

句に基づく機械翻訳

デコード

渡辺太郎

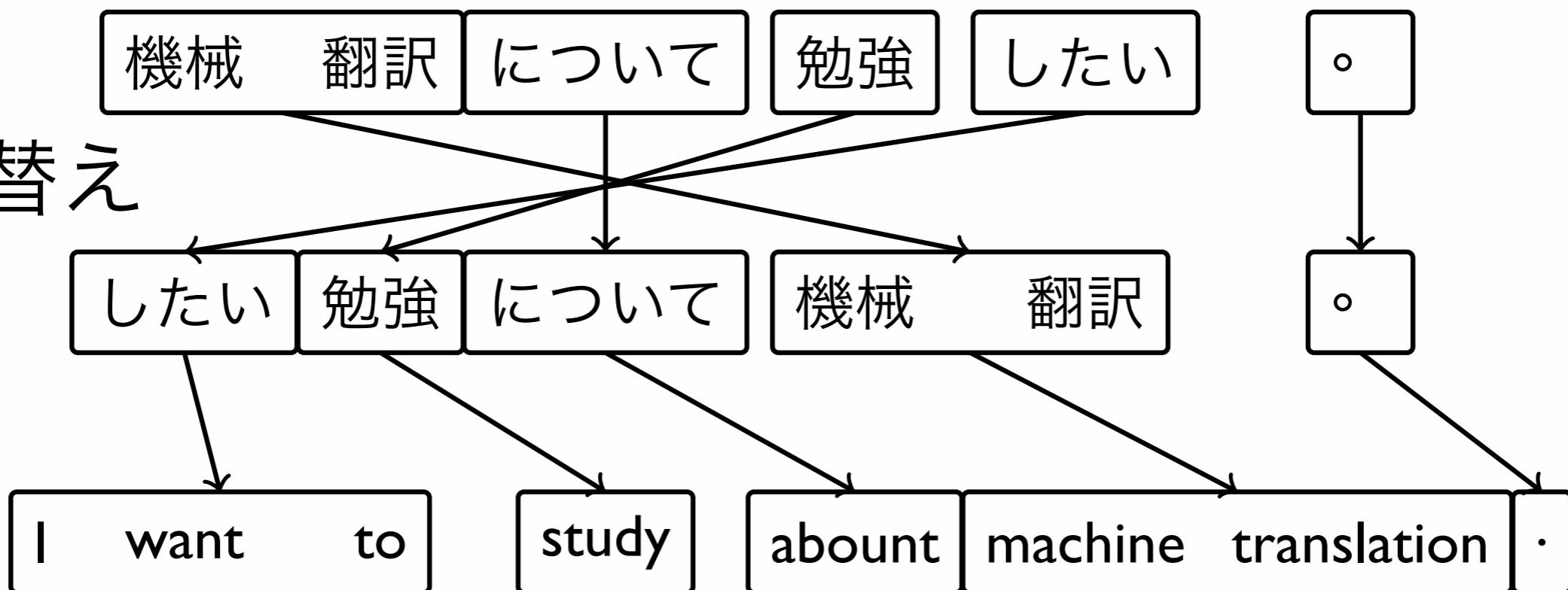
taro.watanabe at nict.go.jp



<https://sites.google.com/site/alaginmt2015/>

句に基づく機械翻訳

並び替え



(Koehn et al., 2003)

句に基づくモデル

$$\begin{aligned} Pr(e|f) \Pr(e) &= \sum_{\phi, \alpha} Pr(f, \phi, \alpha|e) \Pr(e) \\ &\approx \sum_{\phi, \alpha} p_d(f, \alpha|\phi) p_\phi(\phi|e) p_{lm}(e) \end{aligned}$$

句単位の変換

並び替え

言語モデル

- $\Phi: (f, e)$ の句単位の分割、 α : 句の並び替え
- 言語モデルにより、正しい目的言語が生成されることを保証

Viterbi近似

$$\begin{aligned}\hat{e} &= \arg \max_e \sum_{\phi, \alpha} p_d(f, \alpha | \phi) p_\phi(\phi | e) p_{lm}(e) \\ &\approx \arg \max_{e, \phi, \alpha} p_d(f, \alpha | \phi) p_\phi(\phi | e) p_{lm}(e)\end{aligned}$$

- ϕ, α を全て列挙不可能: ϕ, α を最大にする翻訳を探索

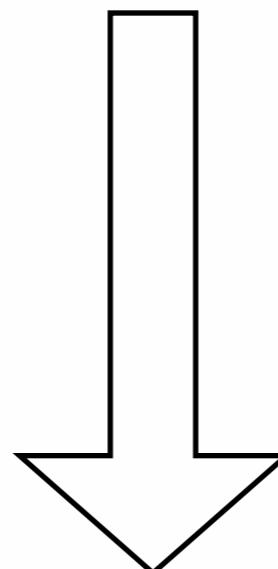
対数線形モデル

$$\begin{aligned}\hat{e} &= \arg \max_e \frac{\sum_d \exp(\mathbf{w}^\top \mathbf{h}(f, d, e))}{\sum_{e', d'} \exp(\mathbf{w}^\top \mathbf{h}(f, d', e'))} \\ &\approx \arg \max_{\langle e, d \rangle} \mathbf{w}^\top \mathbf{h}(f, d, e)\end{aligned}$$

- 複数の素性 $\mathbf{h}(e, d, f)$ をlog-linearに組み合わせ、最大化
- $d = (\Phi, \alpha)$: 導出(句単位の分割、翻訳+並び替え)
- \mathbf{h} の例: $\mathbf{h}(f, d, e) = \begin{pmatrix} \log p_d(f, \alpha | \phi) \\ \log p_\phi(\phi | e) \\ \log p_{lm}(e) \end{pmatrix}$
- \mathbf{w} : 各素性の重み付け

では、翻訳！

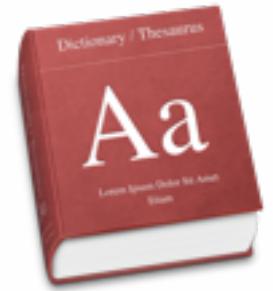
機械翻訳について勉強したい。



どうないしましょう？

I want to study about machine translation.

フレーズペアとは...



		$\log Pr(\bar{f} \bar{e})$
機械	machine mechanism machines	-2.5 -3.5 -3.0
翻訳	translation translate	-1.0 -1.5
勉強したい	wants to study I want to study	-3.5 -4.0
勉強	learn study	-2.0 -1.5

フレーズペアの列举

機械 翻訳 について 勉強 したい。

machine translation

about

勉強

したい

。

mechanism translate

regarding

learn

he wants to

?

machine translation

wants to study

!

about the translation

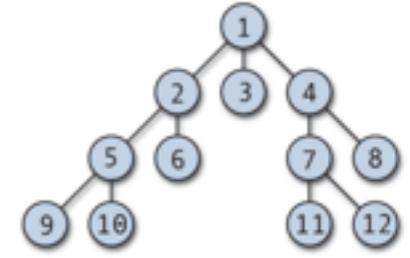
I want to study

a study about

want to do

目的言語(英語側)の順番で選択

探索



機械翻訳について勉強したい。

-2.5

翻譯 → translation - I.0

The diagram illustrates a neural network architecture for machine translation, specifically from Japanese to English. It shows the flow of information from the input Japanese sentence "機械翻訳" through various layers to the final English translation "machine translation".

The process starts with the Japanese input "機械翻訳" (Machine Translation). This input is processed by a layer represented by a green box containing a dashed line. The output of this layer is then passed through a series of hidden states, each represented by a box containing colored dots (purple, teal, blue, red, yellow, green) and dashed lines.

Key components and transitions include:

- Input Layer:** "機械翻訳" (Machine Translation) is shown in green.
- Hidden States:** The input is processed by a sequence of hidden states:
 - A purple box (green background) containing a dashed line.
 - A green box (orange background) containing a purple dot and a dashed line.
 - A green box (light green background) containing two purple dots and a dashed line.
 - A green box (light green background) containing a dashed line.
- Score Calculation:** The final hidden state (green box with a dashed line) is passed through a layer represented by a green box containing a purple dot and a blue dot, followed by a calculation step: $\text{score}(\text{---} \bullet \bullet \text{---}) = -3.0$.
- Output Layer:** The final output is "machine translation" (blue text).

Annotations provide additional context:

- "したい" → "I want to" (red text)
- "勉強" → "study" (yellow text)
- "勉強" → "learn" (yellow text)
- "勉強" → "study" (yellow text)
- "機械" → "machine" (purple text)

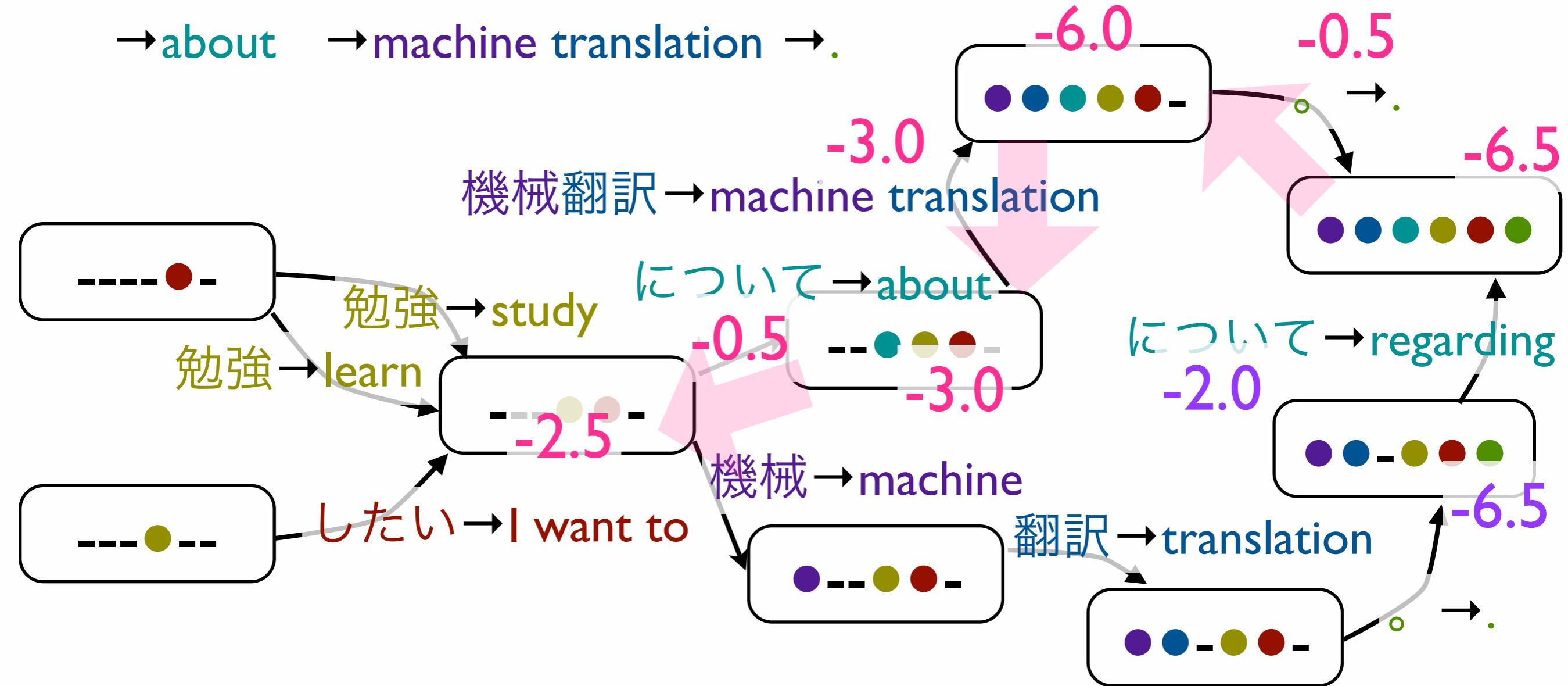
グラフを作った後は?



機械翻訳について勉強したい。

について 機械翻訳

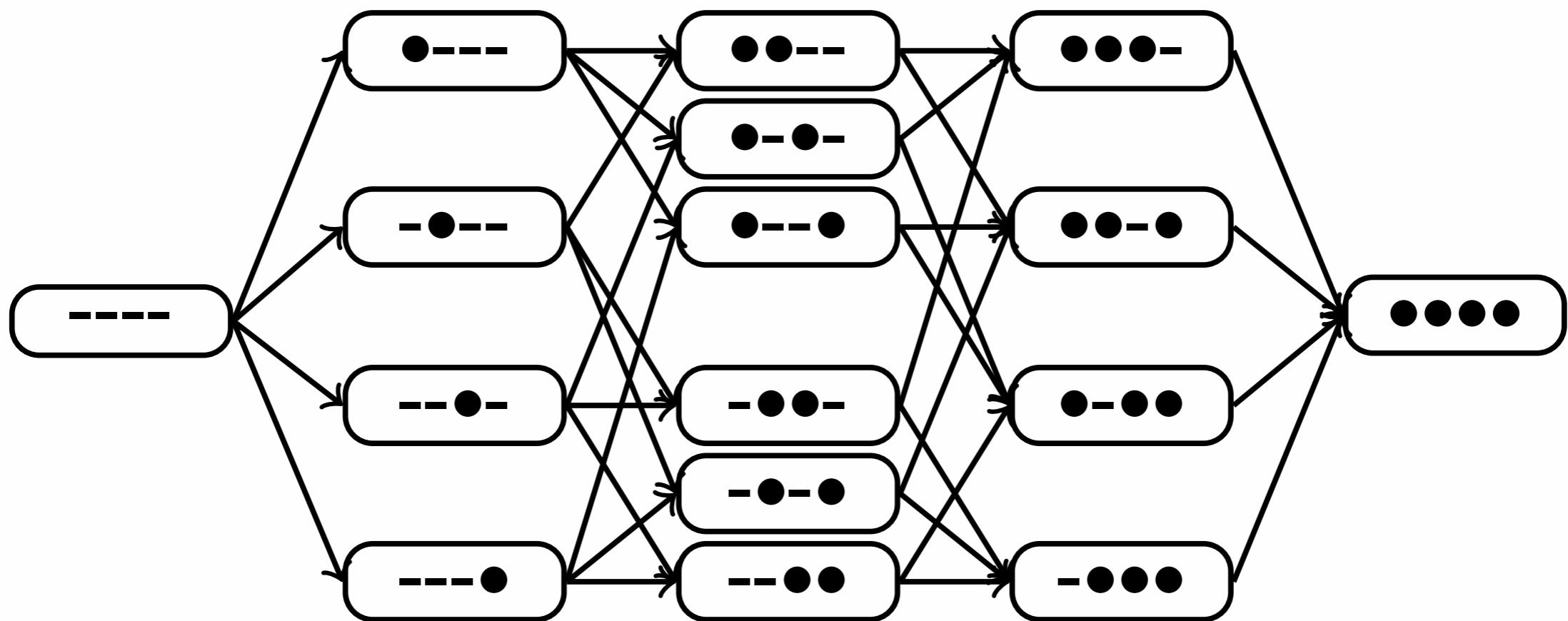
→about →machine translation →.



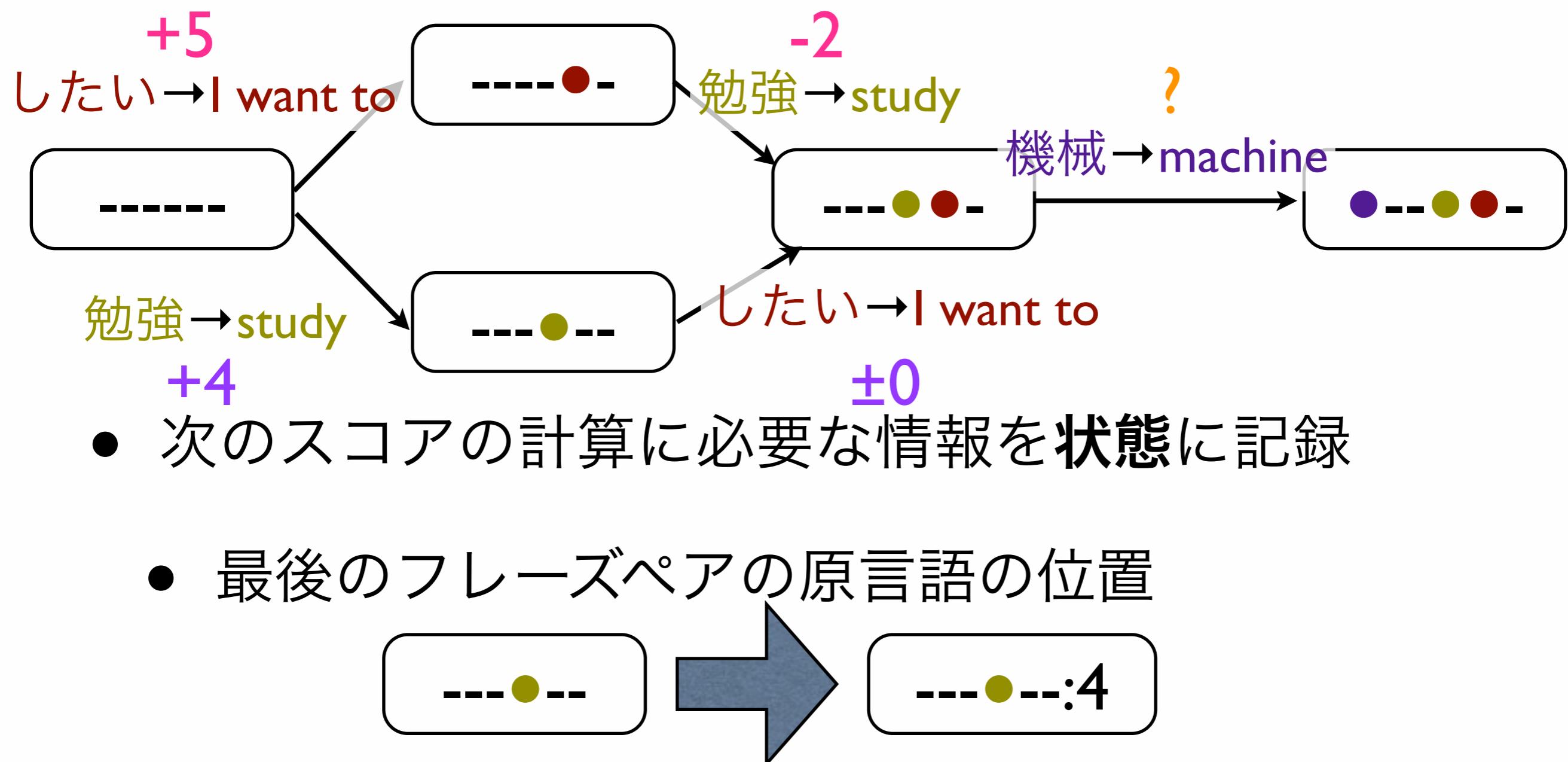
巡回セールスマン問題

- NP困難問題:各都市を一度だけ訪れる
 - 巡回セールスマン問題としてのMT(Knight, 1999)
 - 原言語の各単語 = 都市
 - 動的計画法(DP)による解:
 - 状態: 訪れた都市 (bit-vector)
 - 空間: $\mathcal{O}(2^n)$ 、 時間: $\mathcal{O}(2^n n)$

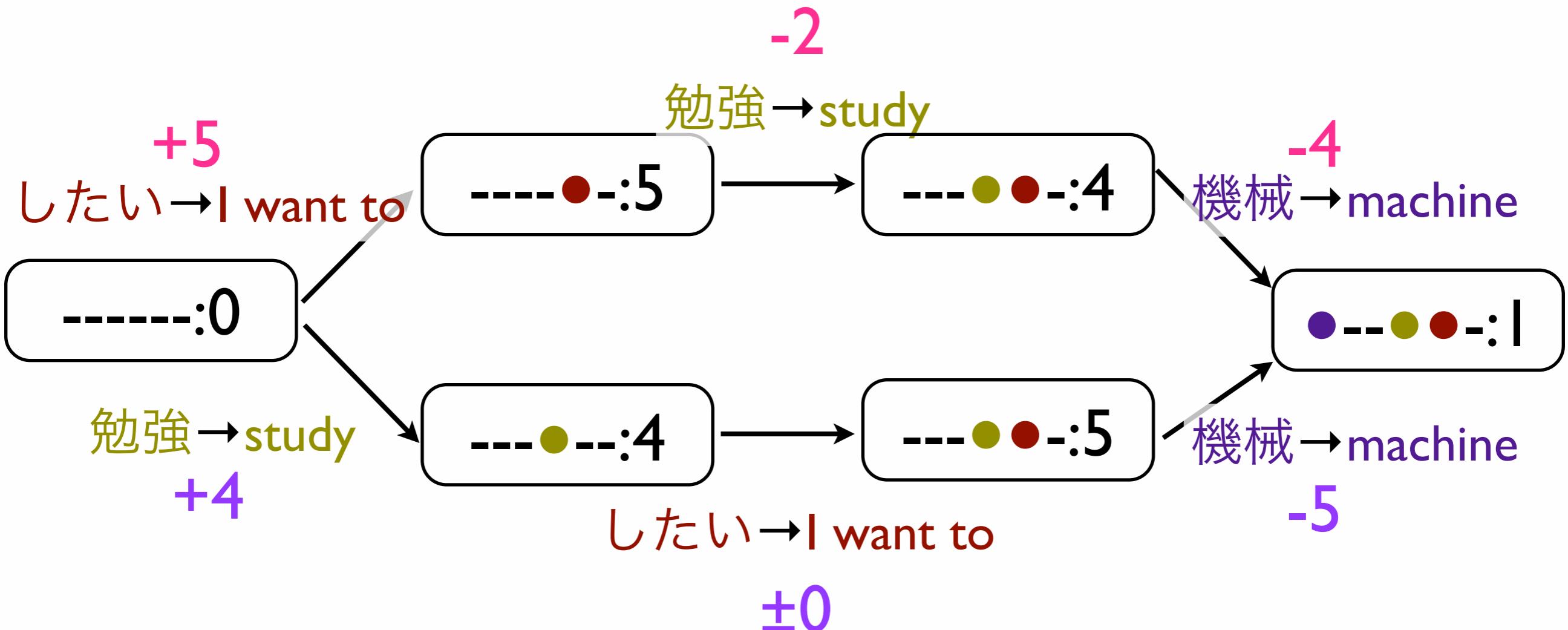
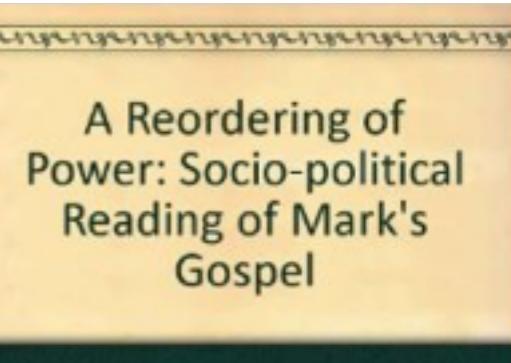




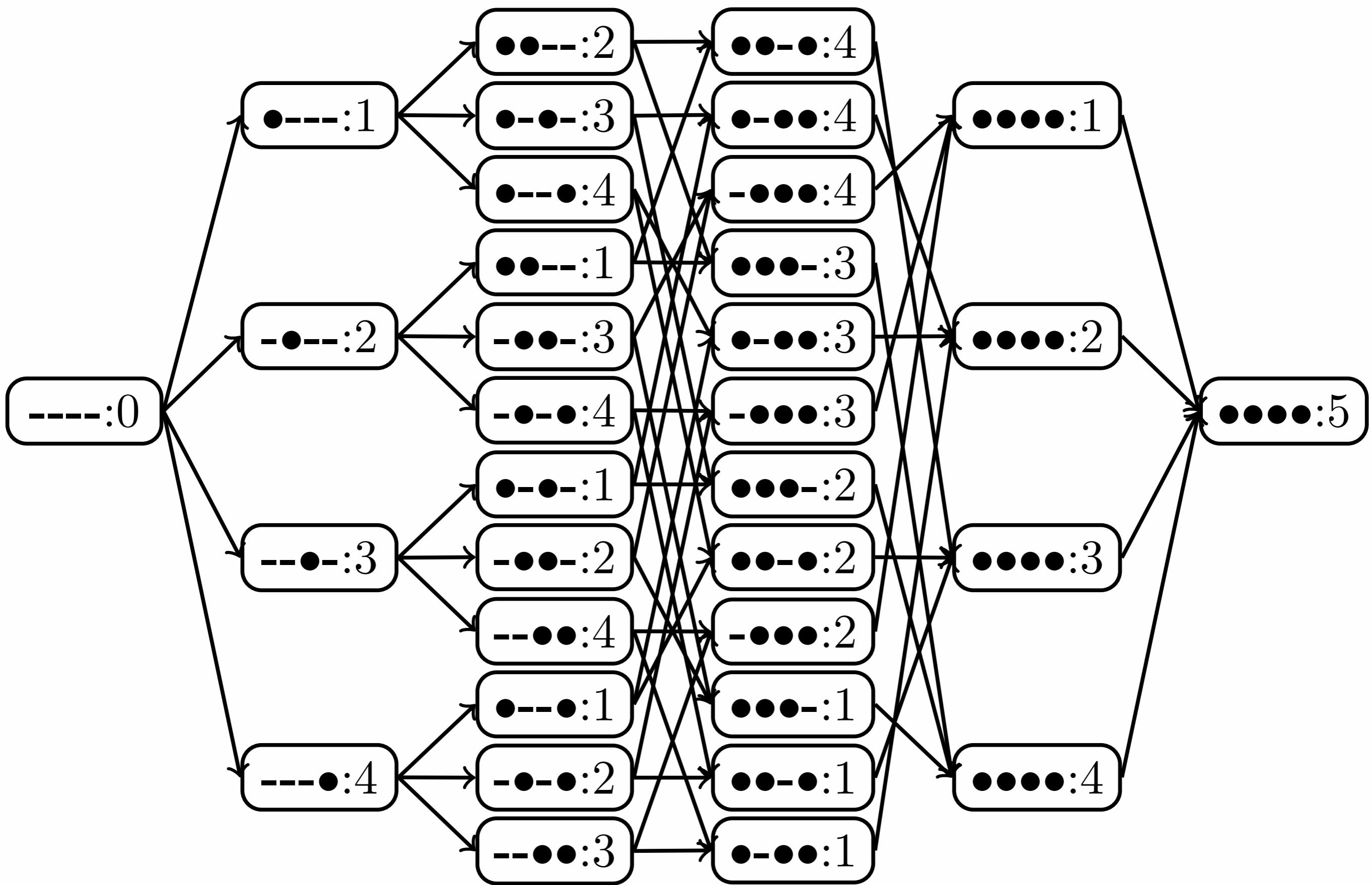
並べ替えモデルは？



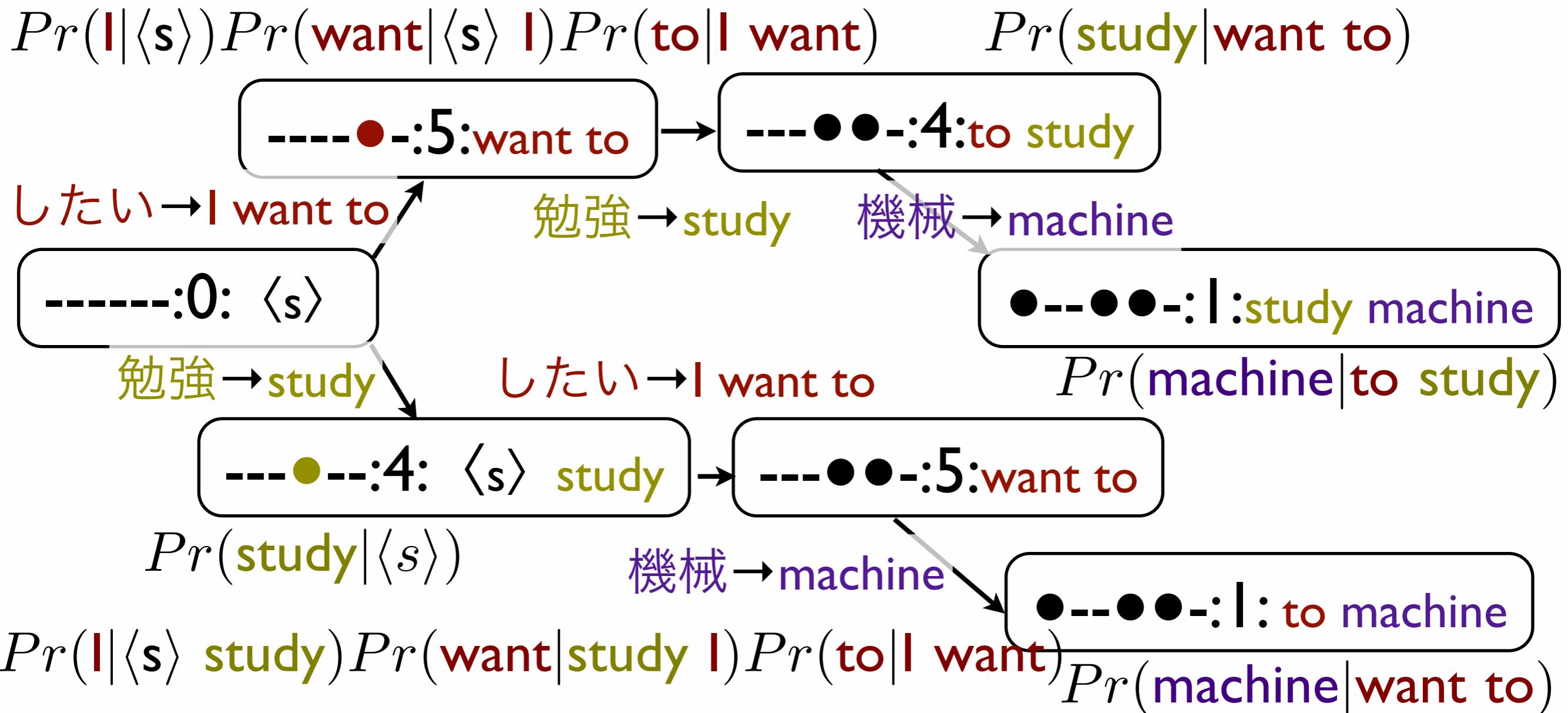
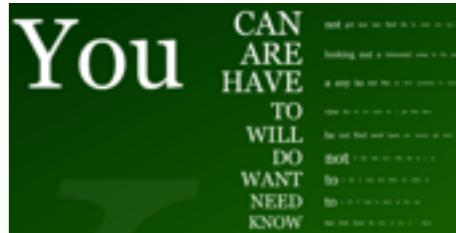
並べ替えモデル



空間: $O(2^n n)$ 、 時間: $O(2^n n^3)$



ngram言語モデル?



同じ考え方: 3-gramの場合、2-gramを記憶

空間: $O(2^n n V^{m-1})$ 、 時間: $O(V^{m-1} 2^n n^3 V^{m-1})$

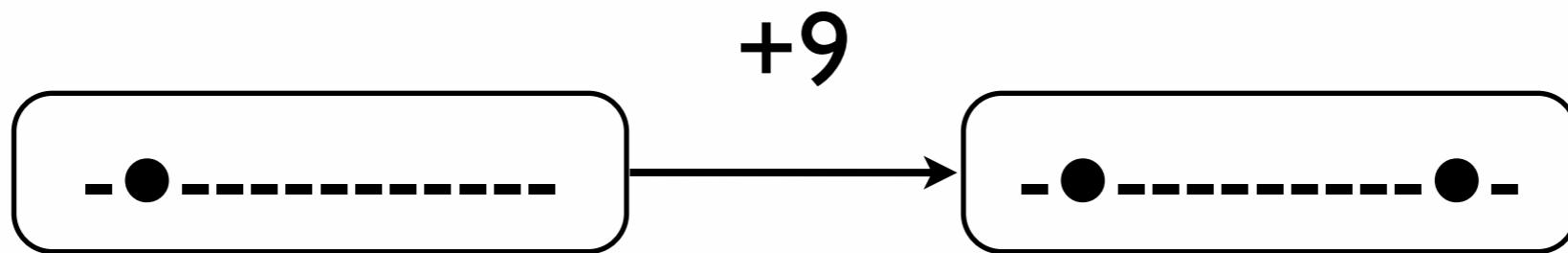
大きすぎて、グラフを描けません

素性の局所性

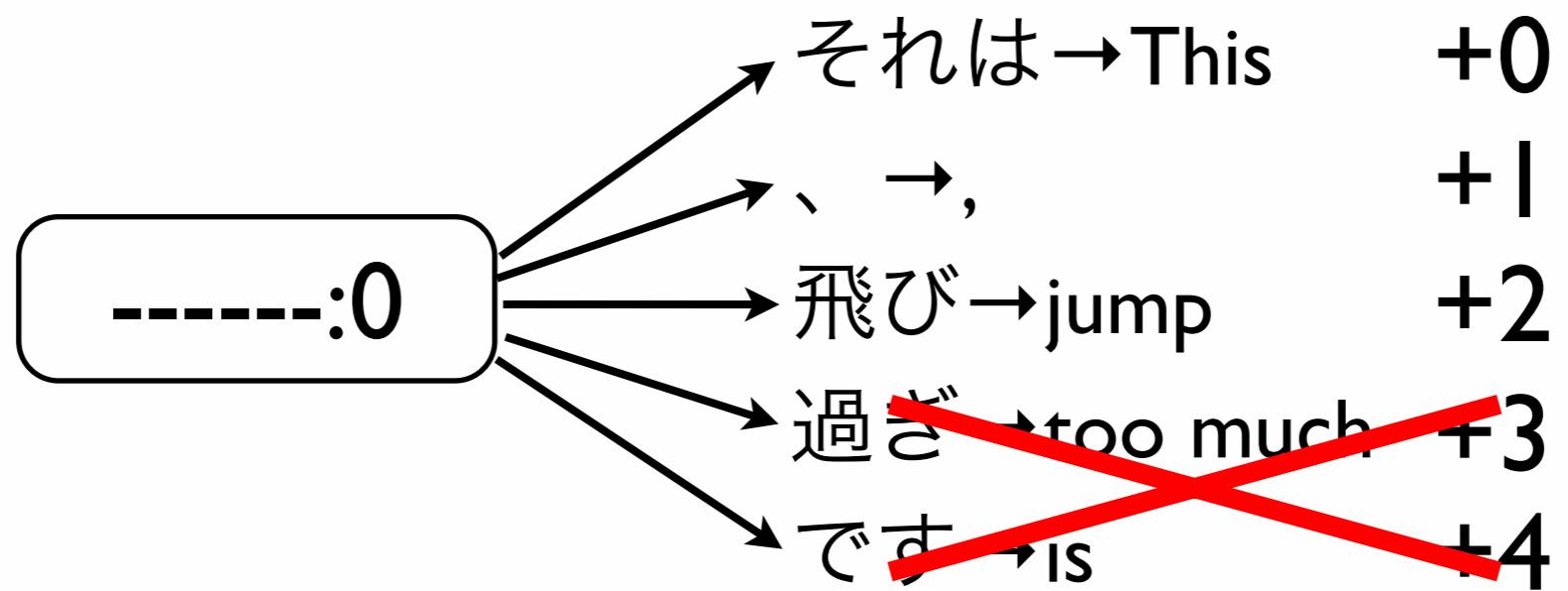
- 局所的な素性: フレーズに閉じた素性
 - フレーズ翻訳確率、語彙化翻訳確率
- 非局所的な素性: 複数のフレーズから計算
 - 並び替え素性、ngram言語モデル

計算量の呪い

- アイデアI: 全部列挙する必要なし
- 歪み制約、並び替え制約

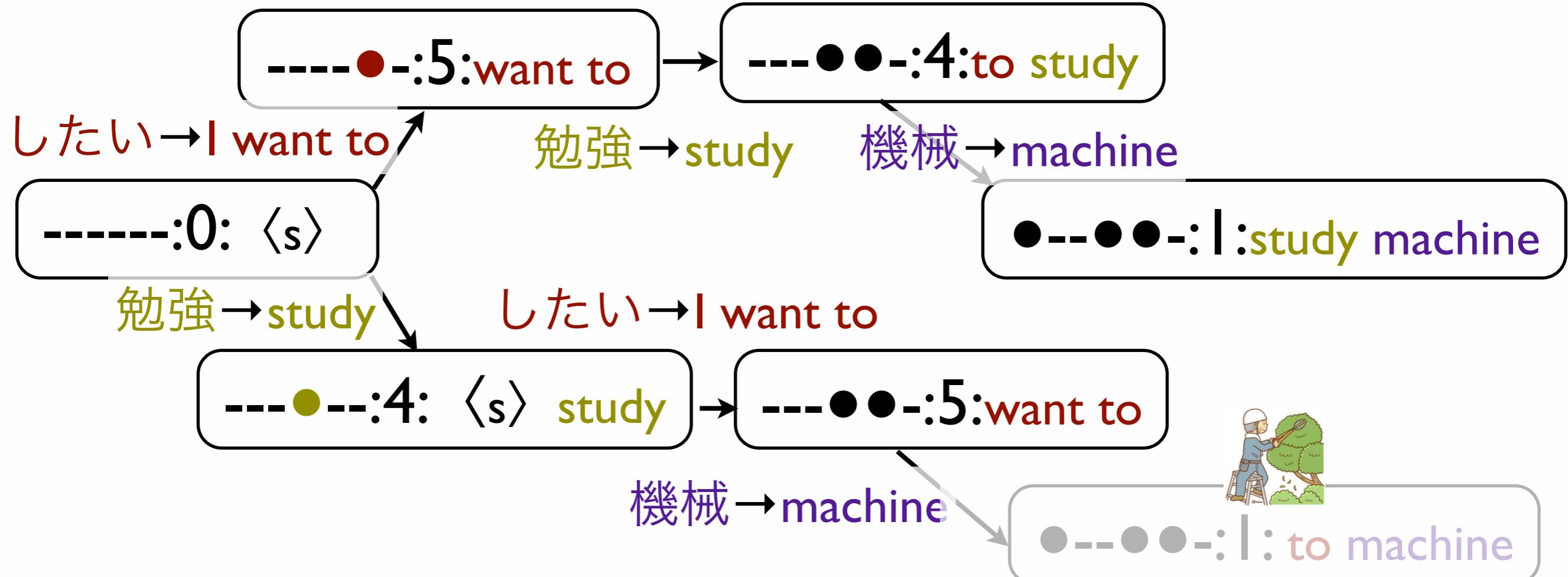


それは、飛び過ぎです。



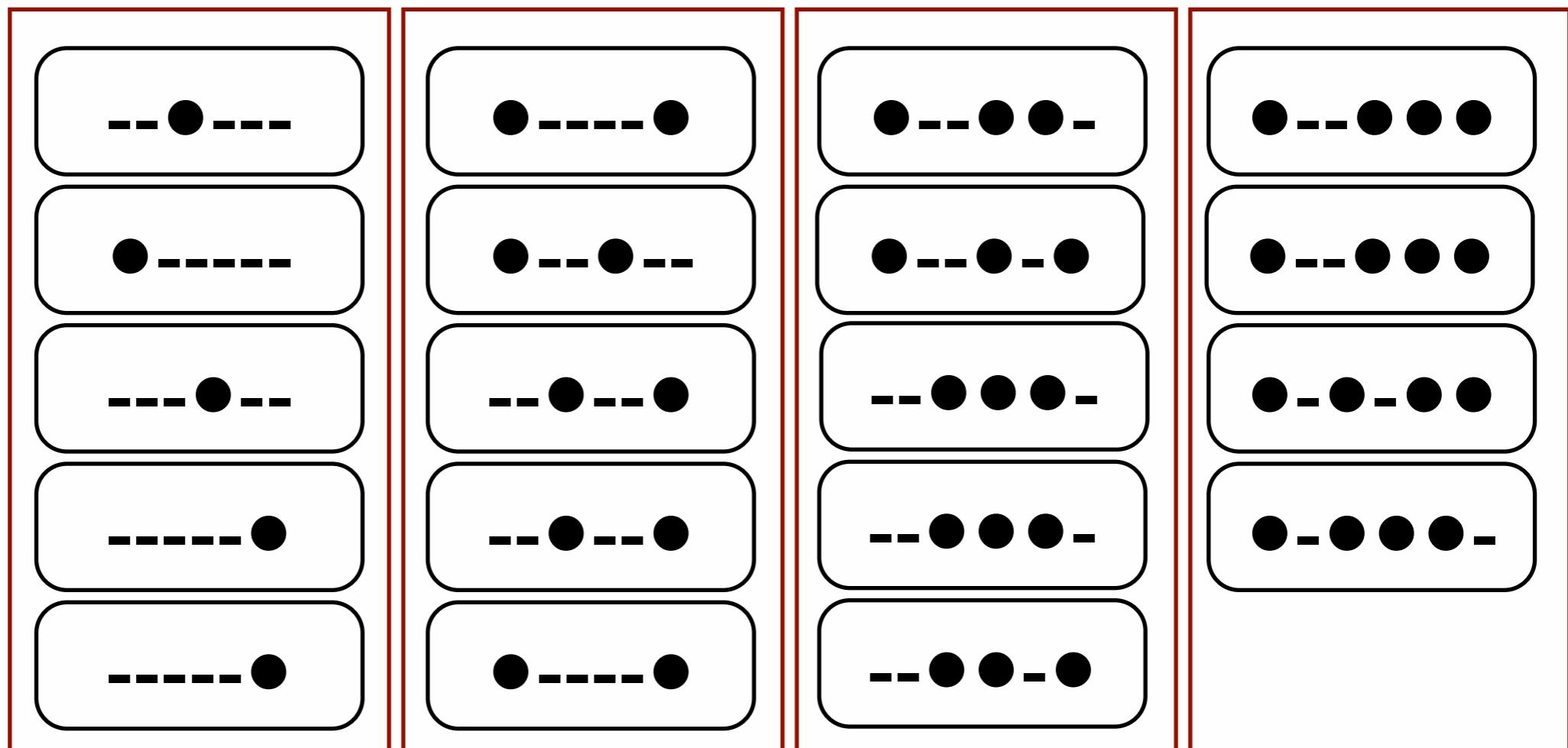


枝刈り



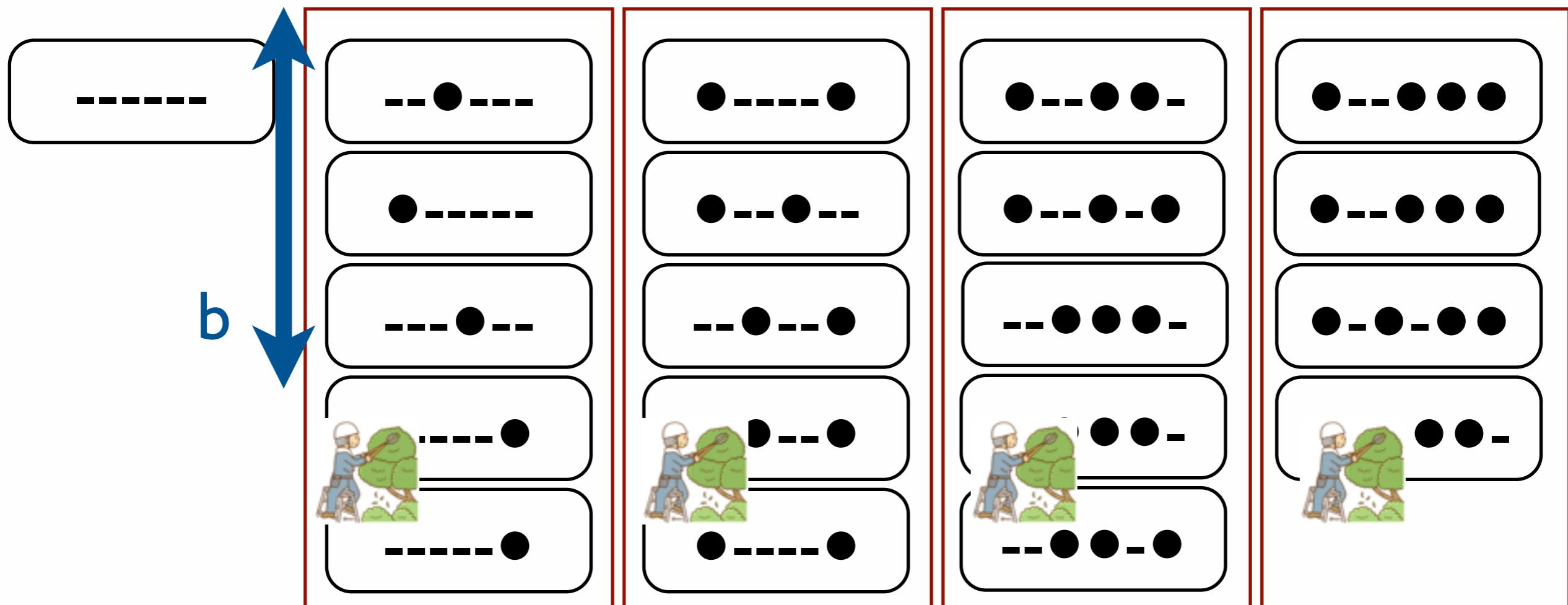
- アイデア2: ヒューリスティックなプルーニング
- ビーム探索: スコアの高そうな状態以外は枝刈り

グループ化



- 「翻訳された単語数 = cardinality」でグループ化
- 小さいcardinalityを持つ仮説から展開

グループ毎の枝刈り

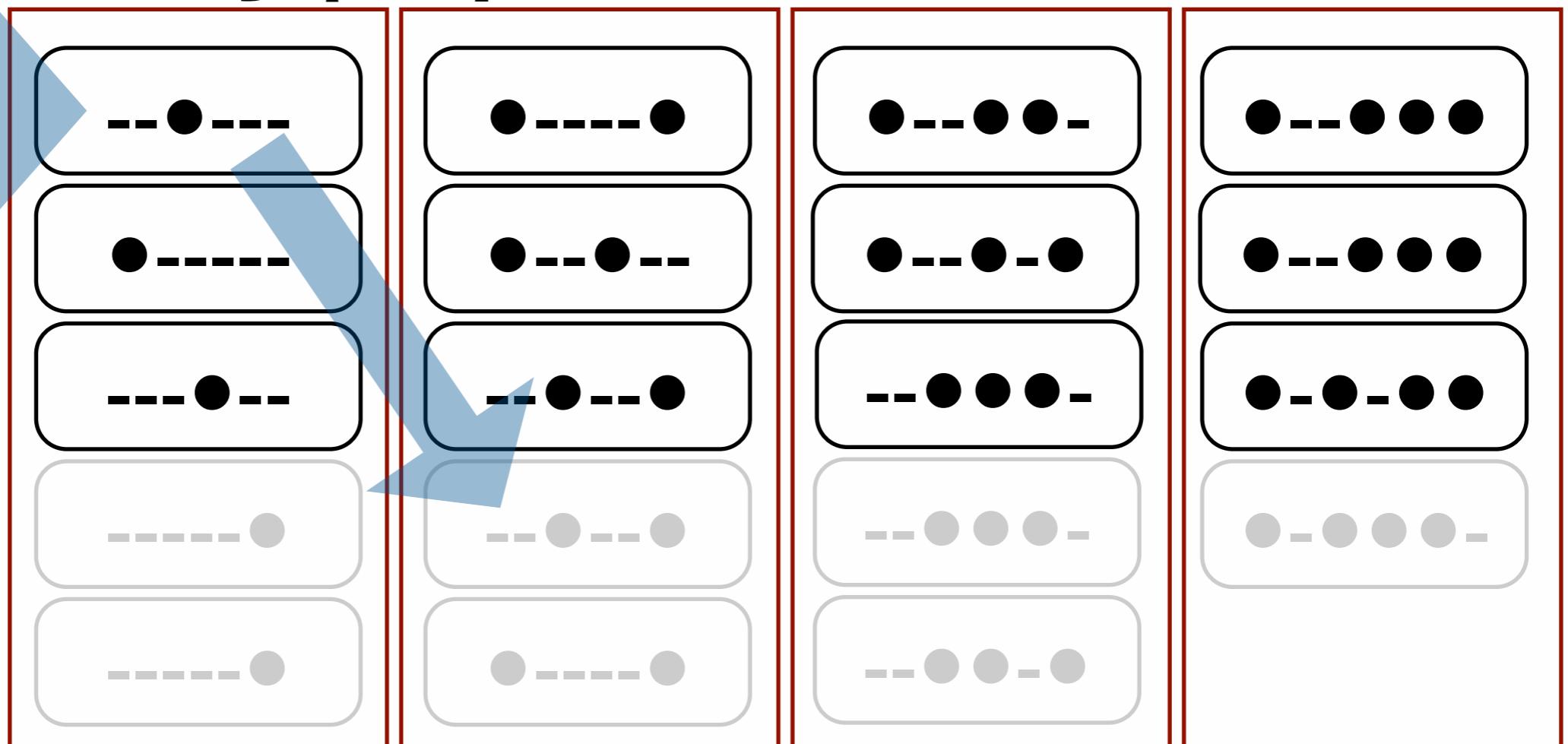


- 同じグループで比較、枝刈り: ビーム探索
- $\mathcal{O}(2^n)$ の項を $\mathcal{O}(nb)$ へ縮小

探索アルゴリズム

```
1: procedure PHRASEDECODE( $f$ )
2:   for  $c = 0 \dots J - 1$  do           ▷ 入力文の長さ  $|f| = J$ 
3:     PRUNE( $\tilde{Q}_c$ )                 ▷ 枝刈り
4:     for  $n \in \tilde{Q}_c$  do           ▷  $\tilde{Q}_c$  のノード  $n$  を列挙
5:       for  $p \in \Phi$  do           ▷ フレーズペア  $p$  を列挙
6:          $\tilde{Q}_{c+|p|} \leftarrow \tilde{Q}_{c+|p|} \cup \{n \circ p\}$  ▷ 新しいノード
7:       end for
8:     end for
9:   end for
10:  return BACKWARD( $\tilde{Q}_J$ )          ▷ 後ろ向きに探索
11: end procedure
```

探索エラー “404 Not Found”



本当は、この経路が良かったのに...

$$\text{score}(\text{---●---●}) + \text{estimate}(\text{●●-●●●-})$$

先読み



機械 翻訳 について 勉強 したい。

machine translation

about

study

I want to

.

mechanism translate

regarding

learn

he wants to

?

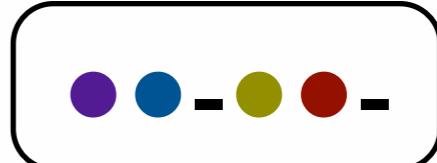
machine translation

wants to study

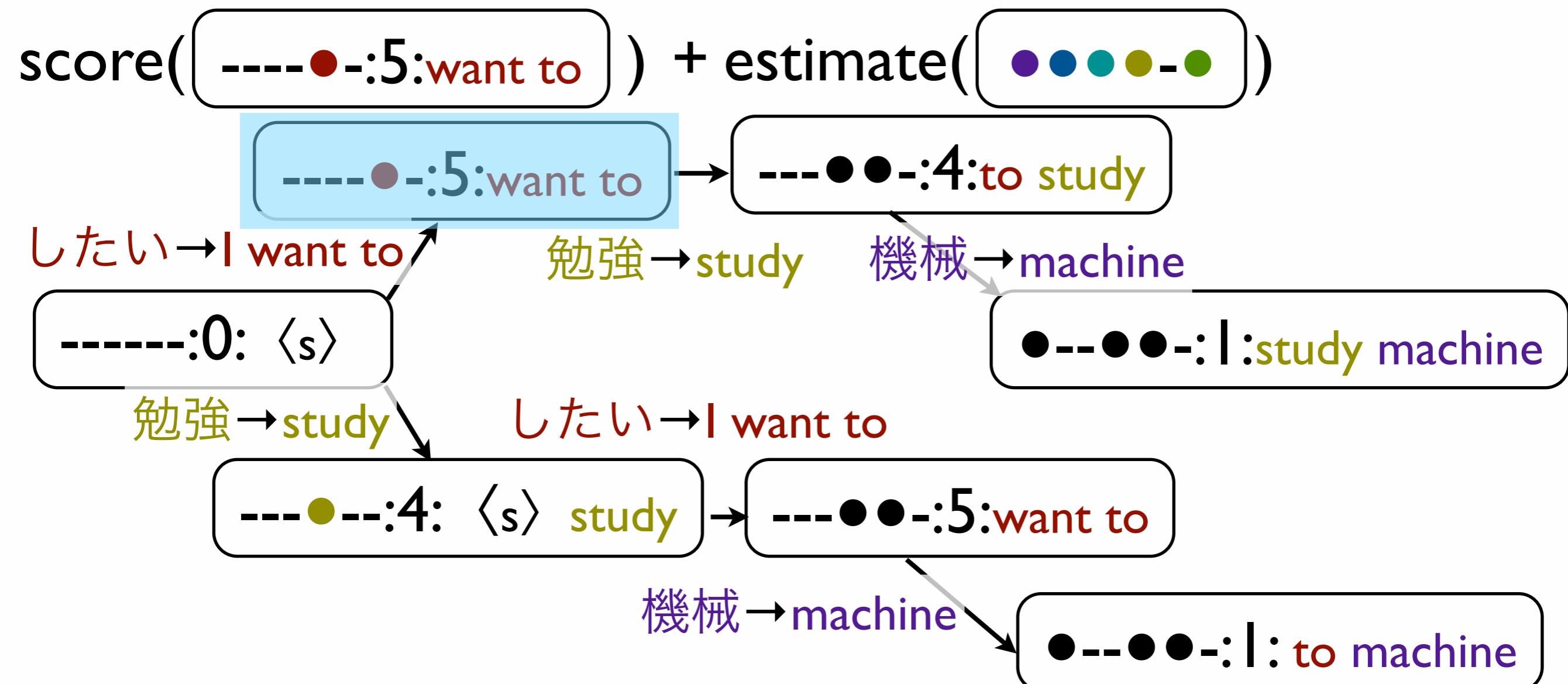
!

$$Pr(\text{machine}) Pr(\text{translation} | \text{machine})$$

$$\times Pr(\text{機械 翻訳} | \text{machine translation})$$

- 列挙されたフレーズからestimate()を計算
- 各フレーズを文と考え、並び替えを無視

先読み付き探索



- score + estimateでスコア計算

まとめ

- フレーズベースなデコード
- 原言語の並び替え+目的言語を文頭から生成
- 「状態」に素性の計算に必要な情報
- ヒューリスティックな探索空間の制約+ヒューム探索による近似

その他の手法

- 貪欲法、線形計画法(Germann et al., 2002)
- ラグランジュ緩和法(Chang and Collins, 2011)

参考文献

- Yin-Wen Chang and Michael Collins. 2011. Exact decoding of phrase-based translation models through lagrangian relaxation. In *Proceedings of the 2011 Conference on Empirica Methods in Natural Language Processing*, pages 26-37, Edinburgh, Scotland, UK., July. Association for Computational Linguistics.
- Ulrich Germann, Michael Jahr, Kevin Knight, Daniel Marcu and Kenji Yamada. 2001. Fast decoding and optimal decoding for machine translation. In *Proceedings of 39th Annua Meeting of the Association for Computational Linguistics*, pages 228-235, Toulouse, France, July. Association for Computational Linguistics.
- Kevin Knight. 1999. Decoding complexity in word-replacement translation models. *Computational Linguistics* 25:607-615, December.
- Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003 Statistical pharse-based translation. In *Proc. of HLT-NAACL 2003*, pages 48-54, Edmonton, May-June.