

機械翻訳

渡辺太郎

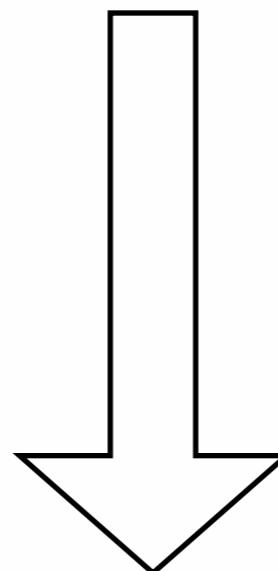
taro.watanabe at nict.go.jp



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

原言語

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.

目的言語

合肥市人大常委会原副主任郭超涉腐上千万获无期

2013年10月25日 09:50:16 来源：中安在线 + 分享到： 15

A- A+

昨日，六安市中级人民法院对合肥市人大常委会原副主任郭超受贿、贪污、滥用职权一案进行公开宣判，郭超涉案上千万一审被判无期徒刑。

法院审理查明，郭超在2001年至2011年十年间，利用职务之便，为他人在减免房产税、土地使用税、土地出让、缓减免交项目建设配套费、工程建设、变更土地用途及工程项目审批等方面谋取利益，非法收受相关企业和个人人民币569万，港币20万元；通过国有公司为其实际占用的公司垫资代建工程并让国有公司高价回购的方式，非法侵吞国有资产675.0117万元人民币；滥用职权造成国有资产损失198.57万元人民币。

六安市中院以受贿罪、贪污罪和滥用职权罪判处被告人郭超无期徒刑，剥夺政治权利终身，没收个人财产人民币130万元，并依法追缴违法所得人民币1244.0117万元、港币20万元。宣判后，郭超当庭表示服判，不上诉。（通讯员 李传丽记者 方荣刚）

+ 分享到： 15

[【关闭】](#) [【打印】](#) [【纠错】](#) [责任编辑：卢俊宇]

Former deputy director of the Standing Committee of Hefei Guo Chao involved in corruption of millions is no period

October 25, 2013 09:50:16

Source: Zhong An Online



Share to:



0



Yesterday, Lu'an, Hefei Intermediate People's Court on the former deputy director of the Standing Committee Guo Chao, bribery, embezzlement, abuse of a case of an open verdict, Guo Chao involving tens of millions of first instance sentenced to life imprisonment.

Court found, Guo Chao in 2001-2011 decade, the use of his position, for others in the property tax relief, land use tax, land transfer, mitigation projects exempt from ancillary fees, construction, land use change and project approval, etc. to seek benefits related businesses and individuals illegally accepting RMB 5.69 million, \$ 200,000; through its state-owned companies on behalf of the actual occupation Loaning construction companies and state-owned companies to buy expensive way illegal misappropriation of state assets 675.0117 ten thousand yuan; abuse causing loss of 1.9857 million yuan of state assets.

Lu'an City Intermediate People's Court for taking bribes, embezzlement and abuse of power and sentenced the defendant Guo Chao life imprisonment and deprivation of political rights for life and confiscation of personal property of 1.3 million yuan, and shall pursue the illegal income RMB 1244.0117 million and HK \$ 0.2 million .

After the verdict, Guo Chao court said the service contract, not to appeal. (Reporter correspondent Li Zhanli Fangrong Gang)

Share to: 0

[\[Close\]](#) [\[Print\]](#) [\[correction\]](#) [Editor: Lu Junyu]

(translated by <http://translate.google.com>)

詳細表示:CN101933412(A) - Mozilla Firefox
 ファイル(E) 編集(E) 表示(V)履歴(S) ブックマーク(B) ツール(T) ヘルプ(H)
 詳細表示:CN101933412(A) +
 105.178/epe/members/fan_show_abstract_app1.php
 ☆ □ Google 検索

Japio 世界特許情報検索サービス

| 著権明細 | マニュアル | しおり表示 | 検索式表示 | ログアウト

検索 > [PO:[20110101 TO 20111231]] > 詳細表示

2,706,626件中1件目

1 2 3 4 5 6 7 8 9 10 次へ>

ダウンロード
EPO Patent Register

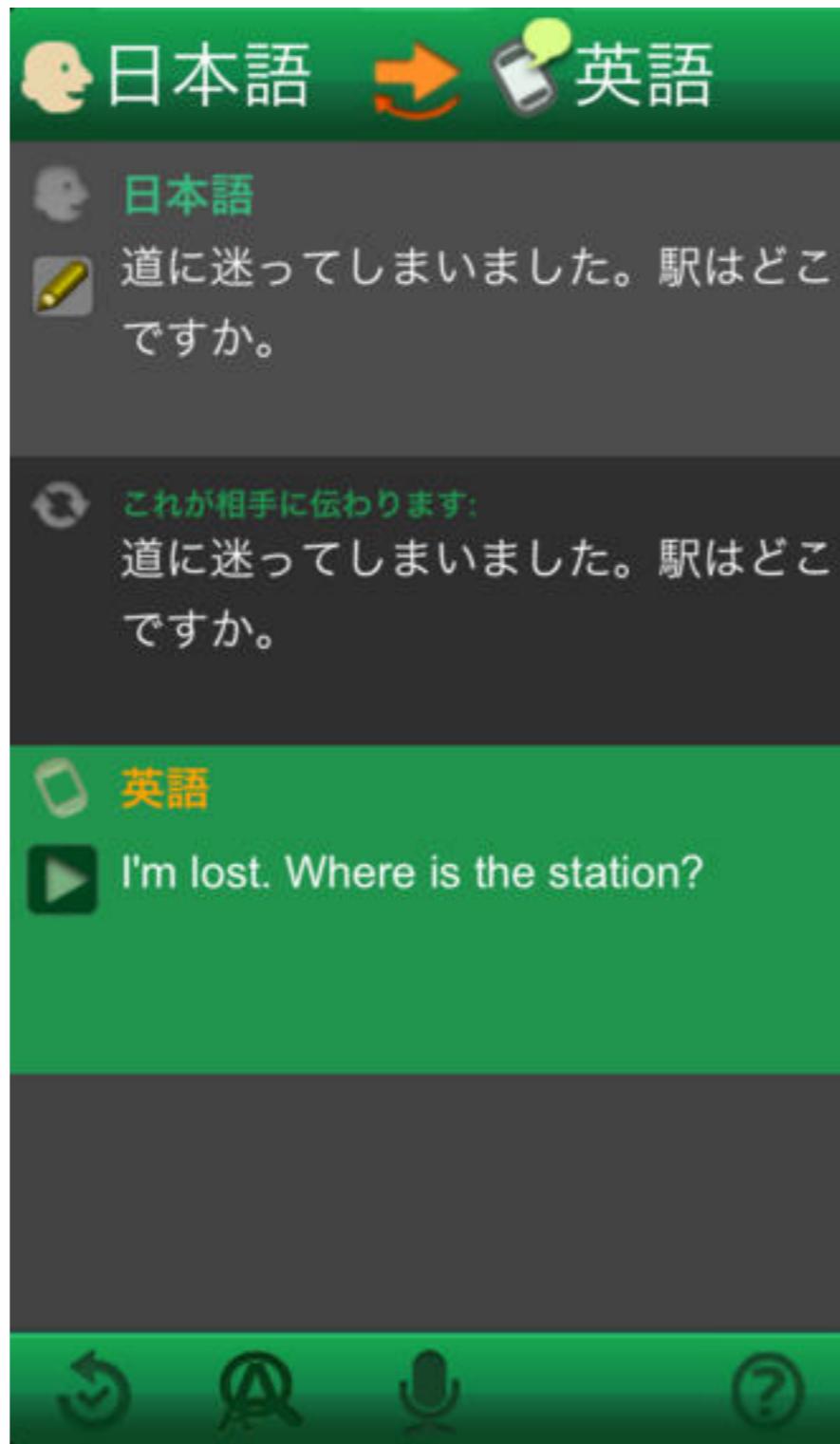
1.
Seedling-raising disk soil compacting method
【MT】 苗木を上げるディスク土圧方法
 出願人: JIANGSU YUNMA AGRICULTURAL MACHINERY CO., LTD.
 発明者: JI SHUNZHONG; LU DONGXING; LI QUN; CHEN SONGHUI
 出願番号: CN20101220302 [2010-07-02]
 文献番号: CN101933412(A) [2011-01-05] 公報PDF フロントPDF
 IPC: A01B29/00
 ECLA:
 優先権主張番号: CN20101220302
 特許 (英):
 The invention discloses a seedling-raising disk soil compacting method. After the method is adopted, because an automatic subsoil compacting device is installed on a seedling-raising disk flow line and comprises a compacting plate corresponding to a seedling-raising disk, the seedling-raising disk on a guide rail of the seedling-raising flow line can pause when running to the bottom of the compacting device, the compacting plate can be downwards pressed, the subsoil of the seedling-raising disk can be compacted by pressure heads on the compacting plate, then the compacting plate is reset, and the seedling-raising disk continues to advance to enter the next working procedure; and meanwhile, because the number of the pressure heads of the compacting plate corresponds to the specification of the seedling-raising disk, the whole seedling-raising disk can be entirely compacted by once downwards pressing, the efficiency is quite high, the structure of the whole device is not quite complicated, and the manufacturing cost is also not high. Manpower is released from manual operation through the compacting device and the application thereof, thereby not only improving the operation efficacy, lowering the agricultural cost and ensuring the basic seedling requirement of the seedling-raising disk, but also enabling the seedling-raising flow line operation to upgrade to a new height and really realizing the target of whole-process automated flow-line seedling-raising operation.

特許 (日):
 【MT】 その発明は苗木を上げるディスク土圧方法を示す。自動の心土を圧縮する装置が苗木を上げるディスク流動線にインストールされ、苗木を上げるディスクに対応する圧縮するプレートを含むので、方法が採用された後、圧縮する装置の底に及ぶ場合、苗木を上げる流動線のガイドレール上の苗木を上げるディスクは停止することができる。圧縮するプレートは押し下げることができる。苗木を上げるディスクの心土は、圧縮するプレート上の圧力ヘッドによって圧縮することができる。その後、圧縮するプレートはリセットされる。また、苗木を上げるディスクは次の動作する処理を入力するために進み続ける;また、一方、圧縮するプレートの圧力ヘッドの数が苗木を上げるディスクの規格に対応するので、完全に全体の苗木を上げるディスクによって一度下方へ圧縮することができる。プレッシング、効率は全く高い。全体の装置の構造は全く複雑ではない。また、生産費はさらに高くない。マンパワーは、手動操作から圧縮する装置およびその適用までそれによって解放される。だけでなく、操作有効性の改善、農業のコストを低下させること、また苗木を上げるディスクの基礎的な苗木要件を保証すること、また、新しい高さにアップグレードする苗木を上げる流動線操作を可能にすること、そして、実際に、全工程の目標の実現、流動線苗木を上げる操作を自動化したこと。

クレーム (英):
クレーム (甲):
 【公報】 一种育苗盘土壤压实方法，其特征在于，该方法包括：A、在育苗流水线导轨上安装压实装置的过程；B、在育苗流水线上设置光电开关，并与压实装置联动的过程；C、育苗盘在育苗流水线导轨上行走至压实装置正下方时，触发光电开关的过程；D、光电开关启动压实装置，同时育苗流水线导轨暂停，压实装置的压实板在驱动机构作用下对停留在导轨上的育苗盘土壤进行压实作业的过程；E、压实作用后，压实装置的压实板在驱动机构的作用下归位，育苗流水线继续运行，育苗盘前行。

クレーム (日):
 【MT】 一種類の育苗トレイ土壤圧密方法、その特徴は、この方法以下を含む：A、育苗パイプラインレール上での圧密装置；B、育苗ライン上で光電スイッチが設けられている、でコンパクション装置と連動した手順；C、【育苗盤】育苗パイプラインレールにコンパクション装置直下に歩行時、光電スイッチのトリガー処理；D、光電スイッチ押下で起動装置、同時に育苗パイプラインレール停止、圧密装置の【圧密板】機構駆動作用下にガイドレールに留まる育苗トレイ土壤を押し固め作業を行う過程；E、圧密化作用後、圧密装置の【圧密板】駆動機構によってリターン、育苗パイプライン運転を継続する、育苗トレイ前進。

パントファミリー: CN20101220302 [2010-07-02] CN101933412(A) [2011-01-05] 公報PDF フロントPDF



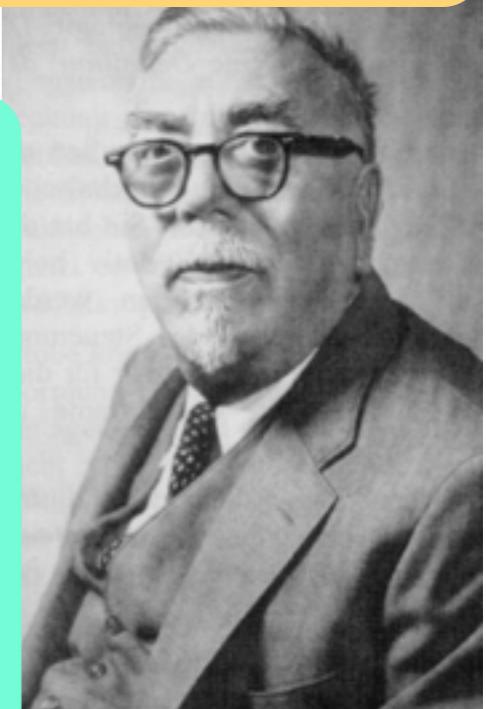
VoiceTra+ from <http://voicetra-plus.jp>

LINE of MT



“ロシア語の文章を見て、思ったのは「これは本当は英語で書かれているが、変なシンボルで暗号化されている。では今から解読しよう。」

... I frankly am afraid the boundaries of
異なる言語間の単語の境界は曖昧すぎ
て ... 擬似機械的な機械翻訳の枠組みでも
捉えられないのではないか、と懸念してい
る。



機械翻訳の歴史

- 1954年: Georgetown-IBMによる機械翻訳デモ
- 1966年: ALPACレポート
- 1970年: Systranによる商用システム
- 1982年: Muプロジェクト

初期の計算機から研究開発: 知識に基づく手法

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

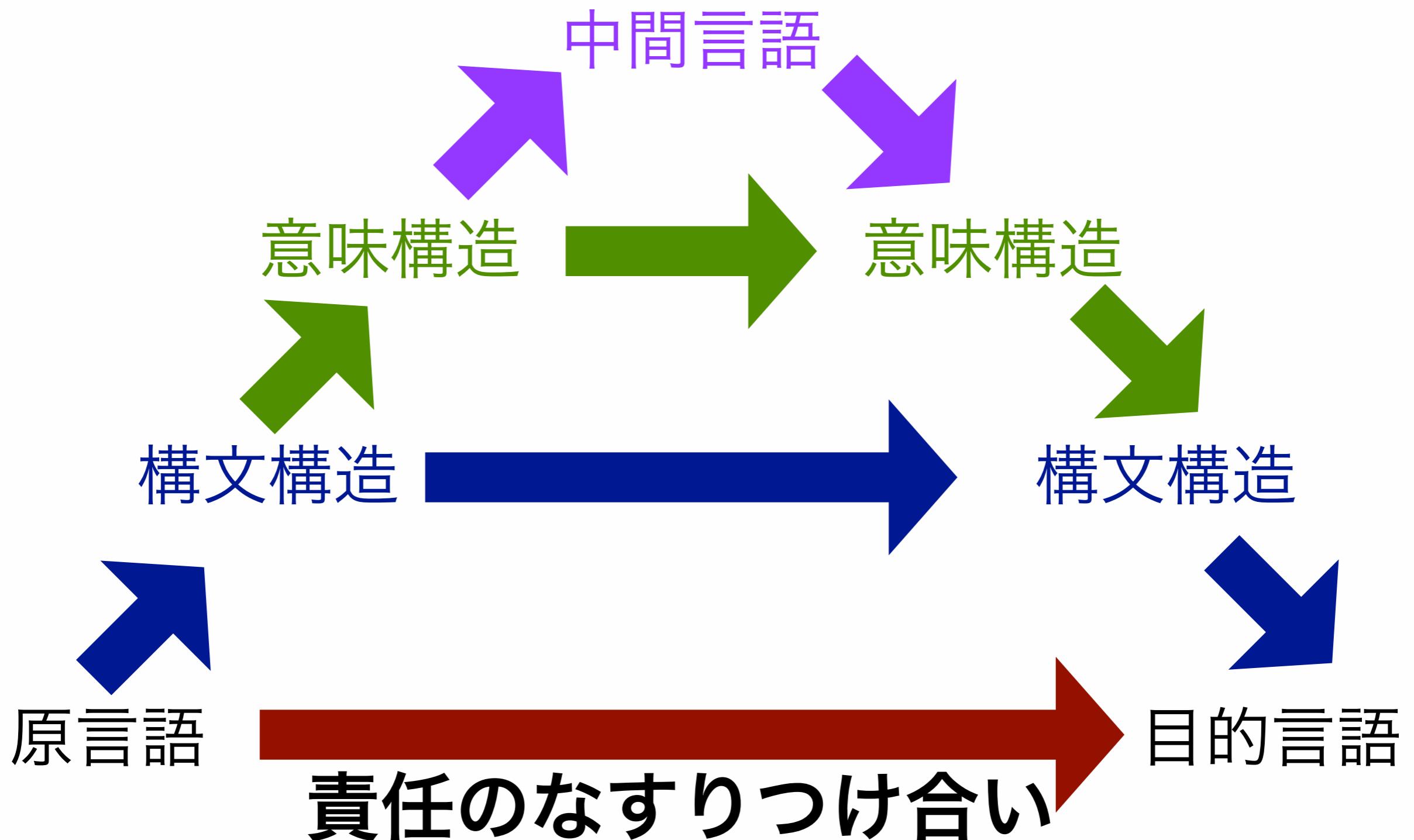
[want] [study] [about] [machine] [translation] 。

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

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

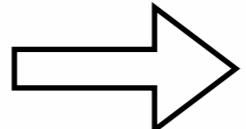
↔ I want taiyaki

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



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などの被害届を警視庁に提出した。【戸上文恵、森下功、岡礼子】

被害が確認されたのは杉並区の11館119冊▽中野区の5館54冊▽練馬区の9館41冊▽新宿区の3館39冊▽豊島区の3館12冊▽武蔵野市の2館9冊▽西東京市の3館10冊▽東久留米市の3館10冊。一方、被害が確認されていない都立中央図書館（港区）と都立多摩図書館（立川市）は、21日から「[アンネの日記](#)」の開架をやめ、希望者の申告に応じて閲覧させることにした。

杉並区によると、今月3日に隣接する練馬区から破損被害の連絡を受けて調査した際には被害を確認できな

<http://mainichi.jp/shimen/news/20140222ddm041040134000c.html>



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

More copies of worldwide bestseller "The Diary of Anne Frank" and related books have been found damaged at public libraries in western Tokyo, bringing the total number to 294 books as of Feb. 21, it has been learned.

The confirmed damages are as follows: 119 books at 11 libraries in Suginami Ward, 54 books at five libraries in Nakano Ward, 41 books at nine libraries in Nerima Ward, 39 books at three libraries in 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

<http://mainichi.jp/english/english/newsselect/news/20140222p2a00m0na001000c.html>



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

拡大写真



| | | | |
|-----|-----------|--|-------|
| 49. | 咕 呀 蝦 | Sweet & Sour Shrimp | 9.25 |
| 50. | 油 爆 明 蝦 | Imperial Shrimp | 9.25 |
| 51. | 魚 香 蝦 | Szechuan Shrimp | 9.25 |
| 52. | 腰 果 蝦 | Cashew Shrimp | 9.25 |
| | | <i>Sauteed with hoisin sauce and crunchy nuts.</i> | |
| 53. | 雪 豆 蝦 | Shrimp with Snow Peas | 9.25 |
| 54. | 宮 保 蝦 | Kung Pao Shrimp ↪ | 9.25 |
| 55. | 大 蒜 蝦 | Garlic Shrimp | 9.25 |
| 56. | 北 京 蝦 | Peking Shrimp ↪ | 9.25 |
| 57. | 蝦 龍 糊 | Shrimp in Lobster Sauce | 9.25 |
| | | <i>This one has a creamy thick tasty succulent lobster sauce this mmm-mmm good!</i> | |
| 58. | 咖 �喱 蝦 | Shrimp in Curry Sauce ↪ | 9.25 |
| 59. | 豆 豉 蝦 | Shrimp in Black Bean Sauce | 9.25 |
| 60. | 炒 三 鮮 | Imperial Feast | 9.25 |
| | | <i>Chicken, Beef, Shrimp Sauteed with Broccoli and Snow Peas in Light White Sauce.</i> | |
| 61. | 宮 保 三 樣 | Kung Pao Three Delights ↪ | 9.25 |
| 62. | 什 錦 雲 吞 菜 | Sub Gum Wonton Deluxe | 10.95 |
| | | <i>Shrimp, tender steak, juicy chicken, B.B.Q. pork sauteed with vegetables, topped with golden fried wonton and a delicious oyster sauce.</i> | |
| 63. | 干 燒 蝦 | Hot Braised Shrimp ↪ | 11.95 |
| 64. | 清 炒 蝦 | Sauteed Shrimp | 11.95 |
| 65. | 溜 炒 蝦 | Sweet & Pungent Shrimp ↪ | 11.95 |
| | | <i>Tender shrimp fried in a light batter and sauteed in a special sauce, makes this dishes one of our most wanted!</i> | |
| 66. | 花 籃 蝦 | Shrimp Basket | 11.95 |
| | | <i>Stir-fried crispy shrimp are served in a Crispy basket All you have to do is eat them.</i> | |
| 67. | 椒 鹽 蝦 | Salt & Pepper Shrimp (with shell) ↪ | 11.95 |
| 68. | 陳 皮 蝦 | Orange Shrimp ↪ | 11.95 |
| 69. | 鐵 板 海 鮮 | Seafood Volcano | 12.95 |
| | | <i>An assortment seafood-like, scallops, shrimp, imitation crabmeat and of course vegetables, served on a hot sizzling plate!</i> | |
| 70. | 花 籃 海 鮮 | Seafood Basket | 12.95 |
| | | <i>Just like the one above but well trade you</i> | |

| | | | |
|-----|-------------|---|------|
| 79. | 芥 蘭 牛 肉 | Beef Broccoli | 7.75 |
| 80. | 宮 保 牛 肉 | Kung Pao Beef ↪ | 7.75 |
| 81. | 雪 豆 牛 肉 | Beef with Snow Peas | 7.75 |
| 82. | 咖 �喱 牛 肉 | Beef in Curry Sauce ↪ | 7.75 |
| 83. | 青 椒 牛 肉 | Beef Pepper Steak | 7.75 |
| | | <i>Glonous green peppers sauteed with slices of juicy steak.</i> | |
| 84. | 冬 菇 牛 肉 | Beef & Chinese Black Mushrooms | 7.95 |
| | | <i>Chinese black mushrooms and broccoli with strips of beef in delicious oyster sauce.</i> | |
| 85. | 蒙 古 牛 肉 | Mongolian Beef | 7.95 |
| | | <i>Slices of juicy filet steak sauteed with scallions.</i> | |
| 86. | 干 扁 牛 絲 | Szechuan Beef ↪ | 7.95 |
| | | <i>Tender strips of beef stir-fried with carrots.</i> | |
| 87. | 陳 皮 牛 肉 | Orange Beef ↪ | 8.95 |
| 88. | 芝 麻 牛 肉 | Sesame Beef ↪ | 8.95 |
| 89. | 鐵 板 牛 肉 | Beef Volcano | 8.95 |
| | | <i>This is a noisy one, slice of juicy tender steak mixed with a assortment of vegetables, served on a sizzling plater.</i> | |
| 90. | 花 籃 牛 肉 | Beef Basket | 8.95 |
| | | <i>Sliced tender beef with an assortment vegetables served in an edible crispy basket.</i> | |
| 91. | 湖 南 牛 肉 | Beef Hunan ↪ | 8.95 |
| | | <i>Hoo-Nun not Hunan another spicy beef for someone who like it hot!</i> | |
| 92. | 蠔 油 牛 肉 | Beef Oyster Sauce | 8.95 |
| 93. | 四 季 豆 牛 肉 | Beef whit String Bean | 8.95 |
| 94. | 生 薑 牛 肉 | Ginger Beef | 7.95 |
| 95. | 鐵 板 黑 椒 牛 柳 | Black Pepper Filet Mignon ↪ | 9.95 |

Pork

| | | | |
|------|---------|---|------|
| 96. | 宮 保 肉 | Kung Pao Pork ↪ | 6.95 |
| 97. | 蘑 菇 肉 | Mushroom Chaw Yuke | 6.95 |
| 98. | 芥 蘭 叉 燒 | Bar-B-Q Pork with Broccoli | 6.95 |
| 99. | 咕 呀 肉 | Sweet & Sour Pork | 6.95 |
| 100. | 大 蒜 肉 絲 | Garlic Pork | 6.95 |
| 101. | 魚 香 肉 絲 | Pork Szechuan ↪ | 6.95 |
| 102. | 雪 豆 叉 燒 | Snow Peas Char-Siu | 6.95 |
| | | <i>This is that famous red barbecued pork</i> | |

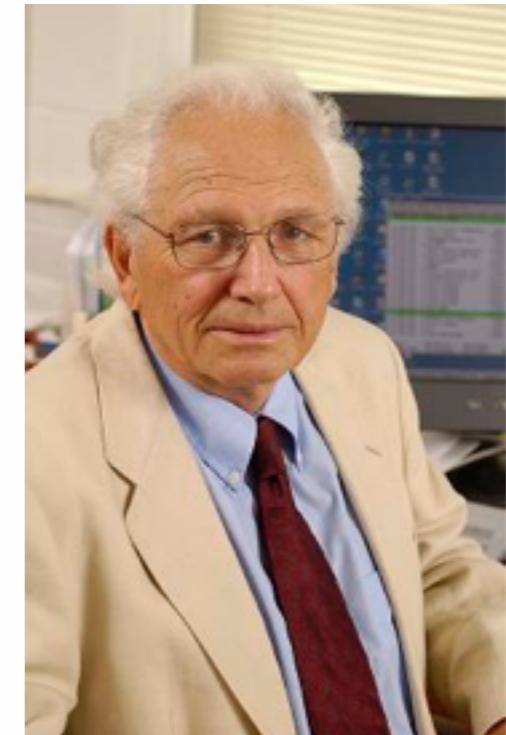
対訳データ

1. 上海浦东开发与法制建设同步
2. 新华社上海二月十日电（记者谢金虎、张持坚）
3. 上海浦东近年来颁布实行了涉及经济、贸易、建设、规划、科技、文教等领域的七十一件法规性文件，确保了浦东开发的有序进行。
4. 浦东开发开放是一项振兴上海，建设现代化经济、贸易、金融中心的跨世纪工程，因此大量出现的是以前不曾遇到过的新情况、新问题。
5. 对此，浦东不是简单的采取“干一段时间，等积累了经验以后再制定法规条例”的做法，而是借鉴发达国家和深圳等特区的经验教训，聘请国内外有关专家学者，积极、及时地制定和推出法规性文件，使这些经济活动一出现就被纳入法制轨道。
6. 去年初浦东新区诞生的中国第一家医疗机构药品采购服务中心，正因为一开始就比较规范，运转至今，成交药品一亿多元，没有发现一例回扣。

1. The development of Shanghai's Pudong is in step with the establishment of its legal system
2. Xinhua News Agency, Shanghai, February 10, by wire (reporters Jinhu Xie and Chijian Zhang)
3. In recent years Shanghai's Pudong has promulgated and implemented 71 regulatory documents relating to areas such as economics, trade, construction, planning, science and technology, culture and education, etc., ensuring the orderly advancement of Pudong's development.
4. Pudong's development and opening up is a century-spanning undertaking for vigorously promoting Shanghai and constructing a modern economic, trade, and financial center. Because of this, new situations and new questions that have not been encountered before are emerging in great numbers.
5. In response to this, Pudong is not simply adopting an approach of "work for a short time and then draw up laws and regulations only after waiting until experience has been accumulated." Instead, Pudong is taking advantage of the lessons from experience of developed countries and special regions such as Shenzhen by hiring appropriate domestic and foreign specialists and scholars, by actively and promptly formulating and issuing regulatory documents, and by ensuring that these economic activities are incorporated into the sphere of influence of the legal system as soon as they appear.
6. Precisely because as soon as it opened it was relatively standardized, China's first drug purchase service center for medical treatment institutions, which came into being at the beginning of last year in the Pudong new region, in operating up to now, has concluded transactions for drugs of over 100 million yuan and hasn't had one case of kickback.

最先端の機械翻訳

- 1994年: Candideシステム
- 1999年: JHUワークショップ
- 2006年: Google翻訳
- 2010年: VoiceTra、 TexTra



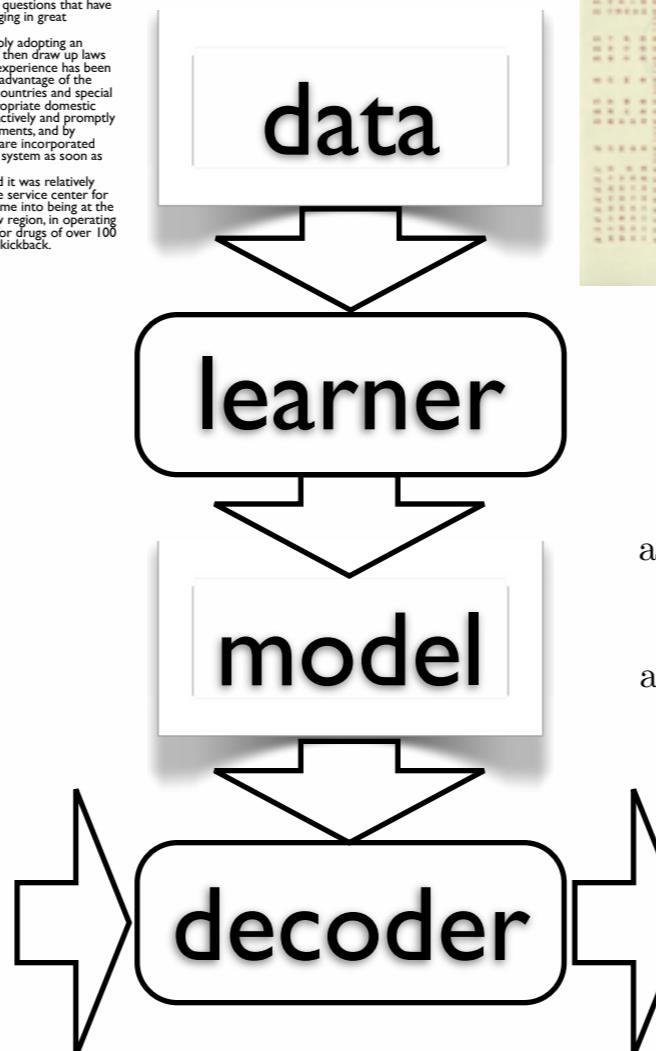
“I fire linguists”

1990年代から、統計的機械翻訳

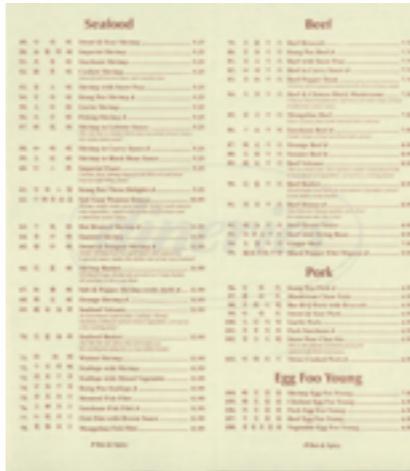
統計的機械翻譯

1. 上海浦东开发与法制建设同步
2. 新华社上海二月十日电 (记者金虎、张持坚)
3. 上海浦东近年来颁布实施了涉及经济、贸易、建设、规划、科技、文教等领域的七十一件法规性文件，确保了浦东开发的有序进行。
4. 浦东开发开放是一项振兴上海、建设现代化经济、贸易、金融中心的跨世纪工程，因此大量出现的是以前不曾遇到的新情况、新问题。
5. 对此，浦东不是简单的采取“干一段时间，等积累了经验以后再制定法规条例”的做法，而是借鉴发达国家和深圳等特区的经验教训，聘请国内外有关专家学者，积极、及时地制定和推出法规性文件，使这些经济活动一出现就被纳入法制轨道。
6. 去年年初浦东新区诞生的中国第一家医疗机构药品采购服务中心，正因为一开始就要比较规范，运转至今，成交药品一亿多元，没有发现一例回扣。

- 1.The development of Shanghai's Pudong is in step with the establishment of its legal system
- 2.Xinhua News Agency, Shanghai, February 10, by wire (reporters Jinhu Xie and Chidian Zhang)
- 3.In recent years Shanghai's Pudong has promulgated and implemented 71 regulatory documents relating to areas such as economics, trade, construction, planning, science and technology, culture and education, etc., ensuring the orderly advancement of Pudong's development.
- 4.Pudong's development and opening up is a century-spanning undertaking for vigorously promoting Shanghai and constructing a modern economic, trade, and financial center. Because of this, new situations and new questions that have not been encountered before are emerging in great numbers.
- 5.In response to this, Pudong is not simply adopting an approach of "work for a short time and then draw up laws and regulations only after waiting until experience has been accumulated." Instead, Pudong is taking advantage of the lessons from experience of developed countries and special regions such as Shenzhen by hiring appropriate domestic and foreign specialists and scholars, by actively and promptly formulating and issuing regulatory documents, and by ensuring that these economic activities are incorporated into the sphere of influence of the legal system as soon as they are carried out.
- 6.Precisely because as soon as it opened it was relatively standardized, China's first drug purchase service center for medical treatment institutions, which came into being at the beginning of last year in the Pudong new region, in operating up to now, has concluded transactions for drugs of over 100 million yuan and hasn't had one case of kickback.

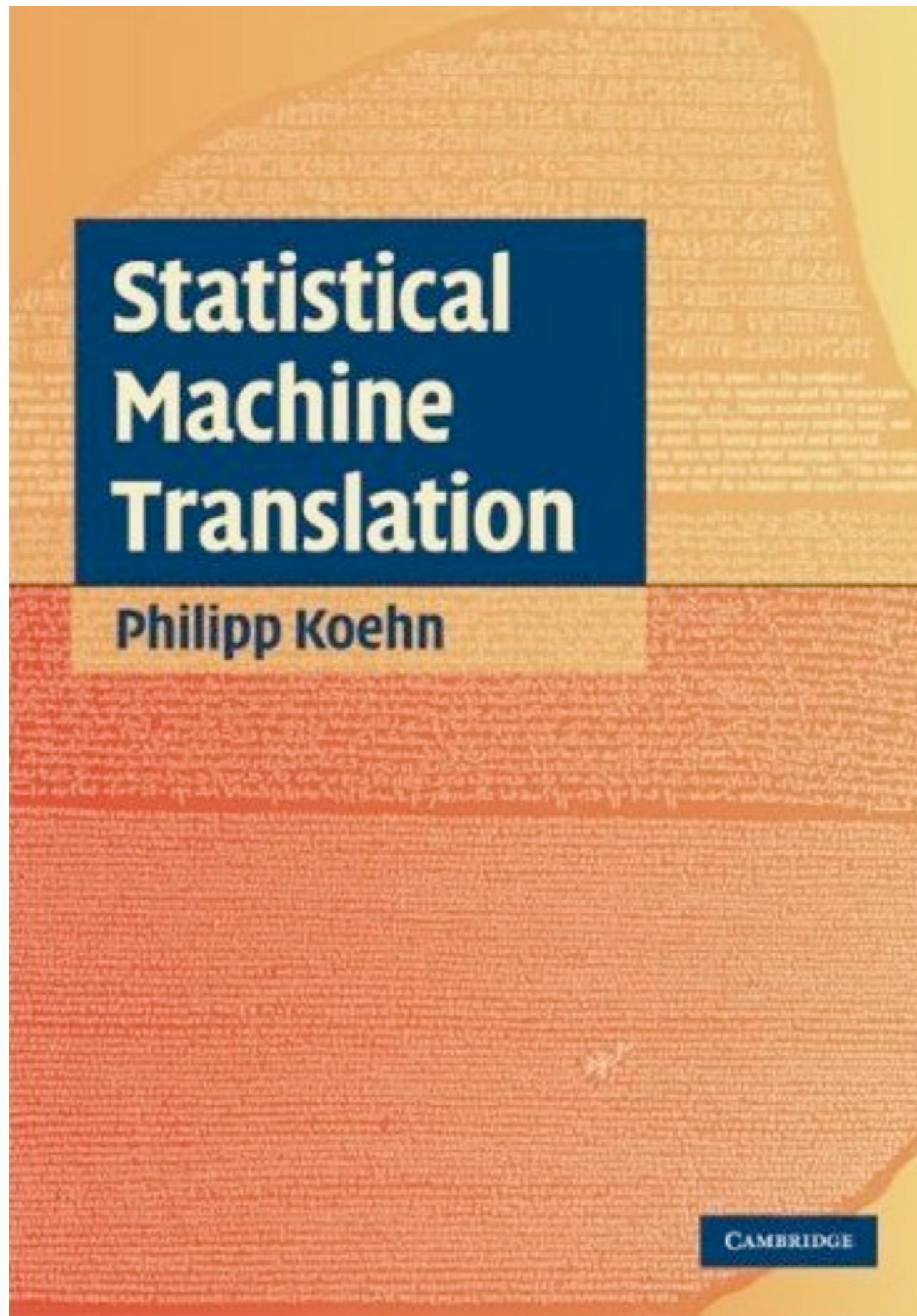


黑山头口岸联检部门将原来要二至三天办完的出入境手续改为一天办完。



$$\begin{aligned} & \arg \max_{\theta} \prod_{\langle f, e \rangle \in \mathcal{D}} \sum_d P_{\theta}(f, d | e) \quad \arg \max_{\theta} \prod_{e \in \mathcal{D}} P_{\theta}(e) \\ & \arg \min_{\boldsymbol{w}} \ell_{\text{error}}\left(\left\{\arg \max_e \boldsymbol{w}^{\top} \cdot \boldsymbol{h}(f^{(s)}, e)\right\}_{s=1}^S, \left\{e^{(s)}\right\}_{s=1}^S\right) \end{aligned}$$

The United Inspection Department of Heishantou Port has shortened the procedures for leaving and entering the territory from originally 2 - 3 days to 1 day.



<http://www.statmt.org/book/>

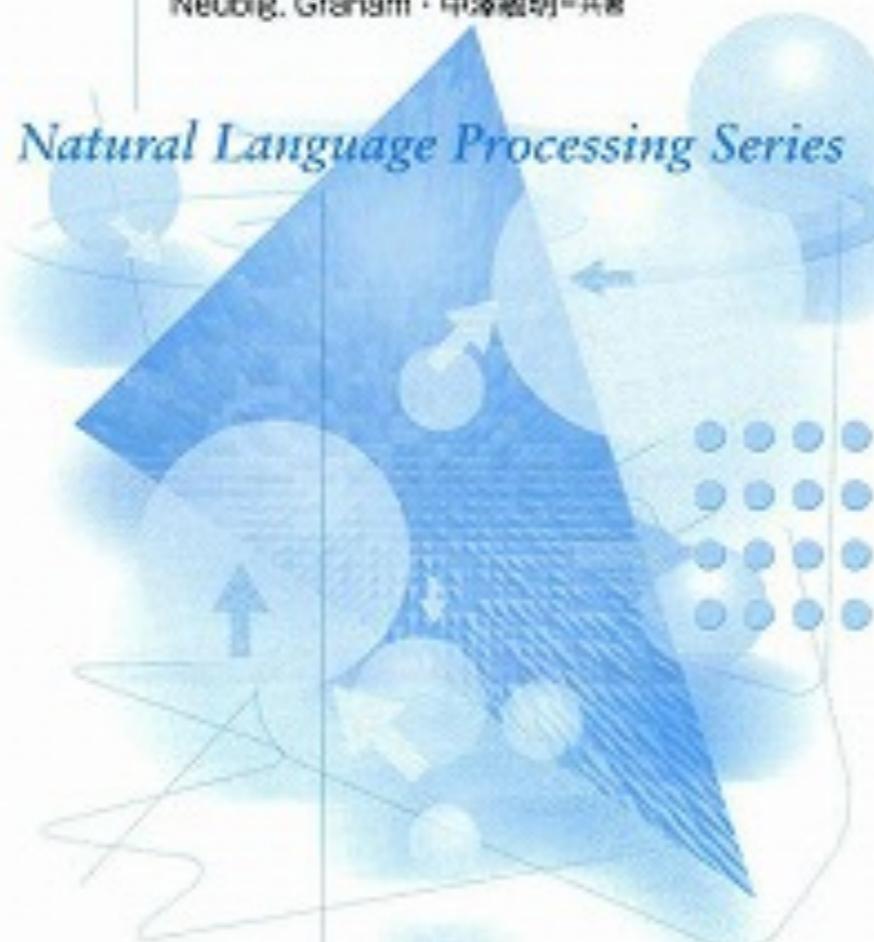
自然言語処理シリーズ 4

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/>

通信路モデル



通信路モデル + 雑音



三菱素材伝送用 HDTV エンコーダ・デコーダ
MH-2700E (左側) MH-2700D (右側)



$$\begin{aligned}\hat{y} &= \arg \max_y Pr(y|x) \\ &= \arg \max_y \frac{Pr(x|y)Pr(y)}{Pr(x)} \\ &= \arg \max_y Pr(x|y)Pr(y)\end{aligned}$$

f = 原言語

e = 目的言語

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



- 応用技術: 音声認識、OCR、機械翻訳...

機械翻訳のモデル



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

デコーダ 翻訳モデル 言語モデル

(Brown et al., 1990)

- 翻訳モデル: 翻訳としての正しさ
- 言語モデル: 「スタイル」の統一、流暢さ
- デコーダ: 翻訳を生成

翻訳 Interpretation
通訳 Translation
モデル Model

確率

f = 機械翻訳について勉強したい

e

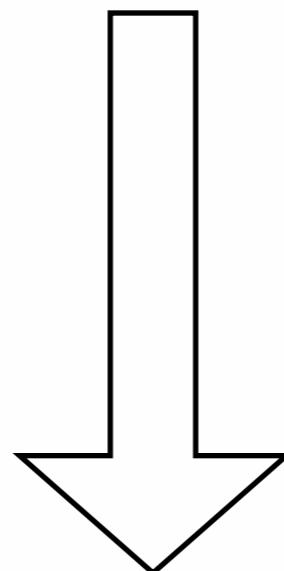
$Pr(f|e)$ $Pr(e)$

| | | |
|---|--------|--------|
| I want to study about machine translation | 0.9 | 0.15 |
| I need to master machine translation | 0.7 | 0.1 |
| machine translation want to study | 0.9 | 0.0001 |
| I don't want to learn anything | 0.0001 | 0.2 |

では...

機械翻訳について勉強したい。

どのように確
率を計算?



最も確率の高い
翻訳を生成?

I want to study about machine translation.

あるステップを経て生成されたと仮定: 導出

近似

機械翻訳について勉強したい。

I want to

$Pr(\text{したい} | \text{I want to})$ study

$Pr(\text{勉強} | \text{study})$ about

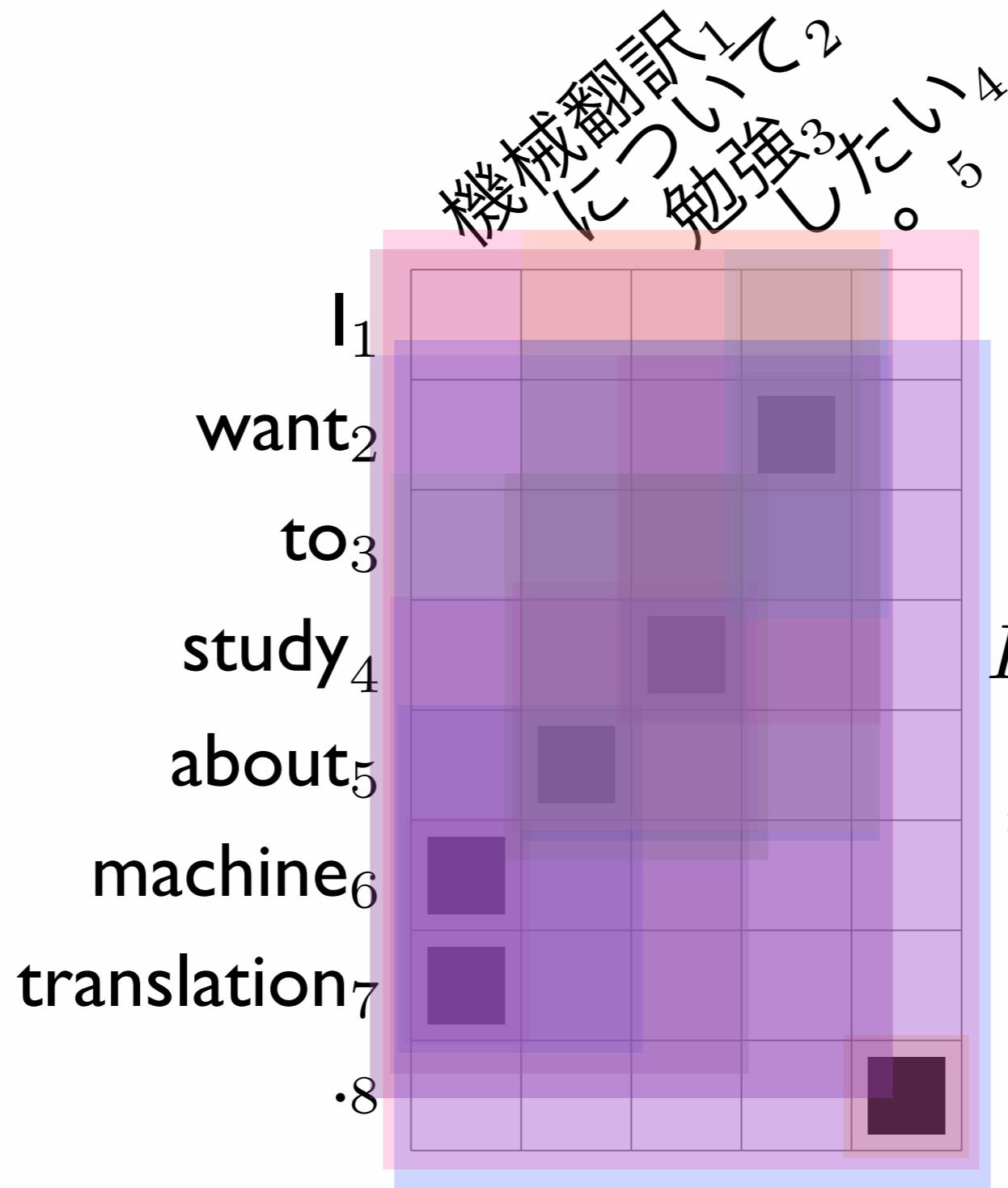
$Pr(\text{について} | \text{about})$ machine translation

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

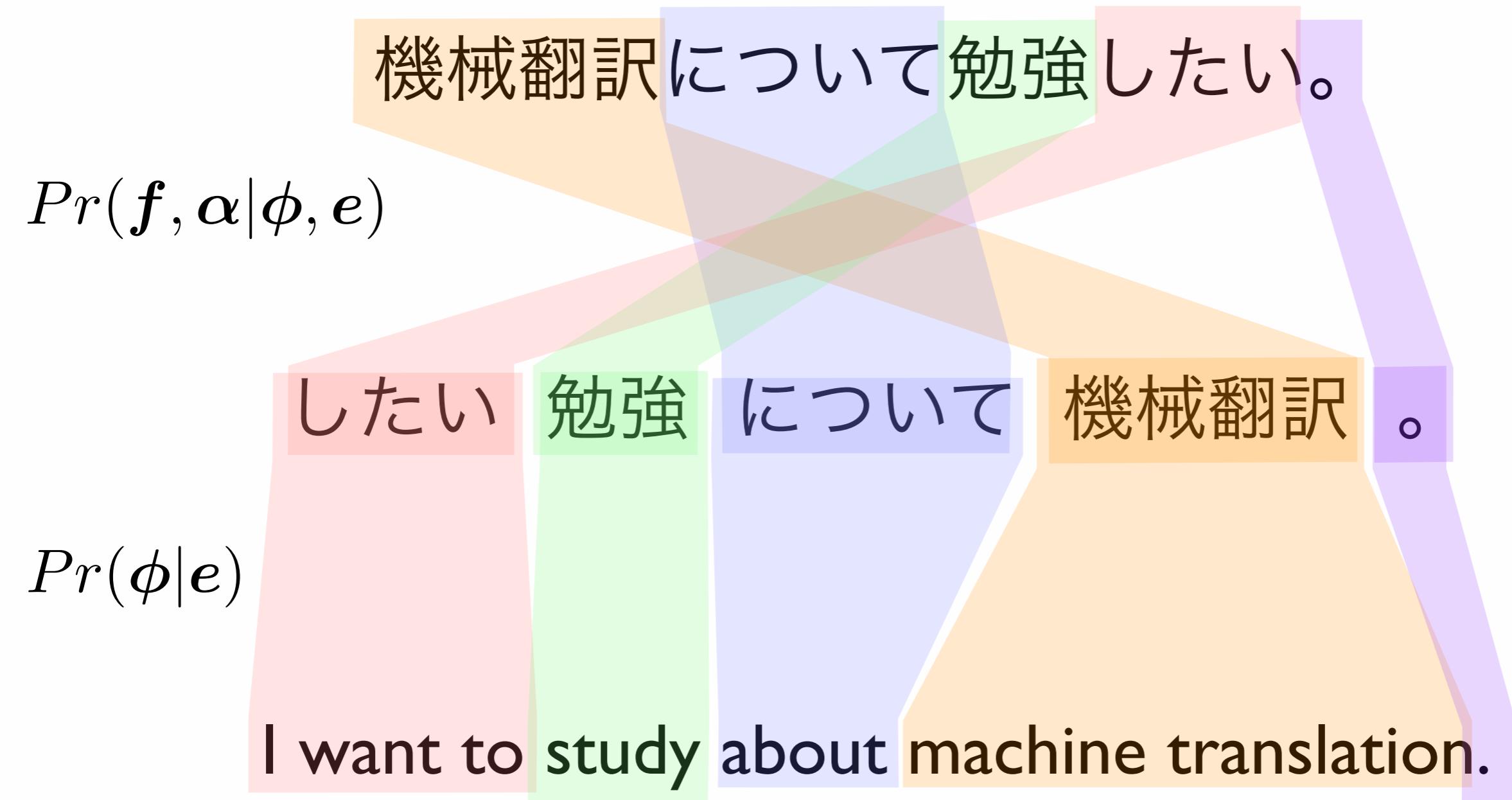
$Pr(\circ | .)$

句単位で計算: 導出 Φ = 句の集合

學習



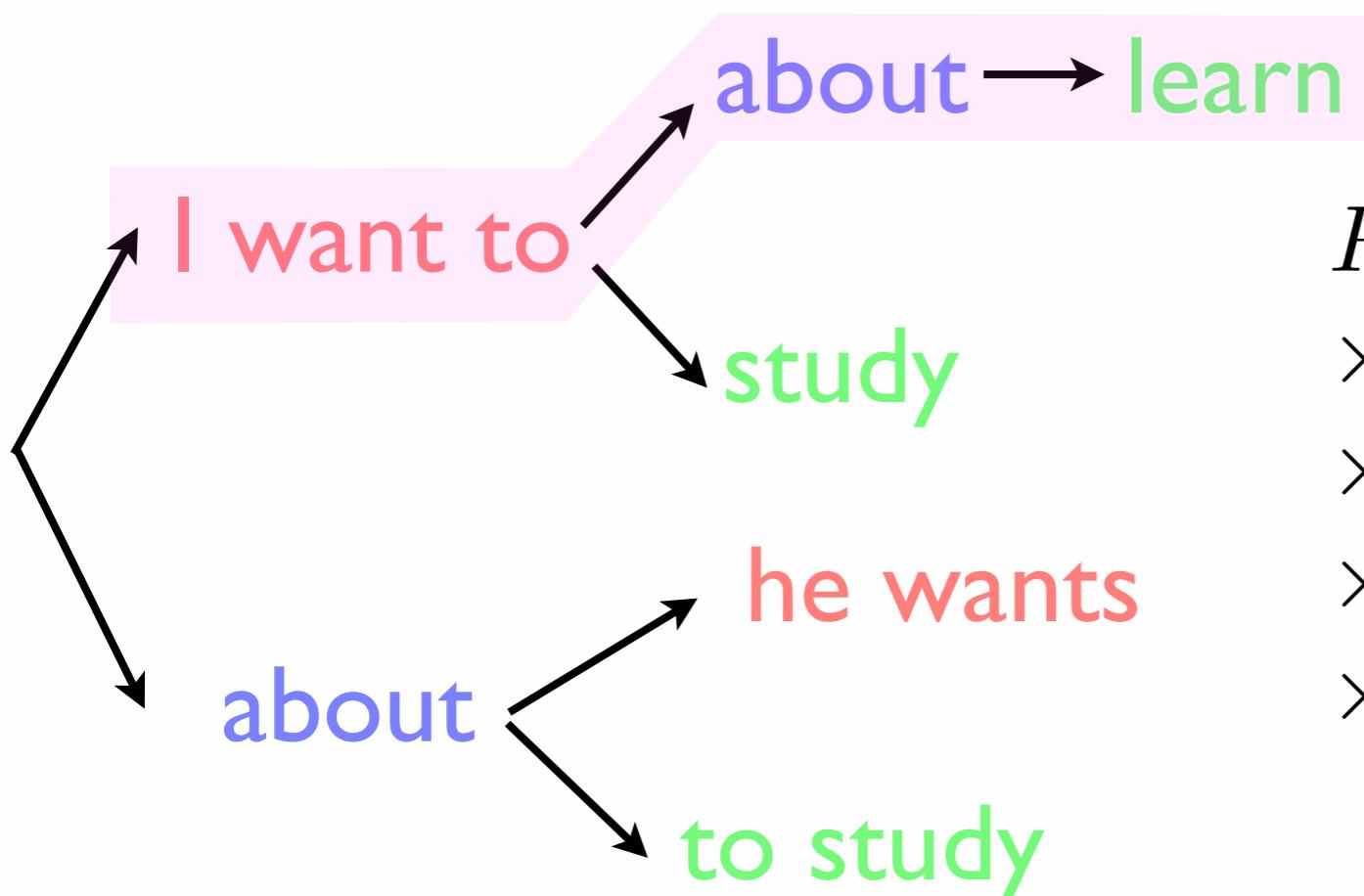
問題分割



句単位の並び替え α +句単位の翻訳 ϕ

探索

機械翻訳について勉強したい。

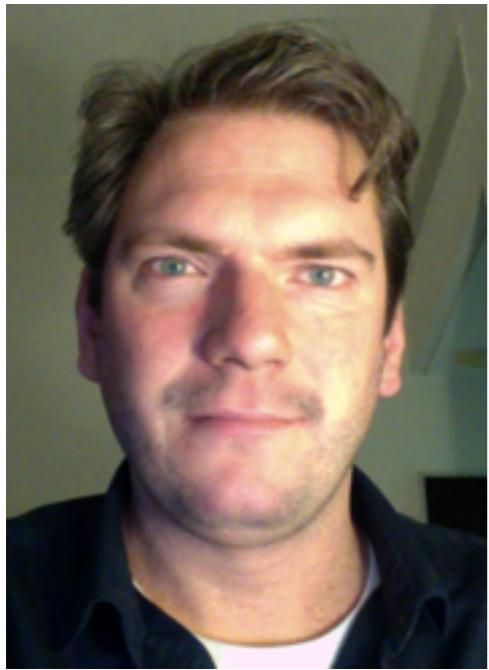


$$\begin{aligned} & Pr(I \text{ want to about learn}) \\ & \times Pr(\text{したい}, \text{について}, \text{勉強}) \\ & \times Pr(\text{したい} | I \text{ want to}) \\ & \times Pr(\text{について} | \text{learn}) \\ & \times Pr(\text{勉強} | \text{learn}) \end{aligned}$$

「部分翻訳」を列举

句に基づく機械翻訳

- 生成モデル: 近似+問題分割+探索により
統一的に表現



なぜ「句」？

- “Words may not be the best candidates for smallest units...” (Koehn, 2010)
- “Phrase-based MT is wrong. It is against my linguistic intuition.” (Mercer, 2013)



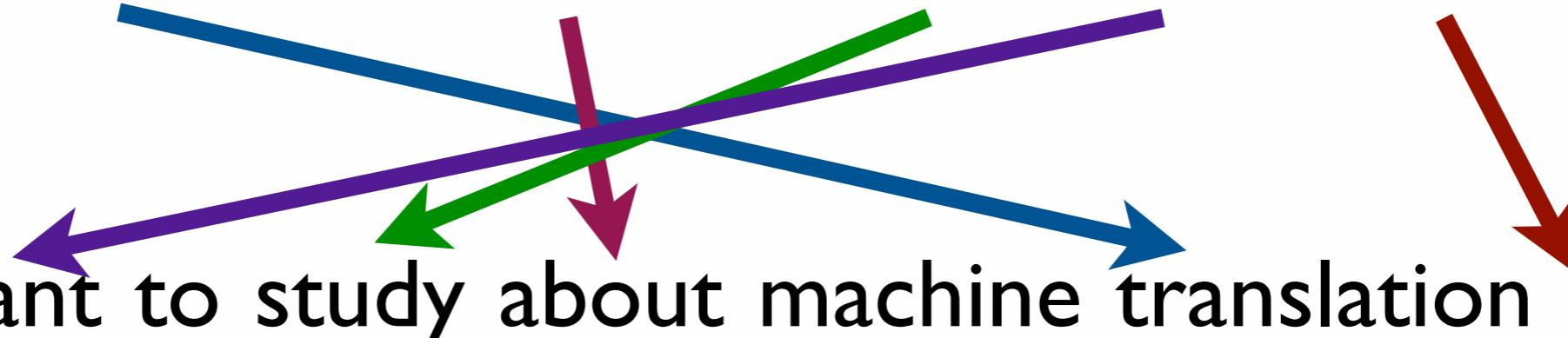
単語アライメント



$f =$ 機械翻訳について勉強したい。

$a =$

$e =$ I want to study about machine translation.



(Brown et al., 1993)

$$\hat{a} = \operatorname{argmax}_a P_\theta(f, a | e)$$

IBM Model I

$$P_{\theta}(\mathbf{f}, \mathbf{a} | \mathbf{e}) \approx \frac{1}{(|\mathbf{e}| + 1)^{|\mathbf{f}|}} \prod_{j=1}^{|\mathbf{f}|} \theta(f_j | e_{a_j})$$

(Brown et al., 1993)

θ (機械翻訳 | translation) = 0.7

機械翻訳について勉強したい。

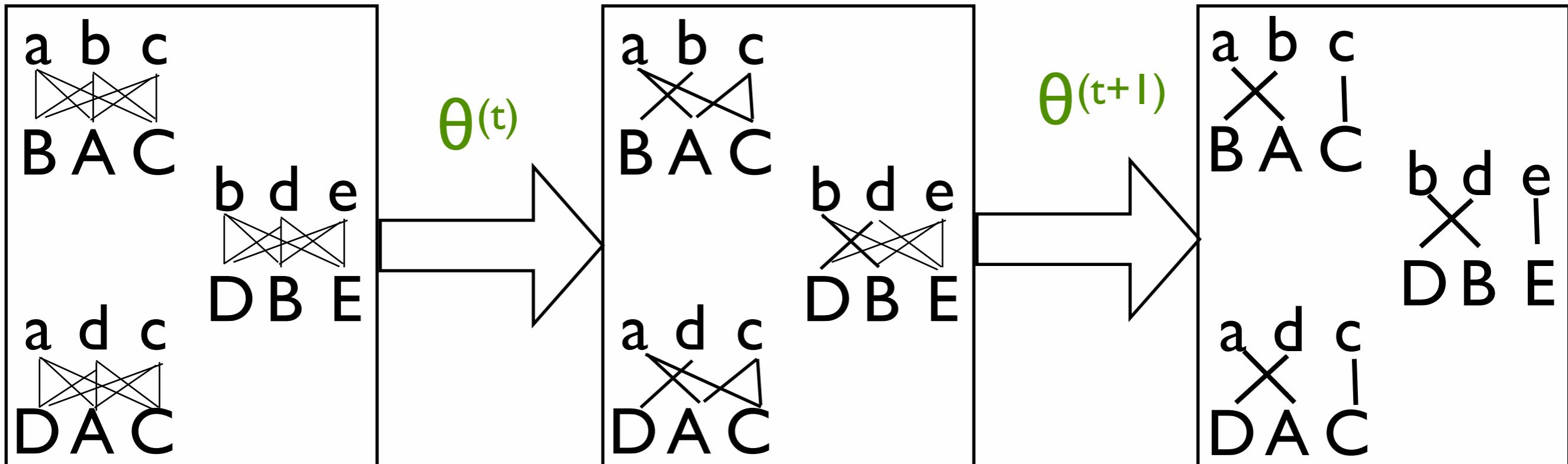
| | | | | | |
|-------------|------|------|------|------|------|
| I | 0.01 | 0.01 | 0.01 | 0.05 | 0.01 |
| want | 0.01 | 0.01 | 0.01 | 0.10 | 0.01 |
| to | 0.01 | 0.01 | 0.05 | 0.05 | 0.01 |
| study | 0.01 | 0.01 | 0.30 | 0.01 | 0.01 |
| about | 0.01 | 0.50 | 0.05 | 0.05 | 0.01 |
| machine | 0.40 | 0.01 | 0.01 | 0.01 | 0.01 |
| translation | 0.70 | 0.01 | 0.01 | 0.01 | 0.01 |
| . | 0.01 | 0.01 | 0.01 | 0.01 | 0.90 |

教師なし学習

$$\hat{\theta} = \operatorname{argmax}_{\theta} \prod_{\langle f, e \rangle \in \mathcal{D}} \sum_a P_{\theta}(f, a | e)$$

(Brown et al., 1993)

E step M step E step M step E step



翻訳評価

- そもそも、人間でも翻訳は難しい：「全自动高品質翻訳を機械翻訳の目的にしてはいけない」(Hillel, 1960)
- 「どの程度使えるのか」知りたいですよね
- 主観評価および客観評価



翻訳評価

機械翻訳について勉強したい。

システムA

machine translation want to study .

システムB

I need to master machine translation .

システムC

I want to learn machine translation .

参照訳

I want to study about machine translation .

(Papineni et al., 2002)

- **BLEU**: 参照訳との近さをn-gramで計算

kbest翻訳候補

機械翻訳について勉強したい。

$$\log Pr(\phi|e) \quad \log Pr(e) \quad \log Pr(f, \alpha|\phi)$$

I want to study about machine translation

| | | | |
|----|----|----|-----|
| -2 | -3 | -4 | -9 |
| -3 | -4 | -4 | -11 |
| -2 | -5 | -1 | -8 |
| -5 | -2 | -3 | -10 |

I need to master machine translation

machine translation want to study

I don't want to learn anything

| | | | |
|--------|--------|--------|------|
| 0.5×-2 | 0.4×-3 | 0.2×-4 | -3.0 |
| 0.5×-3 | 0.4×-4 | 0.2×-4 | -3.9 |
| 0.5×-2 | 0.4×-5 | 0.2×-1 | -3.2 |
| 0.5×-5 | 0.4×-2 | 0.2×-3 | -3.9 |

重み付けにより並び替え

重み付け



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

- より一般化:

$$\begin{aligned}\hat{e} &= \arg \max_e \frac{\sum_d \exp (\boldsymbol{w}^\top \boldsymbol{h}(f, d, e))}{\sum_{e', d'} \exp (\boldsymbol{w}^\top \boldsymbol{h}(f, d', e'))} \\ &\approx \arg \max_{\langle e, d \rangle} \boldsymbol{w}^\top \boldsymbol{h}(f, d, e)\end{aligned}\quad (\text{Och and Ney, 2002})$$

最適化 = 最適な \boldsymbol{w} を識別学習

最適化

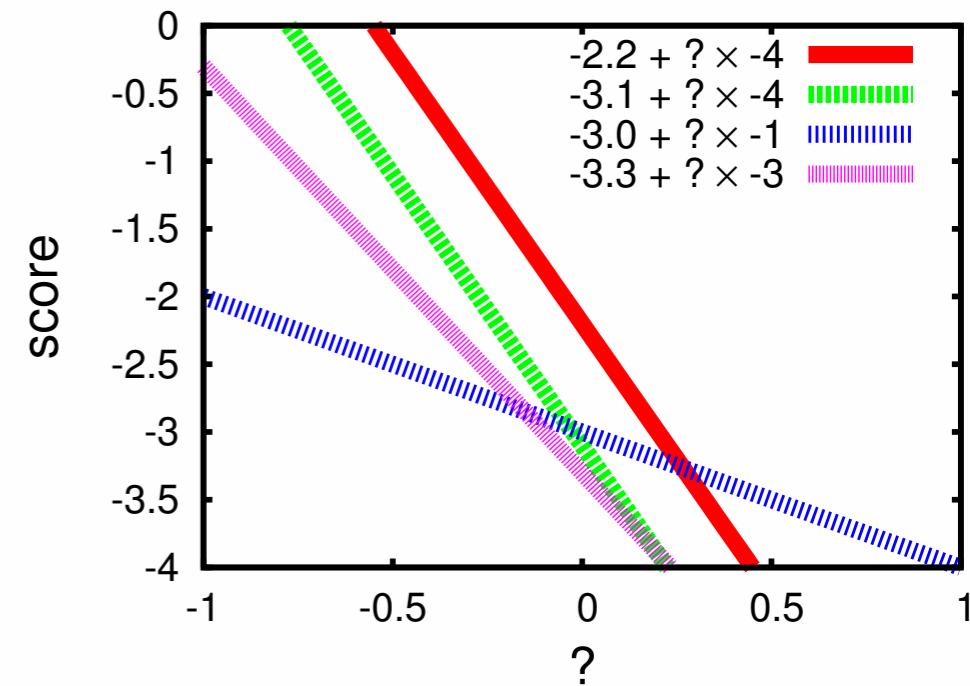


$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \ell_{\text{error}} \left(\left\{ \arg \max_{\mathbf{e}} \mathbf{w}^\top \cdot \mathbf{h}(f^{(s)}, \mathbf{e}) \right\}_{s=1}^S, \left\{ \mathbf{e}^{(s)} \right\}_{s=1}^S \right)$$

(Och, 2003)

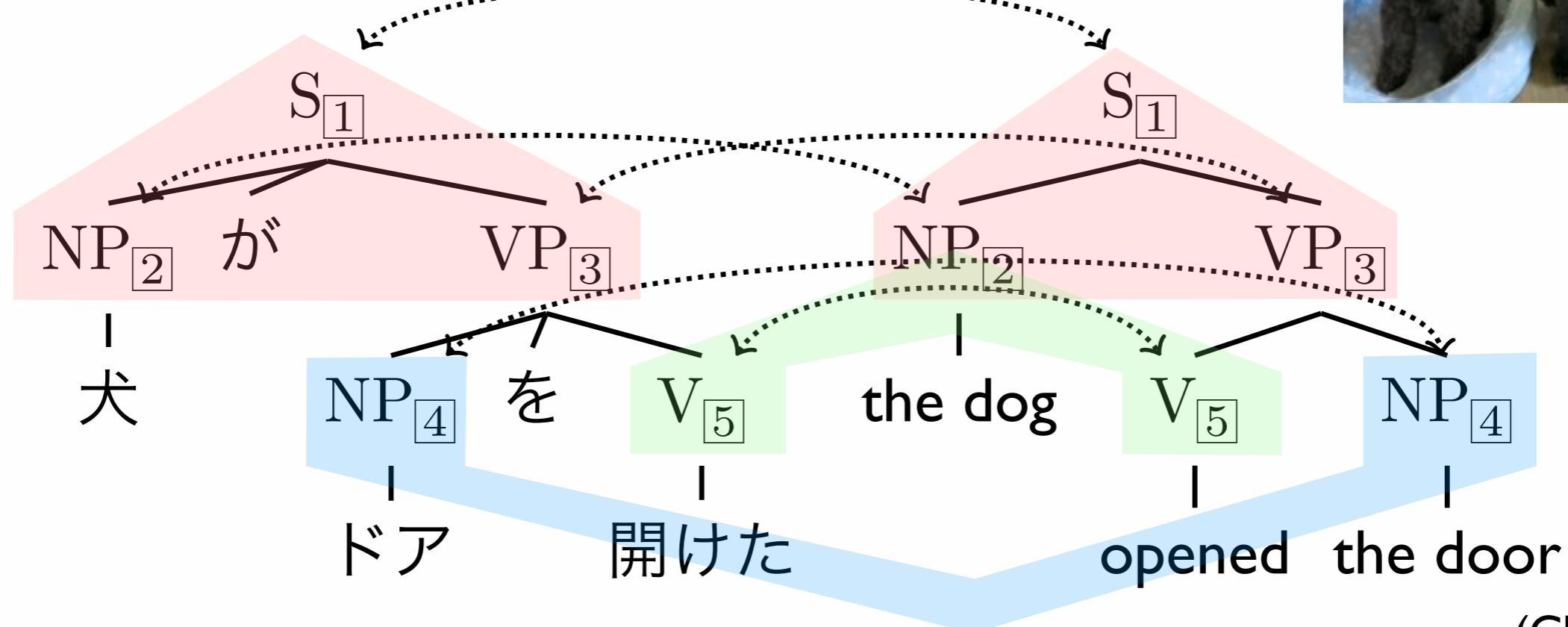
| | | |
|-----------------|-----------------|---------------|
| 0.5×-2 | 0.4×-3 | $? \times -4$ |
| 0.5×-3 | 0.4×-4 | $? \times -4$ |
| 0.5×-2 | 0.4×-5 | $? \times -1$ |
| 0.5×-5 | 0.4×-2 | $? \times -3$ |

| |
|----------------------|
| $-2.2 + ? \times -4$ |
| $-3.1 + ? \times -4$ |
| $-3.0 + ? \times -1$ |
| $-3.3 + ? \times -3$ |



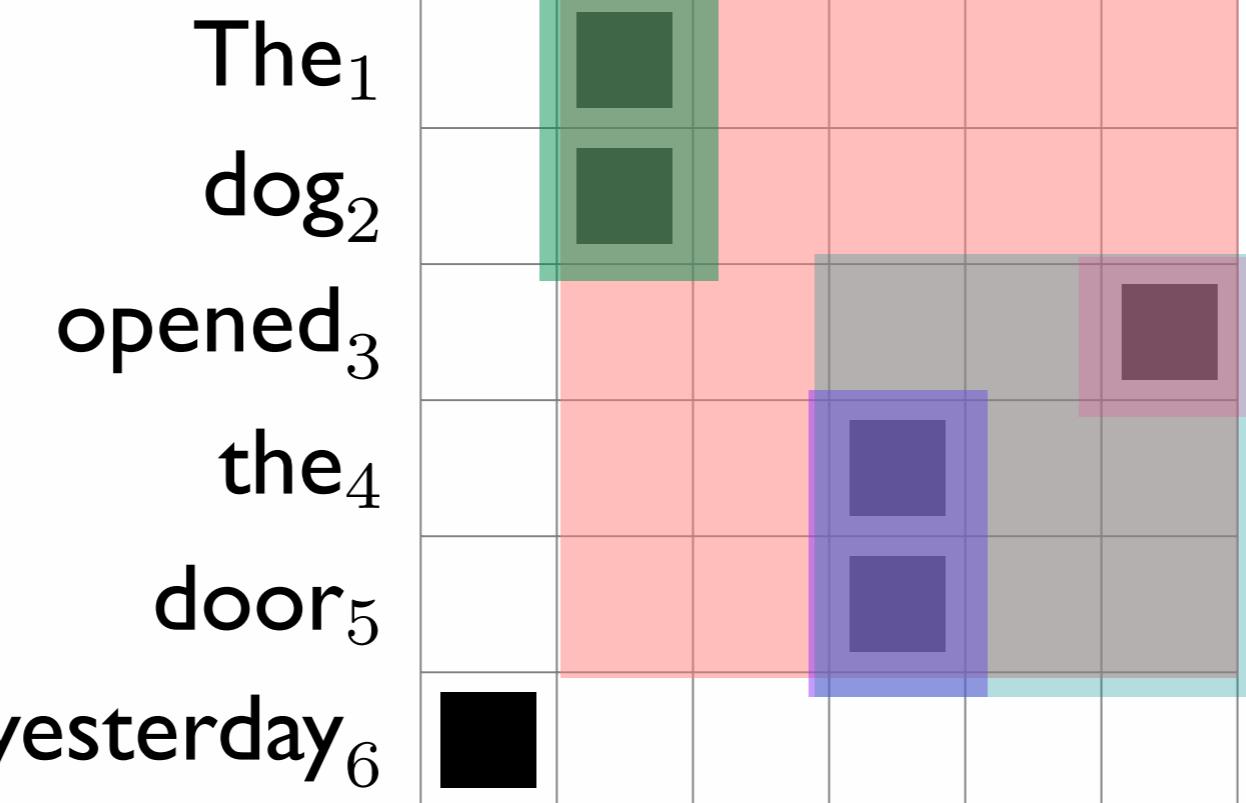
一つの次元に着目、線分探索

木構造



ルール

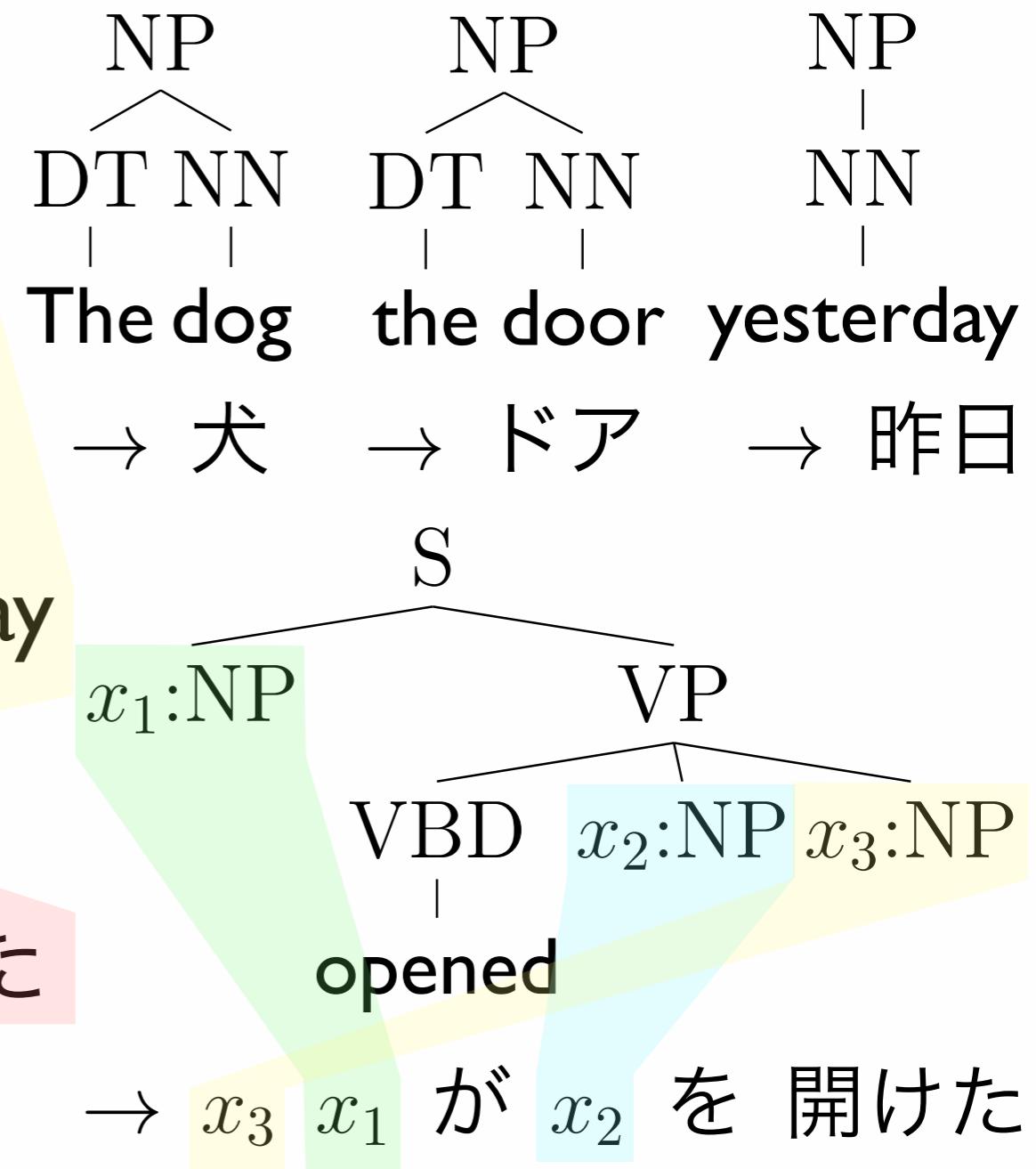
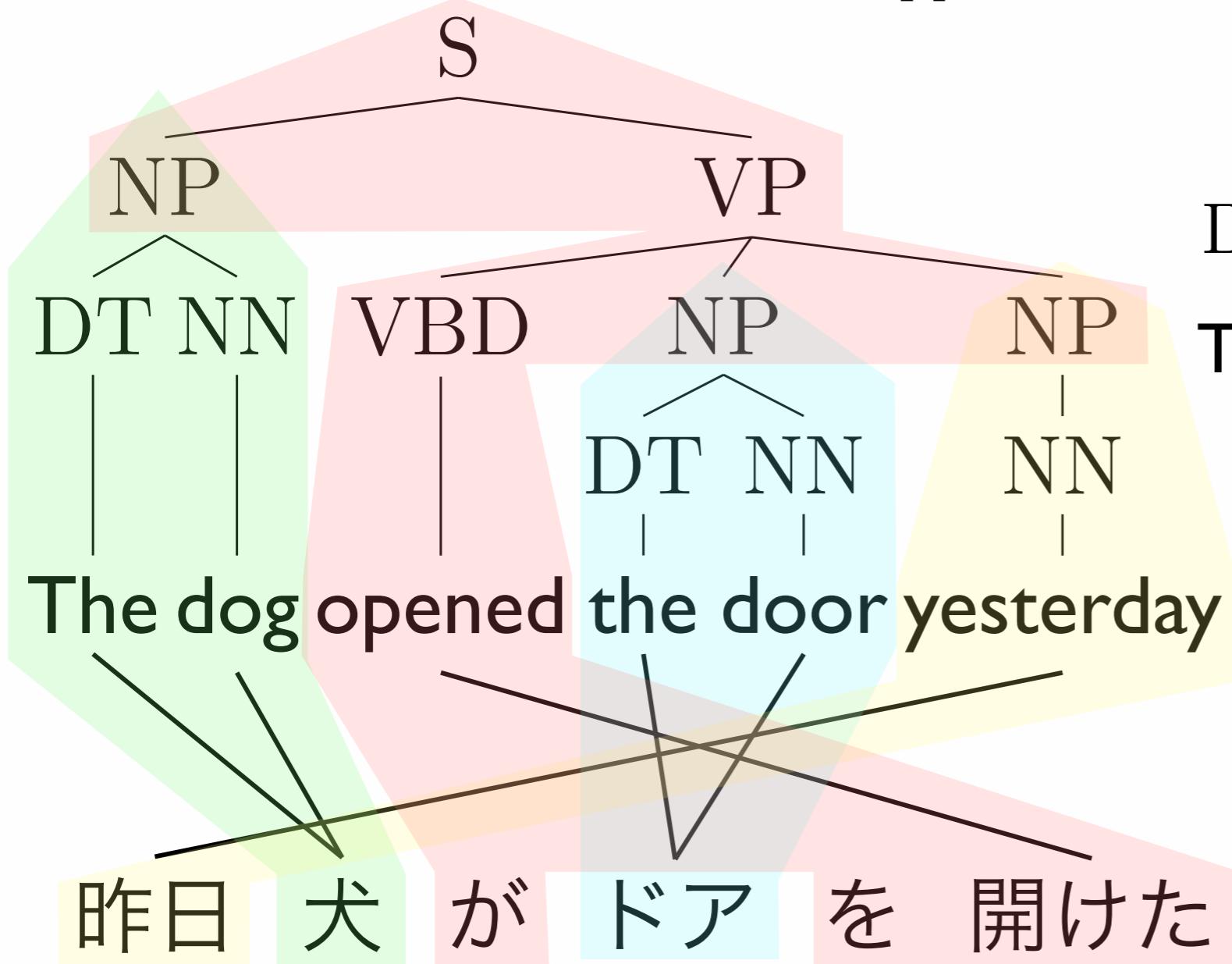
昨日¹ 犬² が³ ドア⁴ を⁵ 開けた⁶



$X \rightarrow \langle X_{[1]} \text{ が } X_{[2]} \text{ を 開けた}, X_{[1]} \text{ opened } X_{[2]} \rangle$

$X \rightarrow \langle X_{[1]} \text{ を } X_{[2]}, X_{[2]} \text{ } X_{[1]} \rangle$

構文



(Galley et al., 2004; Galley et al., 2006)

アウトライン

- フレーズベース機械翻訳: 言語モデル、翻訳モデル、デコード
- 単語アライメント
- 翻訳評価
- 最適化
- 木構造に基づく機械翻訳: 構文解析、階層的フレーズベース機械翻訳、統語論的機械翻訳



渡辺 太郎 Graham Neubig 須藤 克仁 林 克彦