

木構造に基づく機械翻訳

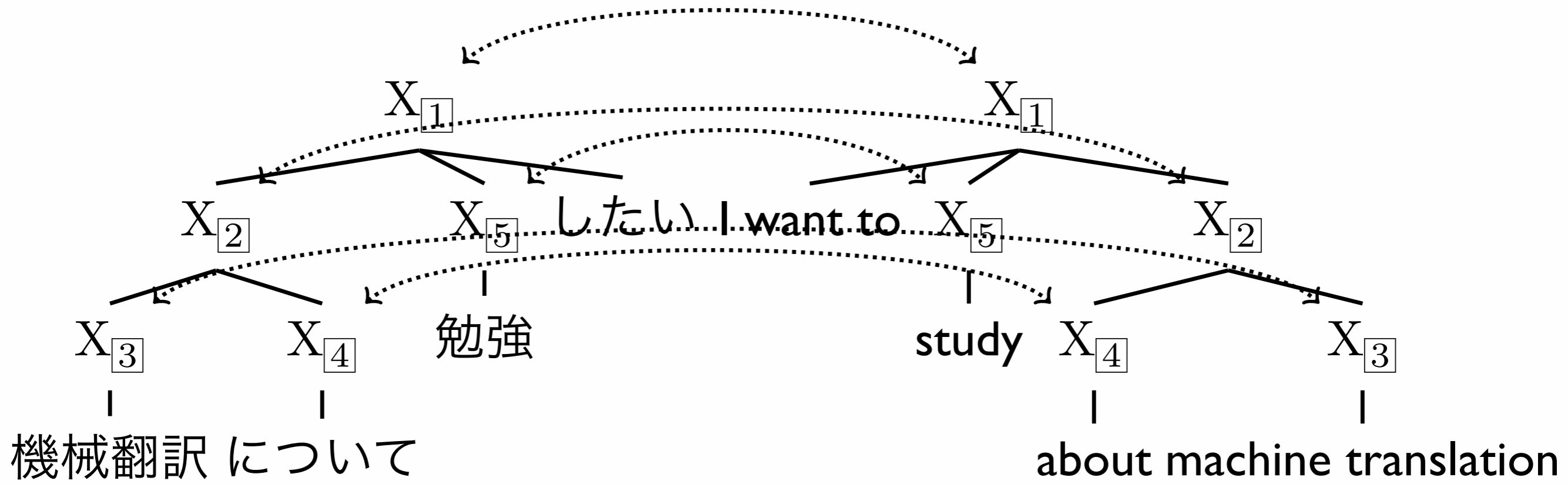
+最近のトピック

渡辺太郎

taro.watanabe at nict.go.jp



<https://sites.google.com/site/alaginmt2014/>



そういえば...

知識に基づく機械翻訳

機械翻訳について勉強したい。

解析

機械 翻訳 について 勉強 したい。

変換

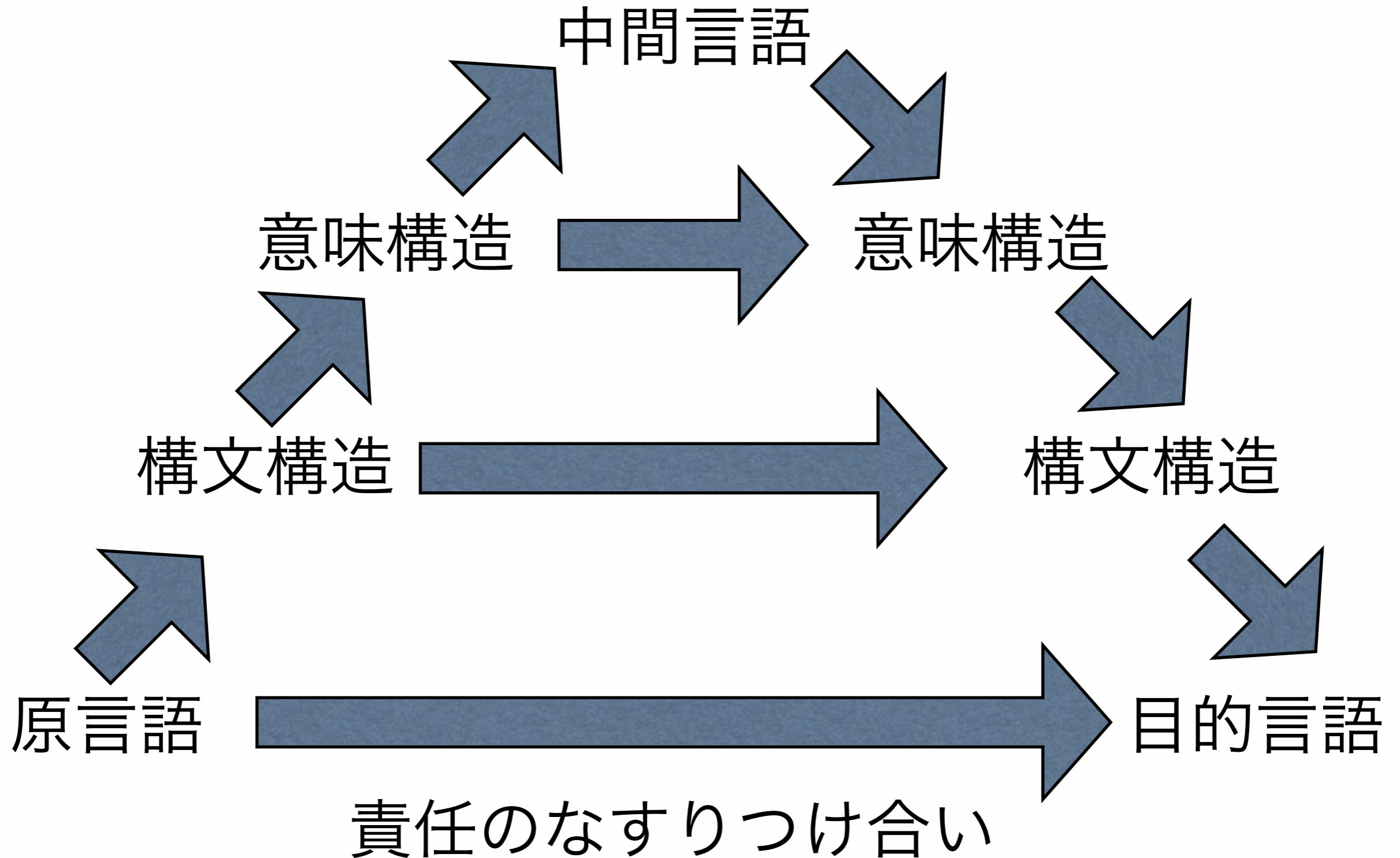
[want] [study] [about] [machine] [translation]。

生成

I want to study about machine translation.

責任分担

三角形



限界

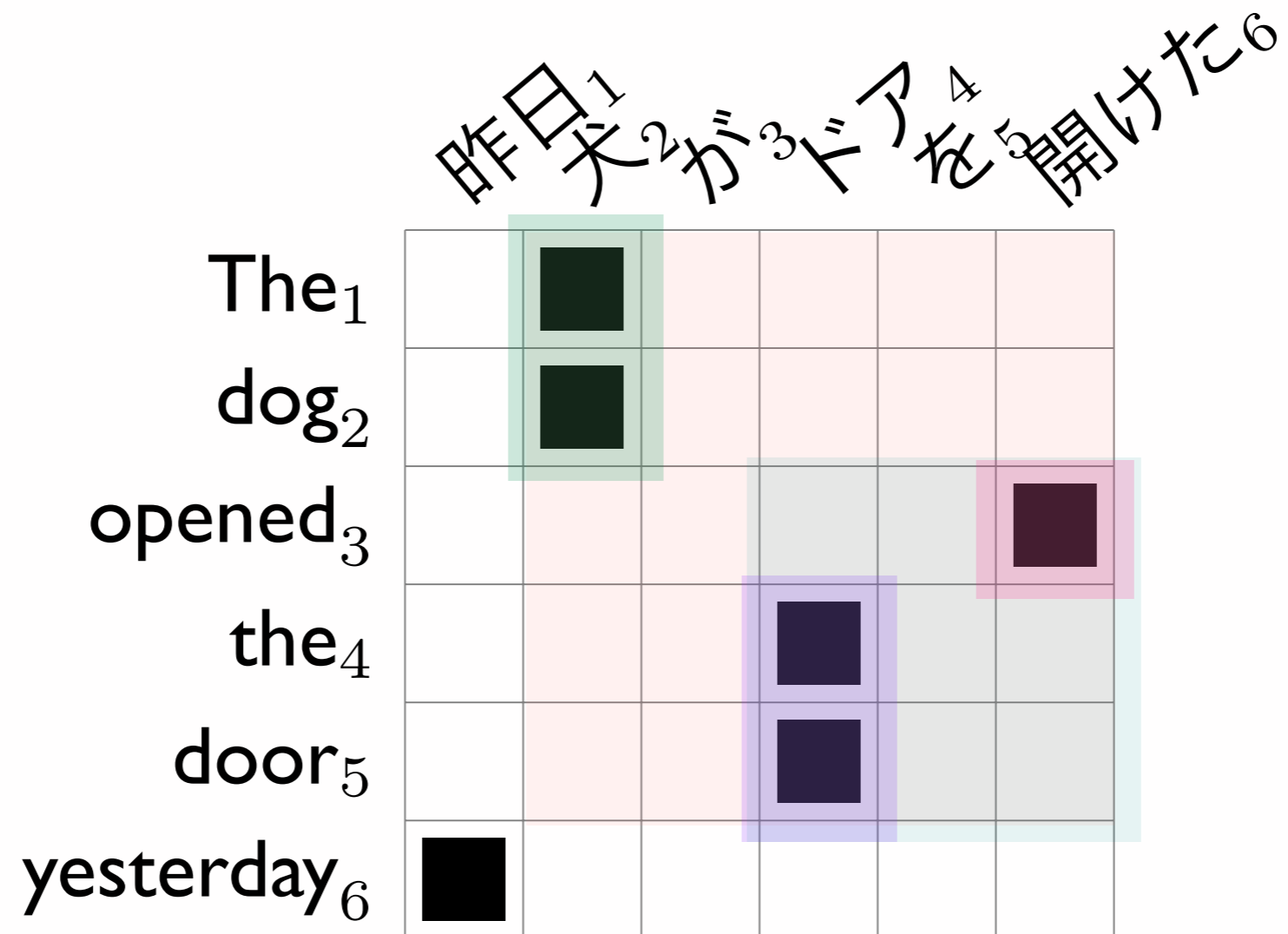
機械	↔	machine
翻訳		translation
勉強		study, studied, studies, learn, etc.
X は Y したい		X want{s, ed} to Y
X したい		{I, he, she}want{s,ed} to X

⋮



網羅できない

自動的に学習



- 単語アライメント、フレーズペア、ルール

他にも...

対訳データの活用



“analogy thinking”

対訳データがたくさんある

↔ There are many bilingual data

たい焼きがほしい

↔ I want taiyaki

対訳データがほしい

↔ I want bilingual data

⋮

たい焼きがたくさんある

➔ There are many taiyaki

用例に基づく機械翻訳: 翻訳メモリとして応用



統計モデル

$$\hat{e} = \arg \max_e Pr(\mathbf{f}|\mathbf{e}) Pr(\mathbf{e})$$

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

- 生成モデルおよび識別モデルにより、
モデル、探索、最適化の各問題へ帰着

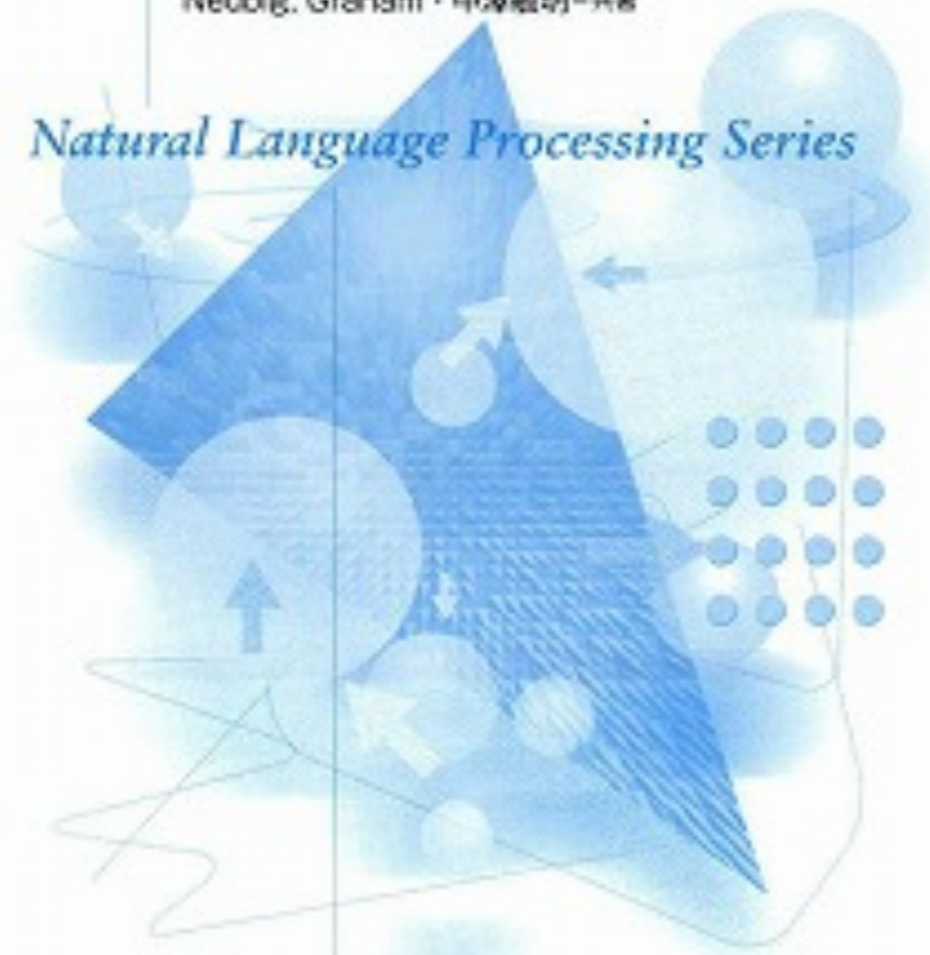
自然言語処理シリーズ 4

機械翻訳

奥村 学=監修

渡辺太郎・今村賢治・眞沢秀人
Neubig, Graham・中澤敏明=共著

Natural Language Processing Series



コロナ社

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

最近のトピックス

並び替えモデル

機械翻訳について勉強したい。

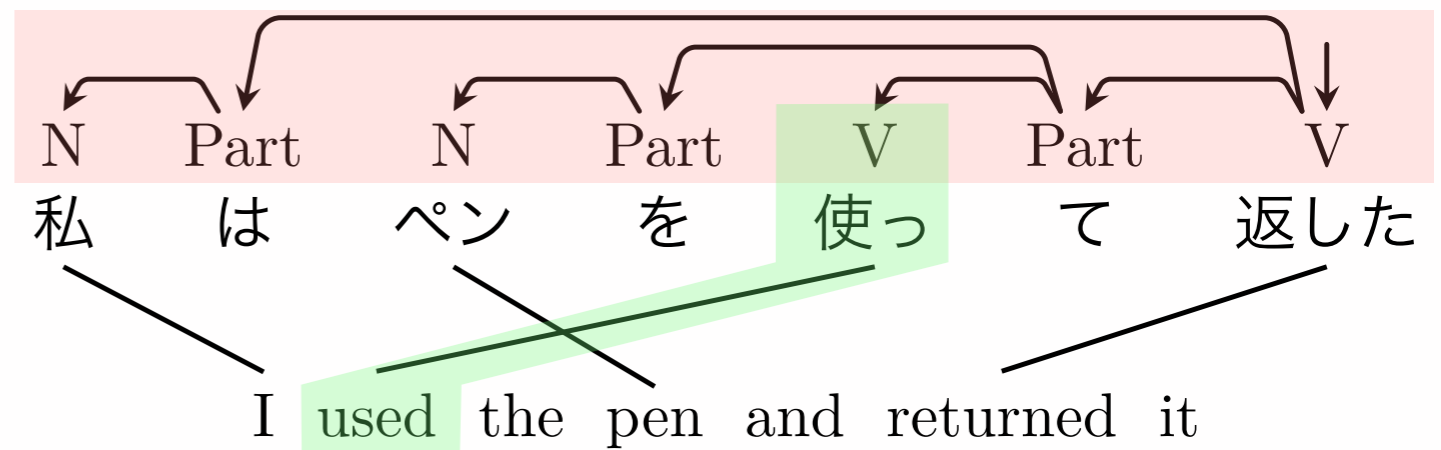
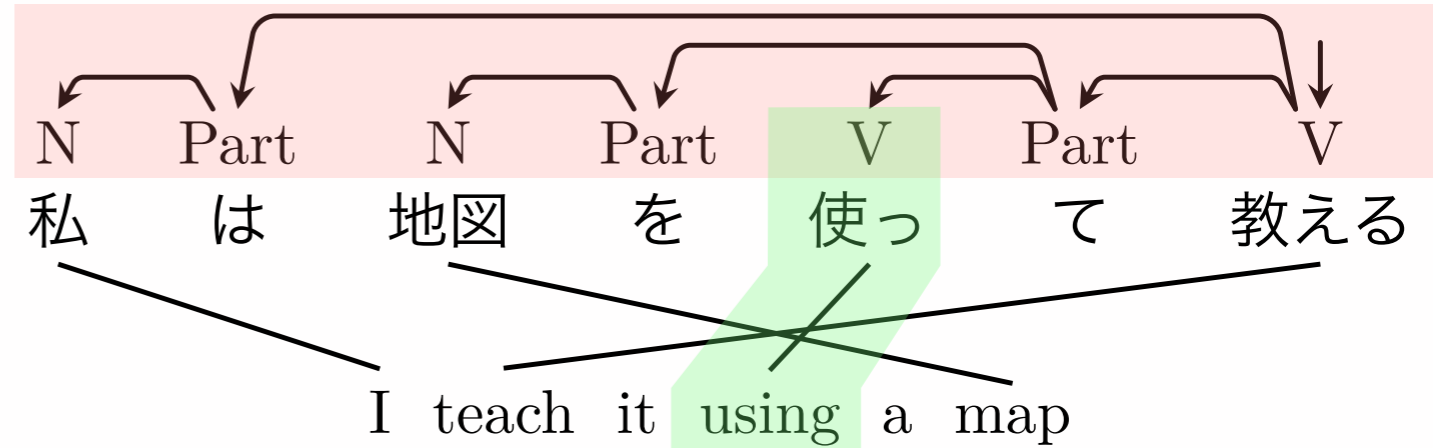
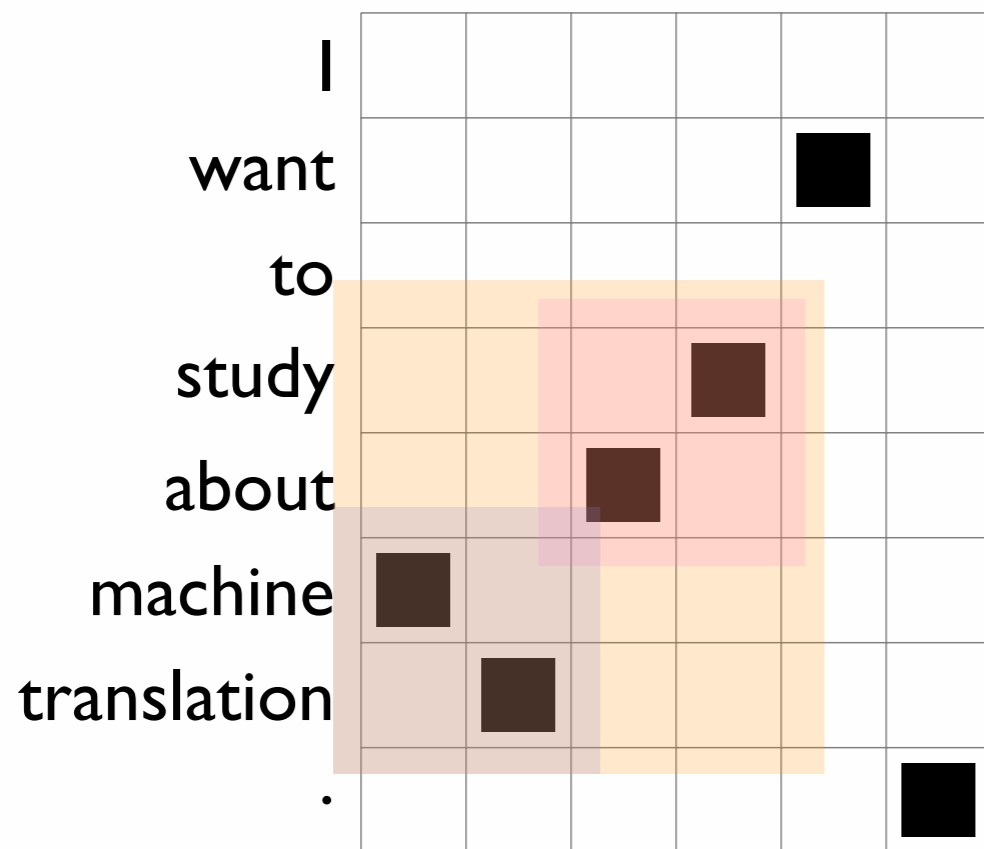
したい 勉強 について 機械翻訳。

I want to study about machine translation.

- 特に、文法的に異なる言語対?

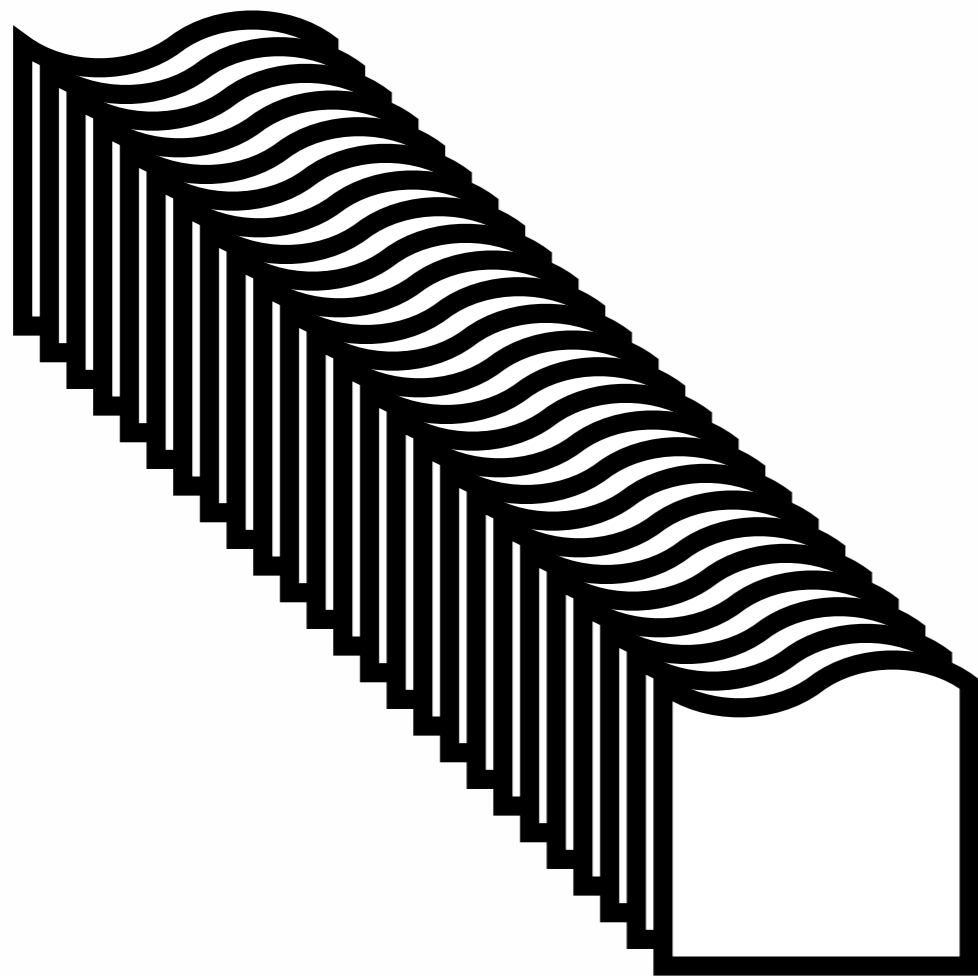
教師なし学習

機械翻訳について勉強したい。

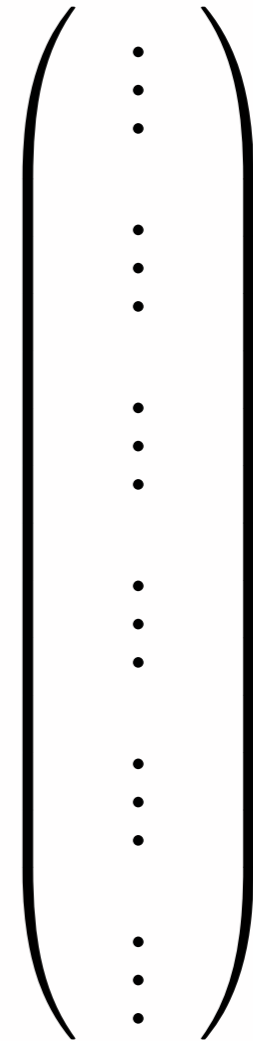


- 「単語アライメント」は必要?
- 構文解析は(翻訳にとって)最適か?

オンライン学習



$$h(f, d, e) =$$



- データが多い、高次元な素性ベクトル
- 効率のよい学習手法

アウトライン

- 木構造に基づく機械翻訳
- 構文解析、同期文脈自由文法、同期木置換文法
- 最近のトピック
- 対訳データ、並び替えモデル、教師なし学習、オンライン学習



渡辺 太郎 Graham Neubig 須藤 克仁 田村 晃裕 林 克彦