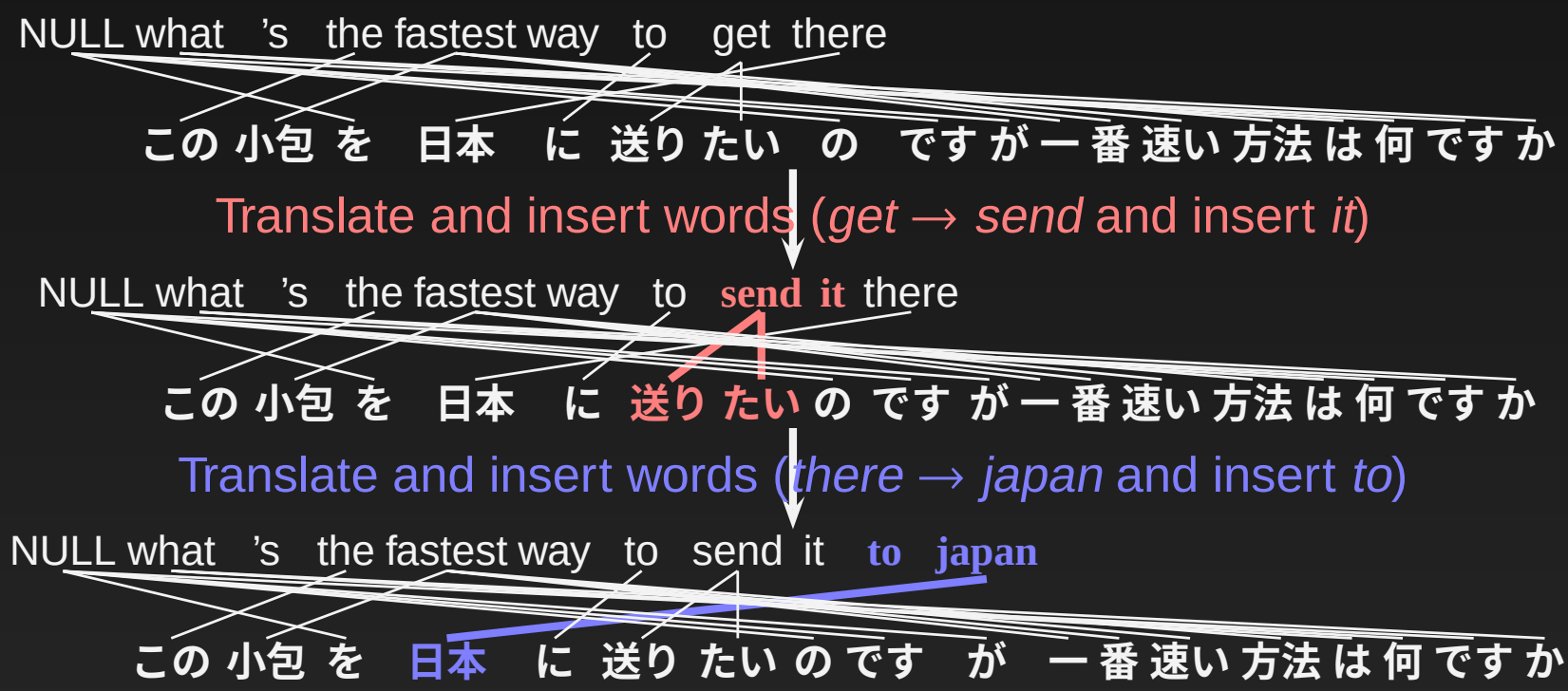
- For each translation example (J_k, E_k),
  1. Compute the viterbi alignment A_k for the pair (J_0, E_k)
  2. Perform greedy decoding algorithm thorough hill-climbing for (J_0, A_k, E_k) to obtain (J_0, A'_k, E'_k) by modifying A_k and E_k.
- Hill-climbing operators (Germann et al. 2001)
  - Translate words
  - Translate and align words
  - Translate and insert words
  - Move alignments
  - Swap segments
  - Remove words
  - Join words

Example of Greedy Process

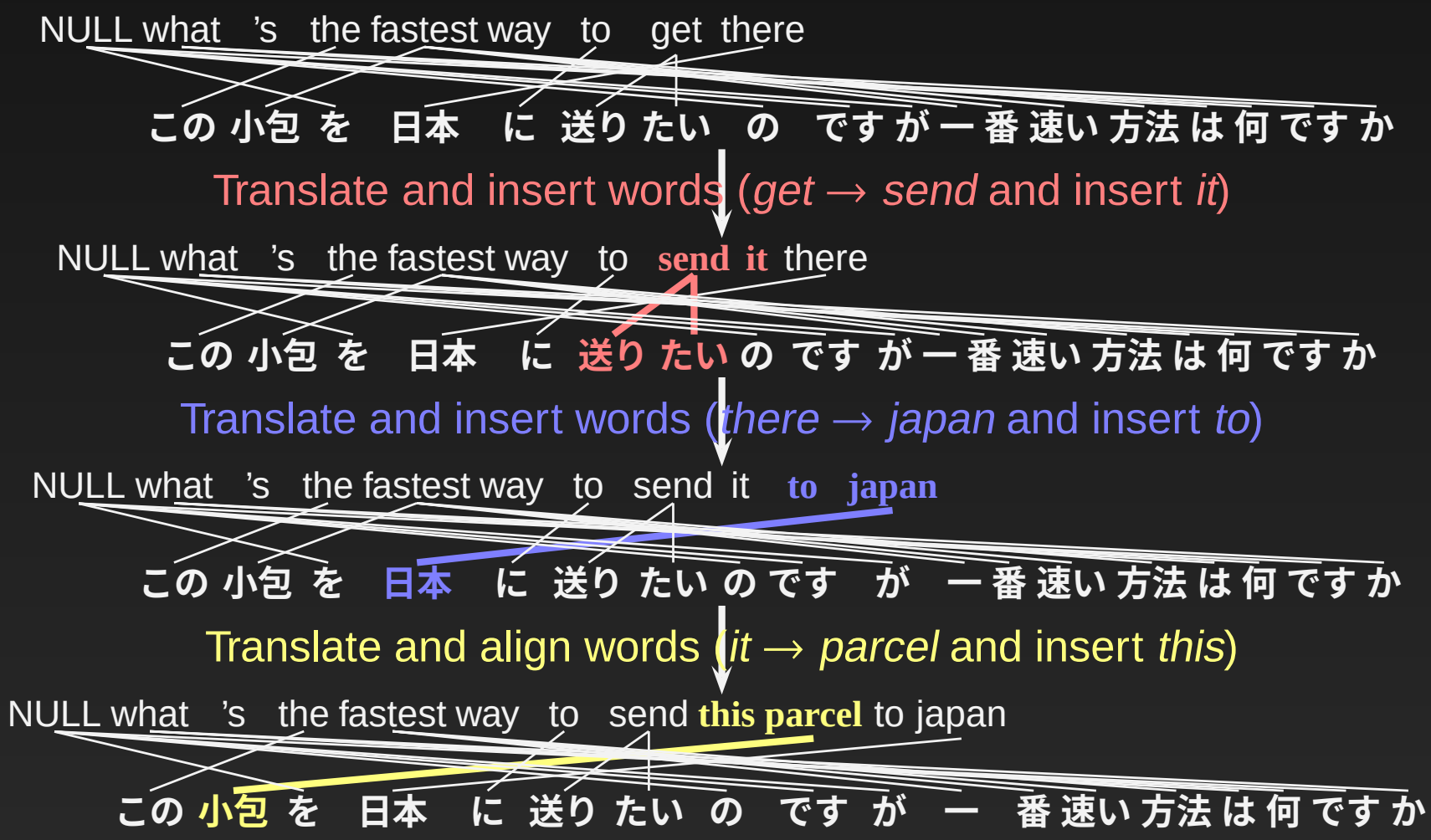
Example of Greedy Process



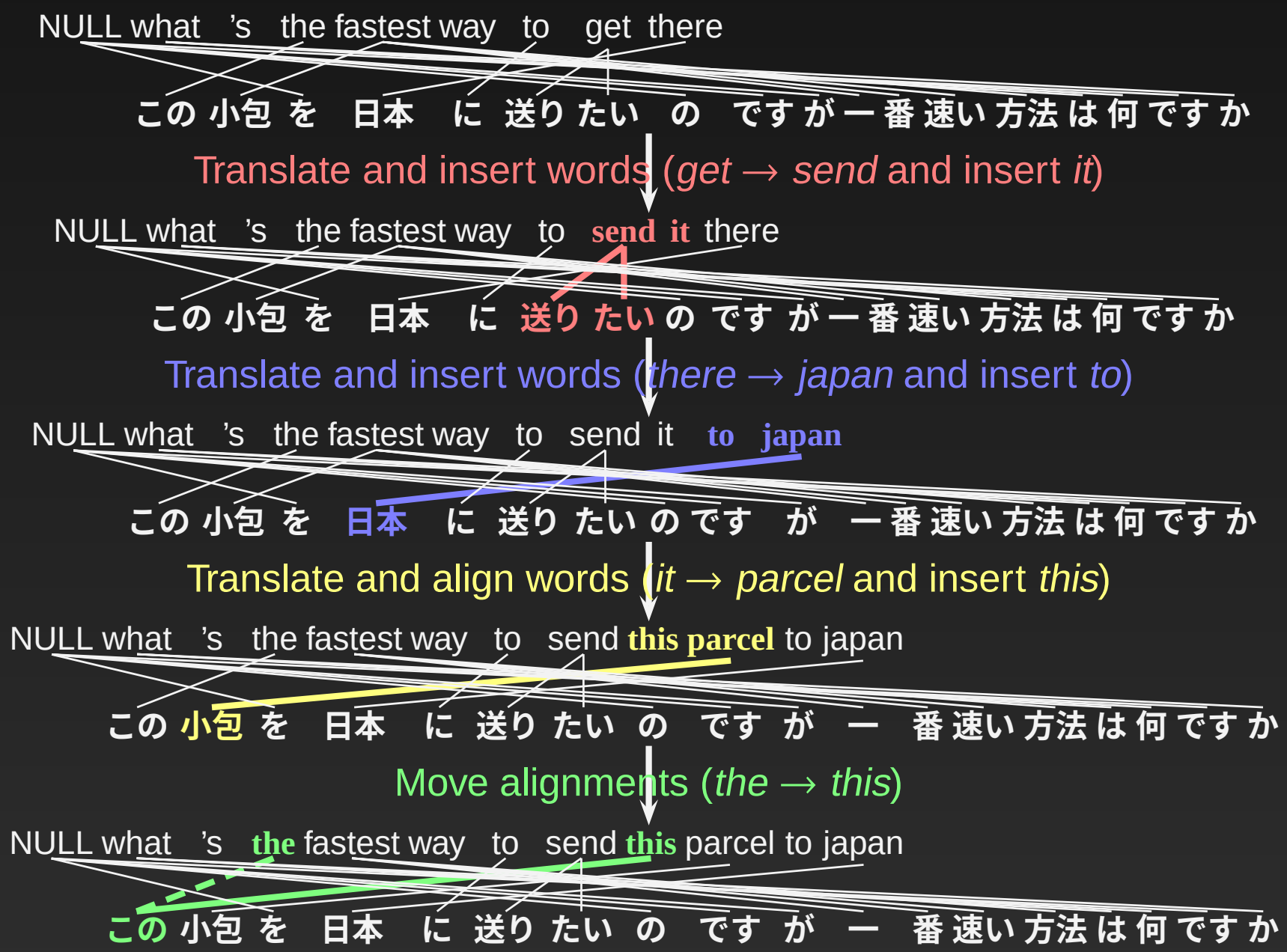
Example of Greedy Process



Example of Greedy Process



Example of Greedy Process



Experiments

Basic Travel Expression Corpus

	Chinese	English	Japanese	Korean
# of sentences		167,163		
# of words	956,732	980,790	1,148,428	1,269,888
vocabulary size	16,411	15,641	21,896	13,395
# of singletons	5,207	5,547	9,220	4,191
3-gram perplexity	45.53	35.35	24.06	20.34

Evaluation

- Two decoders on IBM Model 4 + 3-gram:
  - Left-to-right generation word-by-word beam search decoder
  - Example-based decoder
- Evaluation Metrics

WER: Word-error-rate — edit distance penalty

PER: Position independent WER — WER but ignore positional disfluencies

BLEU: Geometric mean of n-gram precision

SE: Subjective evaluation — ranks ranging from A : perfect, B : fair, C : acceptable and D : non-sense judged by a native speaker

SER: Search error rate — judged by TM+LM scores of outputs from two systems

Remarks: 16 referenses for WER, PER and BLEU

Results

	Exact [%]	WER [%]	PER [%]	BLEU [%]	SE [%] A	SE [%] A+B	SE [%] A+B+C
C-E		45.0/**34.3**	39.8/**30.3**	43.6/**56.7**	48.4/**65.3**	65.9/**76.9**	71.4/**81.0**
C-J	52.7	35.7/**25.5**	31.3/**22.6**	56.9/**67.8**	50.8/**69.0**	59.4/**74.3**	66.9/**80.2**
C-K		38.4/**29.1**	34.2/**26.2**	56.1/**65.0**	-	-	-
E-C		45.0/**38.0**	39.7/**33.4**	42.1/**51.9**	-	-	-
E-J	40.8	34.2/**29.0**	30.5/**26.1**	59.2/**65.7**	55.8/**65.1**	62.4/**71.6**	70.2/**77.8**
E-K		38.7/**35.6**	34.3/**31.6**	57.3/**61.5**	-	-	-
J-C		46.8/**33.0**	38.9/**27.8**	39.7/**57.1**	-	-	-
J-E	33.7	42.9/**35.0**	37.4/**30.3**	47.6/**57.4**	50.8/**63.7**	65.7/**74.5**	70.2/**77.6**
J-K		27.7/**20.8**	25.4/**19.2**	67.2/**73.5**	-	-	-
K-C		41.9/**32.9**	34.4/**27.6**	45.1/**55.5**	-	-	-
K-E	39.2	45.1/**36.4**	38.5/**32.1**	44.3/**56.8**	49.2/**61.6**	65.7/**72.9**	72.2/**78.4**
K-J		26.8/**20.8**	24.6/**19.3**	64.3/**70.8**	56.5/**69.2**	66.5/**77.5**	78.4/**84.7**

Results

	Exact [%]	WER [%]	PER [%]	BLEU [%]	SE [%] A	SE [%] A+B	SE [%] A+B+C
C-E		45.0/**34.3**	39.8/**30.3**	43.6/**56.7**	48.4/**65.3**	65.9/**76.9**	71.4/**81.0**
C-J	52.7	35.7/**25.5**	31.3/**22.6**	56.9/**67.8**	50.8/**69.0**	59.4/**74.3**	66.9/**80.2**
C-K		38.4/**29.1**	34.2/**26.2**	56.1/**65.0**	-	-	-
E-C		45.0/**38.0**	39.7/**33.4**	42.1/**51.9**	-	-	-
E-J	40.8	34.2/**29.0**	30.5/**26.1**	59.2/**65.7**	55.8/**65.1**	62.4/**71.6**	70.2/**77.8**
E-K		38.7/**35.6**	34.3/**31.6**	57.3/**61.5**	-	-	-
J-C		46.8/**33.0**	38.9/**27.8**	39.7/**57.1**	-	-	-
J-E	33.7	42.9/**35.0**	37.4/**30.3**	47.6/**57.4**	50.8/**63.7**	65.7/**74.5**	70.2/**77.6**
J-K		27.7/**20.8**	25.4/**19.2**	67.2/**73.5**	-	-	-
K-C		41.9/**32.9**	34.4/**27.6**	45.1/**55.5**	-	-	-
K-E	39.2	45.1/**36.4**	38.5/**32.1**	44.3/**56.8**	49.2/**61.6**	65.7/**72.9**	72.2/**78.4**
K-J		26.8/**20.8**	24.6/**19.3**	64.3/**70.8**	56.5/**69.2**	66.5/**77.5**	78.4/**84.7**

beam search/example-based decoder

Detailed Results

	SE [%]					
	exactly matched			otherwise		
	A	A+B	A+B+C	A	A+B	A+B+C
C-E	65.4/**92.6**	78.4/**97.0**	82.5/**97.8**	29.5/**34.9**	51.9/**54.4**	58.9/**62.2**
J-E	72.1/**97.1**	80.8/**99.4**	83.7/**99.4**	39.9/**46.7**	58.0/**61.8**	63.3/**66.6**
K-E	69.0/**92.0**	81.0/**97.5**	85.0/**98.0**	36.5/**41.9**	55.8/**57.1**	63.9/**65.8**
C-J	65.4/**97.0**	73.6/**98.1**	78.4/**98.9**	34.4/**37.8**	43.6/**47.7**	53.9/**59.3**
E-J	79.3/**95.2**	81.7/**98.1**	84.6/**98.6**	39.4/**44.4**	49.0/**53.3**	60.3/**63.6**
K-J	74.5/**98.5**	80.0/**99.5**	87.5/**99.5**	44.8/**50.3**	57.7/**63.2**	72.6/**75.2**

beam search/example-based decoder

Search Error

	matched	non-matched	total
C-E	15.2/**58.0**	28.2/**38.6**	21.4/**48.8**
C-J	16.7/**46.8**	30.3/**24.9**	23.1/**24.9**
C-K	14.9/**45.0**	25.7/**29.5**	20.0/**37.6**
E-C	19.2/**46.2**	39.1/**26.8**	31.0/**34.1**
E-J	19.7/**34.1**	33.4/**26.2**	27.8/**29.4**
E-K	14.4/**42.8**	31.8/**32.8**	24.7/**36.9**
J-C	10.5/**63.4**	37.6/**31.4**	28.4/**42.2**
J-E	16.3/**51.7**	35.8/**28.1**	29.2/**36.1**
J-K	14.5/**39.5**	50.3/**10.9**	38.2/**20.6**
K-C	14.0/**51.5**	31.9/**28.7**	24.9/**37.6**
K-E	19.5/**42.5**	31.3/**31.9**	26.7/**36.1**
K-J	15.0/**32.5**	27.7/**12.3**	22.7/**20.2**

beam search/example-based decoder

Search Error

	matched	non-matched	total
C-E	15.2/**58.0**	28.2/**38.6**	21.4/**48.8**
C-J	16.7/**46.8**	30.3/**24.9**	23.1/**24.9**
C-K	14.9/**45.0**	25.7/**29.5**	20.0/**37.6**
E-C	19.2/**46.2**	39.1/**26.8**	31.0/**34.1**
E-J	19.7/**34.1**	33.4/**26.2**	27.8/**29.4**
E-K	14.4/**42.8**	31.8/**32.8**	24.7/**36.9**
J-C	10.5/**63.4**	37.6/**31.4**	28.4/**42.2**
J-E	16.3/**51.7**	35.8/**28.1**	29.2/**36.1**
J-K	14.5/**39.5**	50.3/**10.9**	38.2/**20.6**
K-C	14.0/**51.5**	31.9/**28.7**	24.9/**37.6**
K-E	19.5/**42.5**	31.3/**31.9**	26.7/**36.1**
K-J	15.0/**32.5**	27.7/**12.3**	22.7/**20.2**

beam search/example-based decoder

Some Examples of Translations

input:	銀行 の 前 で バッグ を ひったくら れ まし た
reference:	i was robbed of my bag in front of the bank
beam:	my bag was stolen in the front of the bank
example:	i was robbed of my bag in front of the bank
retrieved:	(i was robbed of my wallet bank on the subway)
input:	今 調べ ており ます ので しばらく お 待ち ください
reference:	would you wait for a moment while i check
beam:	i ’m out so please wait a moment
example:	would you wait a moment while we check
retrieved:	(would you wait for a moment while we check)
input:	お 金 を 入れ た のに 機械 が 動き ませ ん お 金 を 返し てもらい たい の です が
reference:	even though i put in some money the machine doesn’t work i ’d like a refund please
beam:	i ’d like a refund for this machine doesn’t money got heading
example:	although i put in some coins the machine didn’t work i ’d like a refund
retrieved:	(although i put in some coins the machine did’t work i ’d like a refund)

Some Examples of Translations 2

input:	十 時 に ここ で 待っ て い て くれる ん です ね
reference:	you will wait for me here at ten right
beam:	here is your pay wait ten o’clock
example:	are you waiting here at ten o’clock
retrieved:	(why are you waiting here)
input:	いいえ 赤い バッグ の 隣 に ある もの です
reference:	no the one next to the red bag
beam:	no red bag in the room next door is something
example:	no it ’s next to the red one
retrieved:	(it ’s next to that building)
input:	ええ それ と 似 た 手帳 が 届い て い ます
reference:	yes someone did turn in a notebook like that
beam:	yes it is similar there any messages for me
example:	i have a notebook come yes it is similar
retrieved:	(yes we have a japanese speaking guide)

Discussion

- Example-based decoder
  - A method to merge example-based framework and statistical machine translation
    - Retrieve-and-tweak strategy
    - Retrieval of examples — edit distance + *tf/idf*
    - Modification of examples — greedy method, but uses retrieved examples as the initial condition
  - Very strong bias to guide the search, especially suitable for long distance languaga pairs

Discussion, Contd.

- Related Work : Memory-based SMT (Marcu, 2001)
  - Extract phrase translation pattern
  - The greedy decoding process is initiated from the concatenation of phrases found in the translation memory
  - Difference in unit: Phrase vs. Sentence

Discussion, Contd.

- Related Work : Memory-based SMT (Marcu, 2001)
  - Extract phrase translation pattern
  - The greedy decoding process is initiated from the concatenation of phrases found in the translation memory
  - Difference in unit: Phrase vs. Sentence
- Better model, better translation?
  - Some correlation between the probability assigned by the models and the translation quality (Akiba et al. 2002)

Discussion, Contd.

- Related Work : Memory-based SMT (Marcu, 2001)
  - Extract phrase translation pattern
  - The greedy decoding process is initiated from the concatenation of phrases found in the translation memory
  - Difference in unit: Phrase vs. Sentence
- Better model, better translation?
  - Some correlation between the probability assigned by the models and the translation quality (Akiba et al. 2002)
- Future Works
  - *Post*-translation
  - Chunk-based (or Syntax-based) statistical translation model