

木構造に基づく機械翻訳

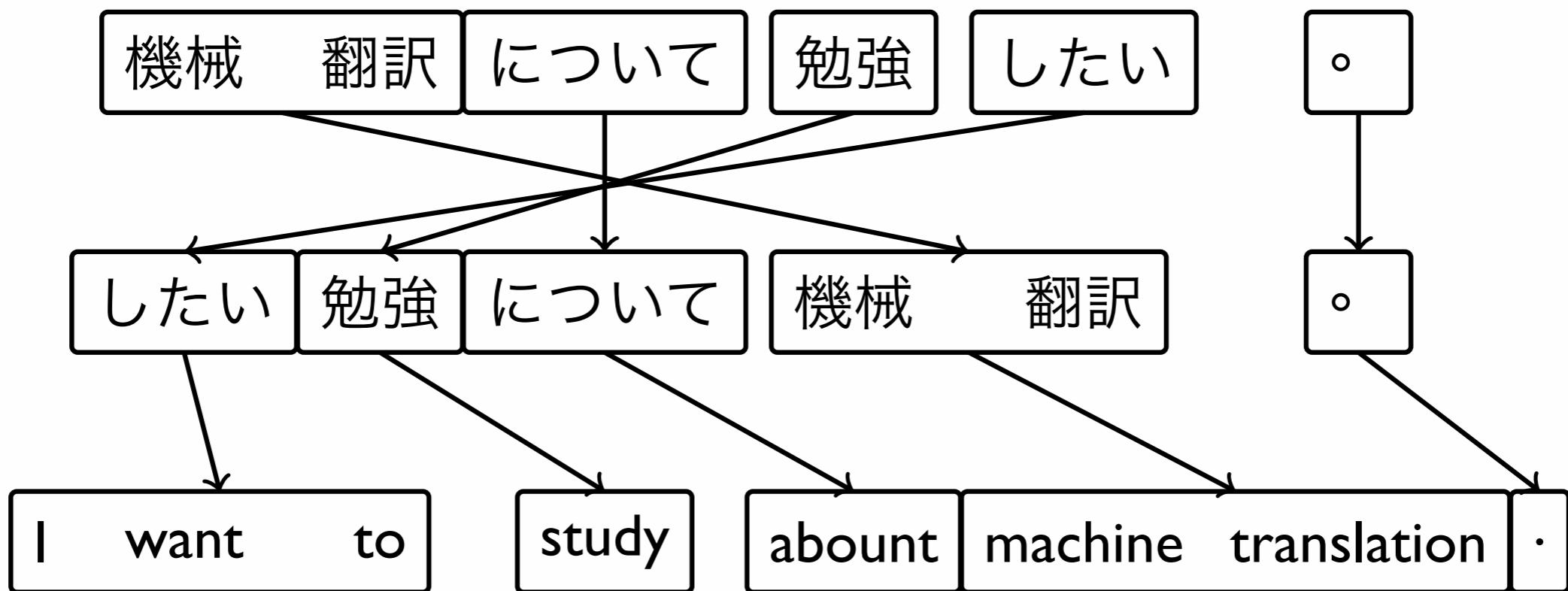
+ 最近のトピック

渡辺太郎

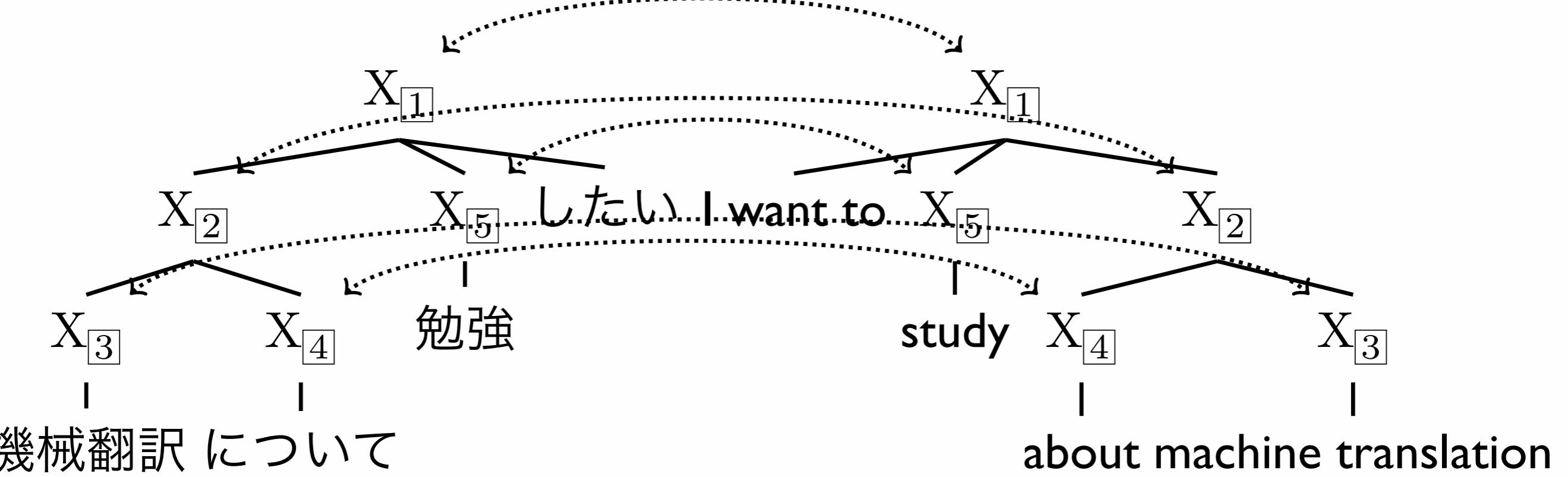
taro.watanabe at nict.go.jp



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



(Koehn et al., 2003)



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

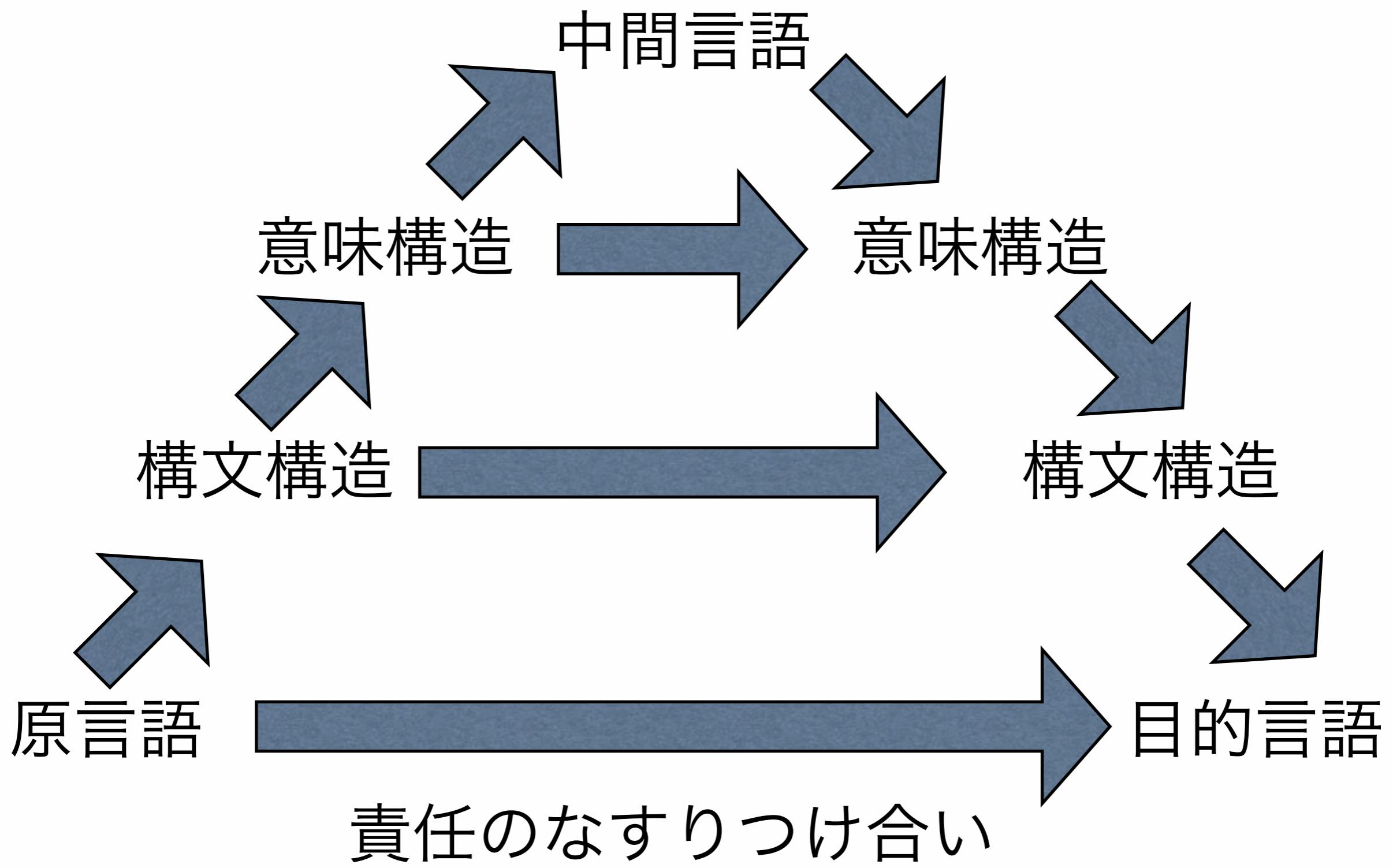
[want] [study] [about] [machine] [translation] 。

生成

I want to study about machine translation.

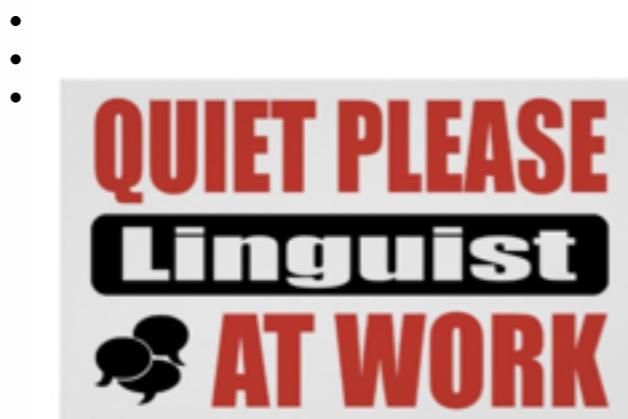
責任分担

三角形



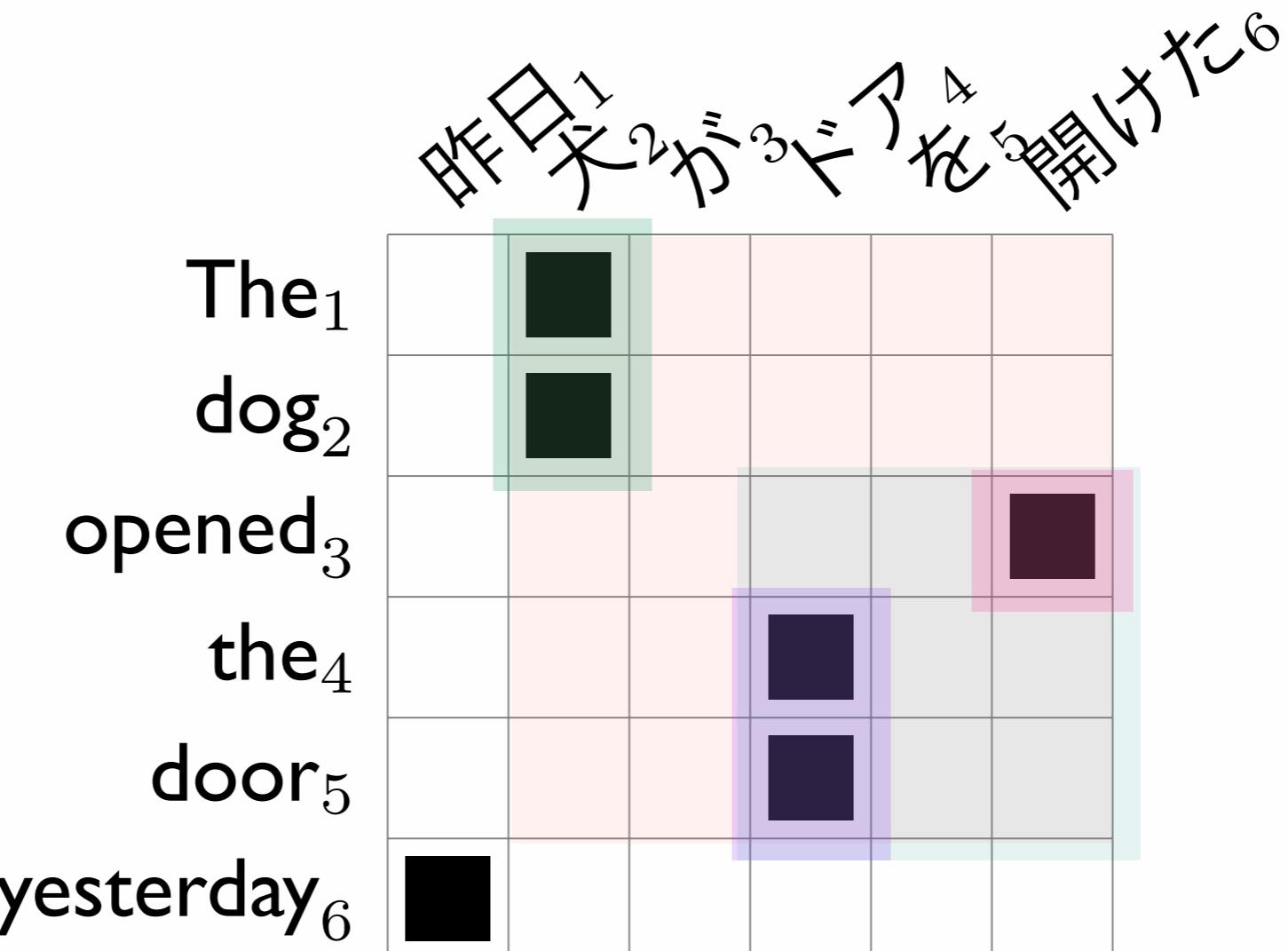
限界

機械	machine
翻訳	translsation
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

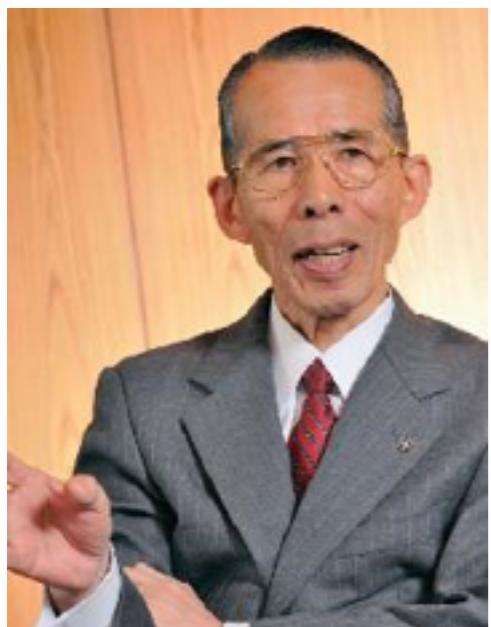
自動的に学習



- 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



“analog
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

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

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

- 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

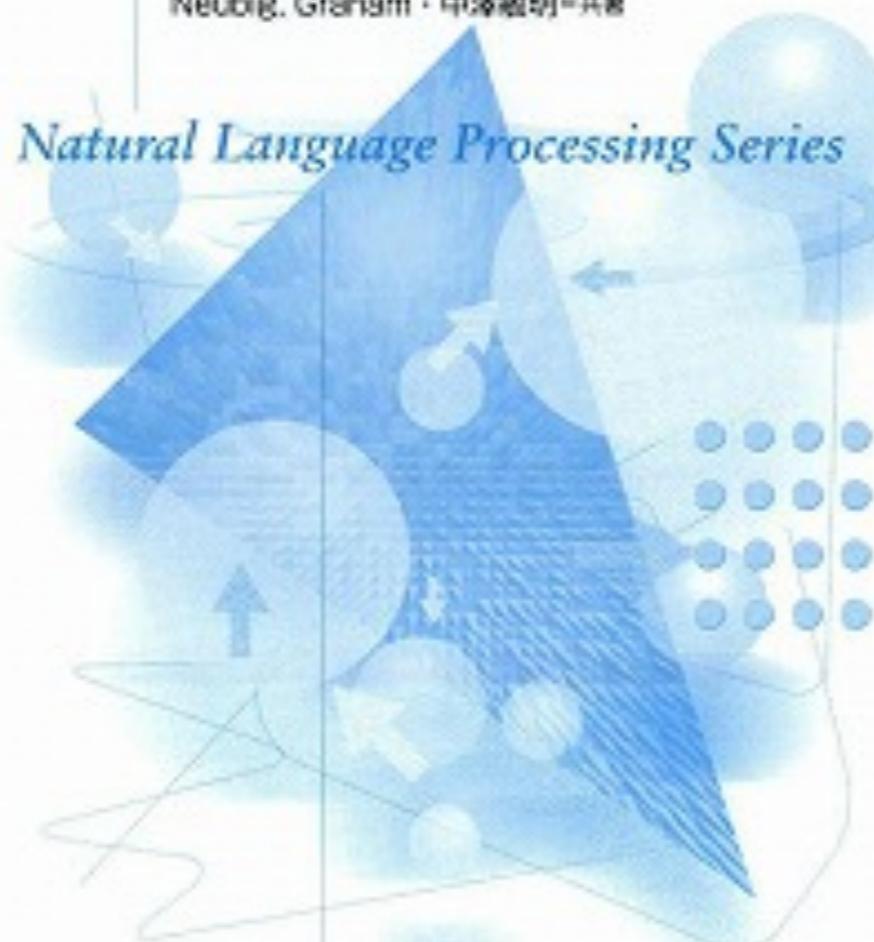
自然言語処理シリーズ 4

機 械 翻 訳

奥村 学=監修

須恵太郎・今村賛治・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

対訳データ

アンネの日記:被害、3市5区に拡大 東京の図書館、関連294冊に

毎日新聞 2014年02月22日 東京朝刊

東京都内の公立図書館で世界的ベストセラー「[アンネ](#)

の日記」と関連図書が相次いで破られた問題で、21日

現在の被害は都西部の3市5区で計294冊に上ること

が分かった。各自治体は器物損壊容疑にて立候補を

視庁に提出した。【戸上文恵、森下

More damaged Anne Frank books found at Tokyo libraries bring total count to 294

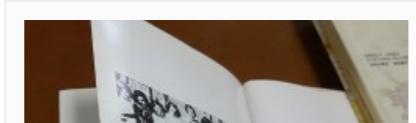
被害が確認されたのは杉並区の1More copies of worldwide bestseller "The Diary of 区の5館54冊▽練馬区の9館41 Anne Frank" and related books have been found 9冊▽豊島区の3館12冊▽武蔵野damaged at public libraries in western Tokyo, 京市の3館10冊▽東久留米市の3bringing the total number to 294 books as of Feb. 害が確認されていない都立中央図書21, it has been learned.

摩図書館（立川市）は、21日から

The confirmed damages are as follows: 119 books 開架をやめ、希望者の申告に応じて at 11 libraries in Suginami Ward, 54 books at five libraries in Nakano Ward, 41 books at nine libraries

杉並区によると、今月3日に隣接In Nerima Ward, 39 books at three libraries in 被害の連絡を受けて調査した際にはShinjuku Ward, 12 books at three libraries in Toshima Ward, nine books at two libraries in the city of Musashino, 10 books at three libraries in the city of Nishitokyo, and 10 books at three libraries in the city of Higashikurume. Affected municipalities have filed for damages with the Metropolitan Police Department.

Meanwhile, as of Feb. 21, the metropolitan Chuo



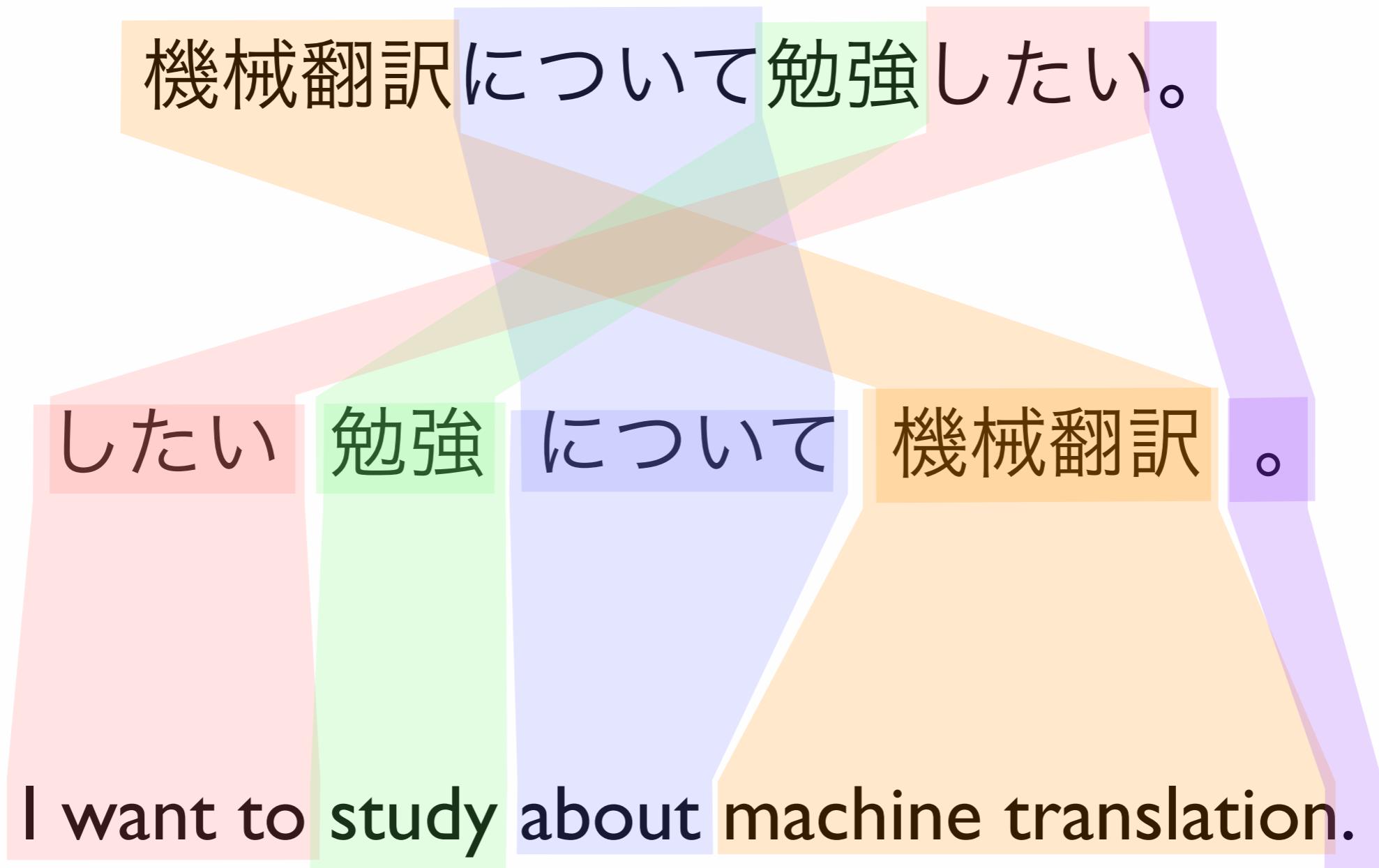
A damaged copy of "The Diary of Anne Frank" is seen at a library in Suginami Ward, Tokyo. (Photo courtesy of the Suginami Ward government)

[拡大写真](#)

Seafood		Beef	
49. 千 場	内	Sweet & Sour Shrimp	9.25
50. 京 道	内	Imperial Shrimp	9.25
51. 京 香	内	Stechuan Shrimp	9.25
52. 京 香	内	Cashew Shrimp	9.25
		<small>shrimp with cashew nuts and crunchy rice</small>	
53. 京 香	内	Shrimp with Snow Peas	9.25
54. 京 香	内	Kung Pao Shrimp	9.25
55. 京 香	内	Garlic Shrimp	9.25
56. 京 香	内	Peking Shrimp	9.25
57. 京 香	内	Shrimp in Lobster Sauce	9.25
		<small>This one has a creamy rich taste because lobster sauce has some fat</small>	
58. 中 場	内	Shrimp in Curry Sauce	9.25
59. 京 香	内	Shrimp in Black Bean Sauce	9.25
60. 中 場	内	Imperial Feast	9.25
		<small>This dish includes Imperial Duck and three pieces of fish fillet</small>	
61. 京 三 里	内	Kung Pao Three Delights	9.25
62. 什 神 金 合	内	Shu Gom Wonton Deluxe	10.95
		<small>Shrimps made meat, juicy chicken, A.J.G. pork mixed with vegetables, served with garlic flavor sauce and topped with wontons</small>	
63. 千 場	内	Hot Braised Shrimp	11.95
64. 京 香	内	Sauted Shrimp	11.95
65. 京 香	内	Sweet & Pungent Shrimp	11.95
		<small>Shrimp stir-fried in a spicy batter and sautéed in a special sauce, makes this dish one of our most wanted</small>	
66. 京 香	内	Shrimp Basket	11.95
		<small>Six-peeled crispy shrimp are served in a Crispy Holder. All you have to do is eat them</small>	
67. 京 香	内	Salt & Pepper Shrimp (with shell)	11.95
68. 京 香	内	Orange Shrimp	11.95
69. 京 香 海	内	Seafood Volcano	12.95
		<small>An assortment seafood like scallops, shrimp, mussels, squid and of course vegetables, served on a bed of lettuce</small>	
70. 京 香 海	内	Seafood Basket	12.95
		<small>Just like the one above but will create your own cooking hot plate for a crisp salad holder</small>	
71. 京 香	内	Walnut Shrimp	12.95
72. 千 場 場	内	Scallop with Shrimp	12.95
73. 京 香 千 場	内	Scallop with Mixed Vegetable	12.95
74. 京 香 千 場	内	Kung Pao Scallop	12.95
75. 京 香 兼 刀	内	Steamed Fish Fillet	12.95
76. 京 香 兼 刀	内	Stechuan Fish Fillet	12.95
77. 京 香 兼 刀	内	Fish Fillet with Brown Sauce	12.95
78. 京 香 兼 刀	内	Mongolian Fish Fillet	12.95
<i>J Hot & Spicy</i>		<i>J Hot & Spicy</i>	
79. 京 香 内	内	Beef Broccoli	7.75
80. 京 香 内	内	Kung Pao Beef	7.75
81. 京 香 内	内	Beef with Snow Peas	7.75
82. 京 香 内	内	Beef in Curry Sauce	7.75
83. 京 香 内	内	Beef Pepper Steak	7.75
		<small>Beef pepper steak is a piece of meat marinated with pepper sauce</small>	
84. 京 香 内	内	Beef & Chinese Black Mussels	7.95
		<small>Beef with black mussels and biscuits with taste of beef in delicious sauce taste</small>	
85. 京 香 内	内	Mongolian Beef	7.95
		<small>Beef stir-fried with onions, served with scallions</small>	
86. 京 香 内	内	Shanghai Beef	7.95
		<small>Beef stir-fried with carrots</small>	
87. 京 香 内	内	Orange Beef	8.95
88. 京 香 内	内	Sesame Beef	8.95
89. 京 香 内	内	Beef Volcano	8.95
		<small>This is a strong one, slice of juicy tender beef mixed with a mixture of vegetables, served on a scalding plate</small>	
90. 京 香 内	内	Beef Basket	8.95
		<small>Beef stir-fried with an assortment vegetables served on an oval plate</small>	
91. 京 香 内	内	Beef Banan	8.95
		<small>Beef stir-fried with an assortment spicy beef for someone who like it hot</small>	
92. 京 香 内	内	Beef Oyster Sauce	8.95
93. 京 香 内	内	Beef with String Bean	8.95
94. 京 香 内	内	Ginger Beef	8.95
95. 京 香 兼 刀	内	Black Pepper Filet Mignon	9.95
<i>Pork</i>		<i>Pork</i>	
96. 京 香 内	内	Kung Pao Pork	6.95
97. 京 香 内	内	Mushroom Chan Taki	6.95
98. 京 香 外 地	内	Bar-B-Q Pork with Broccoli	6.95
99. 京 場 内	内	Sweet & Sour Pork	6.95
100. 京 場 内	内	Garlic Pork	6.95
101. 京 香 兼 刀	内	Pork Sausage	6.95
102. 京 香 兼 刀	内	Snow Peas Char-Siu	6.95
		<small>This is the American red barbecue pork mixed with fresh snow peas</small>	
103. 京 場 内	内	Twin Cooked Pork	6.95
<i>Egg Foo Young</i>		<i>Egg Foo Young</i>	
104. 京 香 家 事	内	Shrimp Egg Foo Young	7.50
105. 京 香 家 事	内	Chicken Egg Foo Young	6.50
106. 京 香 家 事	内	Pork Egg Foo Young	6.50
107. 京 香 家 事	内	Beef Egg Foo Young	6.50
108. 京 香 家 事	内	Vegetable Egg Foo Young	6.50

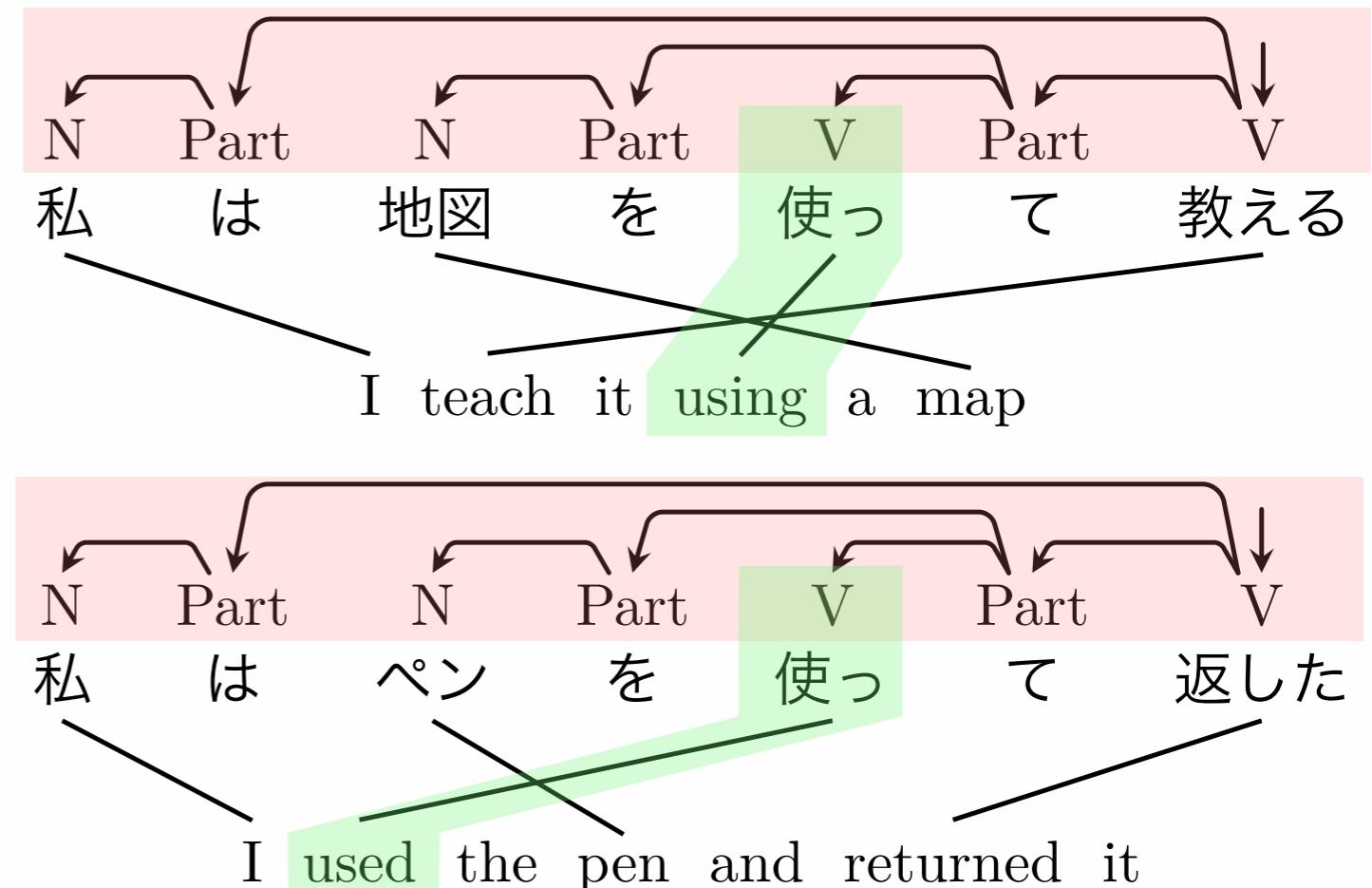
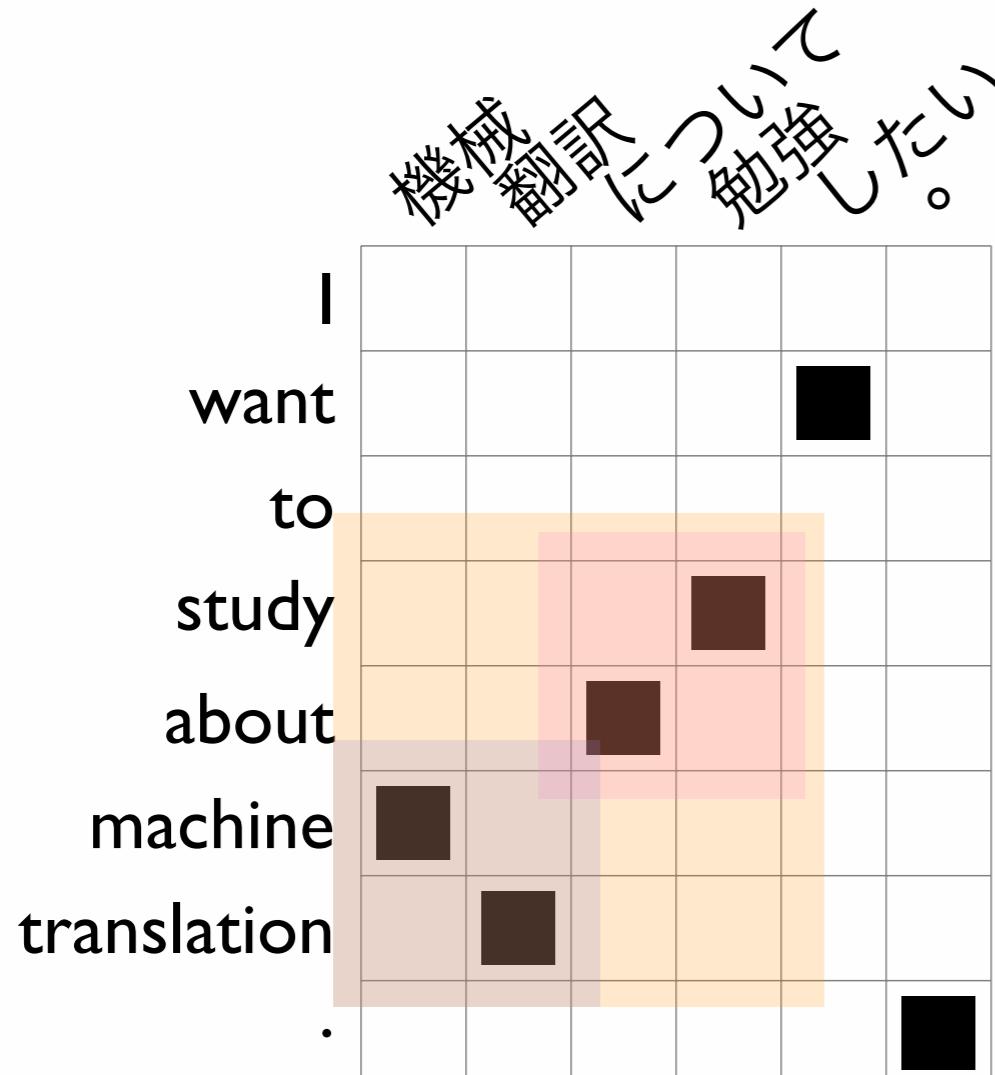
● そもそも、どのように対訳データを収集?

並び替えモデル



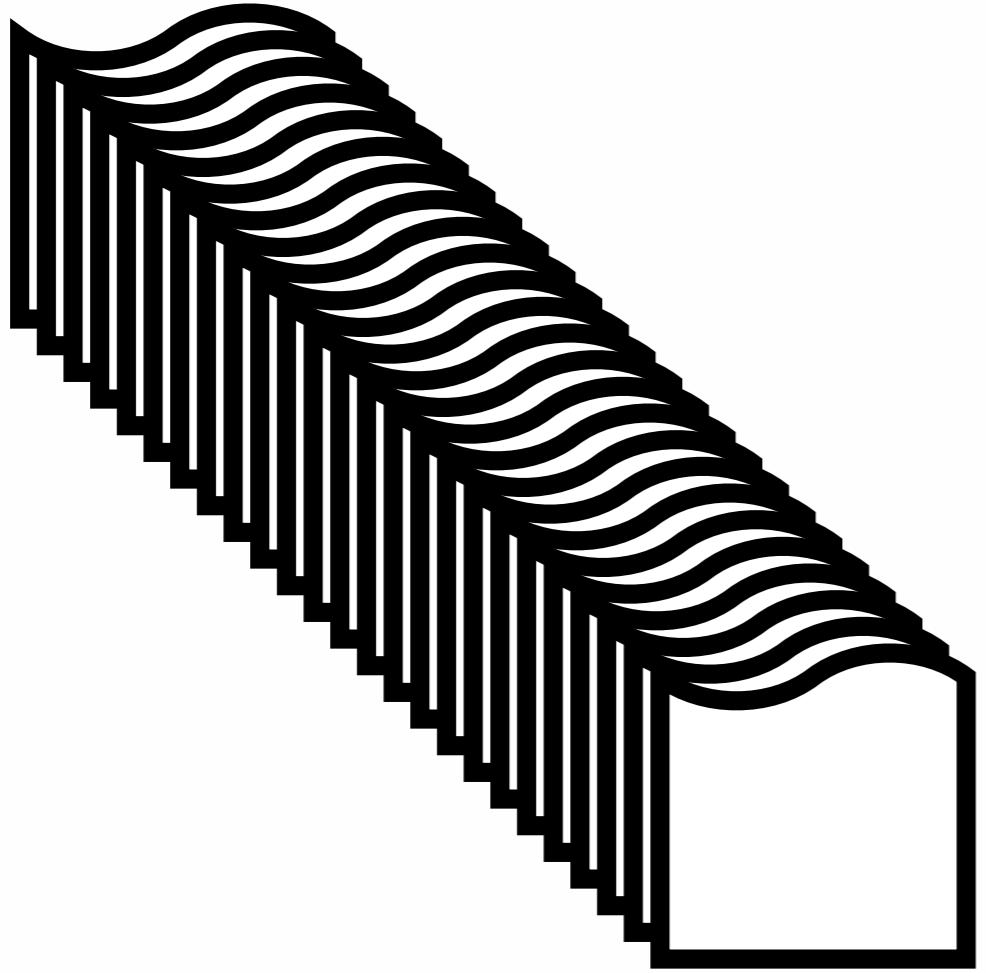
- 特に、文法的に異なる言語対?

教師なし学習



- 「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

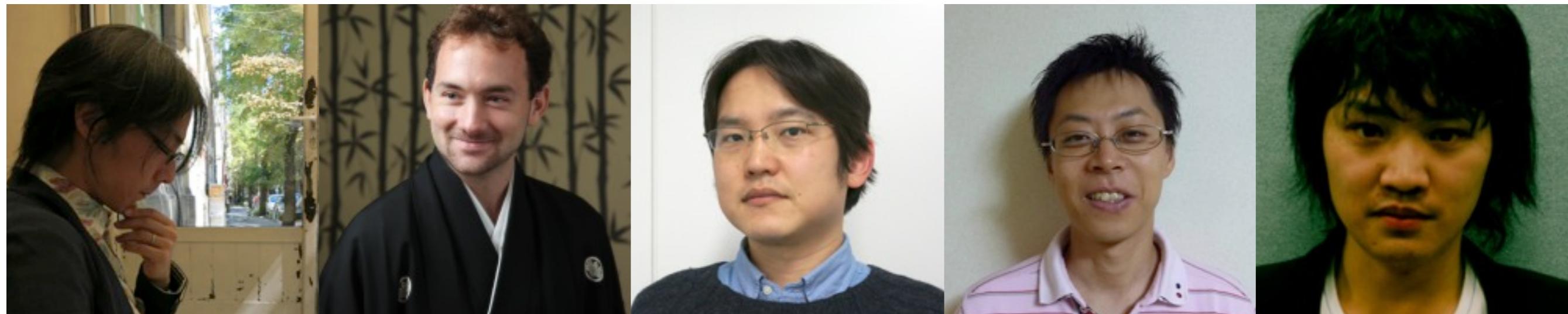


$$h(f, d, e) = \begin{pmatrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{pmatrix}$$

- 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