統計的機械翻訳ことはじめ

渡辺 太郎

taro.watanabe@atr.jp

ATR 音声言語コミュニケーション研究所

目的

- 統計的機械翻訳についてのチュートリアル
  次のような人を対象:
  - 「統計的機械翻訳」を聞いたことがある。あるいは、
  - 関連する論文を読んだ (見た) ことがある。
  - IBM Model 3 まではたどりつけた。
  - etc.
- モデルの構成、確率値の学習手法
- デコーディング
- 最近の研究、および今後の課題

一通目の手紙

Warren Weaver to Nobert Wiener (March 4, 1947)

> *When I look at an article in Russian, I say; "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."*

二通目の手紙

Warren Weaver to Nobert Wiener (March 4, 1947)

> *When I look at an article in Russian, I say; "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."*

N. Wiener to W. Weaver (April 30, 1947)

> *... as to the problem of mechanical translation, I frankly am afraid the boundaries of words in different languages are too vague and the emotion and international connotations are too expensive to make any quasimechanical translation scheme very hopeful.*

三通目の手紙

W. Weaver to N. Wiener (May 9, 1947)

> *... Suppose we take a vocabulary of 2,000 words, and admit for good measure all the two-word combinations as if they were single words. The vocabulary is still only four million: and that is not so formidable a number to a modern computer, is it?*

背景

- 対訳コーパスの増大
- コンピュータの性能向上
- 機械学習アルゴリズムの発達

背景

- 対訳コーパスの増大
- コンピュータの性能向上
- 機械学習アルゴリズムの発達

↓

- コーパスからの対訳知識の自動獲得
- ルール/ヒューリスティックの手作りからの解放

統計的機械翻訳の歴史

- 1947 年:W.Weaver が暗号解読的、情報理論的手法を提案
- 1994 年: Candide システムが実現 (Berger et al., 1994)
- 現在: 飛躍的な性能の向上 (Och et al., 2003)

統計的機械翻訳の歴史

- 1947 年:W.Weaver が暗号解読的、情報理論的手法を提案
- 1994 年: Candide システムが実現 (Berger et al., 1994)
- 現在: 飛躍的な性能の向上 (Och et al., 2003)

問題の論文: *The Mathematics of Machine Translation: Parameter Estimation* (Brown et al., 1993)

- 難解...

統計的機械翻訳の歴史

- 1947 年:W.Weaver が暗号解読的、情報理論的手法を提案
- 1994 年: Candide システムが実現 (Berger et al., 1994)
- 現在: 飛躍的な性能の向上 (Och et al., 2003)

問題の論文: *The Mathematics of Machine Translation: Parameter Estimation* (Brown et al., 1993)

- 難解...
- いろんなチュートリアル
  - Workbook by Knight (1999b)
  - Tutorial by Knight and Koehn (2003)
  - チュートリアル (永田, 2003)

内容

- 統計的機械翻訳
- 多言語へのチャレンジ
- 統計的機械翻訳の現状

内容

- 統計的機械翻訳
- 多言語へのチャレンジ
- 統計的機械翻訳の現状

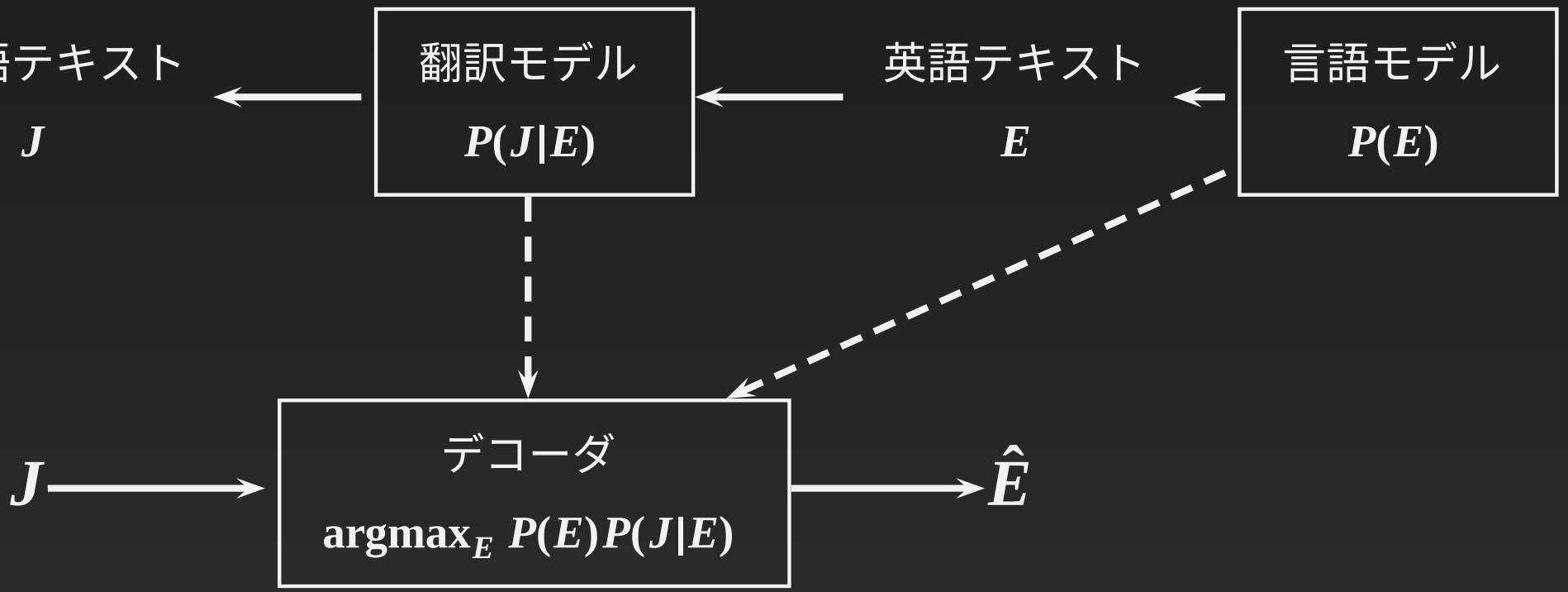
内容

- 統計的機械翻訳
  - 単語アライメント
  - 翻訳モデル
  - 言語モデル
  - デコーディング
  - 評価手法
- 多言語へのチャレンジ
- 統計的機械翻訳の現状

統計的機械翻訳とは?

基本的なアイデア (Brown et al., 1990)

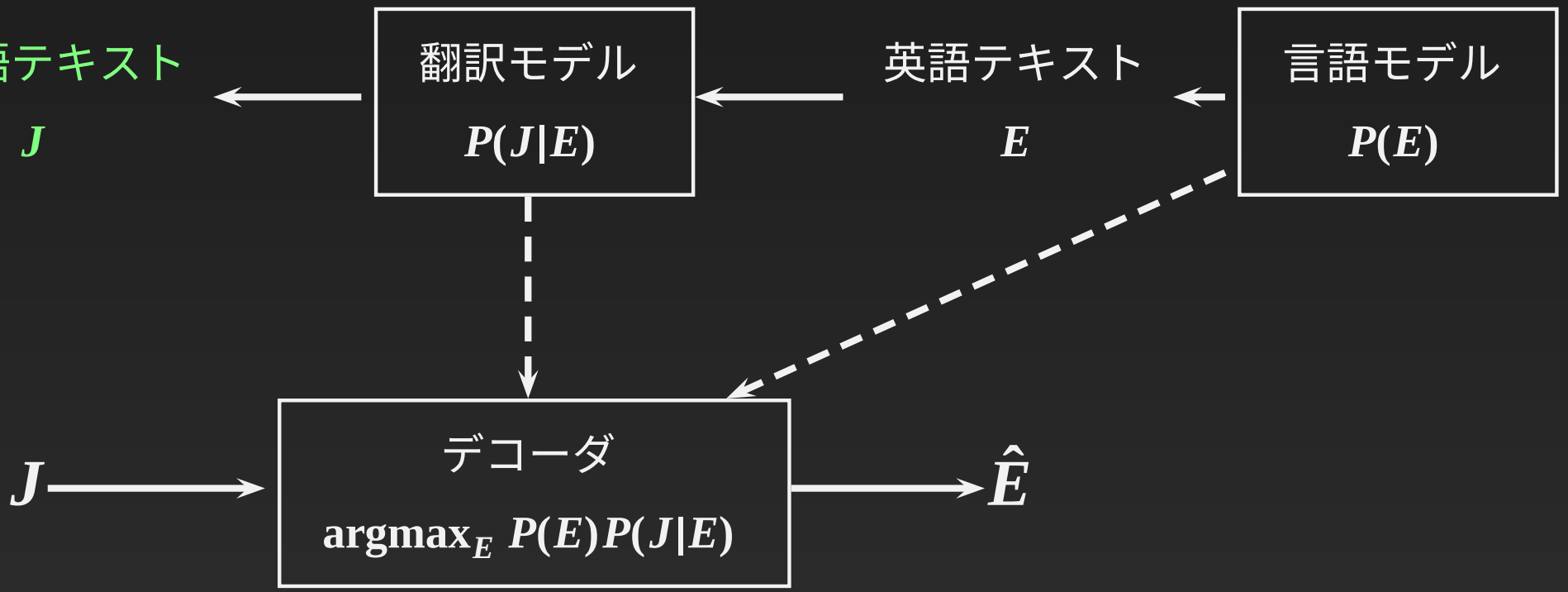
$$\hat{E} = \underset{E}{\operatorname{argmax}}\, P(E|J)$$



統計的機械翻訳とは?

基本的なアイデア (Brown et al., 1990)

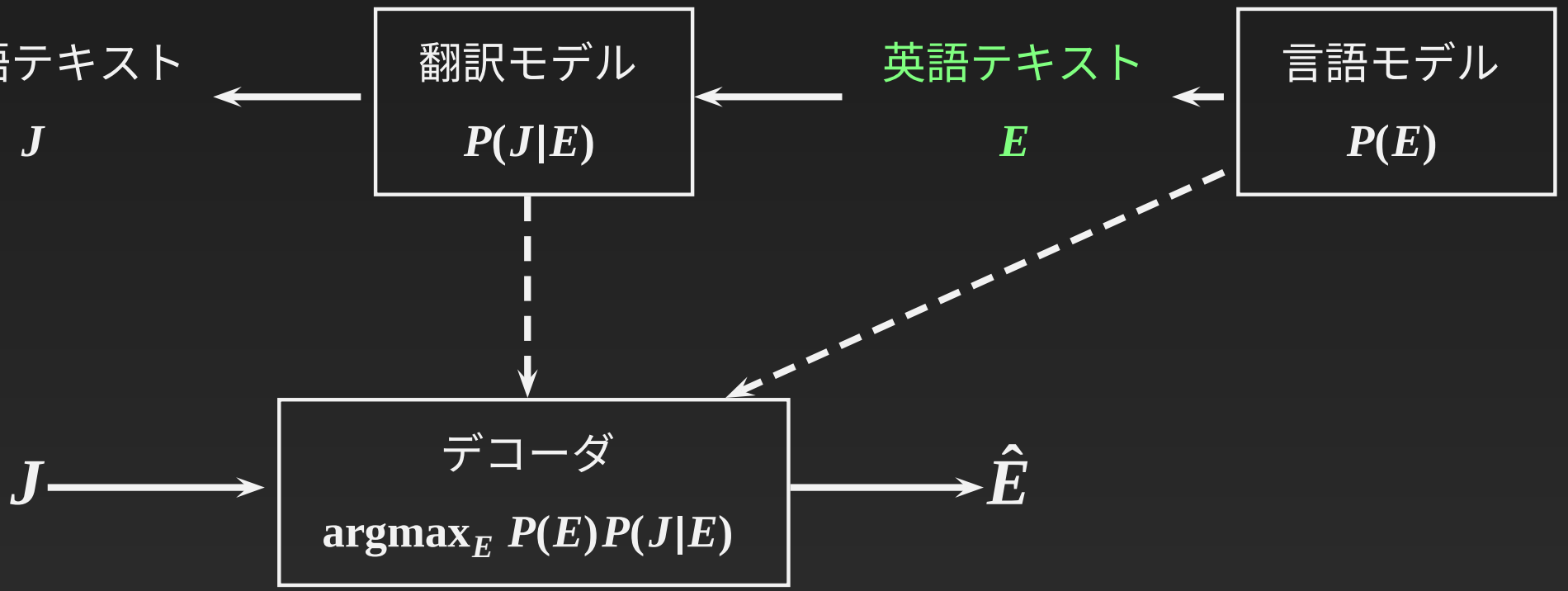
$$\hat{E} = \underset{E}{\operatorname{argmax}}\, P(E|J)$$



統計的機械翻訳とは?

基本的なアイデア (Brown et al., 1990)

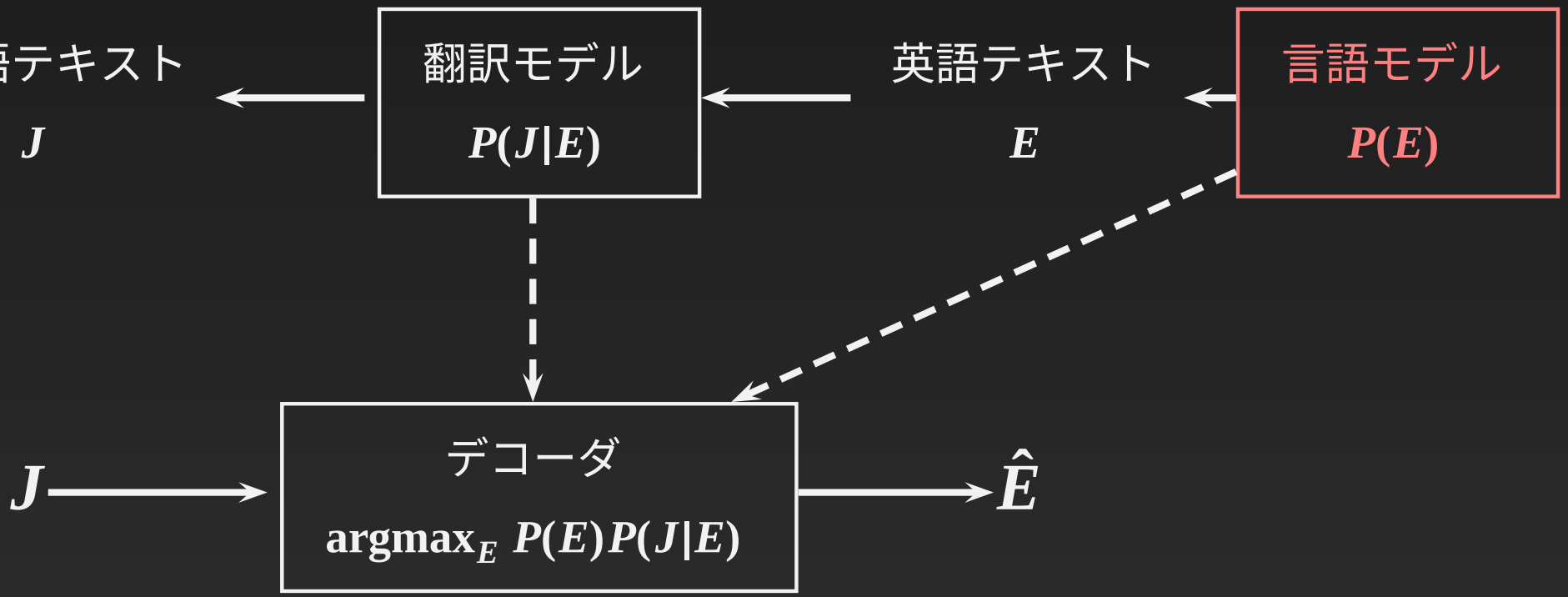
$$\hat{E} = \underset{E}{\operatorname{argmax}}\, P(E|J)$$



統計的機械翻訳とは?

基本的なアイデア (Brown et al., 1990)

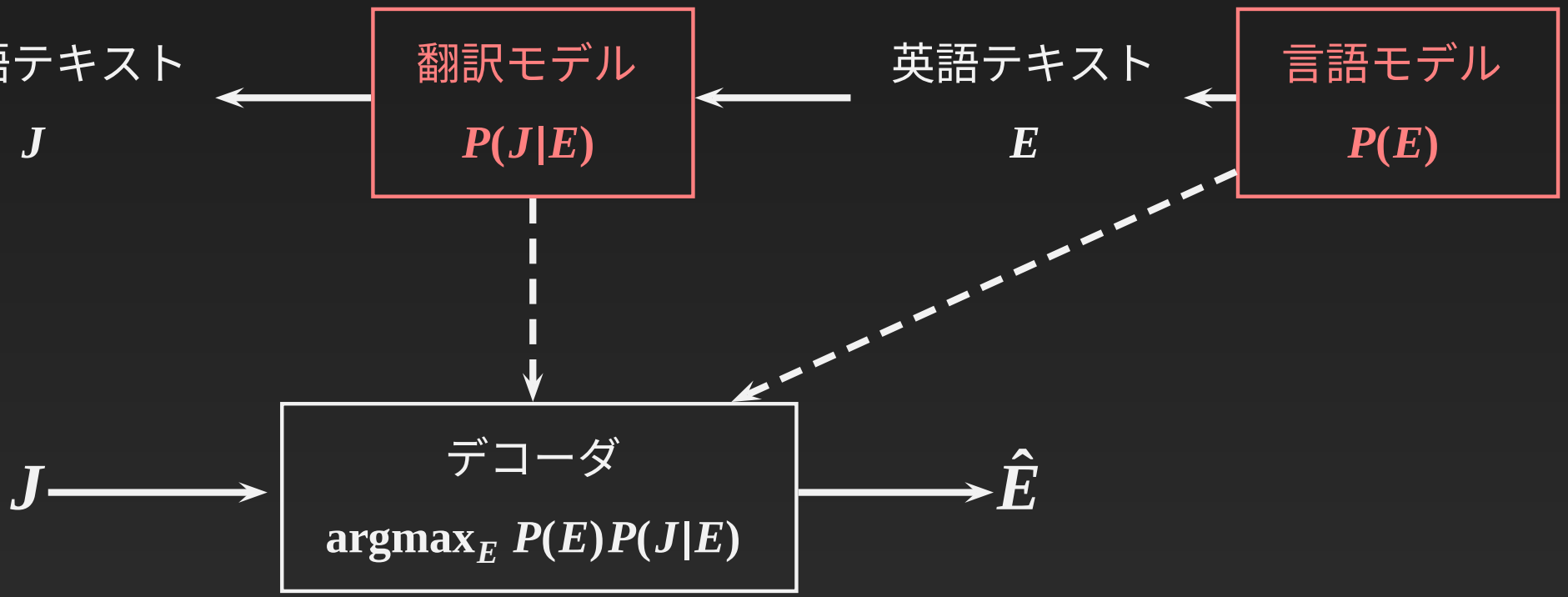
$$\hat{E} = \underset{E}{\operatorname{argmax}}\, P(E|J)$$



統計的機械翻訳とは?

基本的なアイデア (Brown et al., 1990)

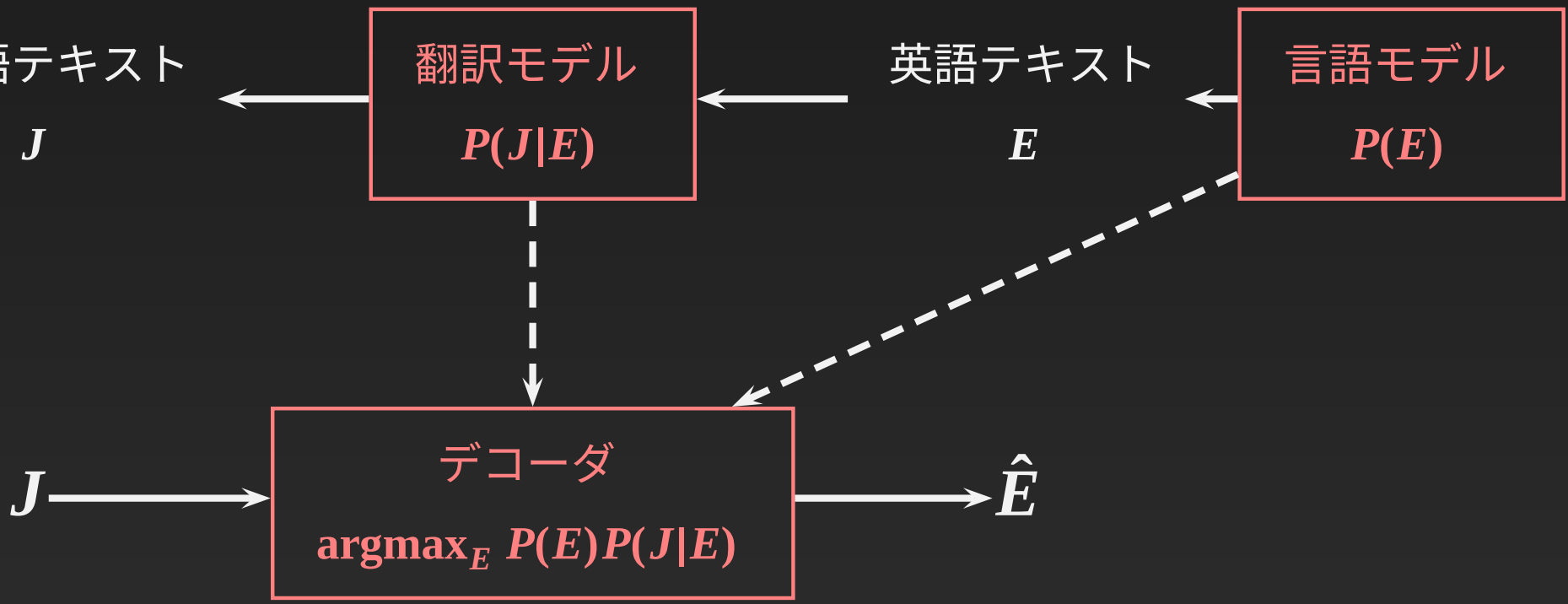
$$\hat{E} = \underset{E}{\operatorname{argmax}}\, P(E|J)$$



統計的機械翻訳とは?

基本的なアイデア (Brown et al., 1990)

$$\hat{E} = \underset{E}{\operatorname{argmax}}\, P(E|J)$$

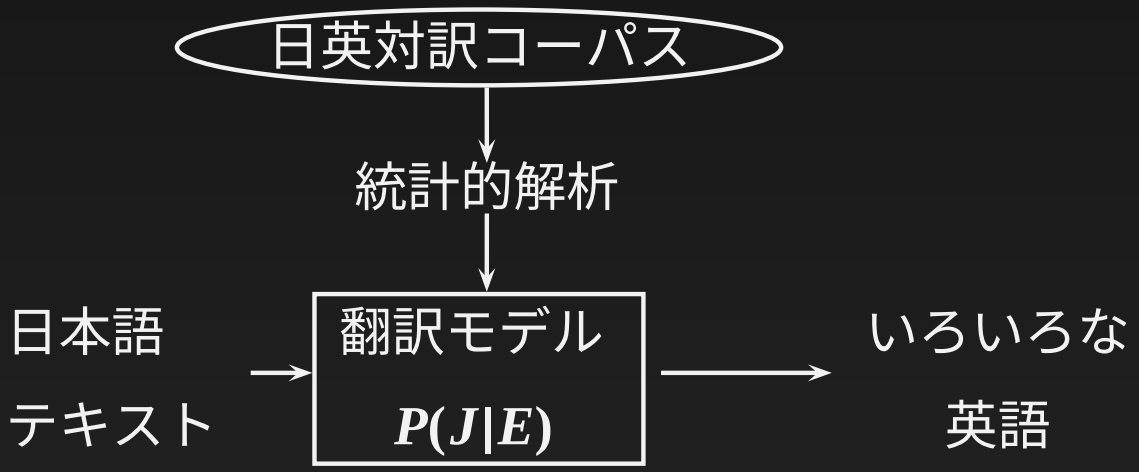


統計的機械翻訳の枠組み

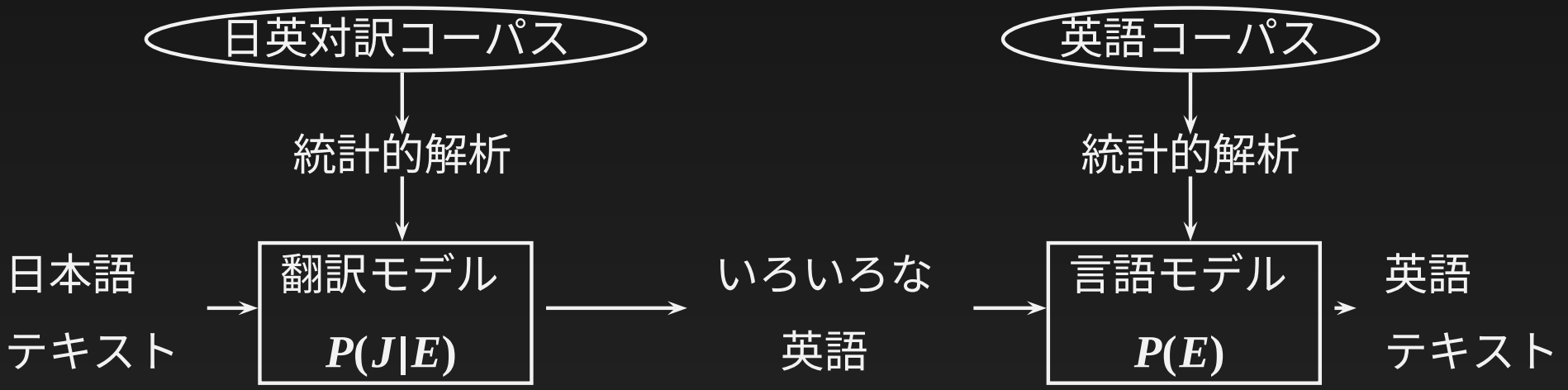
日本語

テキスト

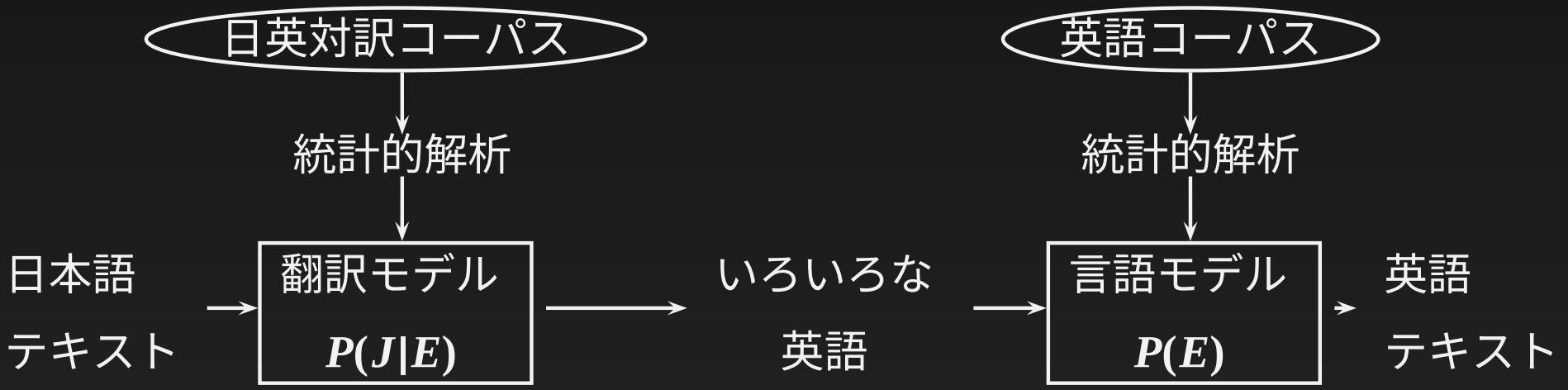
統計的機械翻訳の枠組み



統計的機械翻訳の枠組み



統計的機械翻訳の枠組み



濃いコーヒーが飲みたいのですが

生成

I want strong coffee.

Strong coffee please.

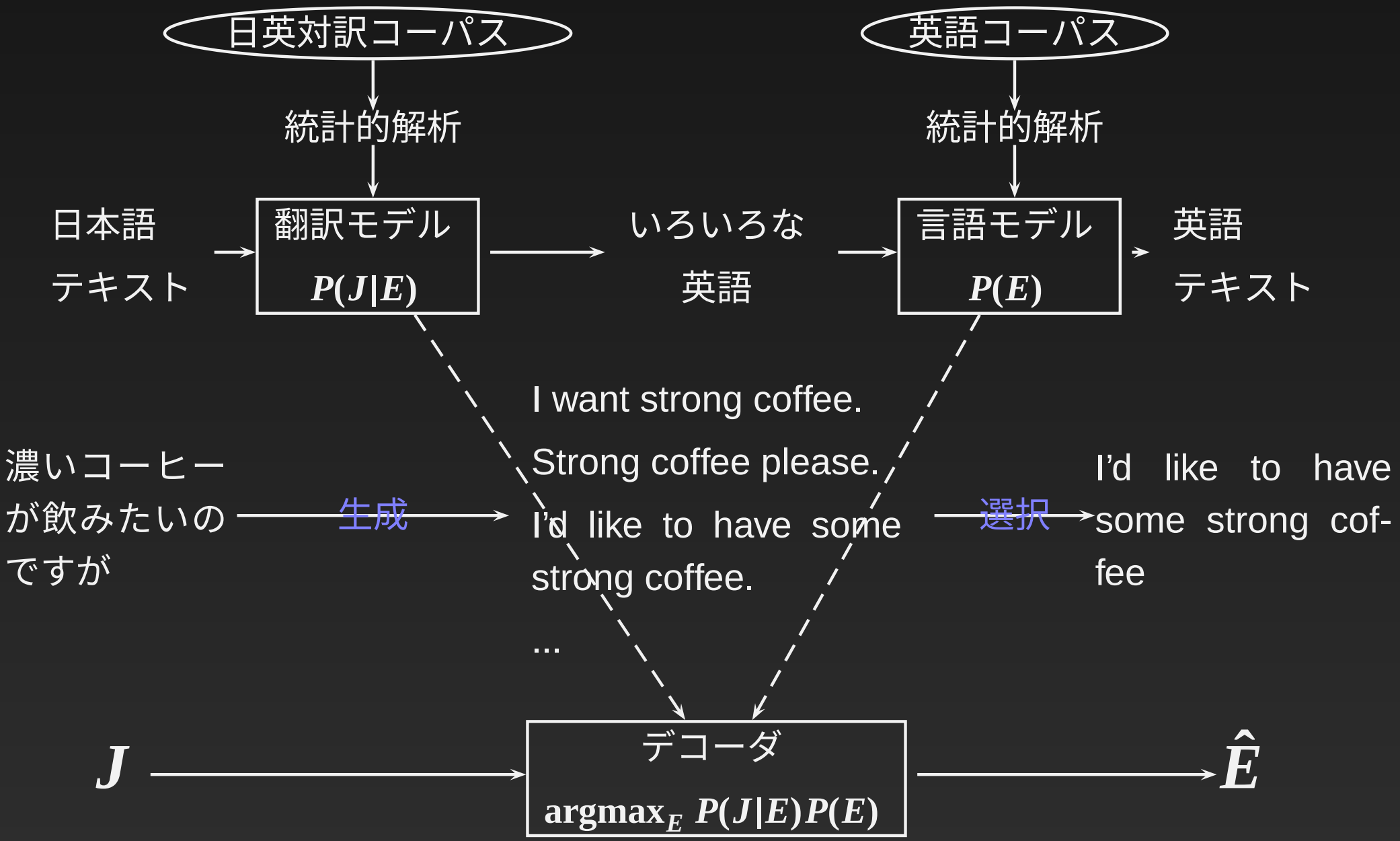
I'd like to have some strong coffee.

...

選択

I'd like to have some strong coffee

統計的機械翻訳の枠組み



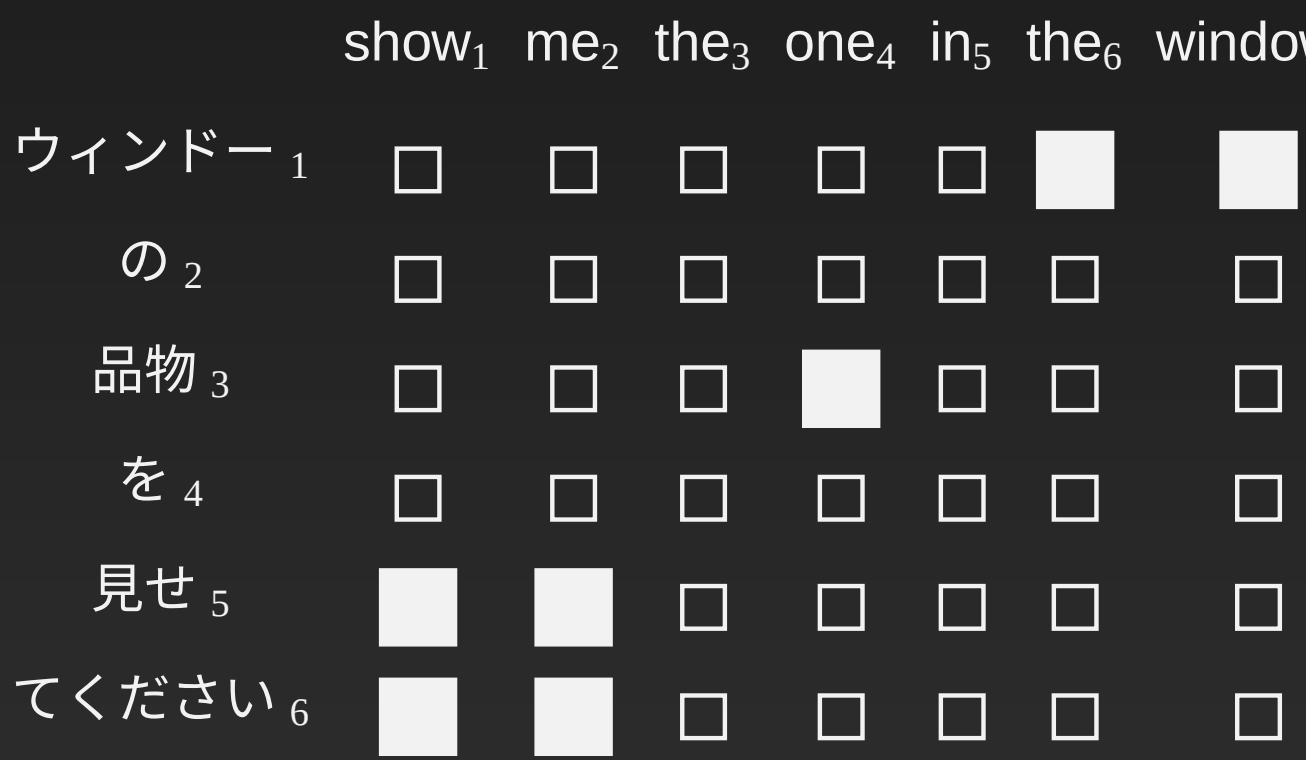
単語アライメントの導入

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

単語アライメントの導入

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

- A: 単語アライメント

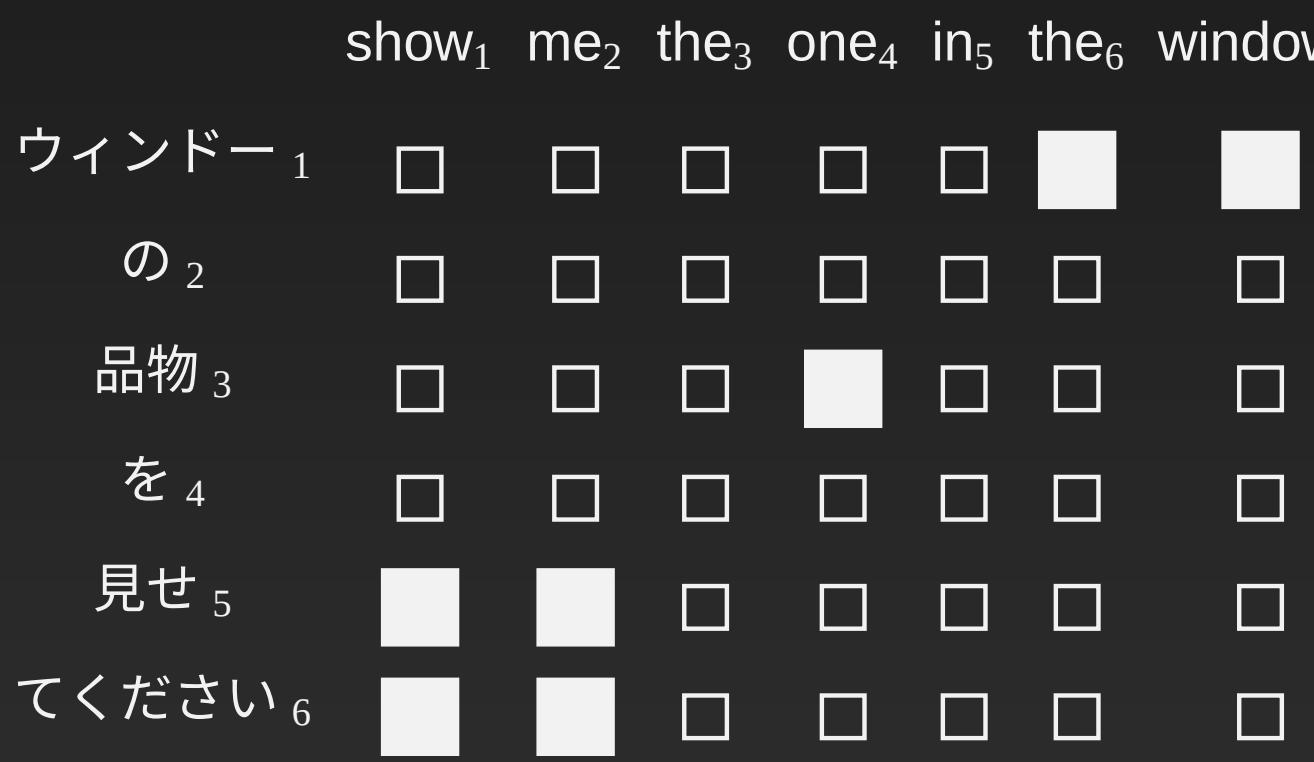


- 全ての単語アライメントの数は?

単語アライメントの導入

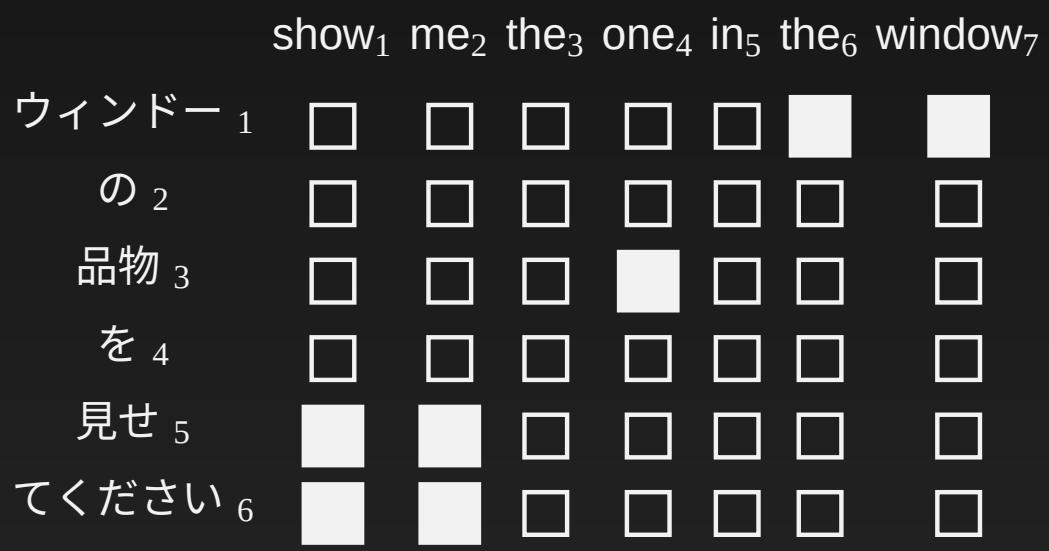
$$P(J|E) = \sum_{A} P(J, A|E)$$

- A: 単語アライメント

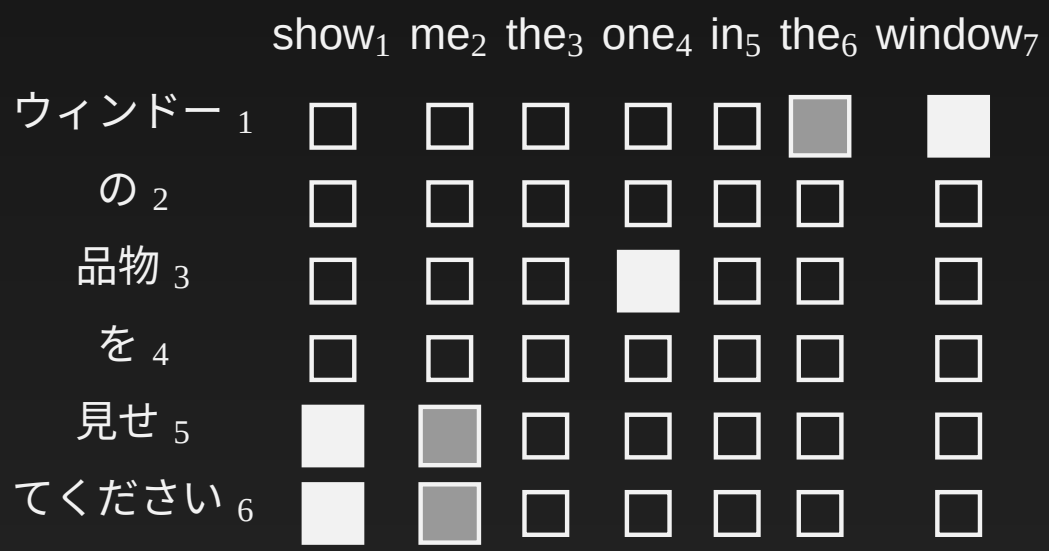


- 全ての単語アライメントの数は? $\longrightarrow 2^{lm}$ ($m = |J|$、$l = |E|$)

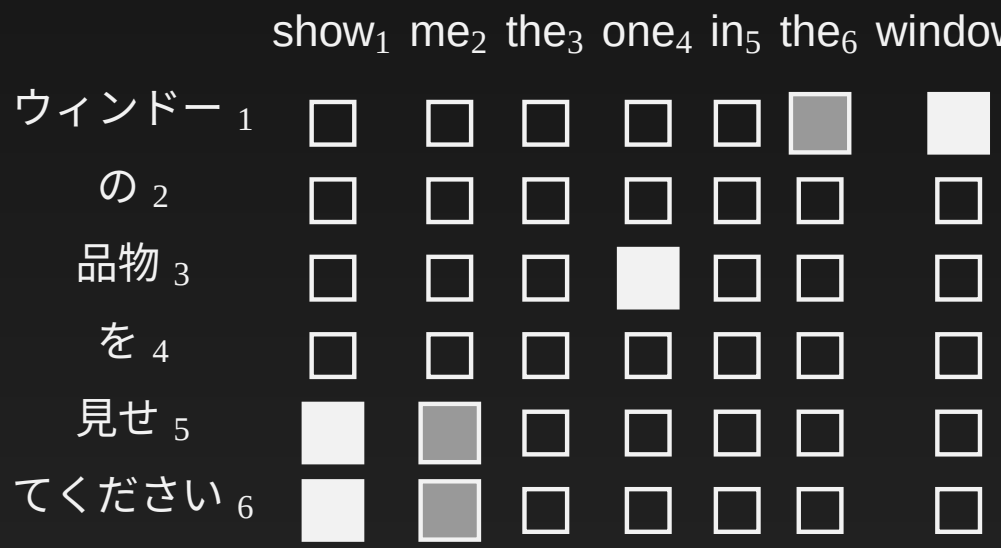
単語アライメントの表現



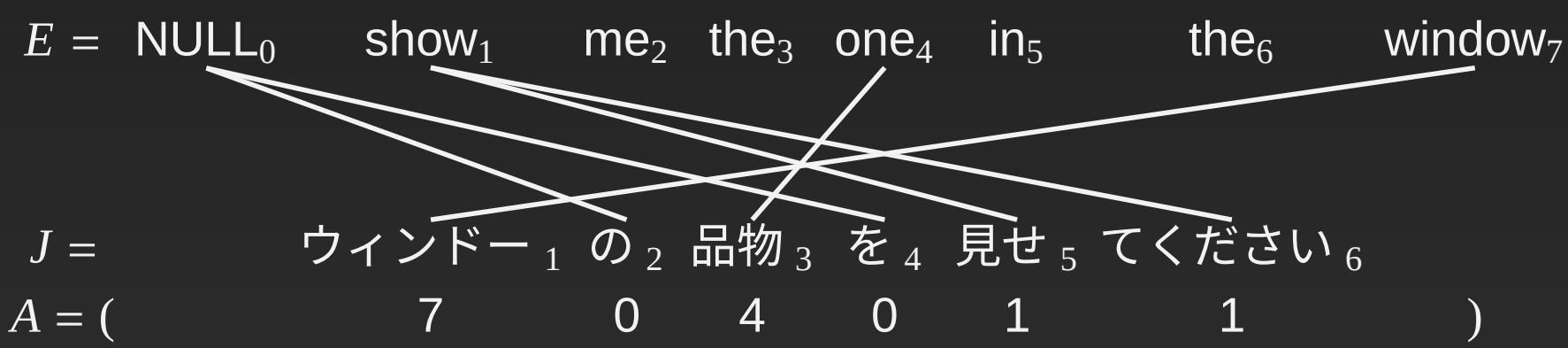
単語アライメントの表現



単語アライメントの表現

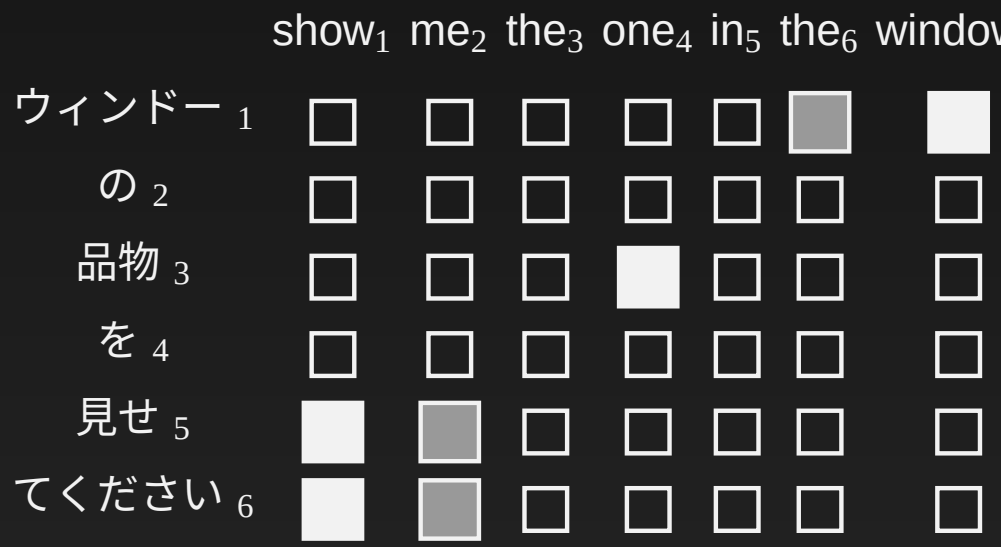


- 単語アライメントのコンパクトな表現

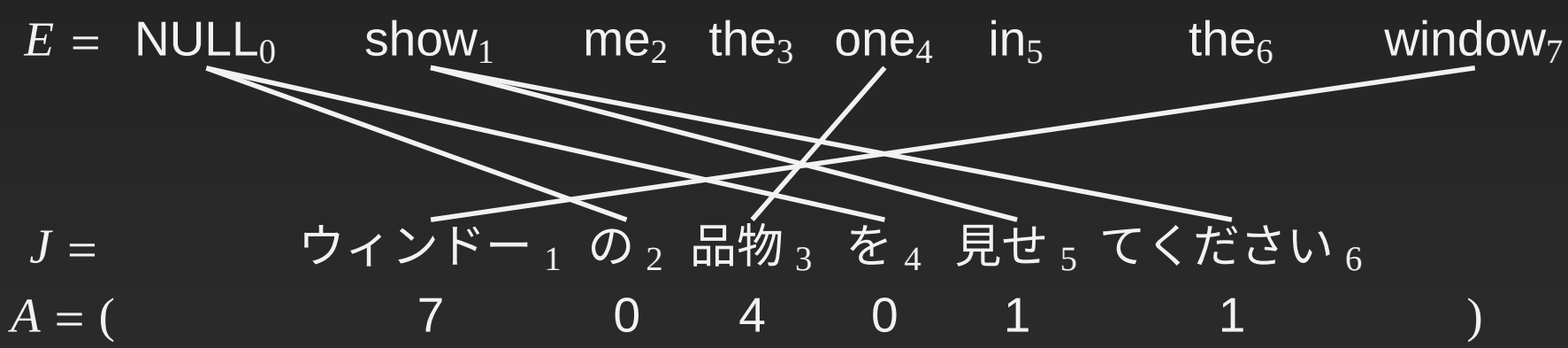


- 全ての単語アライメントの数は?

単語アライメントの表現



- 単語アライメントのコンパクトな表現



- 全ての単語アライメントの数は? $\longrightarrow (l+1)^m$

翻訳モデルの構成 (IBM Model 4)

$P(J, A|E)$

could you recommend another hotel

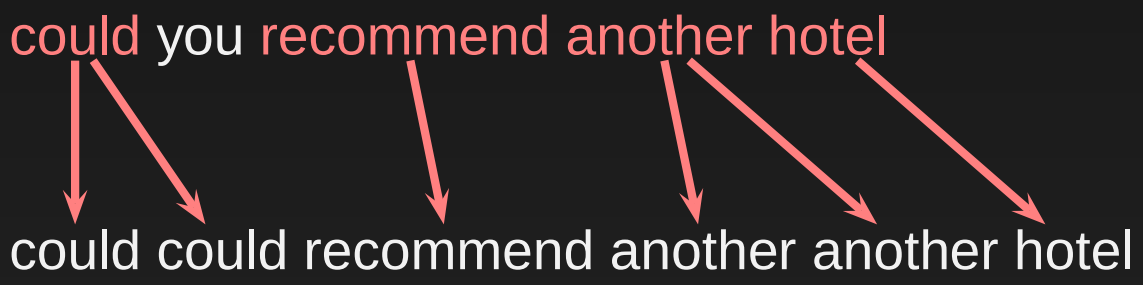
他 の ホテルを 紹介し ていただけます か

翻訳モデルの構成 (IBM Model 4)

$$P(J, A|E)$$

Fertility Model

$$\boxed{\prod n(\phi_i|E_i)}$$



他 の ホテルを 紹介し ていただけます か

翻訳モデルの構成 (IBM Model 4)

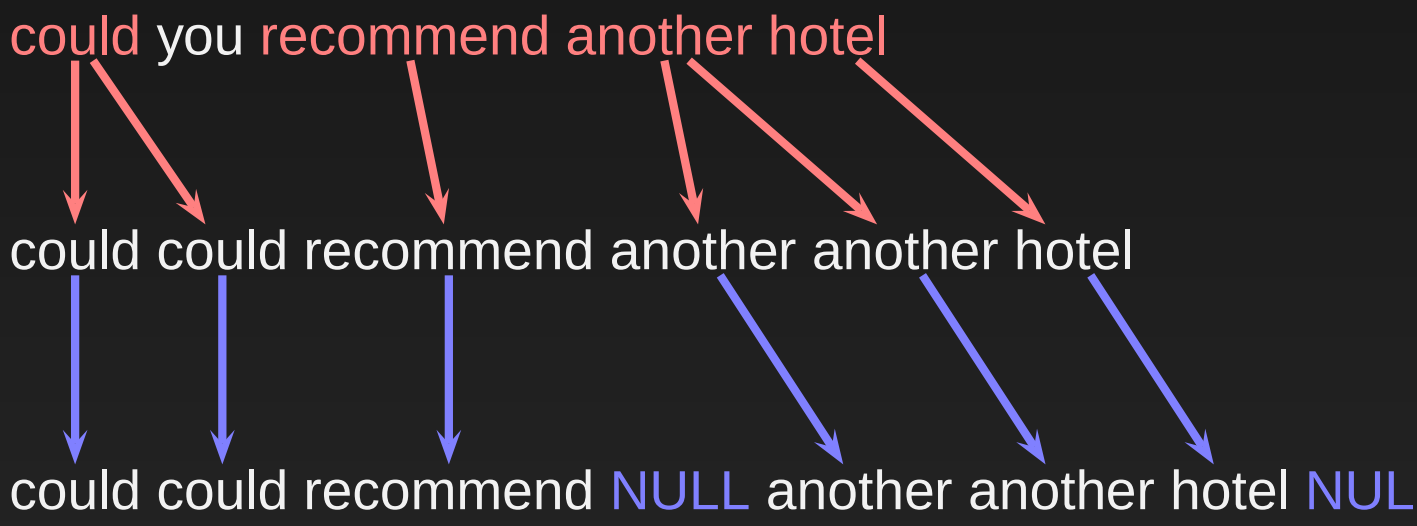
$$P(J, A|E)$$

Fertility Model

$$\prod n(\phi_i|E_i)$$

NULL Generation Model

$$\binom{m-\phi_0}{\phi_0} p_0^{m-2\phi_0} p_1^{\phi_0}$$



他 の ホテルを 紹介し ていただけます か

翻訳モデルの構成 (IBM Model 4)

$$P(J, A|E)$$

Fertility Model

$$\prod n(\phi_i|E_i)$$

NULL Generation Model

$$\binom{m-\phi_0}{\phi_0} p_0^{m-2\phi_0} p_1^{\phi_0}$$

Lexicon Model

$$\prod t(J_j|E_{A_j})$$

他 の ホテルを 紹介し ていただけます か

翻訳モデルの構成 (IBM Model 4)

$$P(J, A|E)$$

Fertility Model

$$\prod n(\phi_i|E_i)$$

NULL Generation Model

$$\binom{m-\phi_0}{\phi_0} p_0^{m-2\phi_0} p_1^{\phi_0}$$

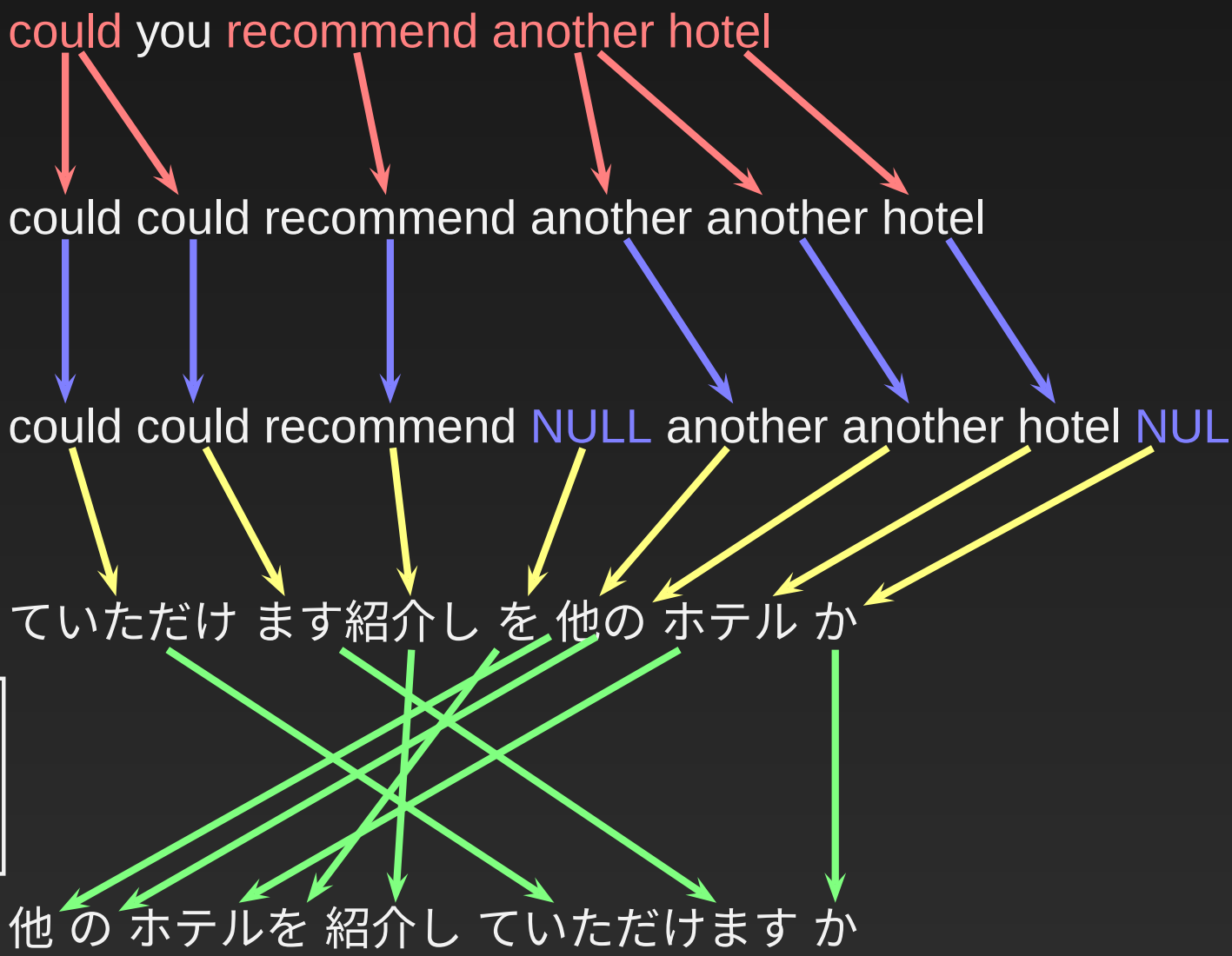
Lexicon Model

$$\prod t(J_j|E_{A_j})$$

Distortion Model

$$\prod d_1(j - k|\mathcal{A}(E_i)\mathcal{B}(J_j))$$

$$\prod d_{1>}(j - j'|\mathcal{B}(J_j))$$



翻訳モデルの学習

- 対訳コーパス $(J_0, E_0), ...(J_k, E_k)...$ からパラメータの推定
- 問題点: アライメントのデータ (J_k, E_k, A_k) がない...

翻訳モデルの学習

- 対訳コーパス $(J_0, E_0), ...(J_k, E_k)...$ からパラメータの推定
- 問題点: アライメントのデータ (J_k, E_k, A_k) がない...

toast and a pot of coffee please

トースト に コーヒー を ポット で ください

coffee or tea

コーヒー それとも 紅茶 に なさい ます か

tea with a slice of lemon please

紅茶 に レモン の 輪切り を 添え てください

which juice shall i bring you

ジュース は 何 を お 持ち し ましょ う

orange juice pelase

オレンジジュース が いいな

shrimp cocktail please

芝エビ の カクテル を 貰い ましょ う

翻訳モデルの学習

- 対訳コーパス $(J_0, E_0), ...(J_k, E_k)...$ からパラメータの推定
- 問題点: アライメントのデータ (J_k, E_k, A_k) がない...

toast and a pot of coffee please
トースト に コーヒー を ポット で ください

coffee or tea
コーヒー それとも 紅茶 に なさい ます か

tea with a slice of lemon please
紅茶 に レモン の 輪切り を 添え てください

which juice shall i bring you
ジュース は 何 を お 持ち し ましょ う

orange juice pelase
オレンジジュース が いいな

shrimp cocktail please
芝エビ の カクテル を 貰い ましょ う

翻訳モデルの学習

- 対訳コーパス $(J_0, E_0), ...(J_k, E_k)...$ からパラメータの推定
- 問題点: アライメントのデータ (J_k, E_k, A_k) がない...

toast and a pot of coffee please
トースト に コーヒー を ポット で ください

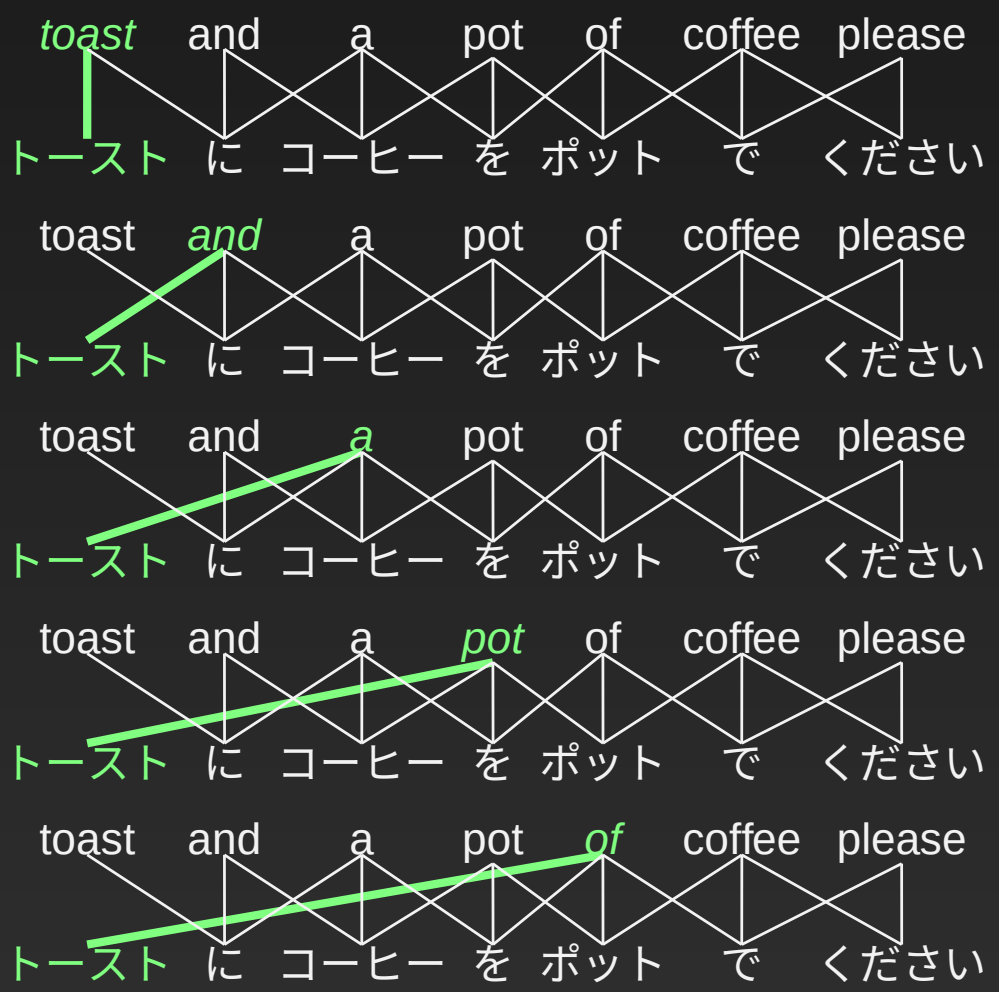
coffee or tea
コーヒー それとも 紅茶 に なさい ます か

tea with a slice of lemon please
紅茶 に レモン の 輪切り を 添え てください

which juice shall i bring you
ジュース は 何 を お 持ち し ましょ う

orange juice pelase
オレンジジュース が いいな

shrimp cocktail please
芝エビ の カクテル を 貰い ましょ う



EMアルゴリズム (Dempster et al., 1977)

1. 各モデルの初期化 (i.e. uniform distribution)
2. 各(J_k, E_k)に対し、完全なデータ(J_k, E_k, A)を生成、生成された各データにつき、モデルを使い、$P(A|J_k, E_k)$を計算。
3. 完全なデータから、モデルの各パラメータを求める。
4. 2と3のステップを繰り返す。

翻訳モデルの学習における問題点

- 局所解
- Overfitting
- すべての $\boldsymbol{A}$ についての足し算

翻訳モデルの学習における問題点

- 局所解
  - 簡単なモデル (IBM Model 1 etc.) による初期値の決定
- Overfitting
- すべての $\boldsymbol{A}$ についての足し算

翻訳モデルの学習における問題点

- 局所解
  - 簡単なモデル (IBM Model 1 etc.) による初期値の決定
- Overfitting
  - テストセットパープレキシティーによる判定
  - 数回のトレーニング
- すべての A についての足し算

翻訳モデルの学習における問題点

- 局所解
  - 簡単なモデル (IBM Model 1 etc.) による初期値の決定
- Overfitting
  - テストセットパープレキシティーによる判定
  - 数回のトレーニング
- すべての A についての足し算

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

  - アライメントの集合のサブセットを使用 (Och and Ney, 2003)

翻訳モデルの学習における問題点

- 局所解
  - 簡単なモデル (IBM Model 1 etc.) による初期値の決定
- Overfitting
  - テストセットパープレキシティーによる判定
  - 数回のトレーニング
- すべての A についての足し算

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

  - アライメントの集合のサブセットを使用 (Och and Ney, 2003)
- 他にも...
  - 句アライメントの制約を用いた学習 (Watanabe et al., 2004)
  - 翻訳フレーズ抽出+再トレーニング (Yamada et al., 2003)

他にも...

- HMM Model (Vogel et al., 2000)
  - 単語アライメントが以前の単語アライメントに依存

$$P(J, A|E) = \prod_j t(J_j|E_{A_j})a(A_j|A_{j-1}, |E|)$$

  - Forward-Backward アルゴリズムによるトレーニング

他にも...

- HMM Model (Vogel et al., 2000)
  - 単語アライメントが以前の単語アライメントに依存

$$P(J, A|E) = \prod_j t(J_j|E_{A_j})a(A_j|A_{j-1}, |E|)$$

  - Forward-Backward アルゴリズムによるトレーニング
- Maximum Entropy Modeling (Foster, 2000a,b)
  - $P(E|J)$ を直接モデル化

$$P(E|J) = \prod_i P(E_i|E_1, E_2, ..., E_{i-1}, J)$$

  - J を bag-of-word として素性を実現

翻訳モデル学習ツール

- GIZA
  - http://www.clsp.jhu.edu/ws99/projects/mt/
  - EGYPT ツールキットの一部
  - 他にも、様々なツール
- GIZA++ (Och and Ney, 2003)
  - http://www.isi.edu/~och/GIZA++.html
  - GIZA のバグフィックス
  - IBM Model 4/5 の学習
  - HMM Model の学習 (Och and Ney, 2003)
  - 自動単語アライメントツールとして使用可能?

言語モデル (n-gram)

$P($ I 'd like to have some strong coffee $) =$

言語モデル (n-gram)

$P($ I 'd like to have some strong coffee $) =$

I $\Rightarrow P($ I $)$

I 'd $\Rightarrow P($ 'd | I $)$

I 'd like $\Rightarrow P($ like | I 'd$)$

'd like to $\Rightarrow P($ to | 'd like $)$

like to have $\Rightarrow P($ have | like to $)$

to have some $\Rightarrow P($ some | to have $)$

have some strong $\Rightarrow P($ strong | have some $)$

some strong coffee $\Rightarrow P($ coffee | some strong $)$

言語モデル (n-gram) の問題点

- 長いコンテキストの制約
- 構文の制約

言語モデル (n-gram) の問題点

- 長いコンテキストの制約
- 構文の制約
- 他にも...
  - Web counts (Koehn et al., 2003)
    - Web 検索エンジンにより検索されたドキュメント数、フレーズ数

      例: 2 google、1 altavista
  - 構文木に基づく言語モデル (Charniak et al., 2003)
  - 音声認識で用いられている言語モデルを使用可能

言語モデルのツール

- The CMU-Cambridge Statistical Language Modeling Toolkit (http://mi.eng.cam.ac.uk/~prc14/toolkit.html)
- The SRI Language Modeling Toolkit (http://www.speech.sri.com/projects/srilm/)
- etc.

デコーディング

- 入力文 J が与えられたときの最適化問題の解

$$\begin{aligned}\hat{E} &= \operatorname*{argmax}_{E} \sum_{A} P(J, A|E)P(E) \\ &= \operatorname*{argmax}_{E} \max_{A} P(J, A|E)P(E)\end{aligned}$$

デコーディング

- 入力文 J が与えられたときの最適化問題の解

$$\begin{aligned}\hat{E} &= \operatorname*{argmax}_{E} \sum_{A} P(J, A|E)P(E) \\ &= \operatorname*{argmax}_{E} \max_{A} P(J, A|E)P(E)\end{aligned}$$

- NP-Complete (Knight, 1999a)
  - 訳語の選択問題 — Minimum Set Cover Problem
  - 並び替えの問題 — Hamilton Circuit Problem

デコーディング

- 入力文 J が与えられたときの最適化問題の解

$$\begin{aligned}\hat{E} &= \operatorname*{argmax}_{E} \sum_{A} P(J, A|E)P(E) \\ &= \operatorname*{argmax}_{E} \max_{A} P(J, A|E)P(E)\end{aligned}$$

- NP-Complete (Knight, 1999a)
  - 訳語の選択問題 — Minimum Set Cover Problem
  - 並び替えの問題 — Hamilton Circuit Problem
- デコーダの実装は簡単? (Foster et al., 2003)
  - 単語/文の境界線がはっきりしている
  - 並び替えの問題

ビームサーチ (Tillmann and Ney, 2003)

ビームサーチ (Tillmann and Ney, 2003)

input: 精算書 を 確認 さ せ てください

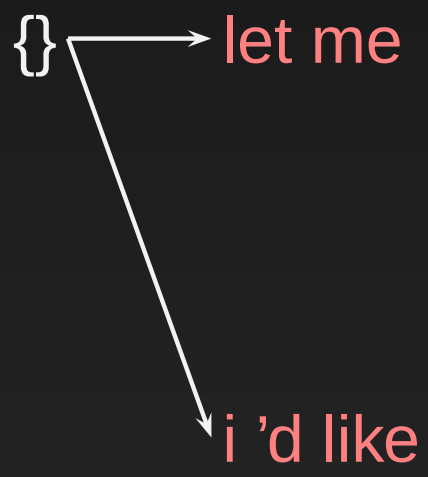
ビームサーチ (Tillmann and Ney, 2003)

input: 精算書 を 確認 さ せ てください

{}

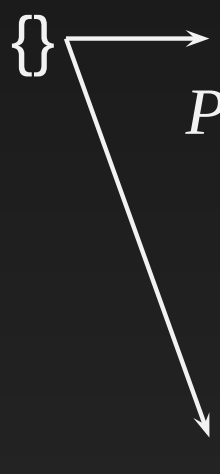
ビームサーチ (Tillmann and Ney, 2003)

input: 精算書 を 確認 さ せ てください



ビームサーチ (Tillmann and Ney, 2003)

input: 精算書 を 確認 さ せ てください

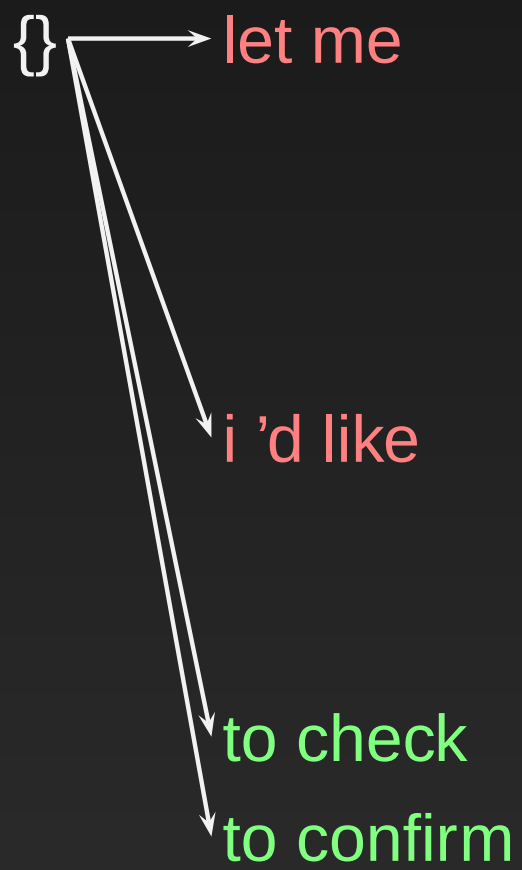


P(てください | let me)P(let me)

P(てください | i 'd like)P(i 'd like)

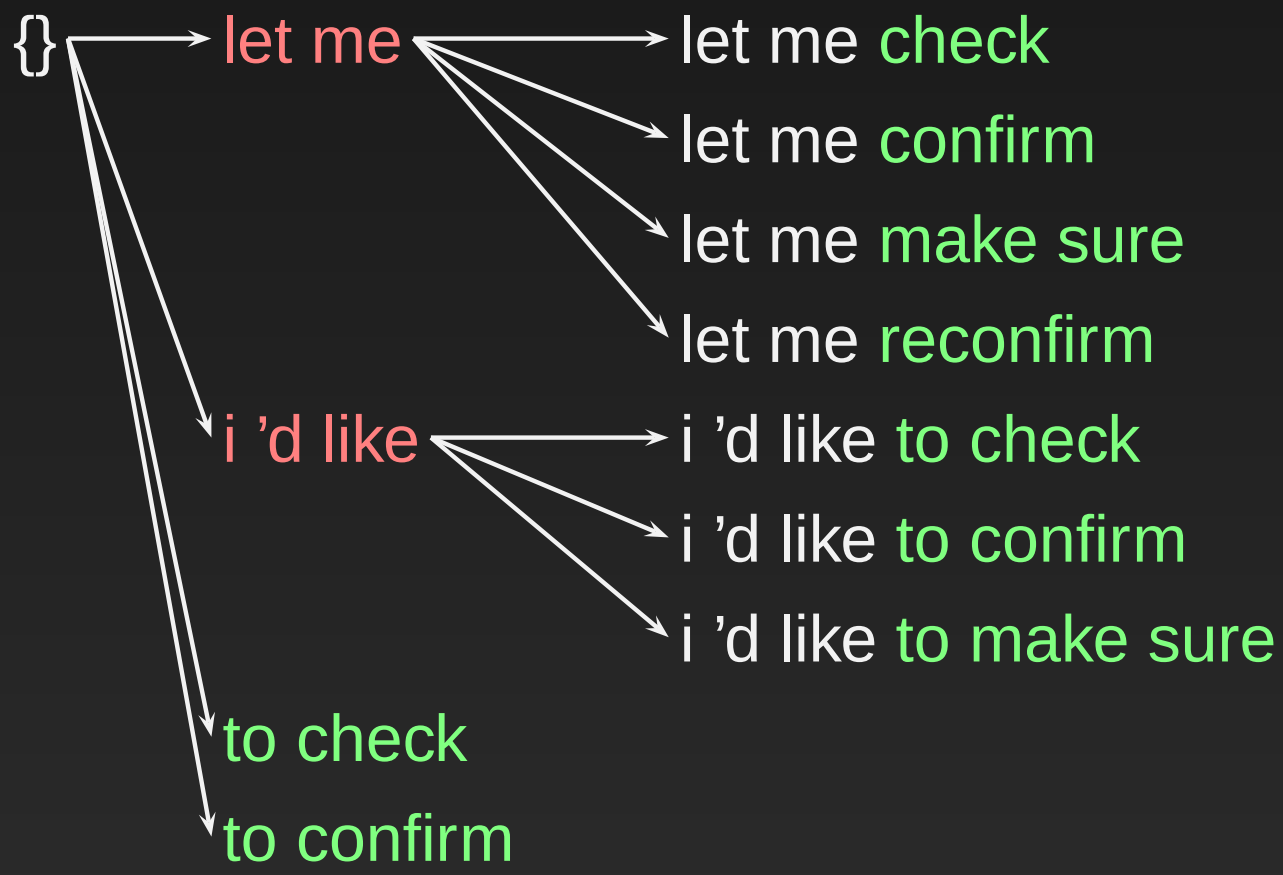
ビームサーチ (Tillmann and Ney, 2003)

input: 精算書 を 確認 さ せ てください



ビームサーチ (Tillmann and Ney, 2003)

input: 精算書 を 確認 さ せ てください



翻訳候補の生成

- 単語対応 — 逆 Lexicon Model (Berger et al., 1996)

$$P(E|J) = \frac{P(J|E)P(E)}{\sum_E P(J|E)P(E)}$$

翻訳候補の生成

- 単語対応 — 逆 Lexicon Model (Berger et al., 1996)

$$P(E|J) = \frac{P(J|E)P(E)}{\sum_E P(J|E)P(E)}$$

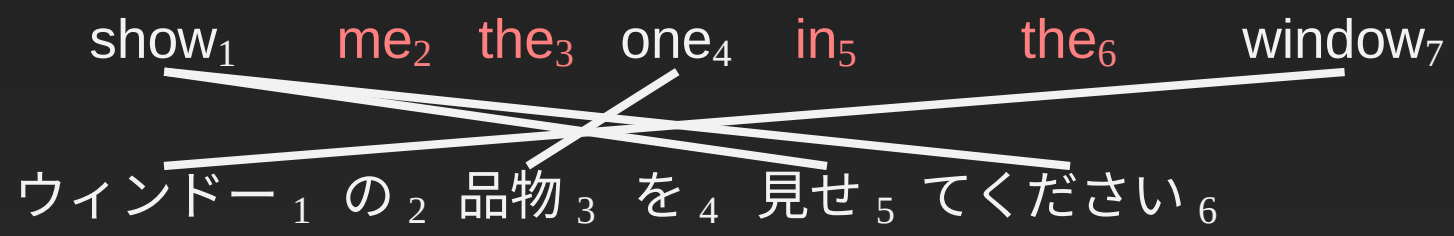
- 挿入 — Viterbi Alignment から抽出 (Watanabe and Sumita, 2002)

翻訳候補の生成

- 単語対応 — 逆 Lexicon Model (Berger et al., 1996)

$$P(E|J) = \frac{P(J|E)P(E)}{\sum_E P(J|E)P(E)}$$

- 挿入 — Viterbi Alignment から抽出 (Watanabe and Sumita, 2002)



- 単語対応+挿入

てください ⟶ let *me*, i 'd *like*, ...

確認 ⟶ to *confirm*, make *sure*, ...

プルーニング

- 仮説空間 — 2^m (m = 入力単語数)

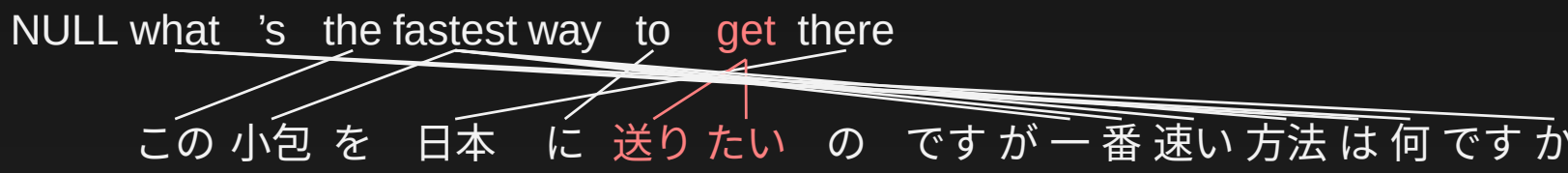
プルーニング

- 仮説空間 — 2^m (m = 入力単語数)
- プルーニング
  - Threshold — 各ビームの仮説のスコアは、
    最大の仮説のスコア × 閾値以上
  - Histogram — 各ビーム毎に、M-best のみを保持

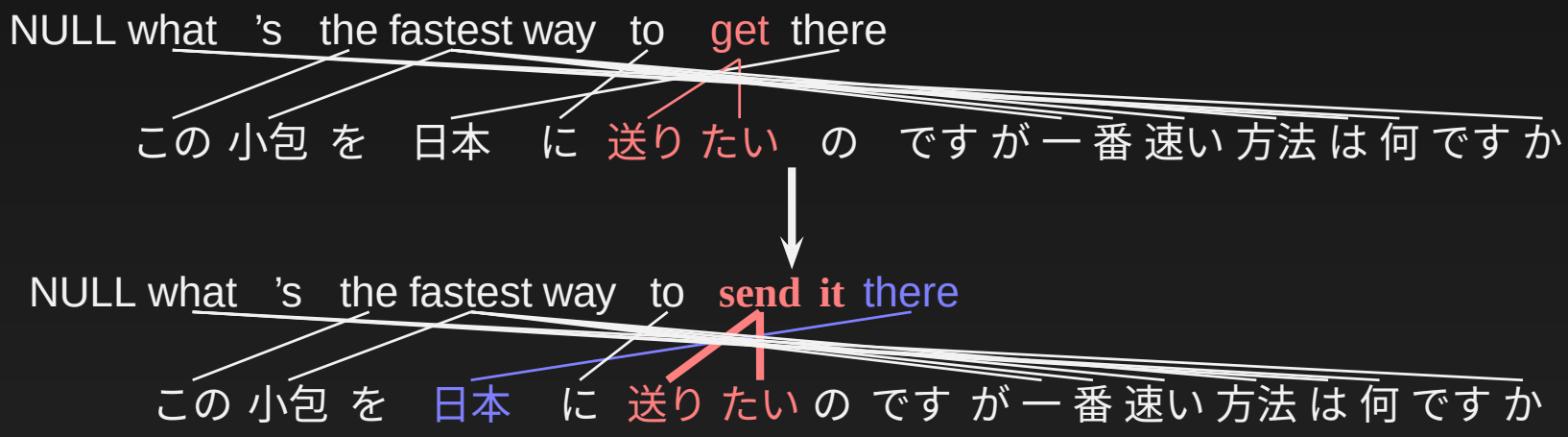
プルーニング

- 仮説空間 — 2^m (m = 入力単語数)
- プルーニング
  - Threshold — 各ビームの仮説のスコアは、
    最大の仮説のスコア × 閾値以上
  - Histogram — 各ビーム毎に、M-best のみを保持
- 制約
  - Skipping — スキップする入力単語の数 (Tillmann and Ney, 2003)
  - Fertility — 最大の fertility (Watanabe and Sumita, 2002)
  - FSA による制約 (Tillmann and Ney, 2003)
  - 構文による制約 (Wu, 1996; Zens and Ney, 2003)

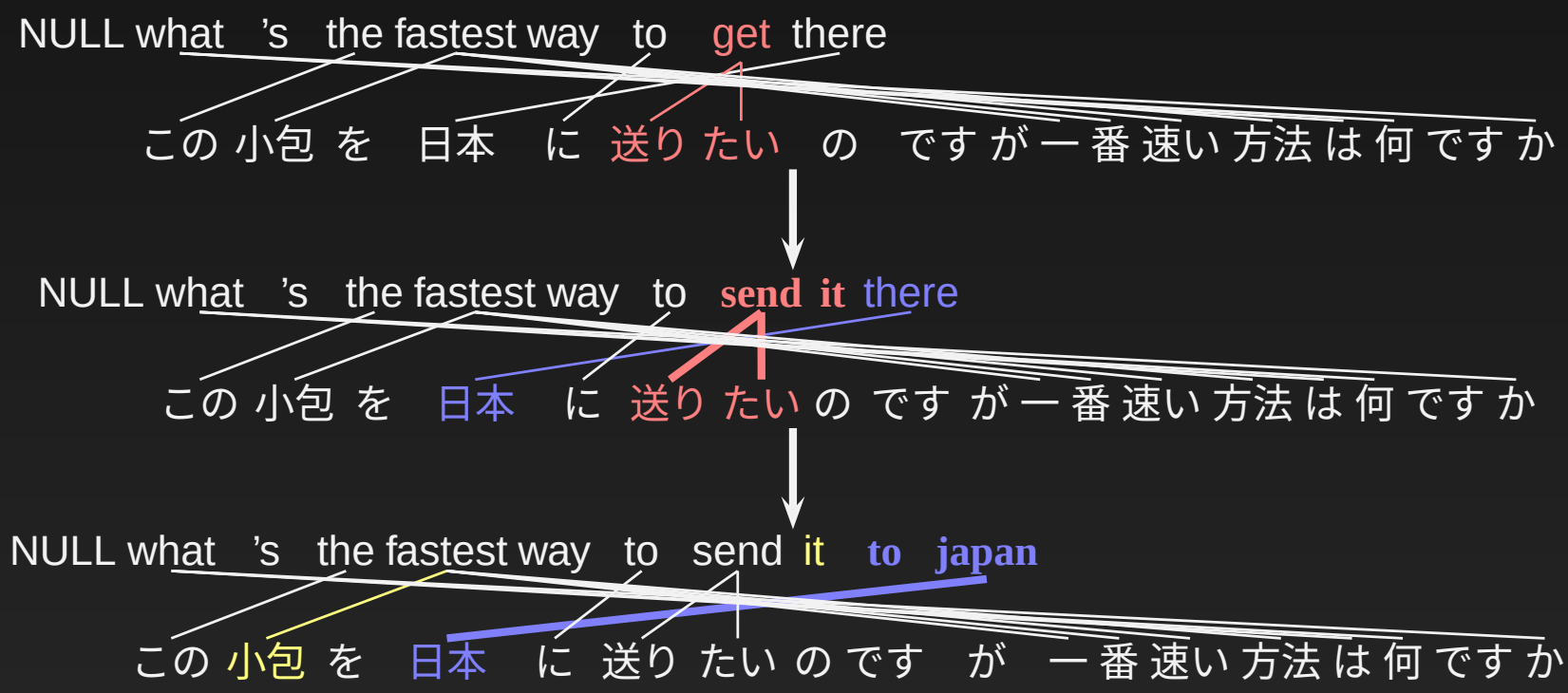
Greedy Decoding (Germann et al., 2001)



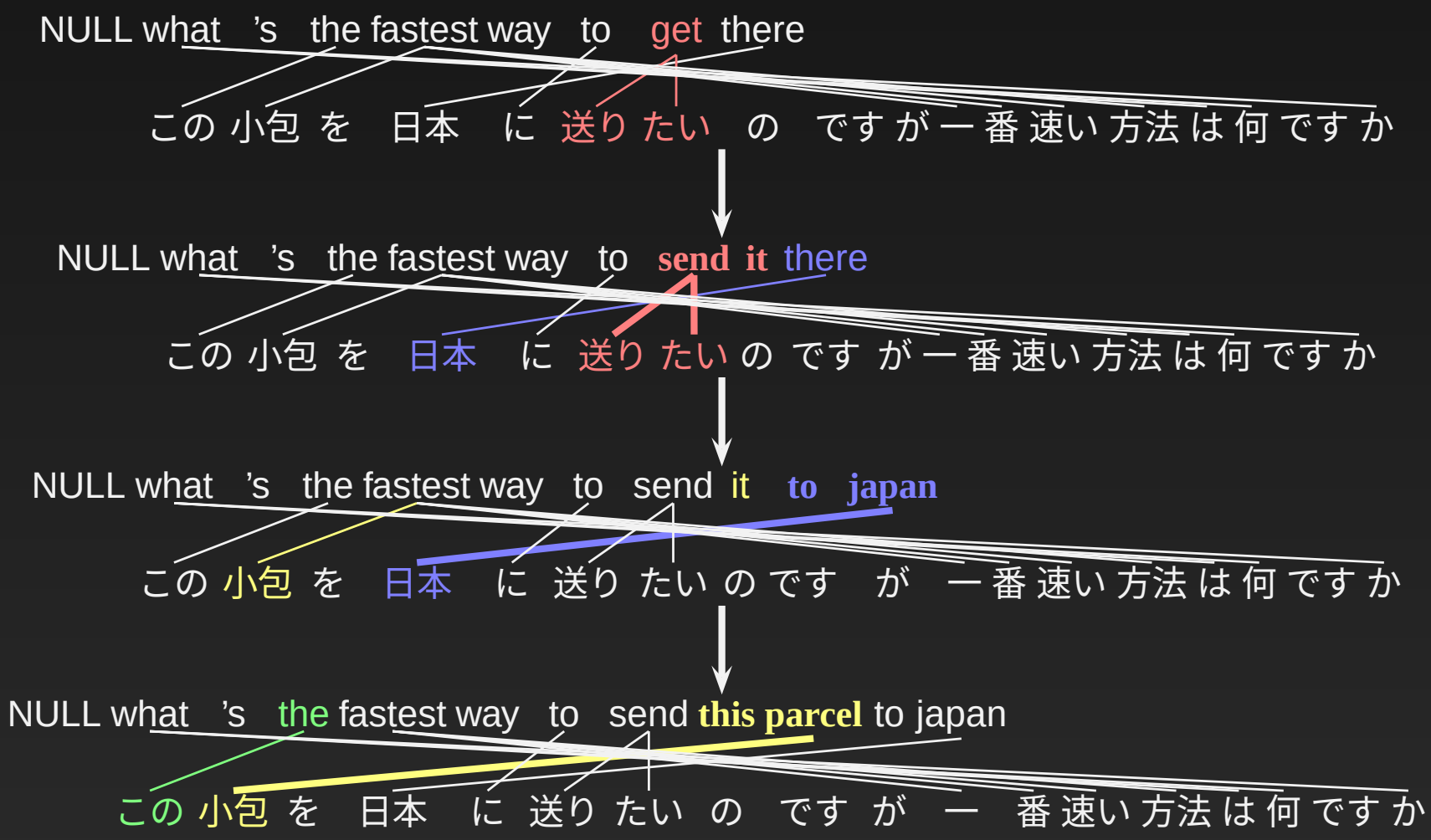
Greedy Decoding (Germann et al., 2001)



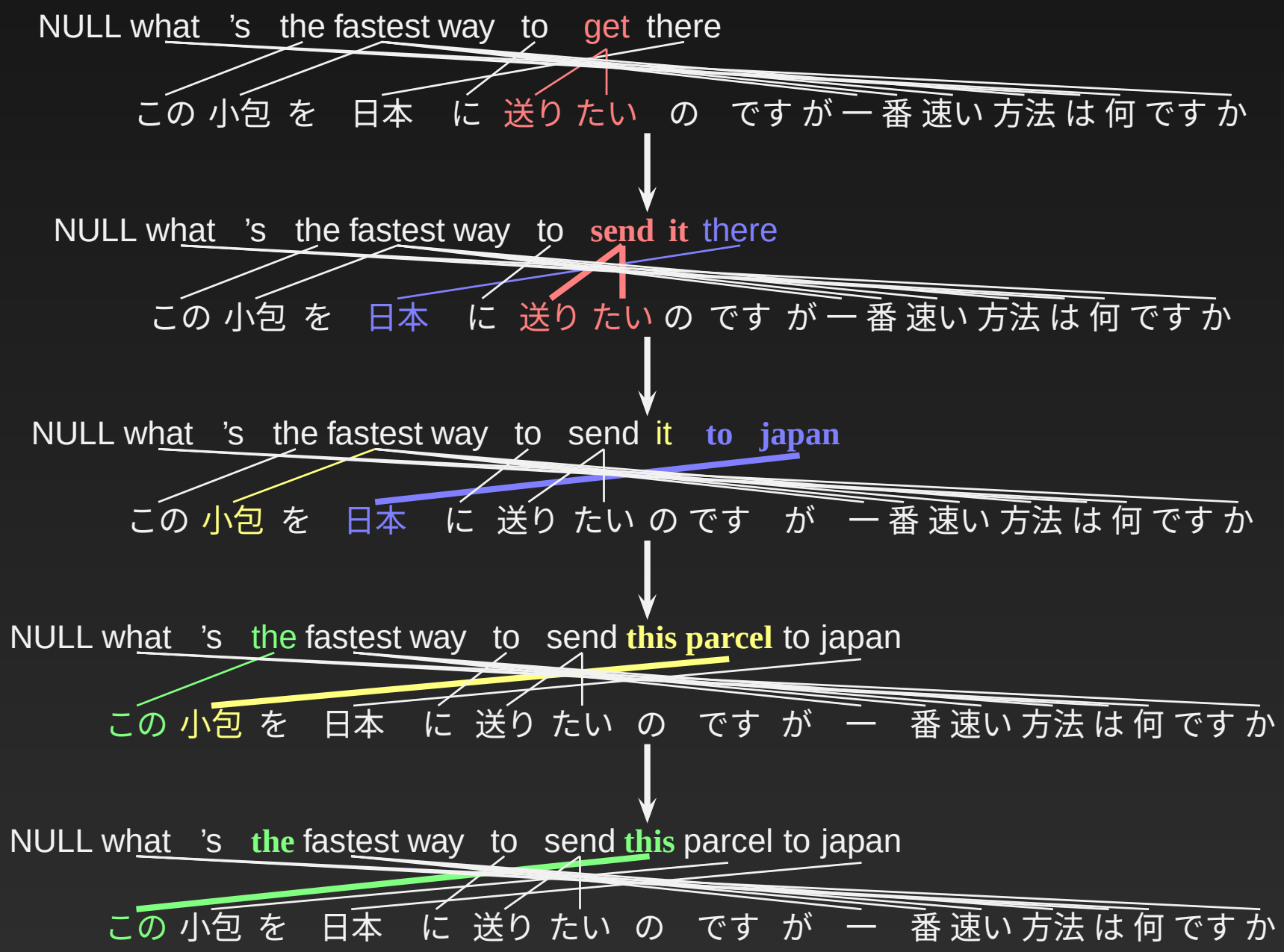
Greedy Decoding (Germann et al., 2001)



Greedy Decoding (Germann et al., 2001)



Greedy Decoding (Germann et al., 2001)



Greedy Decoding の特徴

- 高速
  - ただし、終了の予測が (ほぼ) 不可能

Greedy Decoding の特徴

- 高速
  - ただし、終了の予測が (ほぼ) 不可能
- 複雑な翻訳モデルについてもデコーディングが可能
  - ただし、best な解とは限らない

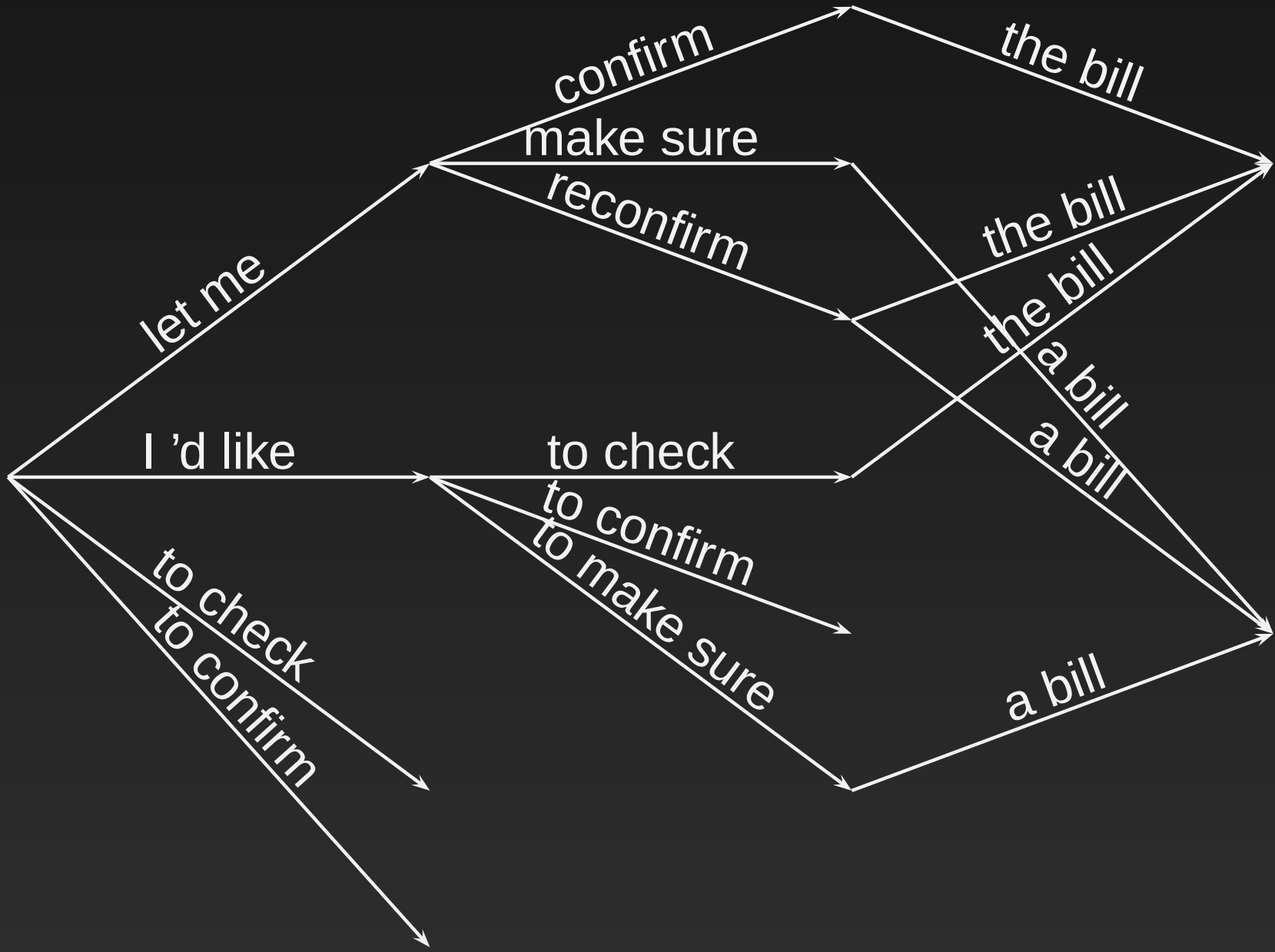
Greedy Decoding の特徴

- 高速
  - ただし、終了の予測が (ほぼ) 不可能
- 複雑な翻訳モデルについてもデコーディングが可能
  - ただし、best な解とは限らない
- 初期値 (種文) の決定
  - あらかじめフレーズ翻訳を抽出、フレーズを結合 (Marcu, 2001)
  - 文用例の検索 (Watanabe and Sumita, 2003a)

Greedy Decoding の特徴

- 高速
  - ただし、終了の予測が (ほぼ) 不可能
- 複雑な翻訳モデルについてもデコーディングが可能
  - ただし、best な解とは限らない
- 初期値 (種文) の決定
  - あらかじめフレーズ翻訳を抽出、フレーズを結合 (Marcu, 2001)
  - 文用例の検索 (Watanabe and Sumita, 2003a)
- 高速化
  - 互いに干渉しない複数の操作を同時に実行 (Germann, 2003)

Word Graph Decoding (Ueffing et al., 2002)



Word Graph Decodingの特徴

- グラフによる簡潔な表現
  - 翻訳候補のマージによる効率よい探索

Word Graph Decodingの特徴

- グラフによる簡潔な表現
  - 翻訳候補のマージによる効率よい探索
- 2-passによるデコーディング
  1. bigramによるビームサーチによりWord graphの生成
  2. trigramによるA*探索によりWord graphを探索、最適解(n-best)の生成

Word Graph Decodingの特徴

- グラフによる簡潔な表現
  - 翻訳候補のマージによる効率よい探索
- 2-passによるデコーディング
  1. bigramによるビームサーチにより Word graph の生成
  2. trigramによる A*探索により Word graph を探索、最適解 (n-best) の生成
- 様々な知識を用い、edgeの生成
  - 単語翻訳
  - フレーズ翻訳
  - Named Entity

他にも...

- Integer Programming (Germann et al., 2001)
  - 各都市 (原言語) にはいくつかのホテル (目的言語) がある
  - 各都市を回り、その中のホテルに滞在する巡回セールスマン問題

他にも...

- Integer Programming (Germann et al., 2001)
  - 各都市 (原言語) にはいくつかのホテル (目的言語) がある
  - 各都市を回り、その中のホテルに滞在する巡回セールスマン問題
- FST によるデコーディング (Kumar and Byrne, 2003)
  - 翻訳モデル/言語モデルの知識を weighted-FST で表現
  - 静的な構造/高速なデコーディング

デコーダのツール

- ISI Machine Translation Decoder
  - http://www.isi.edu/natural-language/people/germann/software/ReWrite-Decoder/
  - Greedy Decoding
  - IBM Model 4
  - フレーズの境界を指定可能

翻訳の評価手法

- 客観評価
  - WER: 単語誤り率
  - PER: 位置独立単語誤り率
  - BLEU: n-gram の適合率の相乗平均 (Papineni et al., 2002)
  - NIST: 情報量で正規化された、n-gram の適合率の相加平均 (Standards and Technology, 2002)
- 主観評価
  - 例、A:perfect、B:fair、C:acceptable、D:nonsense

内容

- 統計的機械翻訳
  - 単語アライメント
  - 翻訳モデル
  - 言語モデル
  - デコーディング
  - 評価手法
- 多言語へのチャレンジ
- 統計的機械翻訳の現状

内容

- 統計的機械翻訳
- 多言語へのチャレンジ
- 統計的機械翻訳の現状

内容

- 統計的機械翻訳
- 多言語へのチャレンジ
- 統計的機械翻訳の現状

内容

- 統計的機械翻訳
- 多言語へのチャレンジ
  - 対訳コーパス
  - 句に基づく翻訳
  - 言語の構造を考慮した翻訳
  - 最適化手法
- 統計的機械翻訳の現状

多言語翻訳へのチャレンジ

- 翻訳モデルは言語対非依存?
  - 非常に近い言語 — 英語、仏語、独語
  - 遠い言語 — 英語 vs 中国語、日本語、韓国語

多言語翻訳へのチャレンジ

- 翻訳モデルは言語対非依存?
  - 非常に近い言語 — 英語、仏語、独語
  - 遠い言語 — 英語 vs 中国語、日本語、韓国語
- 言語学的に異なる言語対への対応
  - 単語 ⟶ 句
  - 構造をとらえる翻訳モデル

多言語翻訳へのチャレンジ

- 翻訳モデルは言語対非依存?
  - 非常に近い言語 — 英語、仏語、独語
  - 遠い言語 — 英語 vs 中国語、日本語、韓国語
- 言語学的に異なる言語対への対応
  - 単語 $\longrightarrow$ 句
  - 構造をとらえる翻訳モデル
- 対訳データ
  - 文単位 (あるいはセグメント単位) の対応が必要
  - そのためには.... ドキュメント単位の対応が必要

対訳データ

- 様々なリソース
  - 新聞記事 (読売新聞、Xinhua、Hong Kong News)
  - 多言語でニュースを提供している Web(!?)
  - LDC etc.

対訳データ

- 様々なリソース
  - 新聞記事 (読売新聞、Xinhua、Hong Kong News)
  - 多言語でニュースを提供している Web(!?)
  - LDC etc.
- ツール
  - Gale and Church (1993)
    (http://www.research.att.com/~kwc/publications.html)
  - Melamed (1996)
    (http://www.cs.nyu.edu/~melamed/GMA/docs/README.htm)
  - etc.

対訳データ

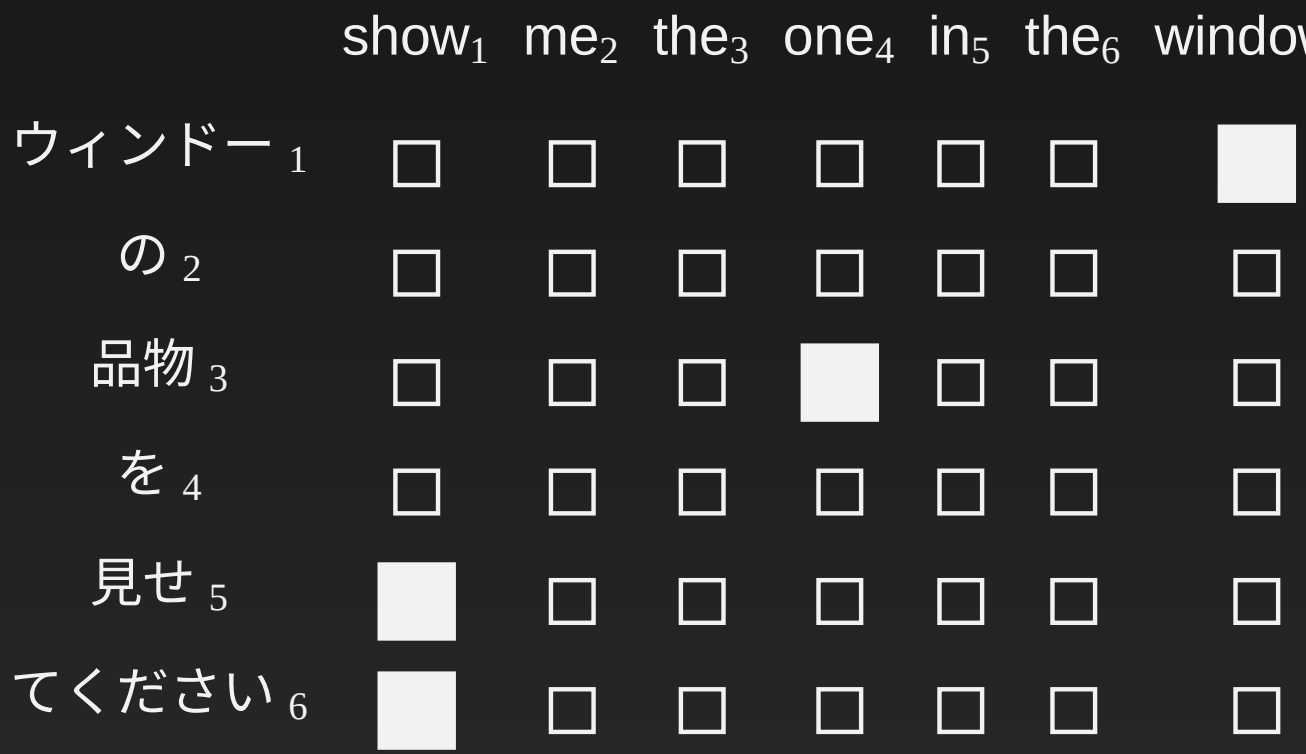
- 様々なリソース
  - 新聞記事 (読売新聞、Xinhua、Hong Kong News)
  - 多言語でニュースを提供している Web(!?)
  - LDC etc.
- ツール
  - Gale and Church (1993)
    (http://www.research.att.com/~kwc/publications.html)
  - Melamed (1996)
    (http://www.cs.nyu.edu/~melamed/GMA/docs/README.htm)
  - etc.
- 日英対訳コーパス
  - 日英新聞記事対応付けデータ
    (http://www2.crl.go.jp/jt/a132/members/mutiyama/jea/index.html)
  - BTEC (ATR、非公開)

フレーズに基づく翻訳モデル

“all the two-word combinations as if they were single words.”

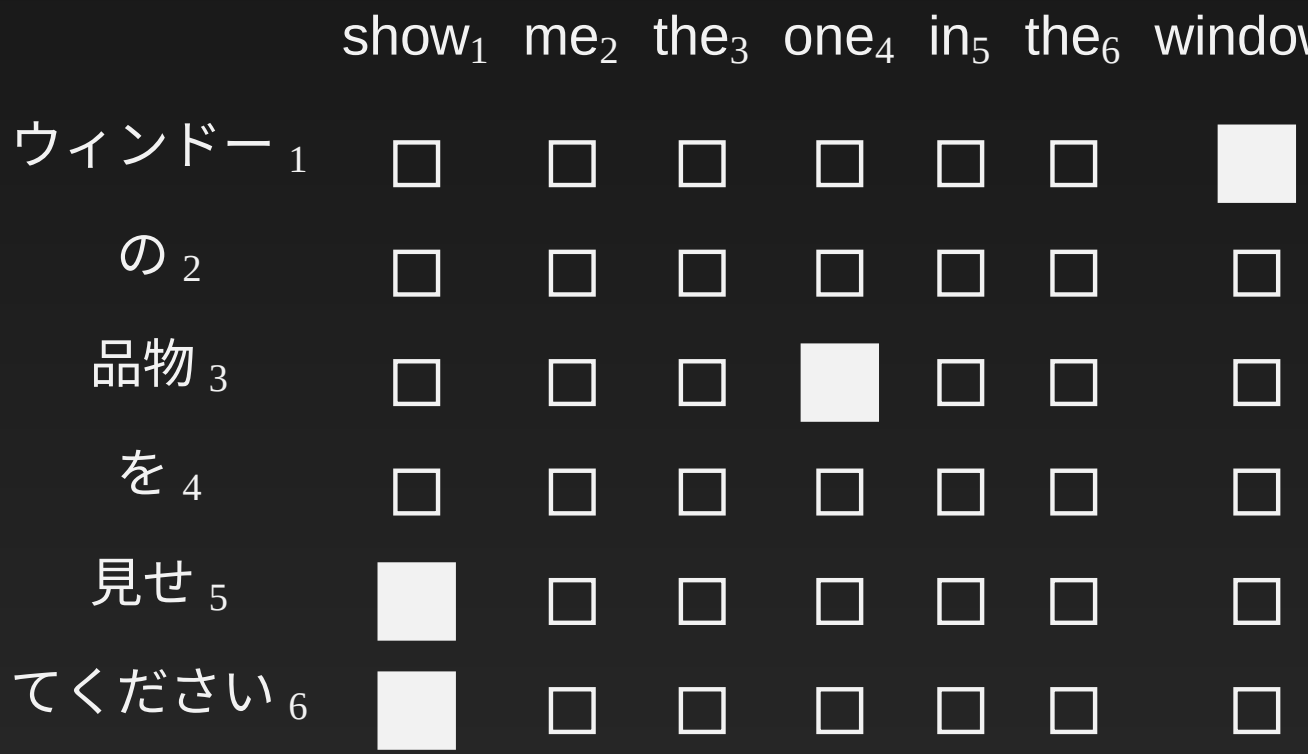
フレーズに基づく翻訳モデル

“all the two-word combinations as if they were single words.”



フレーズに基づく翻訳モデル

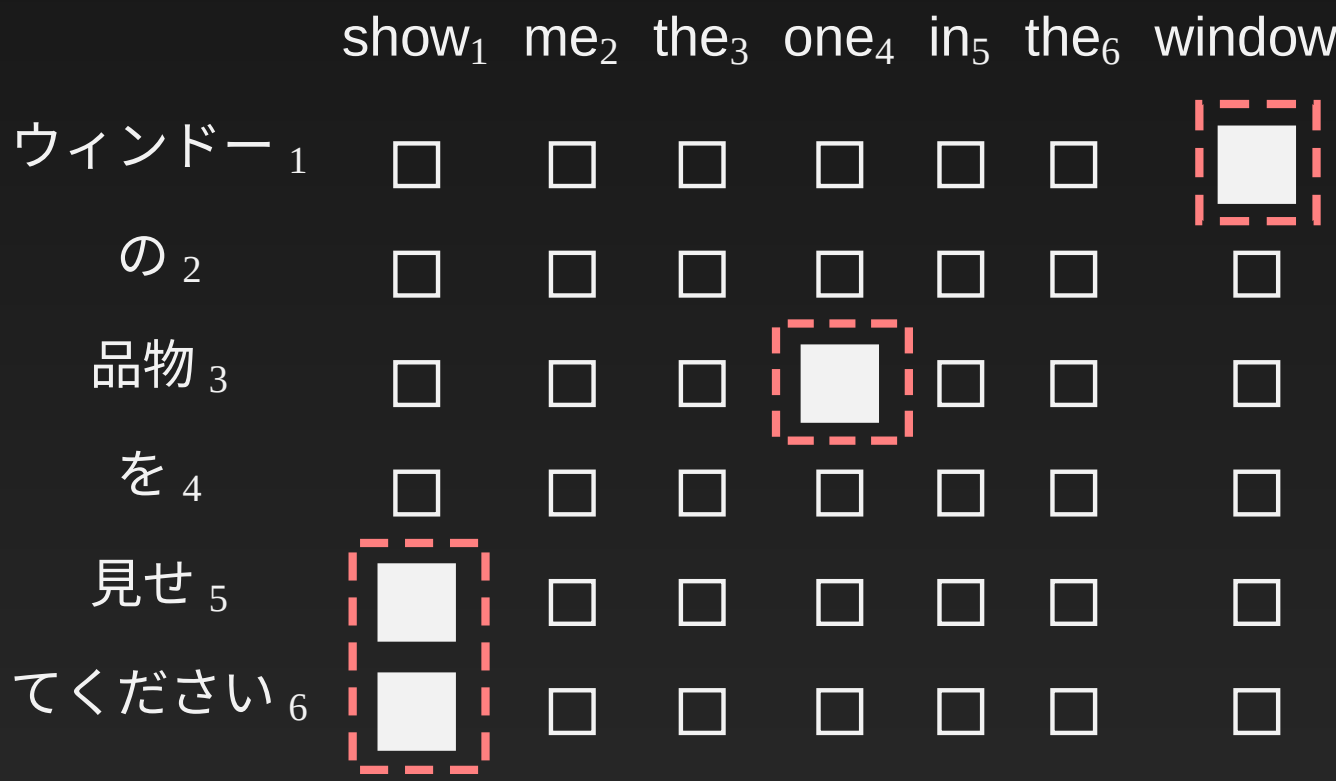
“all the two-word combinations as if they were single words.”



- フレーズの抽出

フレーズに基づく翻訳モデル

“all the two-word combinations as if they were single words.”

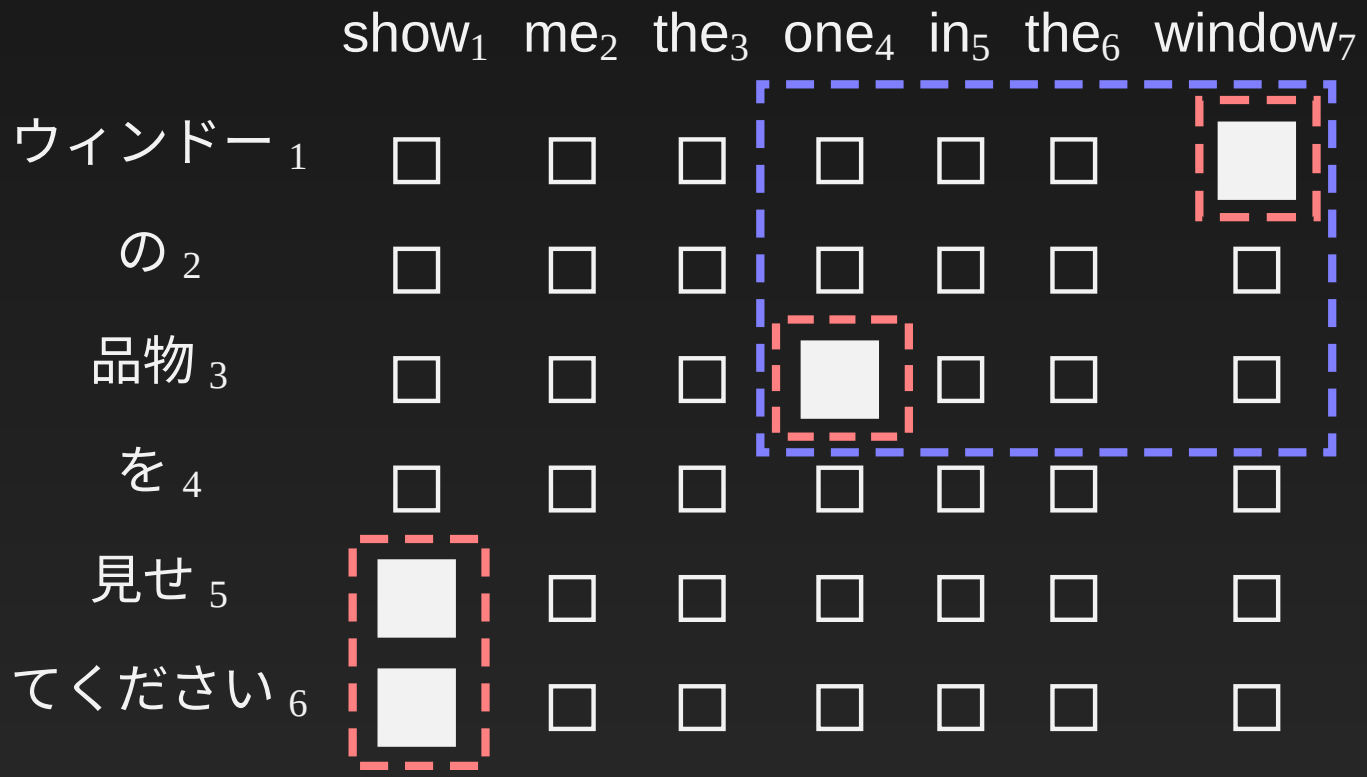


- フレーズの抽出

  ウィンドー ↔ window　品物 ↔ one　見せ てください ↔ show

フレーズに基づく翻訳モデル

“all the two-word combinations as if they were single words.”



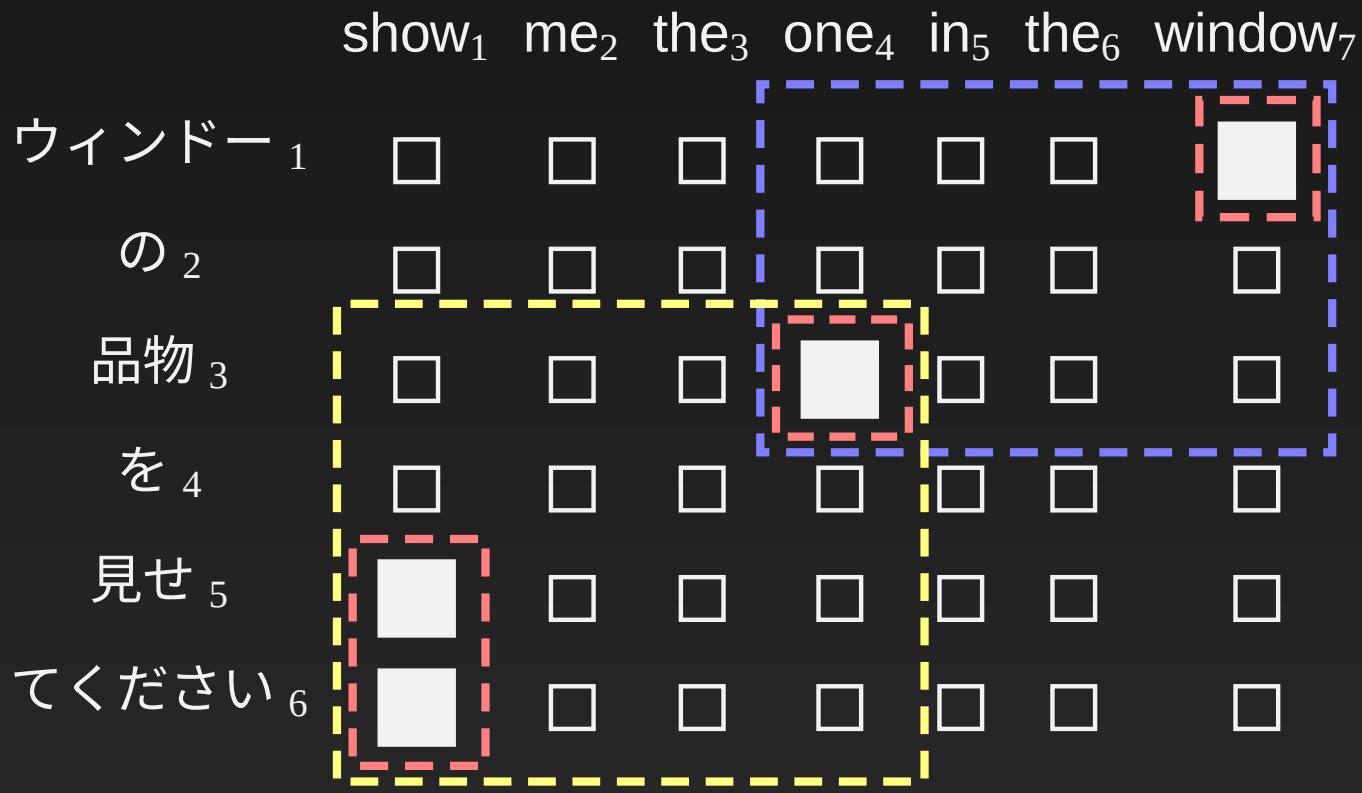
- フレーズの抽出

ウィンドー ↔ window　品物 ↔ one　見せ てください ↔ show

ウィンドー の 品物 ↔ one in the window ?

フレーズに基づく翻訳モデル

“all the two-word combinations as if they were single words.”



- フレーズの抽出

ウィンドー ↔ window　品物 ↔ one　見せ てください ↔ show

ウィンドー の 品物 ↔ one in the window ?

品物 を 見せ てください ↔ show me the one ?

Phrase Transaltion (Koehn et al., 2003; Vogel et al., 2003)

- $J \longrightarrow E$、$E \longrightarrow J$ の両方向のトレーニング
- Intersection により信頼度の高い単語アライメント
- 連続している句の抽出

Phrase Transaltion (Koehn et al., 2003; Vogel et al., 2003)

- $J \longrightarrow E$、$E \longrightarrow J$ の両方向のトレーニング
- Intersection により信頼度の高い単語アライメント
- 連続している句の抽出
- 頻度による確率値 (Koehn et al., 2003)

$$P(\bar{J}|\bar{E}) = \frac{\mathbf{count}(\bar{J}, \bar{E})}{\sum_{\bar{J}} \mathbf{count}(\bar{J}, \bar{E})}$$

Phrase Transaltion (Koehn et al., 2003; Vogel et al., 2003)

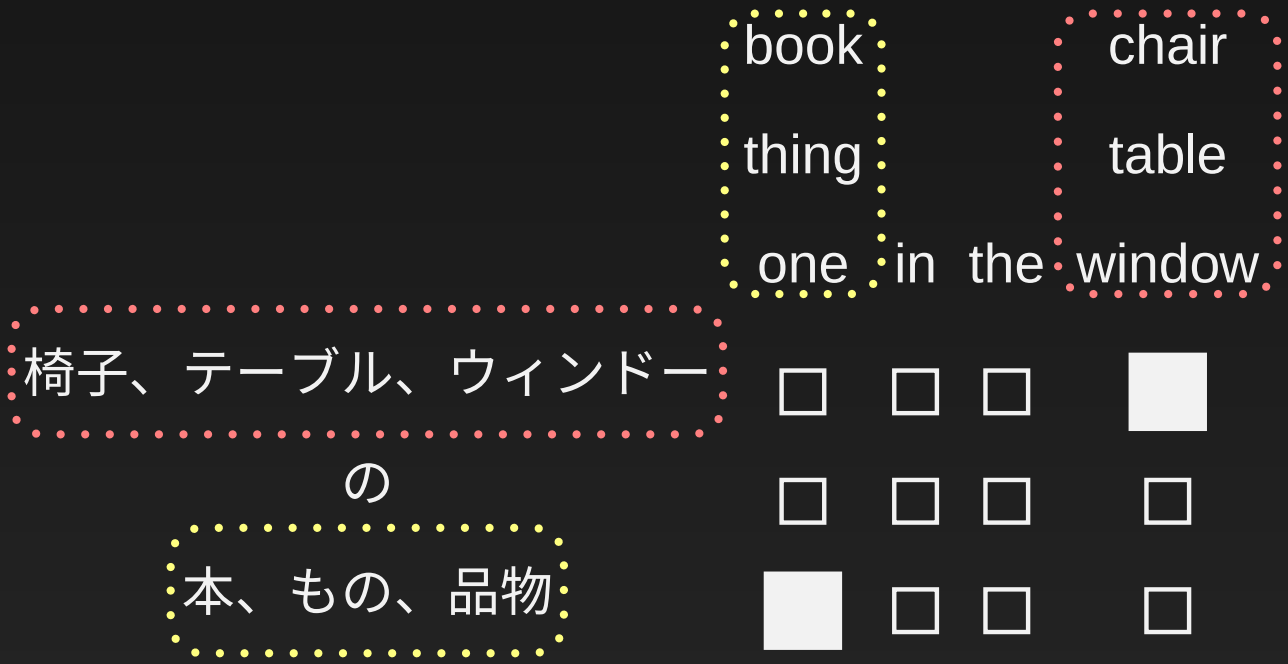
- $J \longrightarrow E$、$E \longrightarrow J$ の両方向のトレーニング
- Intersection により信頼度の高い単語アライメント
- 連続している句の抽出
- 頻度による確率値 (Koehn et al., 2003)

$$P(\bar{J}|\bar{E}) = \frac{\mathbf{count}(\bar{J}, \bar{E})}{\sum_{\bar{J}} \mathbf{count}(\bar{J}, \bar{E})}$$

- Lexicon モデルによる確率値 (Vogel et al., 2003)

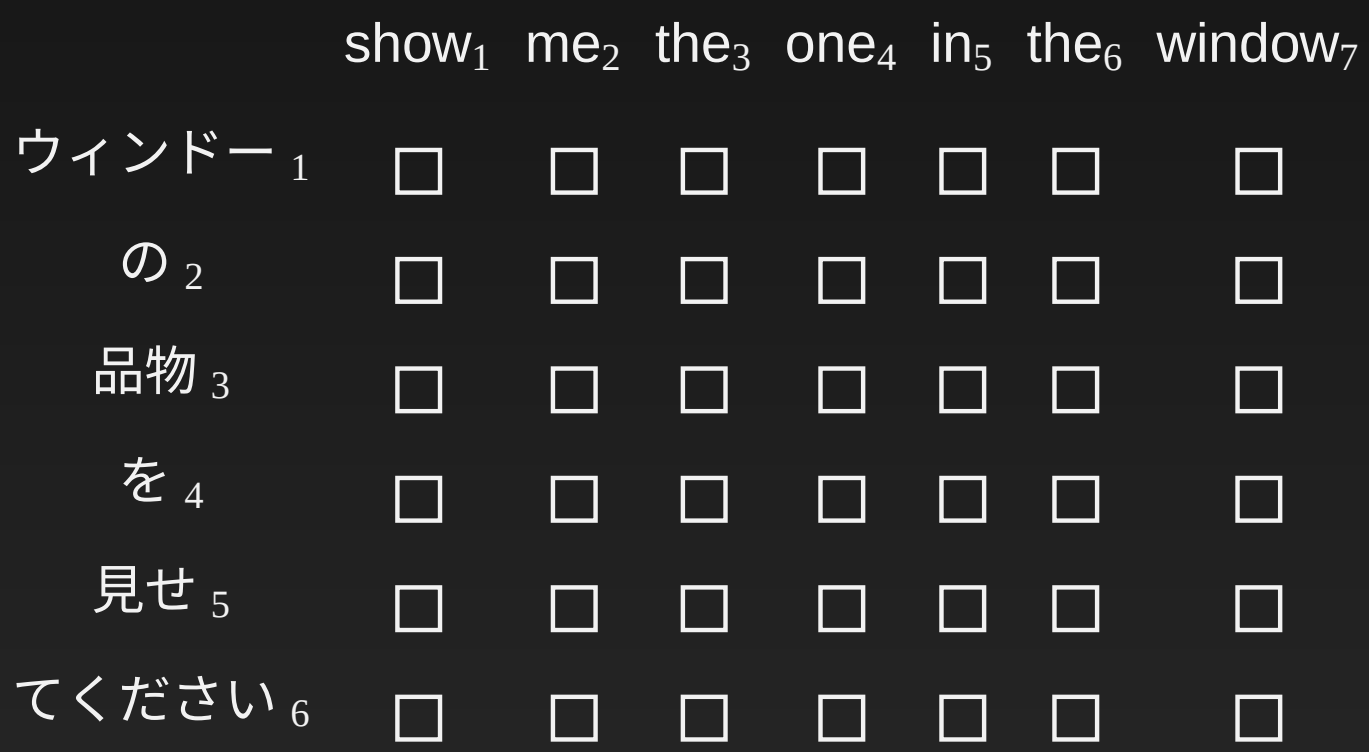
$$P(\bar{J}|\bar{E}) = \prod_{j} \sum_{i} t(\bar{J}_j|\bar{E}_i)$$

Alignment Template (Och et al., 1999)



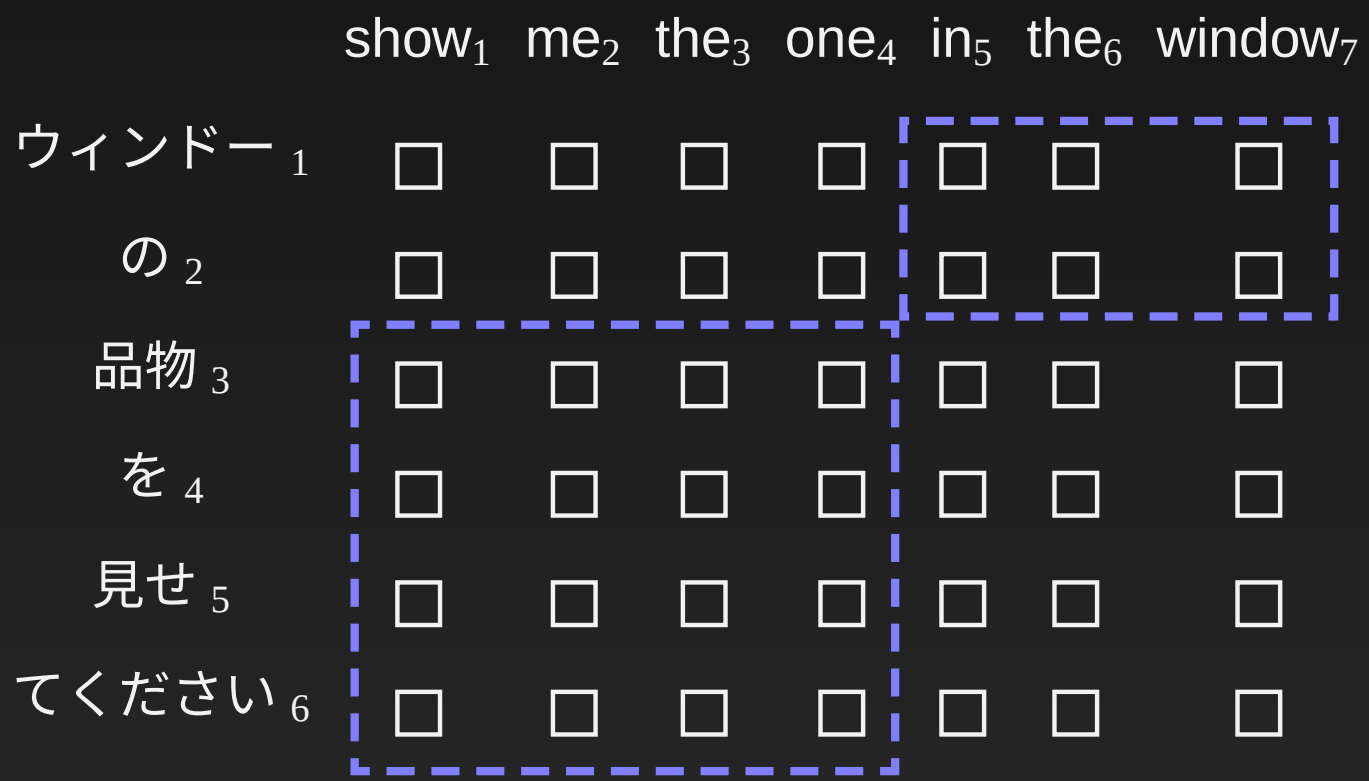
- 単語クラスによる一般化
- 頻度による確率値
- デコーディング時にインスタンシエーション

Phrase Induction (Tillmann, 2003; Marcu and Wong, 2002)



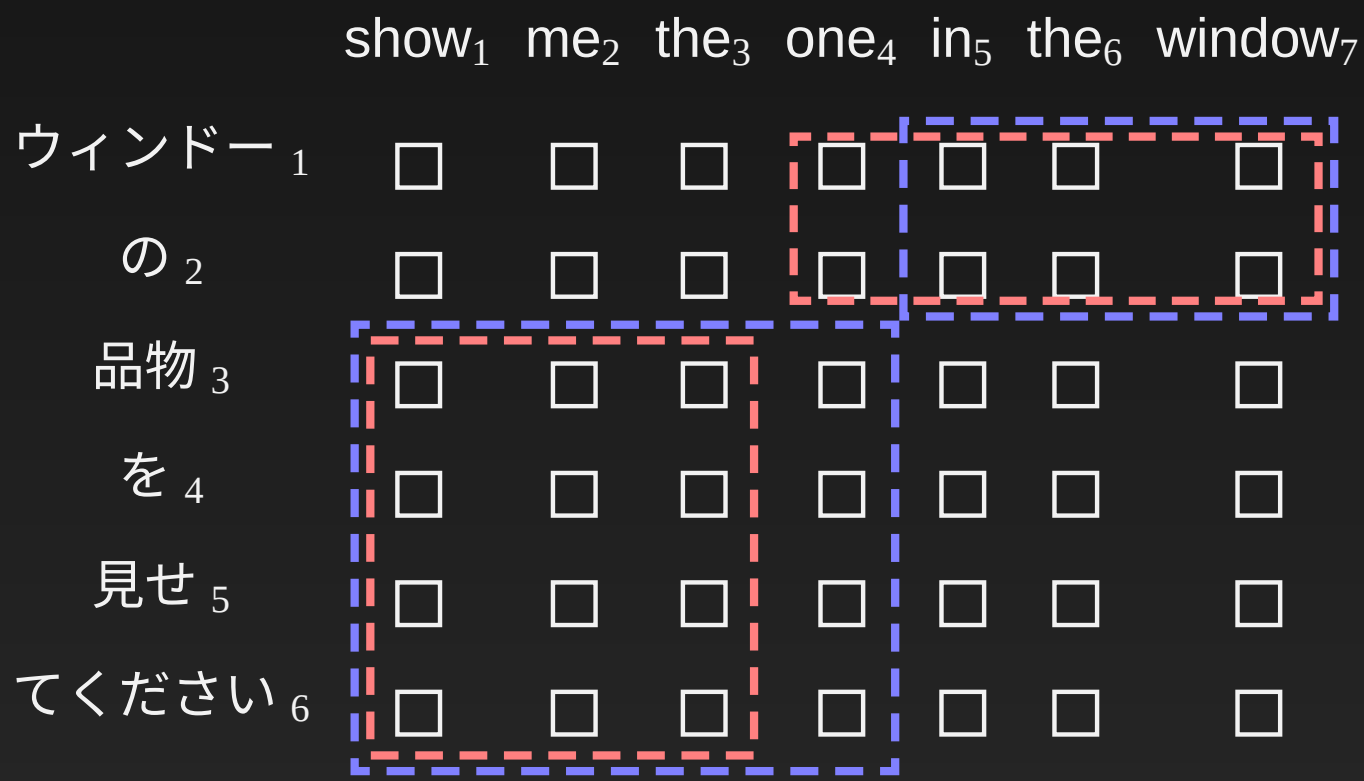
- 単語アライメントは信頼できない

Phrase Induction (Tillmann, 2003; Marcu and Wong, 2002)



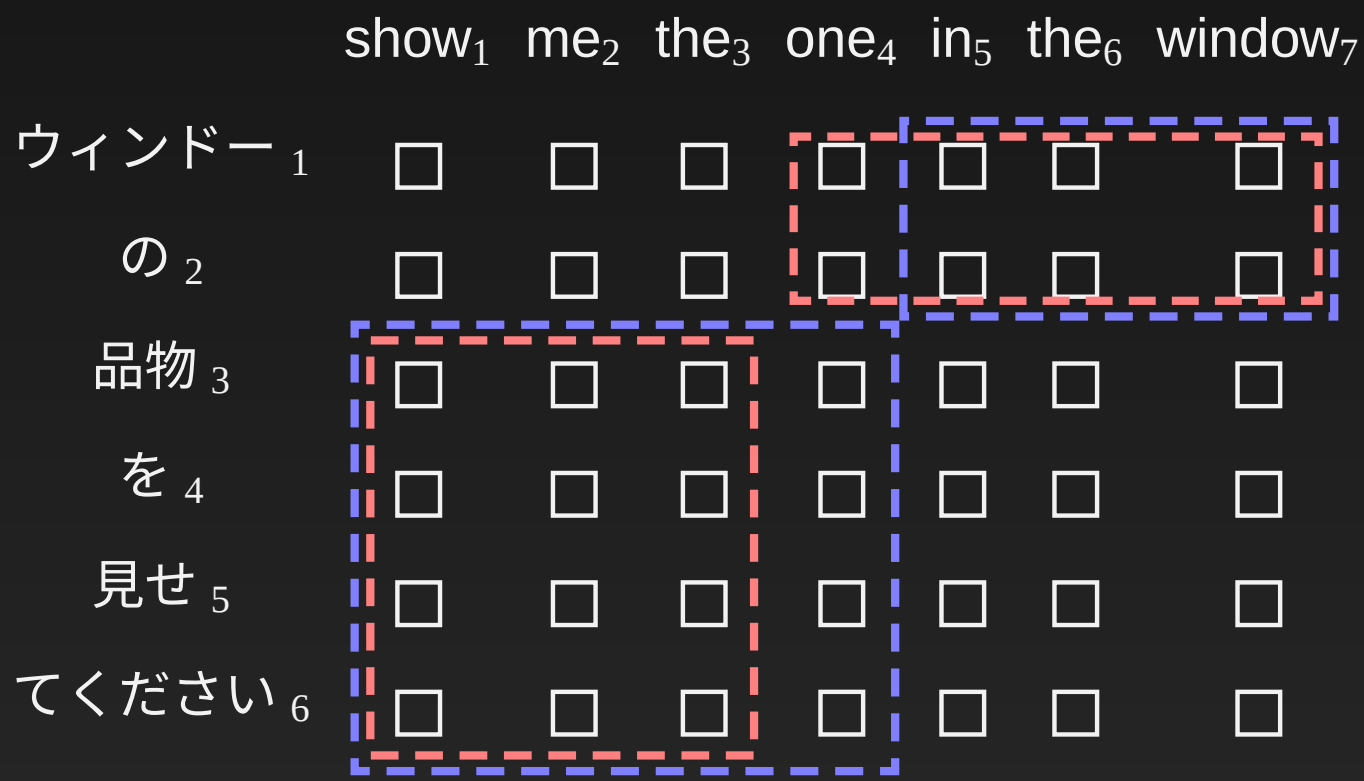
- 単語アライメントは信頼できない

Phrase Induction (Tillmann, 2003; Marcu and Wong, 2002)



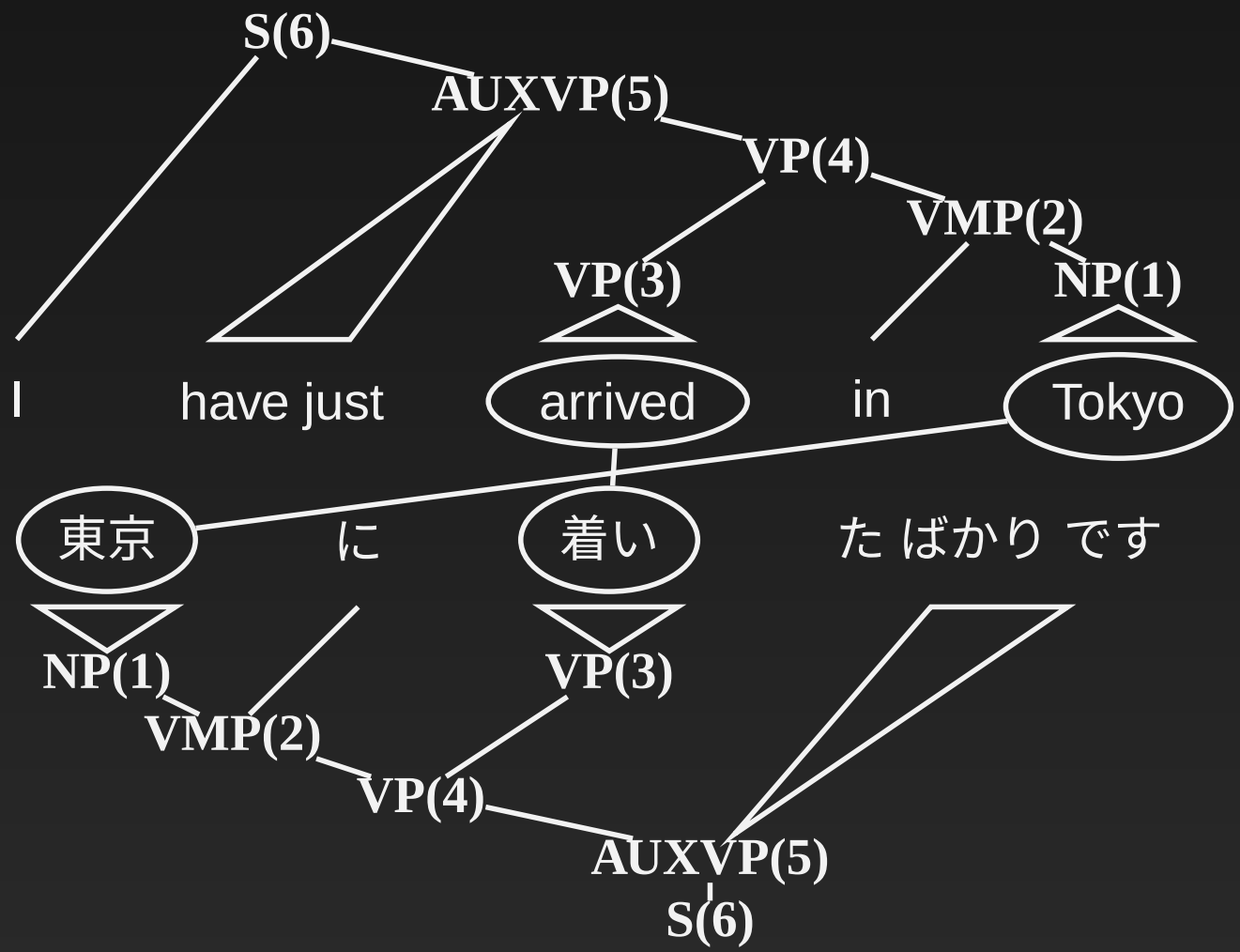
- 単語アライメントは信頼できない

Phrase Induction (Tillmann, 2003; Marcu and Wong, 2002)



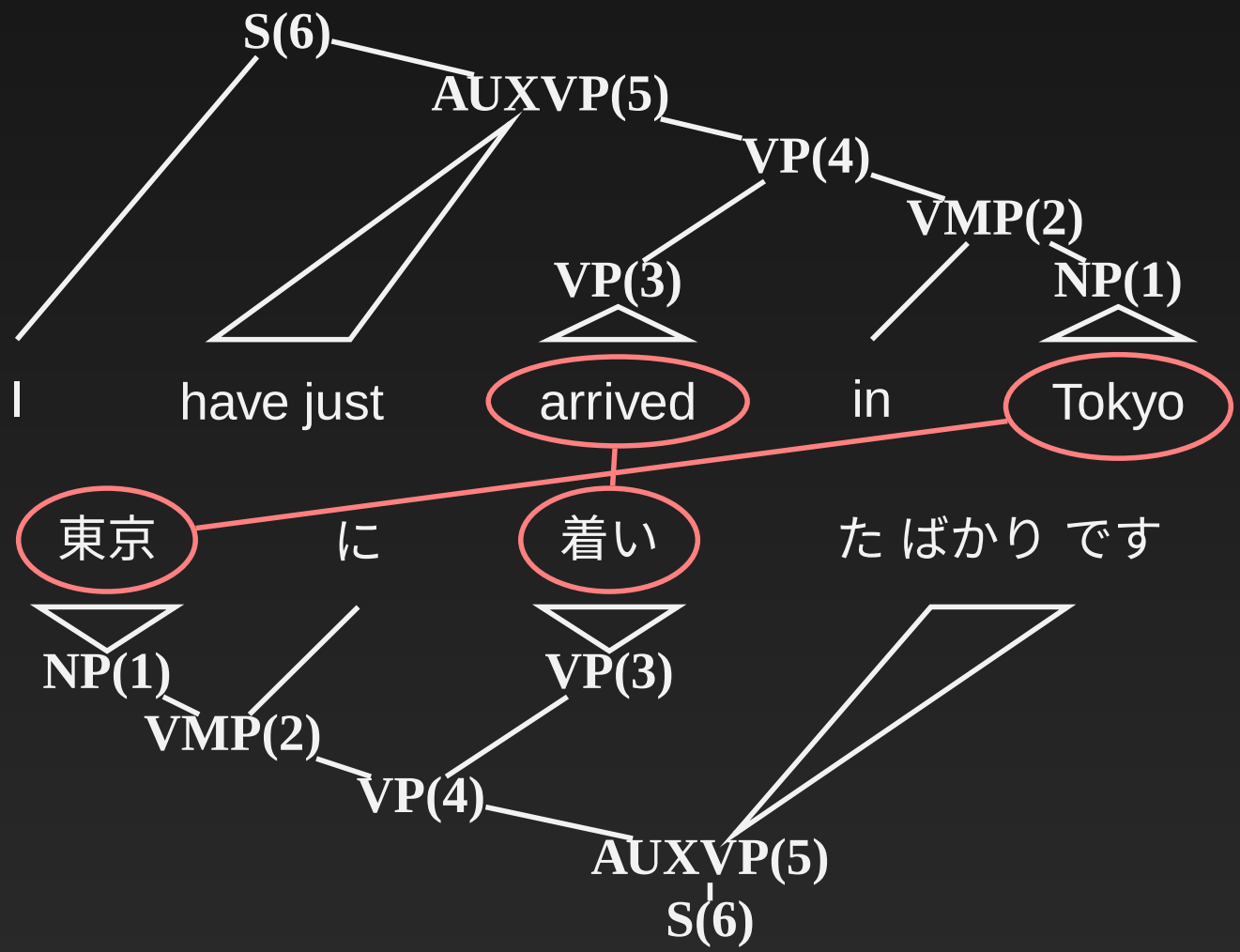
- 単語アライメントは信頼できない
- 最初からブロックによる対応関係の抽出
  - EM Algorithm (Marcu and Wong, 2002)
  - Projection Extention (Tillmann, 2003)

Syntactical Phrase (Watanabe et al., 2002; Koehn et al., 2003



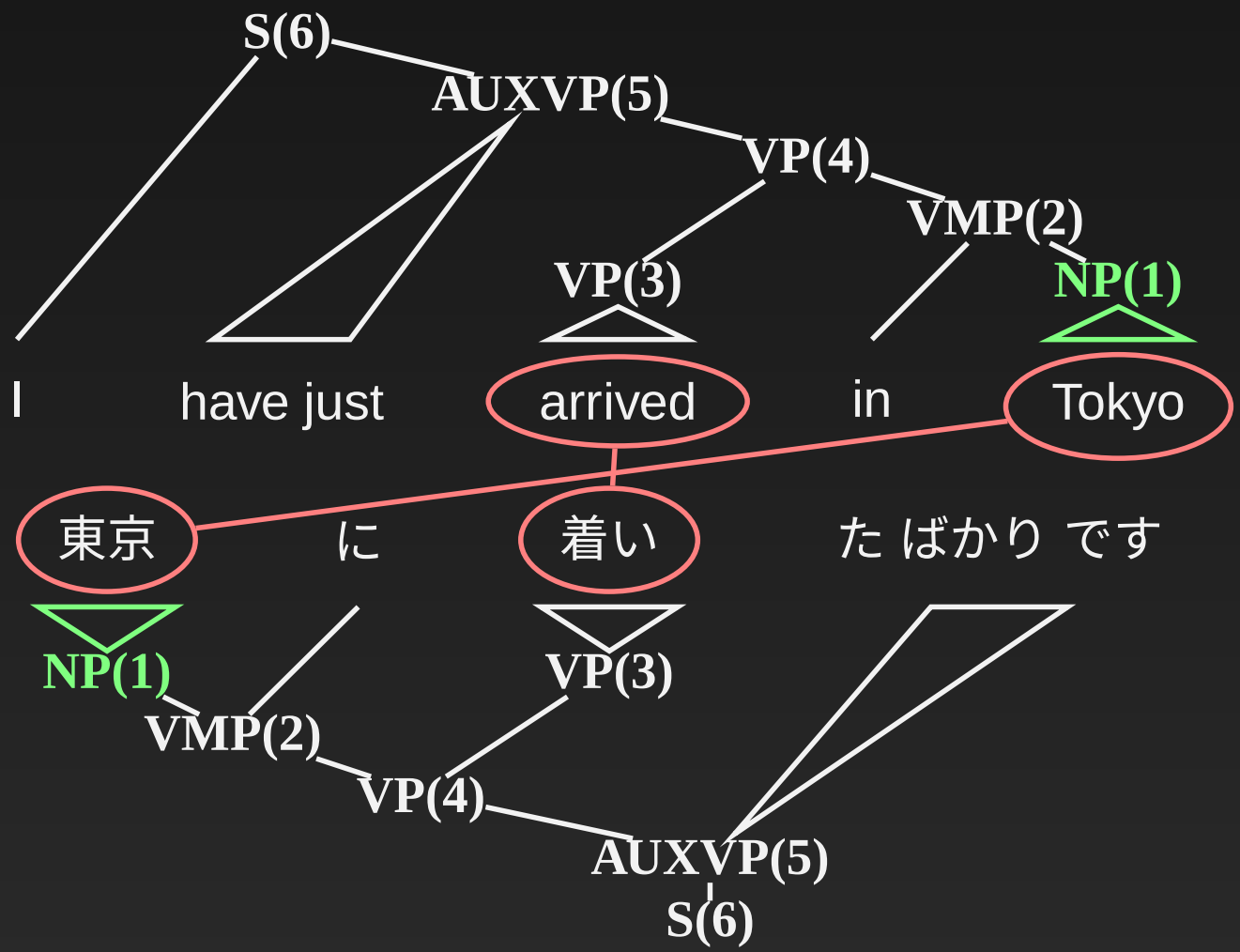
- 階層的句アライメントにより句単位の対応を抽出

Syntactical Phrase (Watanabe et al., 2002; Koehn et al., 2003



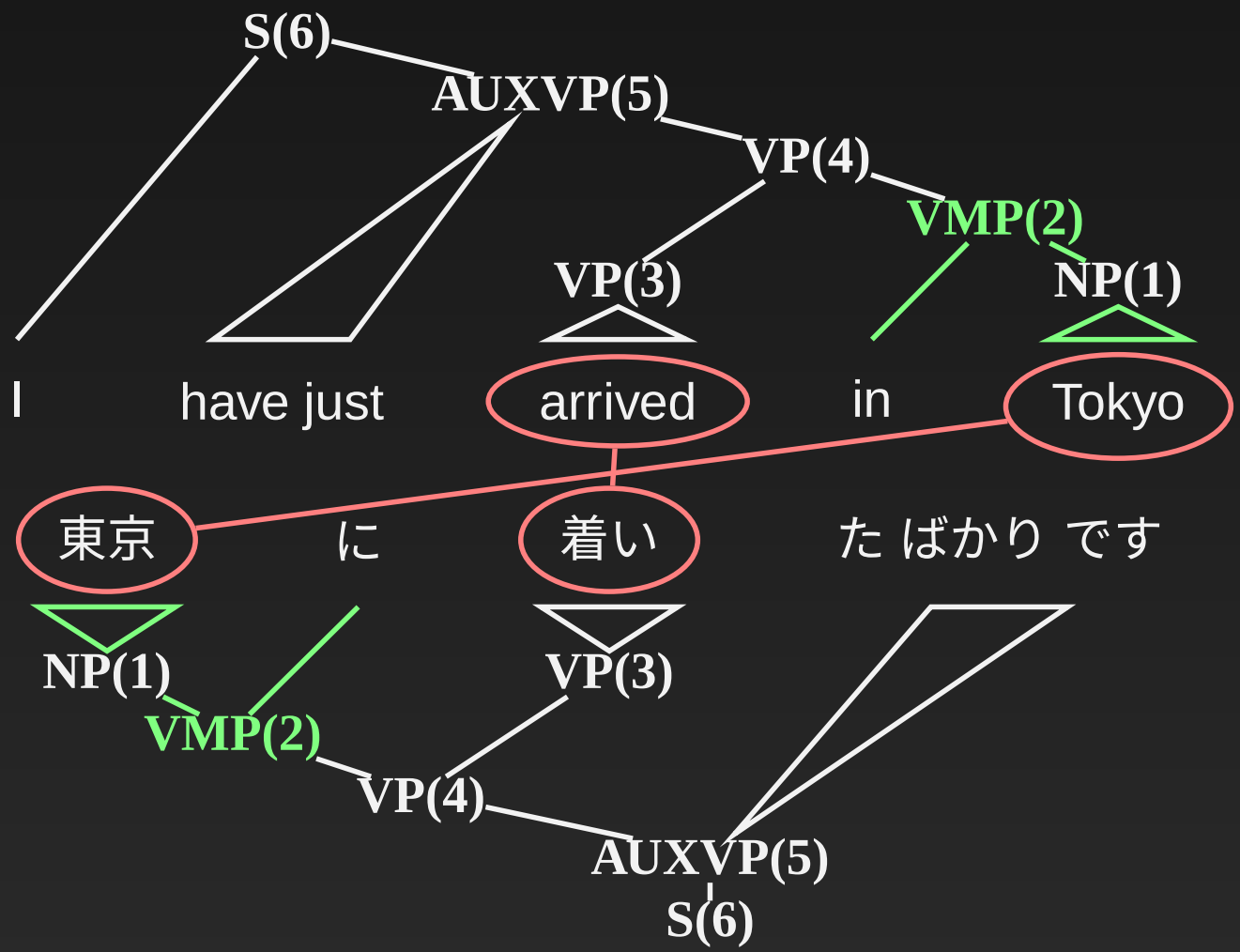
- 階層的句アライメントにより句単位の対応を抽出

Syntactical Phrase (Watanabe et al., 2002; Koehn et al., 2003



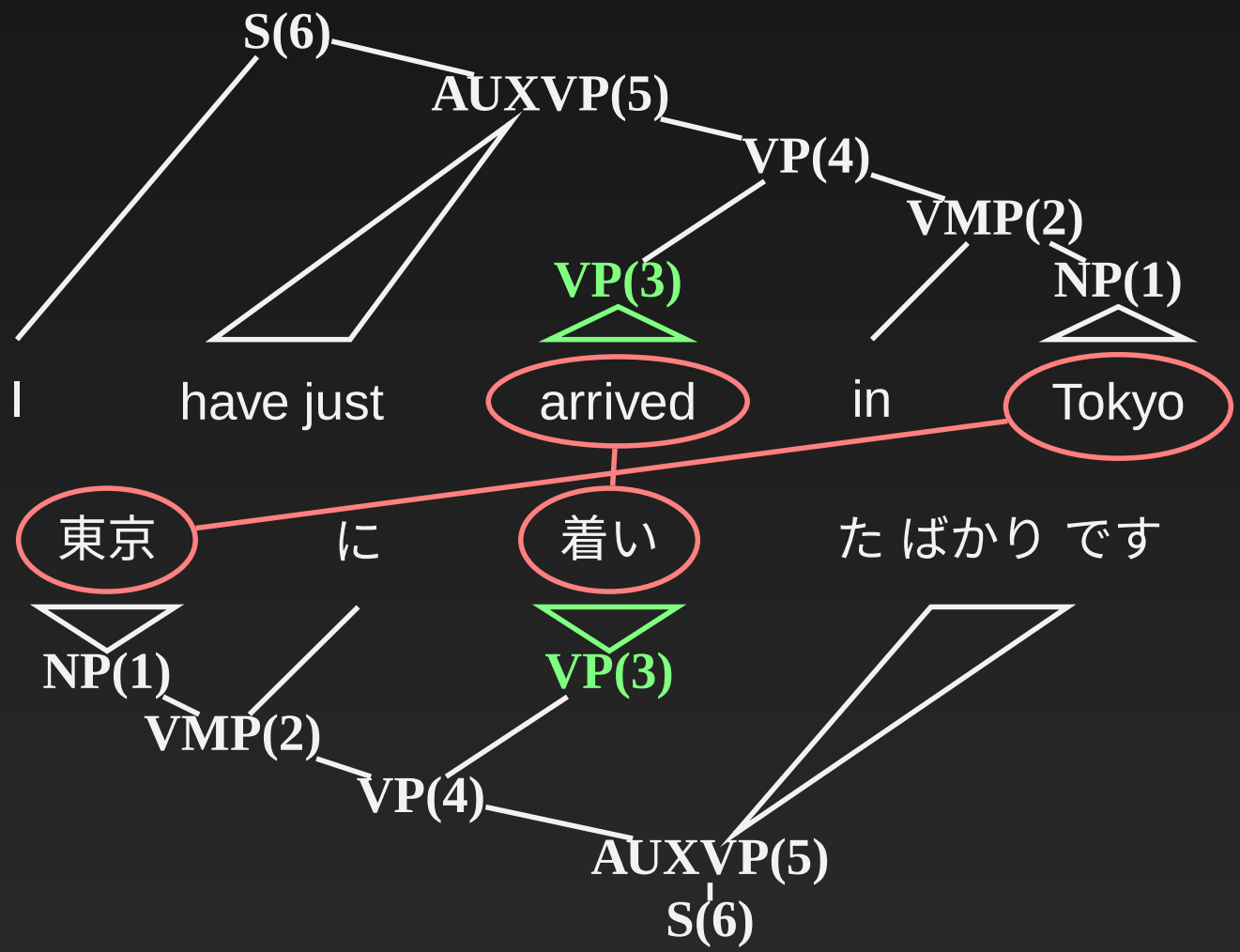
- 階層的句アライメントにより句単位の対応を抽出

Syntactical Phrase (Watanabe et al., 2002; Koehn et al., 2003



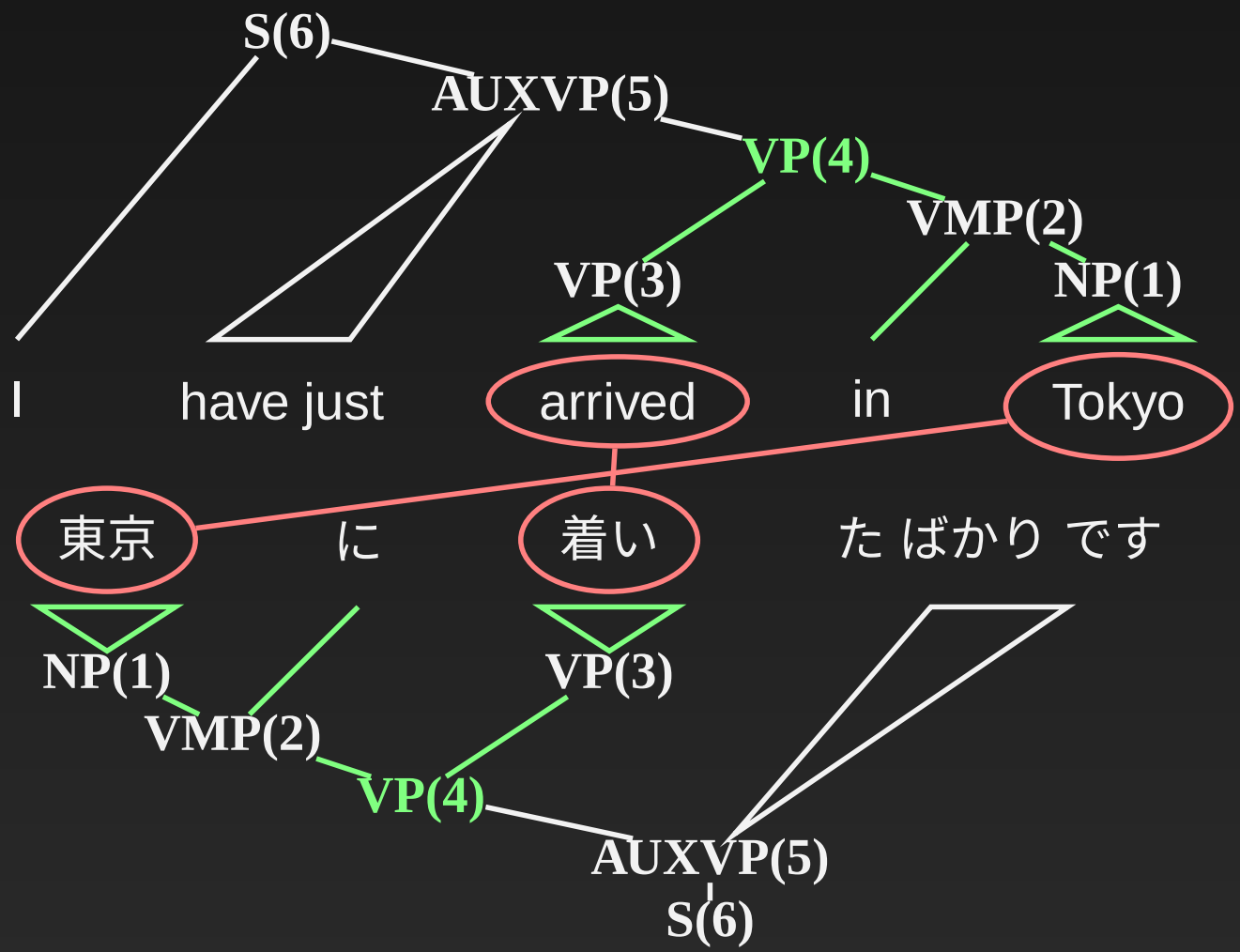
- 階層的句アライメントにより句単位の対応を抽出

Syntactical Phrase (Watanabe et al., 2002; Koehn et al., 2003



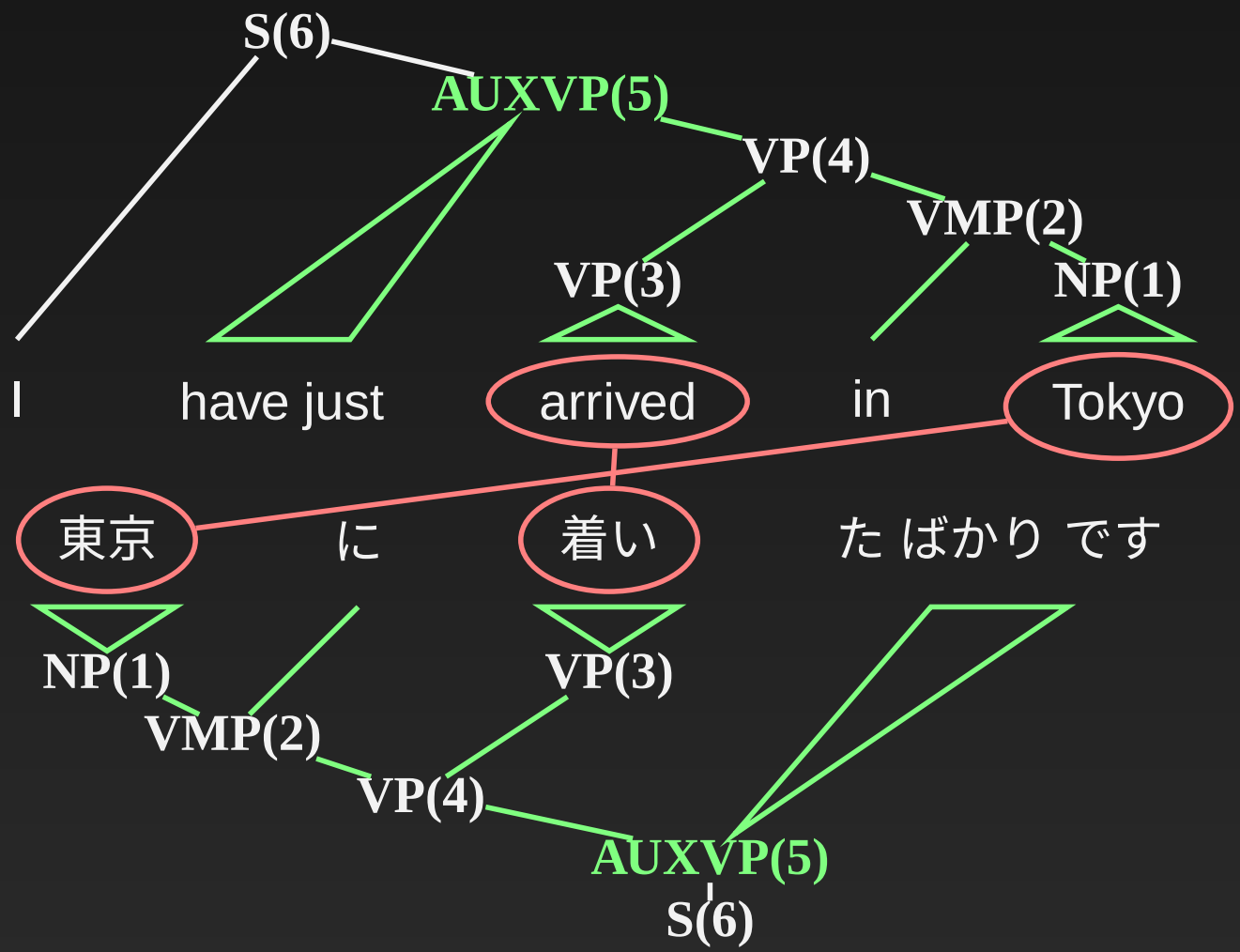
- 階層的句アライメントにより句単位の対応を抽出

Syntactical Phrase (Watanabe et al., 2002; Koehn et al., 2003



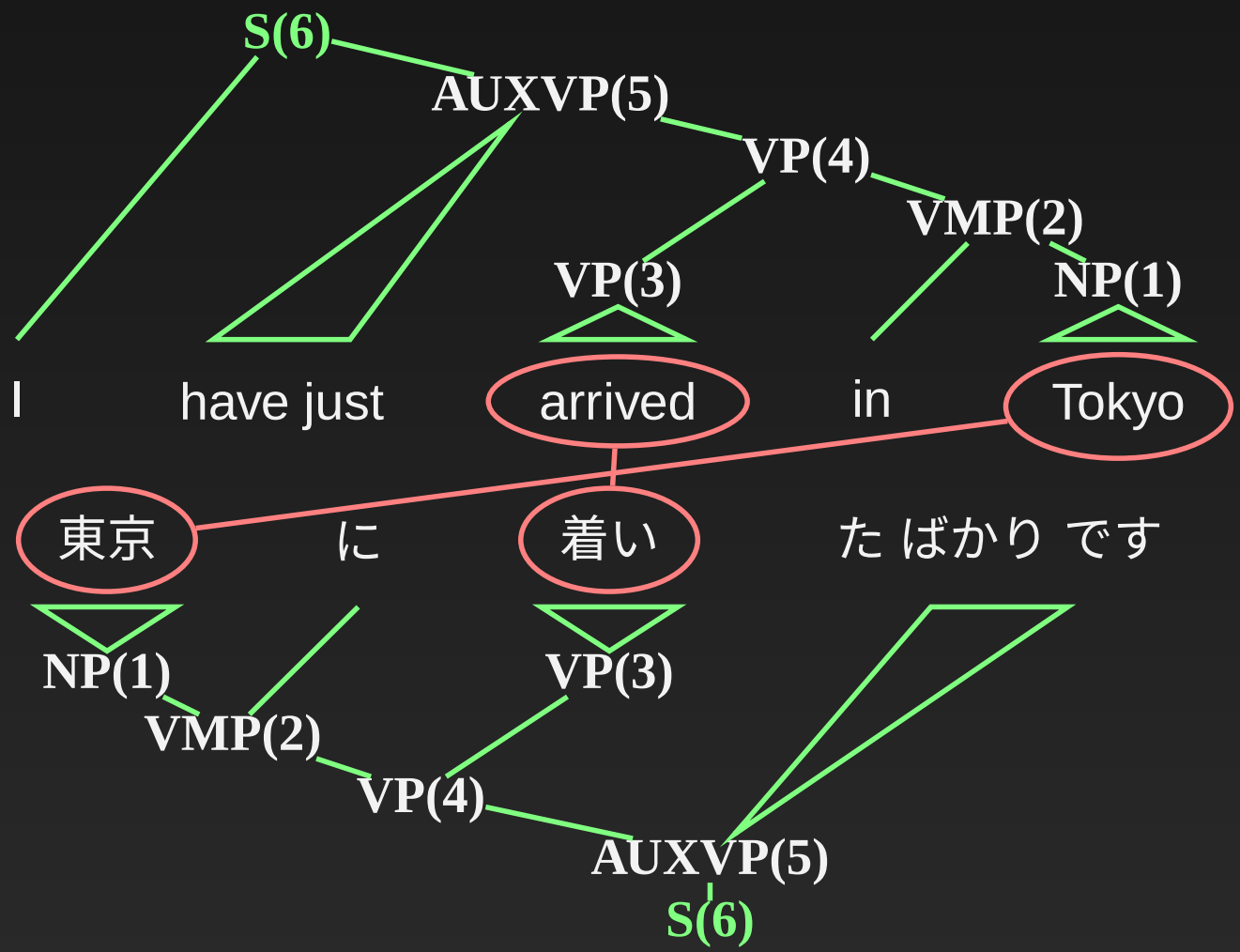
- 階層的句アライメントにより句単位の対応を抽出

Syntactical Phrase (Watanabe et al., 2002; Koehn et al., 2003



- 階層的句アライメントにより句単位の対応を抽出

Syntactical Phrase (Watanabe et al., 2002; Koehn et al., 2003



- 階層的句アライメントにより句単位の対応を抽出

フレーズに基づく翻訳モデルの問題点

“The vocabulary is still only four million”

フレーズに基づく翻訳モデルの問題点

“The vocabulary is still only four million”

- 句の抽出手法
- スパースネス
- 確率値の割り当て
- アライメントのモデル化

Named Entity との統合 (Vogel et al., 2003)

- 統計的機械翻訳では、トークンの意味を考えない...
- 語彙が爆発的に増大

Named Entity との統合 (Vogel et al., 2003)

- 統計的機械翻訳では、トークンの意味を考えない...
- 語彙が爆発的に増大 ⟶ NE の導入
  - 3 月 15 日 ⟶ $DATE
  - 渡辺 太郎 ⟶ $PERSON
  - ATR 音声言語コミュニケーション研究所 ⟶ $ORGANIZATION

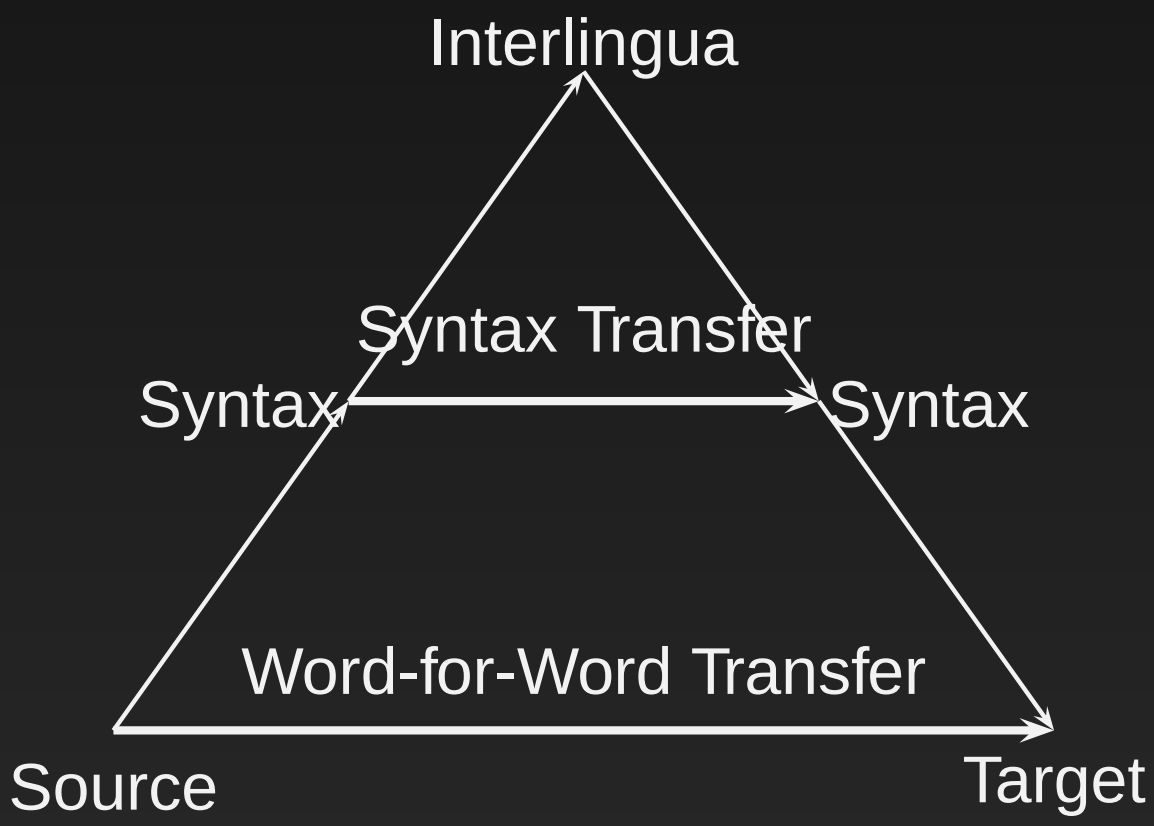
Named Entity との統合 (Vogel et al., 2003)

- 統計的機械翻訳では、トークンの意味を考えない...
- 語彙が爆発的に増大 ⟶ NE の導入
  - 3 月 15 日 ⟶ $DATE
  - 渡辺 太郎 ⟶ $PERSON
  - ATR 音声言語コミュニケーション研究所 ⟶ $ORGANIZATION
- カタカナ語の Transliteration
  - ジョン ⟶ John
  - セバスチャン ⟶ Sebastien

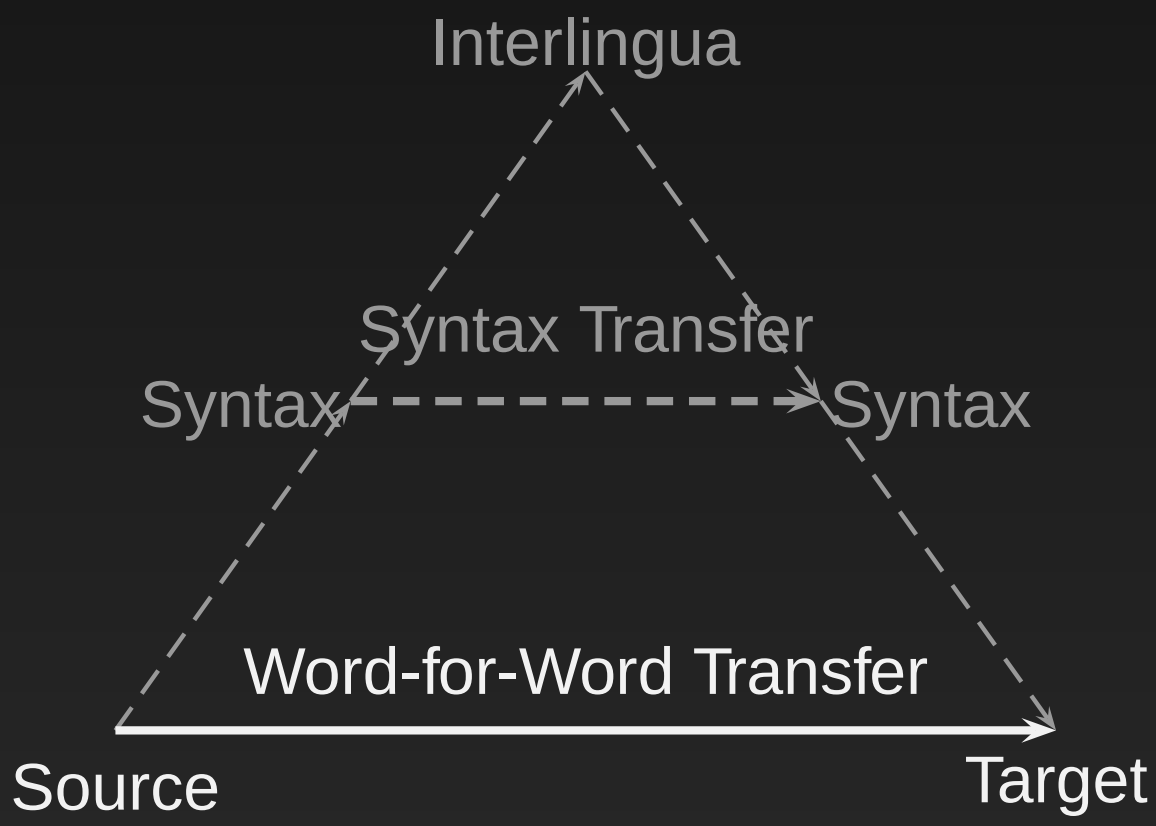
Named Entity との統合 (Vogel et al., 2003)

- 統計的機械翻訳では、トークンの意味を考えない...
- 語彙が爆発的に増大 ⟶ NE の導入
  - 3 月 15 日 ⟶ $DATE
  - 渡辺 太郎 ⟶ $PERSON
  - ATR 音声言語コミュニケーション研究所 ⟶ $ORGANIZATION
- カタカナ語の Transliteration
  - ジョン ⟶ John
  - セバスチャン ⟶ Sebastien
- 問題点
  - 原言語、目的言語の NE 抽出の精度に依存

言語の構造を表現した翻訳モデル



言語の構造を表現した翻訳モデル

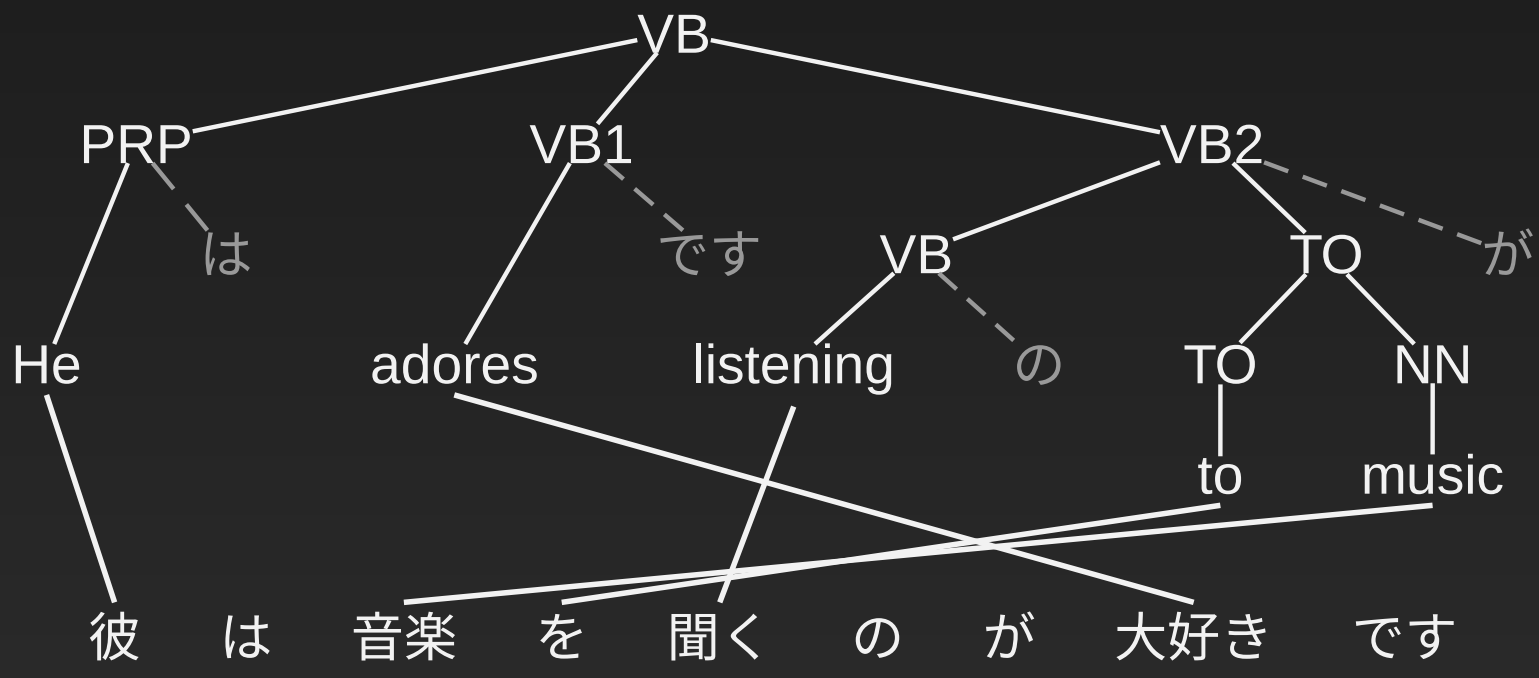


- 統計的機械翻訳は、底辺をさ迷っている...
- 言語の「構造」あるいは「意味」をとらえる翻訳モデルの実現

解析木に基づく翻訳モデル (Yamada and Knight, 2001)

$$P(J|\mathcal{E}) = P(J|\mathcal{E})P(\mathcal{E})$$

$\mathcal{E}$: 英語 E の木構造



チャンクに基づく翻訳モデル (Watanabe et al., 2003)

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

$\mathcal{J}, \mathcal{E}$: チャンクの並び ($|\mathcal{J}| = |\mathcal{E}|$)

チャンクに基づく翻訳モデル (Watanabe et al., 2003)

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

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

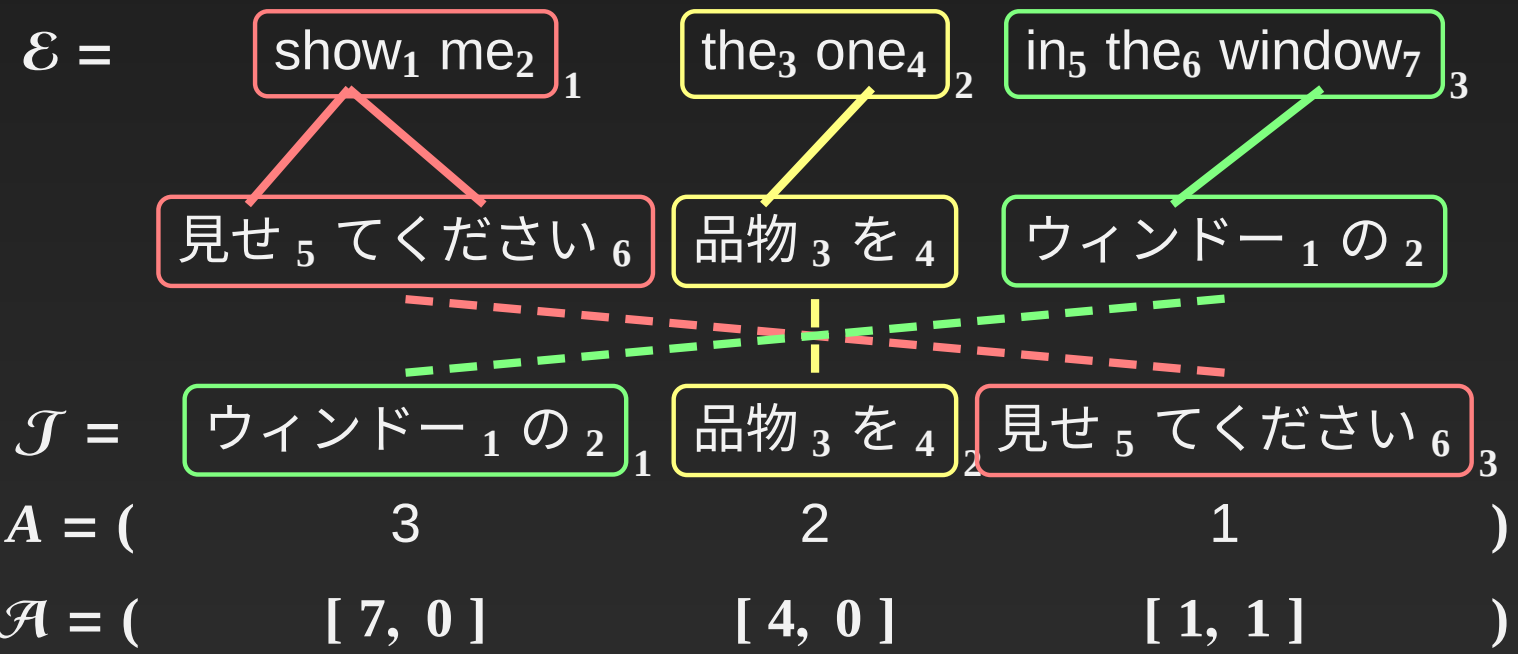
A: チャンクアライメント

$\mathcal{A}$: 単語アライメント

チャンクに基づく翻訳モデル (Watanabe et al., 2003)

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

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



言語の構造を表現した翻訳モデルの問題点

- 学習の複雑さ
  - Inside-Outside アルゴリズムによる学習
  - Approximation
- スパースなコーパス
  - 構造の導入により、いっそうスパースに
  - 解析木の精度に依存

翻訳モデル/言語モデルの最適化

$$P_\lambda(E|J) = \frac{\exp\left(\sum_i \lambda_i f_i(J,E)\right)}{Z_\lambda(J)}$$

$$Z_\lambda(J) = \sum_E \exp\left(\sum_i \lambda_i f_i(J,E)\right)$$

- 素性: $f_i(J,E)$
  - $f_0(J,E) = \log Pr(E)$
  - $f_1(J,E) = l \equiv |E|$
  - $f_2(J,E) = \log \prod t(J_j|E_i)$ etc.

翻訳モデル/言語モデルの最適化

$$P_\lambda(E|J) = \frac{\exp\left(\sum_i \lambda_i f_i(J,E)\right)}{Z_\lambda(J)}$$

$$Z_\lambda(J) = \sum_E \exp\left(\sum_i \lambda_i f_i(J,E)\right)$$

- 素性: $f_i(J,E)$
  - $f_0(J,E) = \log Pr(E)$
  - $f_1(J,E) = l \equiv |E|$
  - $f_2(J,E) = \log \prod t(J_j|E_i)$ etc.
- デコーディング: $Z_\lambda(J)$ は必要なし (理由は?)。
- 翻訳モデル (の各構成要素) と言語モデルとの結合の重みを学習

翻訳モデル/言語モデルの最適化

$$P_\lambda(E|J) = \frac{\exp\left(\sum_i \lambda_i f_i(J,E)\right)}{Z_\lambda(J)}$$

$$Z_\lambda(J) = \sum_E \exp\left(\sum_i \lambda_i f_i(J,E)\right)$$

- 素性: $f_i(J,E)$
  - $f_0(J,E) = \log Pr(E)$
  - $f_1(J,E) = l \equiv |E|$
  - $f_2(J,E) = \log \prod t(J_j|E_i)$ etc.
- デコーディング: $Z_\lambda(J)$ は必要なし (理由は?)。
- 翻訳モデル (の各構成要素) と言語モデルとの結合の重みを学習
- 翻訳モデルの構成問題 $\longrightarrow$ 素性の発見の問題

最適化の手法

- Maximum Likelihood (Och and Ney, 2002)
- Minimum Error Rate (Och, 2003)

最適化の手法

- Maximum Likelihood (Och and Ney, 2002)
  - デコーダの出力の n-best リストにより $Z_\lambda(J)$ の近似
  - GIS などのアルゴリズムにより最適化
- Minimum Error Rate (Och, 2003)

最適化の手法

- Maximum Likelihood (Och and Ney, 2002)
  - デコーダの出力の n-best リストにより $Z_\lambda(J)$ の近似
  - GIS などのアルゴリズムにより最適化
- Minimum Error Rate (Och, 2003)

$$\hat{\lambda} = \underset{\lambda}{\mathbf{argmin}} \left\{ \sum \boldsymbol{ERROR}_\lambda(J, E) \right\}$$

  - $\boldsymbol{ERROR} = \boldsymbol{BLEU}$ または、$\boldsymbol{NIST}$、$\boldsymbol{WER}$、$\boldsymbol{PER}$...
  - 学習時に $\boldsymbol{Z_\lambda(E)}$ を無視してもあまり変らない。つまり、

$$\boldsymbol{P(E|J)} \propto \mathbf{exp}\left(\sum_i \lambda_i f_i(J, E) \right)$$

エラー最小トレーニング

- 学習アルゴリズム — 制約なし最小化問題 (Press et al., 2002)
  $\mathbf{argmin}_{\lambda}\ f(\lambda)$
  - Direction Set Method (Powell's method)
  - Downhill Simplex Method
  - Simulated Annealing etc.

エラー最小トレーニング

- 学習アルゴリズム — 制約なし最小化問題 (Press et al., 2002)
  $\mathbf{argmin}_{\lambda}\ f(\lambda)$
  - Direction Set Method (Powell's method)
  - Downhill Simplex Method
  - Simulated Annealing etc.
- 学習法
  - λ_i に基づき、開発セットのデコード、$ERROR$ による評価
    $\longrightarrow$ 非常に時間がかかる

エラー最小トレーニング

- 学習アルゴリズム — 制約なし最小化問題 (Press et al., 2002)
  $\mathbf{argmin}_{\lambda} f(\lambda)$
  - Direction Set Method (Powell's method)
  - Downhill Simplex Method
  - Simulated Annealing etc.
- 学習法
  - λ_i に基づき、開発セットのデコード、$ERROR$ による評価
    $\longrightarrow$ 非常に時間がかかる
- 近似による学習
  - N-best リストの作成
    $\lambda_i = 1$ として開発セットを用いてデコーディング
  - N-best リストの λ_i による重みづけスコアによりソーティング
  - 各 N-best から 1-best の選択、$ERROR$ により評価

内容

- 統計的機械翻訳
- 多言語へのチャレンジ
  - 対訳コーパス
  - 句に基づく翻訳
  - 言語の構造を考慮した翻訳
  - 最適化手法
- 統計的機械翻訳の現状

内容

- 統計的機械翻訳
- 多言語へのチャレンジ
- 統計的機械翻訳の現状

内容

- 統計的機械翻訳
- 多言語へのチャレンジ
- 統計的機械翻訳の現状

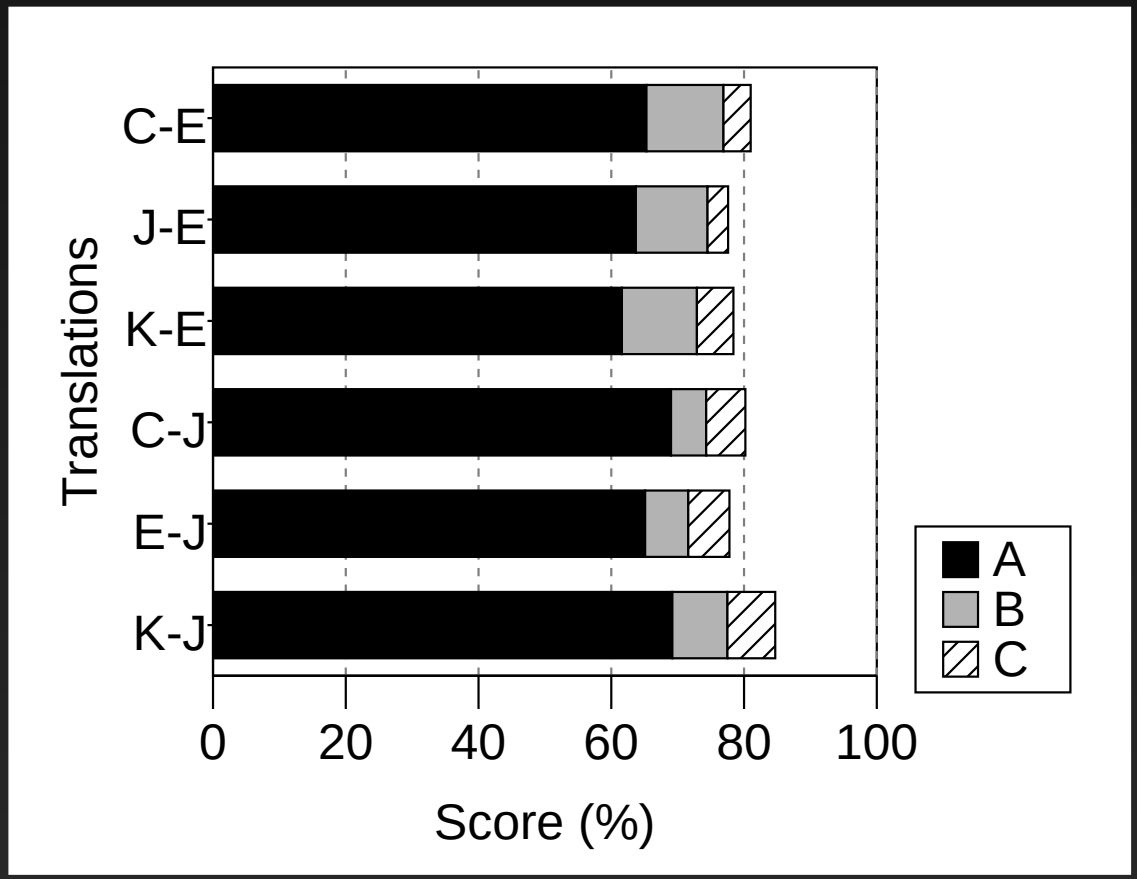
内容

- 統計的機械翻訳
- 多言語へのチャレンジ
- 統計的機械翻訳の現状
  - 多言語翻訳での評価
  - 他のシステムとの比較
  - 今後の課題

統計的機械翻訳の現状

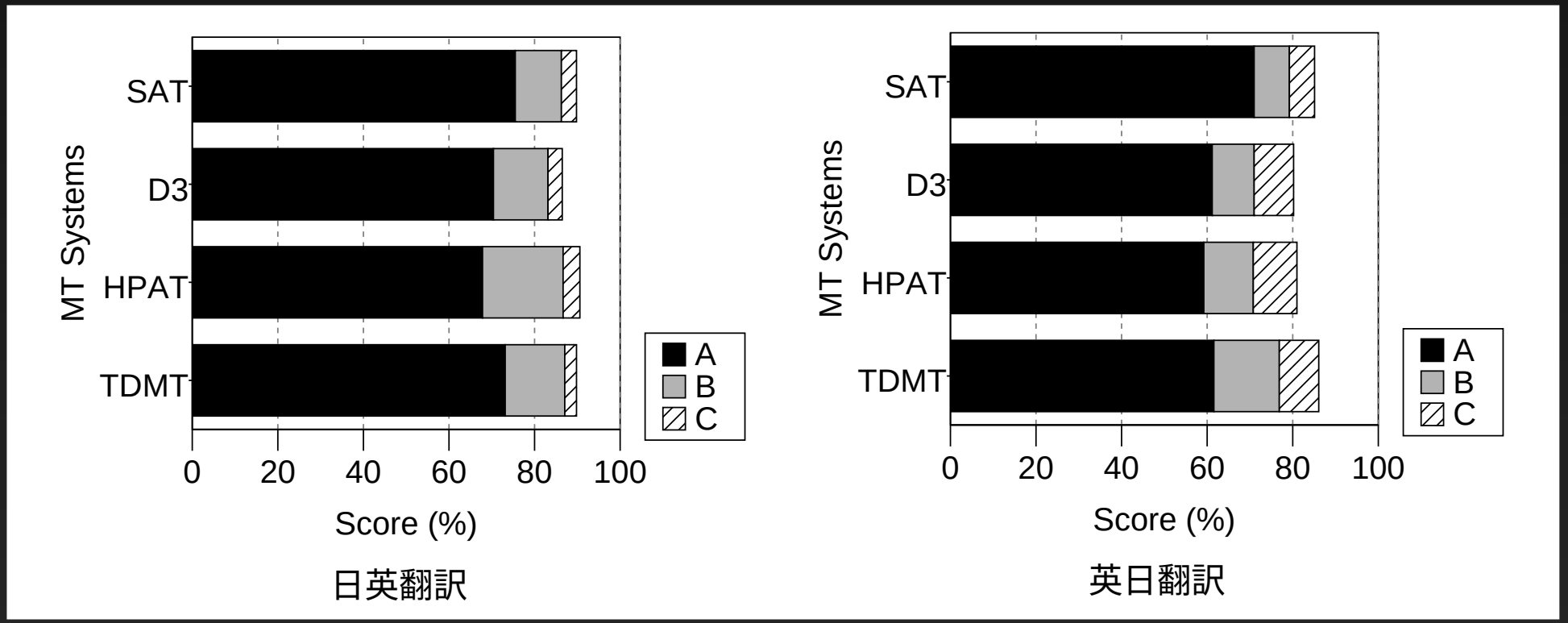
- 実際の性能は?
  - 中国語、英語、日本語、韓国語での比較
  - 他のシステムとの比較
- 使用システム: 用例検索に基づいた統計的機械翻訳 (Watanabe and Sumita, 2003a)
  - tf/idf による用例検索により種文の決定
  - Greedy Decoding
- コーパス: 旅行会話コーパス

多言語翻訳の性能



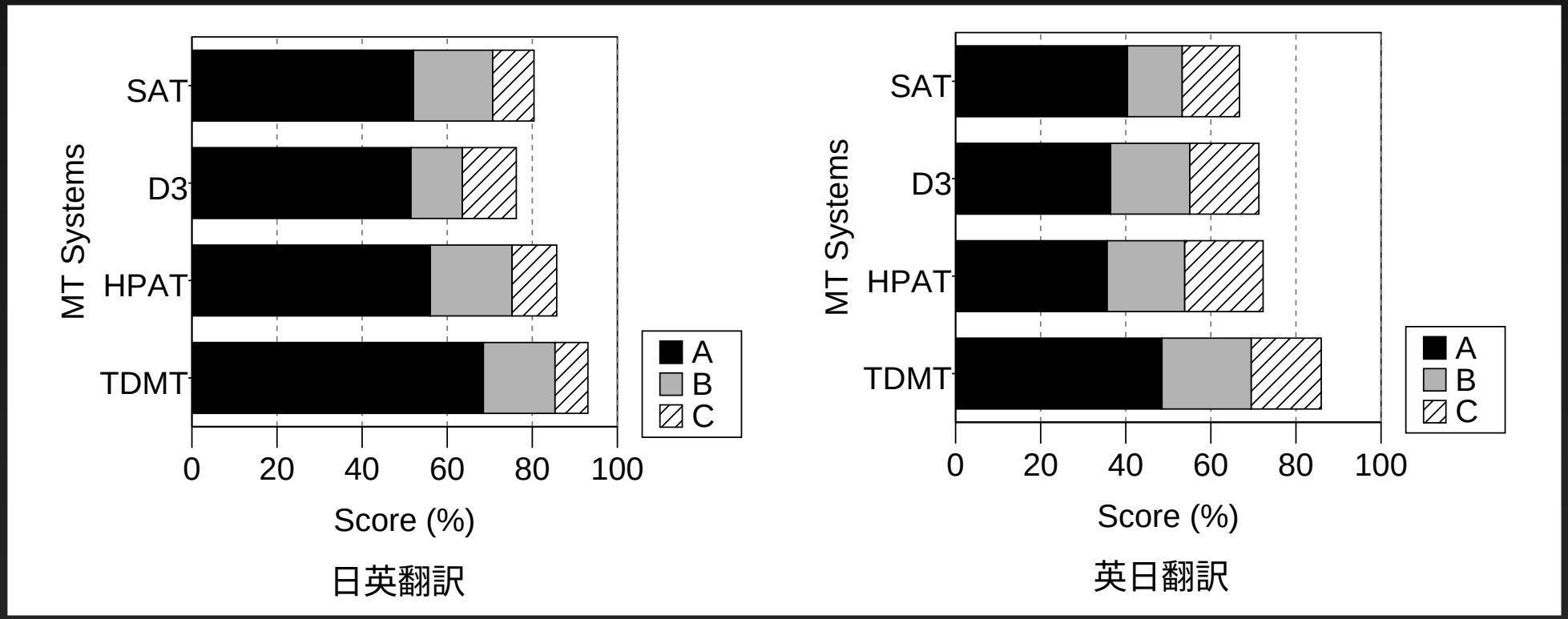
- コーパス — 旅行会話基本表現集 (BTEC) (Takezawa et al., 2002)、約 17 万文
- テストセット — 約 500 文

他のシステムとの比較



- コーパス — 旅行会話など、約 50 万文
- テストセット — 旅行会話基本表現集 (BTEC)、約 500 文
- システム
  - SAT: 用例検索に基づく統計的機械翻訳
  - D3: 文単位の用例翻訳
  - HPAT: 句単位の用例翻訳
  - TDMT: ルールに基づく機械翻訳

他のシステムとの比較 2



- テストセット — 機械翻訳介在バイリンガル音声対話コーパス (Kikui et al., 2003)、約 500 文

今後の課題

- 長文の翻訳
  - 複文などの処理
- コーパスの量
  - 多ければ多いほどよい
  - ゴミがあったとしても関係ない?
- 効率の良い翻訳候補の生成
  - 句単位の翻訳単位の抽出
  - 未知語への対処
- 構造を考慮した素性の開発

協調/競争の時代

- 協調
- 競争

協調/競争の時代

- 協調
  - Statistical Machine Translation Workshop
    - *The Mathematics of Machine Translation: Parameter Estimation* (Brown et al., 1993) の分析
    - A Statistical MT Tutorial Workbook (Knight, 1999b)
    - EGYPT Toolkit
  - Syntax for Statistical Machine Translation
    - 構造を考慮した素性の開発
    - Final Presentation (Och et al., 2003)
- 競争

協調/競争の時代

- 協調
  - Statistical Machine Translation Workshop
    - *The Mathematics of Machine Translation: Parameter Estimation* (Brown et al., 1993) の分析
    - A Statistical MT Tutorial Workbook (Knight, 1999b)
    - EGYPT Toolkit
  - Syntax for Statistical Machine Translation
    - 構造を考慮した素性の開発
    - Final Presentation (Och et al., 2003)
- 競争
  - TIDES — DARPA 主導、closed な workshop
  - IWSLT — CSTAR 主導、open な workshop

TIDES

- http://www.nist.gov/speech/tests/mt/index.htm
- {中国語、アラビア語} ⟶ 英語
- コーパス — 膨大な量/様々なソース、コーディング/低品質?

UN: 3,755,456 文	Xinhua News: 109,792 文
Hong Kong News: 683,305 文	Sinorama Magazine: 103,252 文
Hong Kong Hansard: 351,514 文	etc.

- だれでも参加可能 (義務: 結果の送信、Workshop への参加)
- 予定

04/30/04	Registration to participate deadline.
05/10/04	Evaluation test data E-mailed to participants.
05/14/04	System translations due at NIST by 12 noon.
June 23-24, 2004	The evaluation workshop.

IWSLT 2004

- http://www.slt.atr.jp/IWSLT2004/
- Evaluation Campaign
  - {日本語、中国語} ⟶ 英語
  - 話し言葉コーパス (BTEC の一部、約 2 万文)/高品質?
- Technical Papers
- 予定

Application submission:	April 15, 2004
Training Corpus Release:	May 21, 2004
Test Corpus Release:	August 9, 2004
Run Submission:	August 12, 2004
Result Feedback to Participants:	September 10, 2004
Camera-ready Paper Submission:	September 17, 2004
Workshop:	September 30 - October 1, 2004

参考文献

Berger, A., P. Brown, S. Pietra, V. Pietra, J. Lafferty, H. Printz, and L. Ures (1994) “The Candide System for Machine Translation.” In *Proc. of the ARPA Conference on Human Language Technology.*

Berger, Adam L., Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Andrew S. Kehler, and Robert L. Mercer (1996) “Language Translation Apparatus and Method of Using Context-Based Translation Models.” Technical report United States Patent, Patent Number 5510981.

Brown, Peter F., John Cocke, Stephen Della Pietra, Vincent J. Della Pietra, Frederick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin (1990) “A Statistical Approach to Machine Translation.” *Computational Linguistics.* Vol. 16. No. 2. pp. 79–85.

Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer (1993) “The Mathematics of Statistical Machine Translation: Parameter Estimation.” *Computational Linguistics.* Vol. 19. No. 2. pp. 263–311.

Charniak, Eugene, Kevin Knight, and Kenji Yamada (2003) “Syntax-based Language Models for Statistical Machine Translation.” In *Proceedings of MT Summit IX.* New Orleans, LA.

Dempster, A. P., N.M. Laird, and D.B.Rubin (1977) “Maximum Likelihood from Incomplete Data via the EM Algorithm.” *Journal of the Royal Statistical Society.* Vol. B. No. 39. pp. 1–38.

Foster, George, Simona Gandrabur, Philippe Langlais, Pierre Plamondon, Graham Russell, and Michel Simard (2003) “Statistical Machine Translation: Rapid Deployment with Limited Resources.” In *Proceedings of MT Summit IX.* New Orleans, LA.

参考文献

Foster, George (2000a) “Incorporating Position Information into a Maximum Entropy/Minimum Divergence Translation Model.” In *Proc. of CoNLL-2000 and LLL-2000.* Lisbon, Portugal.

Foster, George (2000b) “A Maximum Entropy/Minimum Divergence Translation Model.” In *Proc. of ACL 2000.* Hong Kong.

Gale, William A. and Kenneth Ward Church (1993) “A Program for Aligning Sentences in Bilingual Corpora.” *Computational Linguistics.* Vol. 19. No. 1. pp. 75–102.

Germann, Ulrich, Michael Jahr, Kevin Knight, Daniel Marcu, and Kenji Yamada (2001) “Fast Decoding and Optimal Decoding for Machine Translation.” In *Proc. of ACL 2001.* Toulouse, France.

Germann, Ulrich (2003) “Greedy Decoding for Statistical Machine Translation in Almost Linear Time.” In Hearst, Marti and Mari Ostendorf. eds. *HLT-NAACL 2003: Main Proceedings.* Edmonton, Alberta, Canada Association for Computational Linguistics.

Kikui, Genichiro, Eiichiro Sumita, Toshiyuki Takezawa, and Seiichi Yamamoto (2003) “Creating Corpora for Speech-to-Speech Translation.” In *Proc. of Eurospeech 2003.* Geneva, Switzerland.

Knight, Kevin and Philipp Koehn (2003) “What’s New in Statistical Machine Translation.” In *Tutorial at HLT/NAACL/MT-Summit IX.* http://www.isi.edu/~koehn/publications/tutorial2003.pdf.

Knight, Kevin (1999a) “Decoding Complexity in Word-Replacement Translation Models.” *Computational Linguistics.* Vol. 25. No. 4. pp. 607–615.

参考文献

Knight, Kevin (1999b) “A Statistical MT Tutorial Workbook.” http://www.clsp.jhu.edu/ws99/projects/mt/mt-workbook.htm.

Koehn, Philipp, Franz Josef Och, and Daniel Marcu (2003) “Statistical Pharse-Based Translation.” In *Proc. of HLT-NAACL 2003.* Edmonton.

Kumar, Shankar and William Byrne (2003) “A Weighted Finite State Transducer Implementation of the Alignment Template Model for Statistical Machine Translation.” In *Proc. of HLT-NAACL 2003.* Edmonton.

Marcu, Daniel and William Wong (2002) “A Phrase-Based, Joint Probability Model for Statistical Machine Translation.” In *Proc. of EMNLP-2002.* Philadelphia, PA.

Marcu, Daniel (2001) “Towards a Unified Approach to Memory- and Statistical-Based Machine Translation.” In *Proc. of ACL 2001.* Toulouse, France.

Melamed, I. Dan (1996) “A Geometric Approach to Mapping Bitext Correspondence.” In Eric Brill and Kenneth Church. eds. *Proceedings of the Conference on Empirical Methods in Natural Language Processing.* Somerset, New Jersey: Association for Computational Linguistics. pp. 1–12.

Och, Franz Josef and Hermann Ney (2002) “Discriminative Training and Maximum Entropy Models for Statistical Machine Translation.” In *Proc. of ACL 2002.* Philadelphia, PA.

参考文献

Och, Franz Josef and Hermann Ney (2003) “A Systematic Comparison of Various Statistical Alignment Models.” *Computational Linguistics.* Vol. 29. No. 1. pp. 19–51.

Och, Franz Josef, Christoph Tillmann, and Hermann Ney (1999) “Improved Alignment Models for Statistical Machine Translation.” In *Proc. of EMNLP/WVLC.* University of Maryland, College Park, MD.

Och, Franz Josef, Daniel Gildea, Anoop Sarkar, Kenji Yamada, Sanjeev Khudanpur, Dragomir Radev, Alex Fraser, Shankar Kumar, David Smith, Libin Shen, Viren Jain, Katherine Eng, and Zhen Jin (2003) “Syntax for Statistical Machine Translation.” http://www.clsp.jhu.edu/ws03/groups/translate/MT_final.pdf.

Och, Franz Josef (2003) “Minimum Error Rate Training in Statistical Machine Translation.” In *Proc. of ACL 2003.* Sapporo, Japan.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu (2002) “Bleu: a Method for Automatic Evaluation of Machine Translation.” In *Proc. of ACL 2002.*

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery (2002) *Numerical Recipes in C++.* Cambridge, UK: Cambridge University Press.

of Standards, National Institute and Technology (2002) “Automatic Evaluation of Machine Translation Quality using N-ram Co-Occurrence Statistics.” http://www.nist.gov/speech/tests/mt/doc/ngram-study.pdf.

Takezawa, Toshiyuki, Eiichiro Sumita, Fumiaki Sugaya, Hirofumi Yamamoto, and Seiichi Yamamoto (2002) “Toward a Broad-coverage Bilingual Corpus for Speech Translation of Travel Conversations in the Real World.” In *Proc. of LREC 2002.* Las Palmas, Canary Islands, Spain.

参考文献

Tillmann, Christoph and Hermann Ney (2003) “Word Reordering and a Dynamic Programming Beam Search Algorithm for Statistical Machine Translation.” *Computational Linguistics.* Vol. 29. No. 1. pp. 97–133.

Tillmann, Christoph (2003) “A Projection Extension Algorithm for Statistical Machine Translation.” In Collins, Michael and Mark Steedman. eds. *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing.*

Ueffing, Nicola, Franz Josef Och, and Hermann Ney (2002) “Generation of Word Graphs in Statistical Machine Translation.” In *Proc. Conference on Empirical Methods for Natural Language Processing (EMNLP02).* Philadelphia, PA.

Vogel, Stephan, Franz Josef Och, Chistof Tillman, Sonja Nie''sen, Hassan Sawaf, and Hermann Ney (2000) “Statistical Methods for Machine Translation.” *Wahlster.* pp. 377–393.

Vogel, Stephan, Ying Zhang, Fei Huang, Alicia Tribble, Ashish Venogupal, Bing Zhao, and Alex Waibel (2003) “The CMU Statistical Translation System.” In *Proceedings of MT Summit IX.* New Orleans, LA.

Watanabe, Taro and Eiichiro Sumita (2002) “Bidirectional Decoding for Statistical Machine Translation.” In *Proc. of COLING 2002.* Vol. 2 Taipei, Taiwan.

Watanabe, Taro and Eiichiro Sumita (2003a) “Example-based Decoding for Statistical Machine Translation.” In *Proceedings of MT Summit IX.* New Orleans, LA.

参考文献

Watanabe, Taro, Kenji Imamura, and Eiichiro Sumita (2002) “Statistical Machine Translation Based on Hierarchical Phrase Alignment.” In *Proc. of TMI 2002.* Keihanna, Japan.

Watanabe, Taro, Eiichiro Sumita, and Hoiroshi G. Okuno (2003) “Chunk-based Statistical Translation.” In *Proc. of ACL 2003.* Sapporo, Japan.

Watanabe, Taro, Kenji Imamura, Eiichiro Sumita, and Hiroshi G. Okuno (2004) “Statistical Machine Translation Using Hierarchical Phrase Alignment.” *IEICE Transactions on Informaitn and Systems (to appear).*

Wu, Dekai (1996) “A Polynomial-Time Algorithm for Statistical Machine Translation.” In Joshi, Arivind and Martha Palmer. eds. *Proc. of ACL 1996.* San Francisco Morgan Kaufmann Publishers.

Yamada, Kenji and Kevin Knight (2001) “A Syntax-based Statistical Translation Model.” In *Proc. of ACL 2001.* Toulouse, France.

Yamada, Setsuo, Masaaki Nagata, and Kenji Yamada (2003) “Improving Translation Models by Applying Asymmetric Learning.” In *Proceedings of MT Summit IX.* New Orleans, LA.

Zens, Richard and Hermann Ney (2003) “A Comparative Study on Reordering Constraints in Statistical Machine Translation.” In *Proc. of ACL 2003.* Sapporo, Japan.

永田昌明 (2003) 統計的自然言語処理, 言語情報アクセス基盤技術としての機械学習 (FIT 2003)