

# 統計的手法による 機械翻訳

渡辺太郎

taro.watanabe at nict.go.jp



[http://www2.nict.go.jp/univ-com/multi\\_trans/member/  
t\\_watana/papers/statmt-tutorial2013.pdf](http://www2.nict.go.jp/univ-com/multi_trans/member/t_watana/papers/statmt-tutorial2013.pdf)

# 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



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 Zhuanli 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/eps/members/fam\_show\_abstract\_app.php ☆ Google

# Japio 世界特許情報検索サービス

| 蓄積時間 | マニュアル | しおり表示 | 検索式表示 | ログアウト

検索 > (EPO: [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鎮圧方法**  
 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] [公開PDF](#) [フロントPDF](#)

IPC: A01B29/00  
 ECLA:  
 優先権主張番号: CN20101220302  
 要約(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.

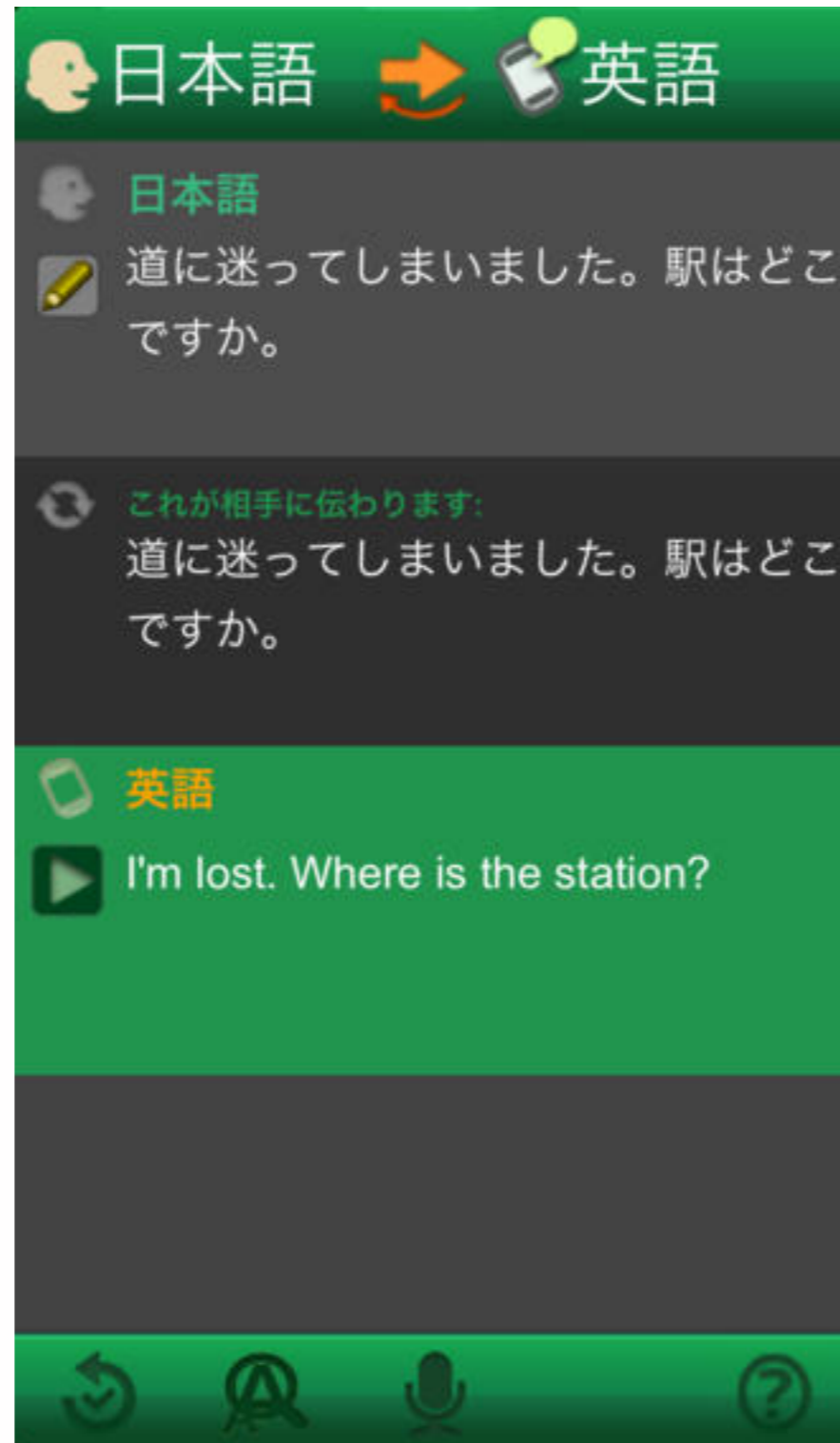
要約(日):  
 【MT】その発明は苗木を上げるディスク土鎮圧方法を開示する。自動の心土を圧縮する装置が苗木を上げるディスク流動線にインストールされ、苗木を上げるディスクに対応する圧縮するプレートを含むので、方法が採用された後、圧縮する装置の底に及ぶ場合、苗木を上げる流動線のガイドレール上の苗木を上げるディスクは休止することができる、圧縮するプレートは押し下げることができる、苗木を上げるディスクの心土は、圧縮するプレート上の圧力ヘッドによって圧縮することができる、その後、圧縮するプレートはリセットされる。また、苗木を上げるディスクは次の動作する処理を入力するために進み続ける;また、一方、圧縮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、農具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。

クレーム(英):  
 クレーム(中):  
 【公報】一種育苗土鎮圧方法，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、光电开关触动压实装置，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土镇压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鎮圧方法，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縮装置；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行き、光電スイッチのトリガー処理；D、光電スイッチ押し込み駆動装置，同時に育苗パイプラインレール停止，圧縮装置の【压实板】機構駆動作用下にガイドレールに留まる育苗トレイ土鎮圧を押し込み作業を行う過程；E、圧縮作用後，圧縮装置の【压实板】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。

パテントファミリー: CN20101220302 [2010-07-02] [CN101933412\(A\)](#) [2011-01-05] [公開PDF](#) [フロントPDF](#)





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

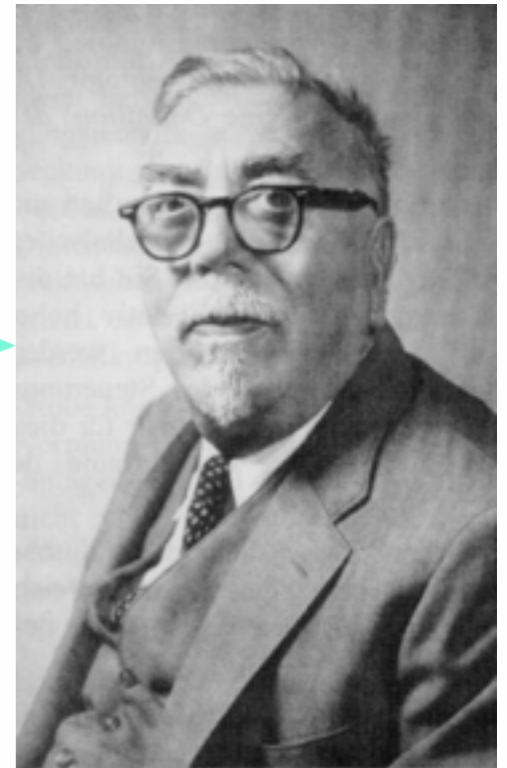
# LINE of MT



W. Weaver

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.”

... I frankly am afraid **the boundaries of  
words in different languages are  
too vague ... to make any  
quasimechanical translation  
scheme very hopeful.**



N. Wiener

# 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**



# 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

# 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。

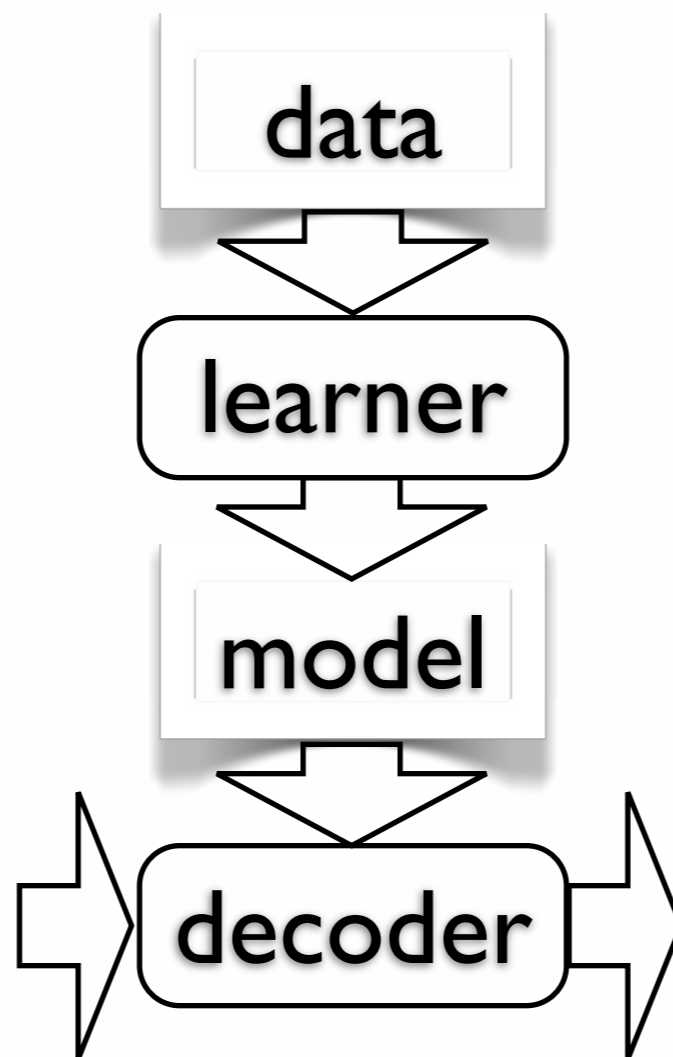
- 1.The development of Shanghai's Pudong is in step with the establishment of its legal system
- 2.Xinhua News Agency, Shanghai, February 10, by wire (reporters Jinhu Xie and Chijian Zhang)
- 3.In recent years Shanghai's Pudong has promulgated and implemented 71 regulatory documents relating to areas 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.

(part of LDC2007T02, English translation of Chinese treebank)

# 統計的機械翻訳



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。



“I'm not a linguist”

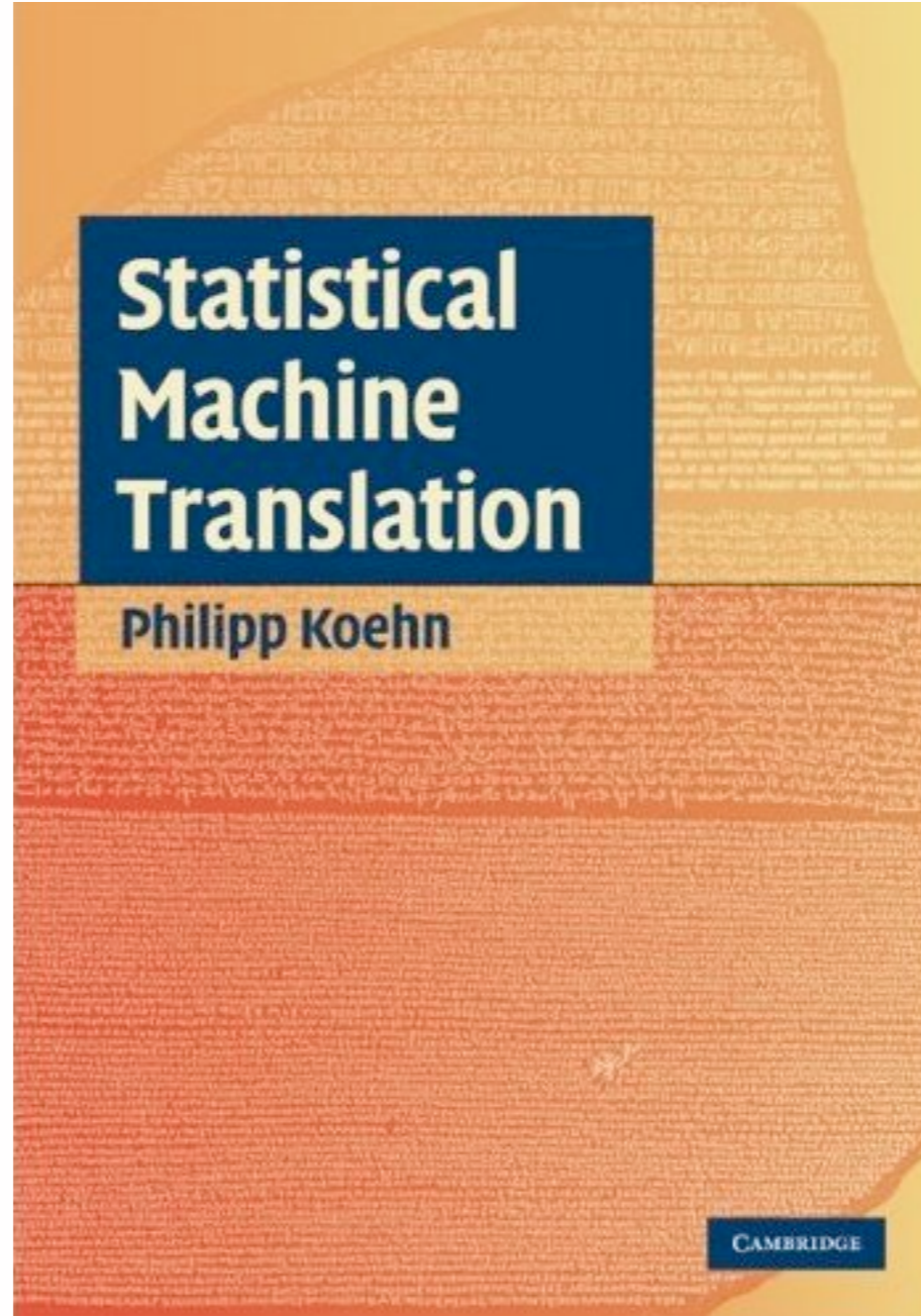


The United Inspection Department of Heishantou Port has shortened the procedures for leaving and entering the territory from originally 2 - 3 days to 1 day.

- モデルを仮定、データからパラメータを学習
- 学習されたモデルでデコード



# “SMT”



Written by Philipp Koehn

# 「機械翻訳」

自然言語処理シリーズ 4

?

**渡辺太郎、今村賢治、賀沢秀人、Graham Neubig、**

**中澤敏明（コロナ社より、来年発刊予定）**

# 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...

# 翻訳モデル

$$\hat{e} = \operatorname{argmax}_e Pr(\mathbf{f}|\mathbf{e}) Pr(\mathbf{e})$$

翻訳モデル

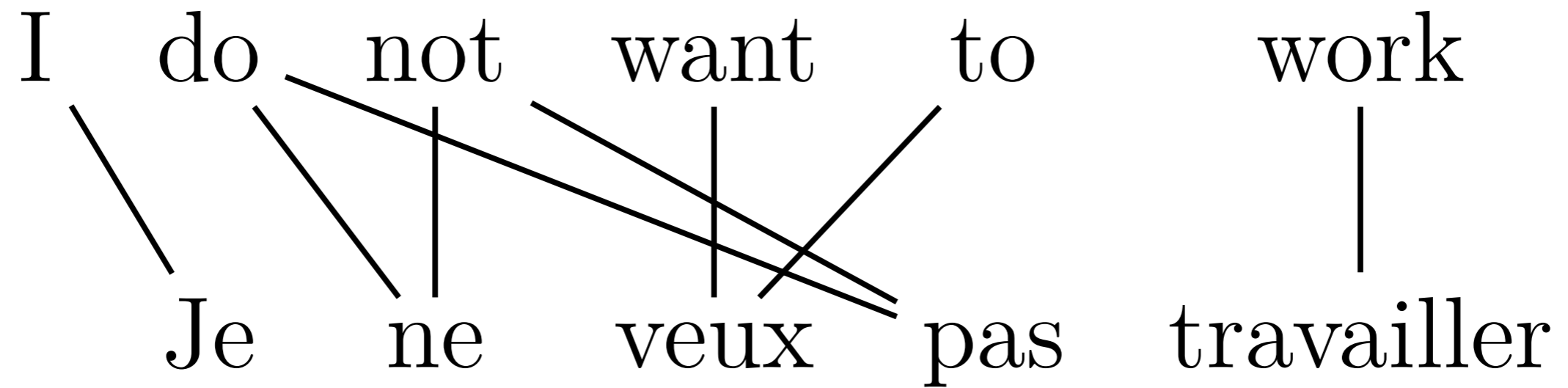
言語モデル

(Brown et al., 1990)

- 翻訳モデル: 翻訳としての正しさ (adequacy)
- 言語モデル: 文法エラーの修正、「スタイル」の統一、流暢さ (fluency)

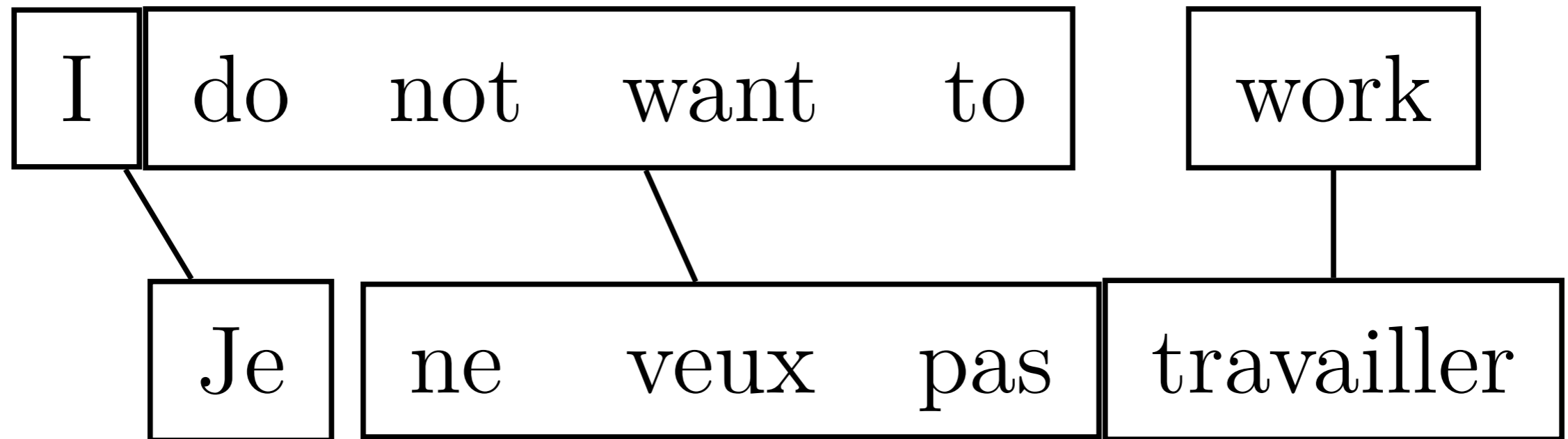
翻訳 Interpretation  
通訳 Translation  
外国人 Model

# 翻訳モデル: 単語



(Brown et al., 1993)

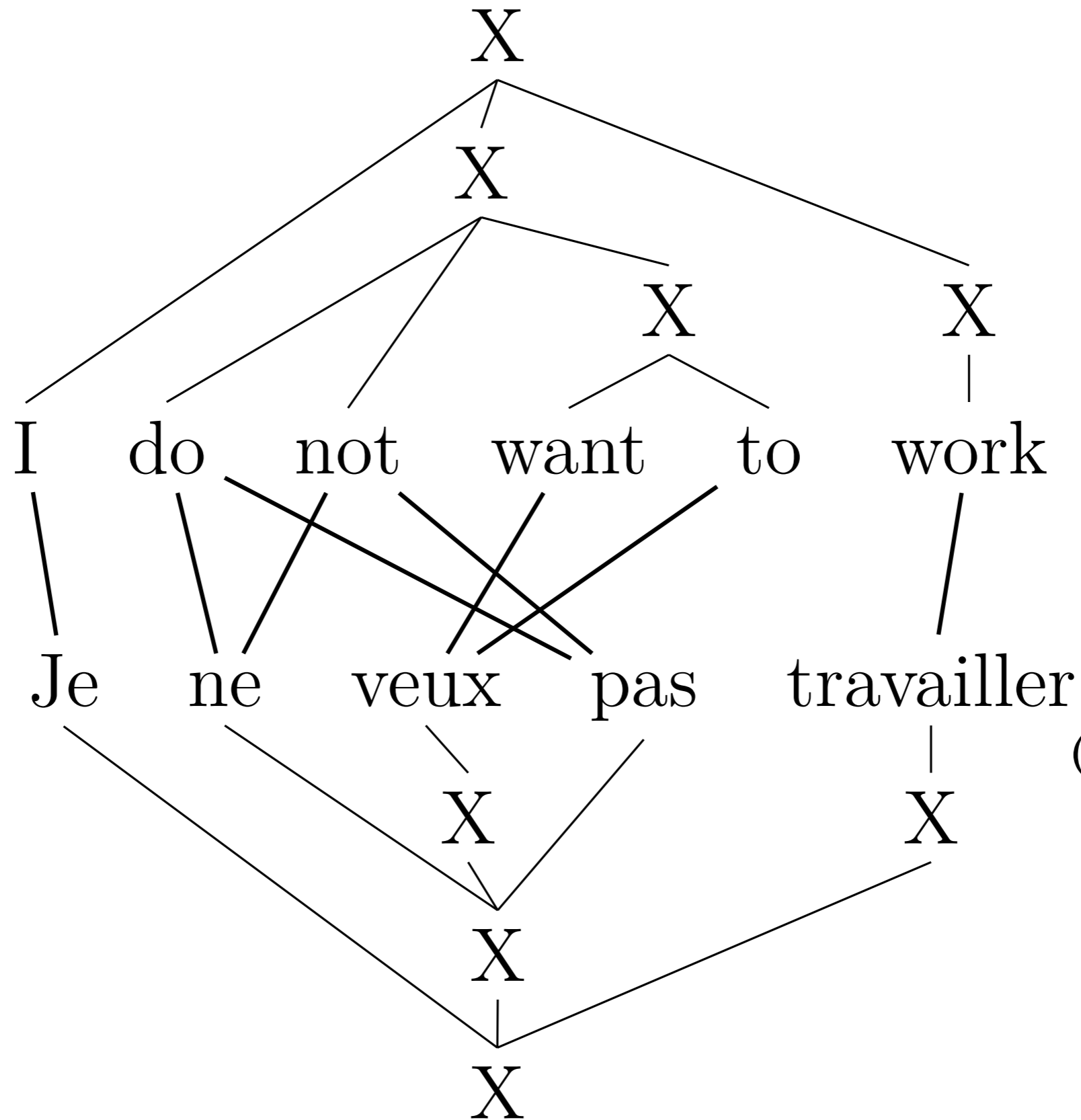
# 翻訳モデル: 句



(Koehn et al., 2003)

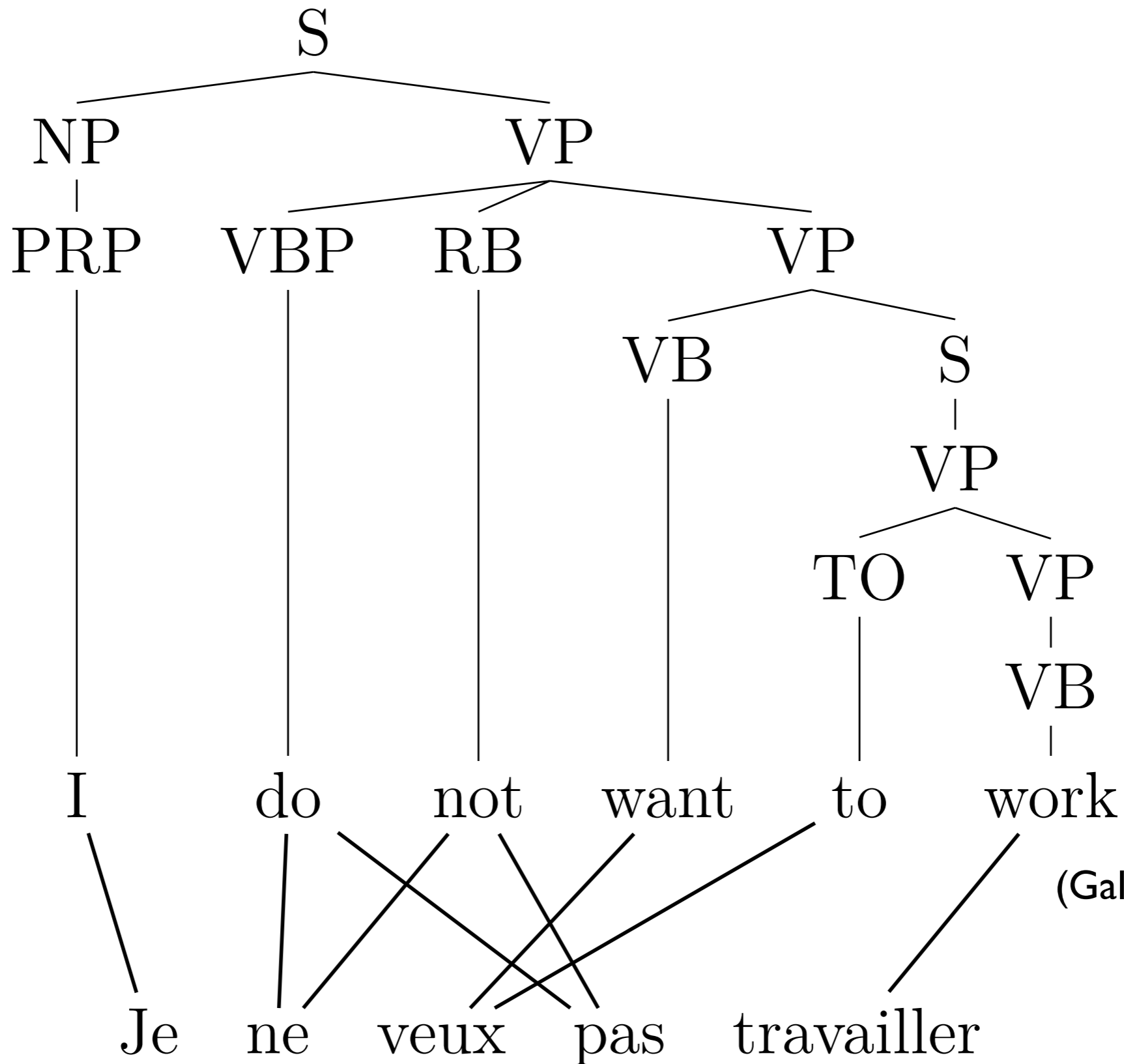


# 翻訳モデル: 階層的な句



(Chiang, 2007)

# 翻訳モデル: 構文



(Galley et al., 2004)

# 言語モデル

# 言語モデル

$$Pr(\text{I do not know}) = ?$$

$$Pr(\text{I not do know}) = ?$$

- 目的言語の文の尤度
- ngramで表現

$$W = w_1, w_2, w_3, \dots, w_N$$

$$p(W) = p(w_1, w_2, w_3, \dots, w_N)$$

$$= p(w_1)p(w_2|w_1)p(w_3|w_1, w_2) \dots$$

$$p(w_N|w_1, w_2, w_3, \dots, w_{N-1})$$



# ngram 言語モデル



- マルコフな仮定:n単語だけ覚えましょう
- Bigram:

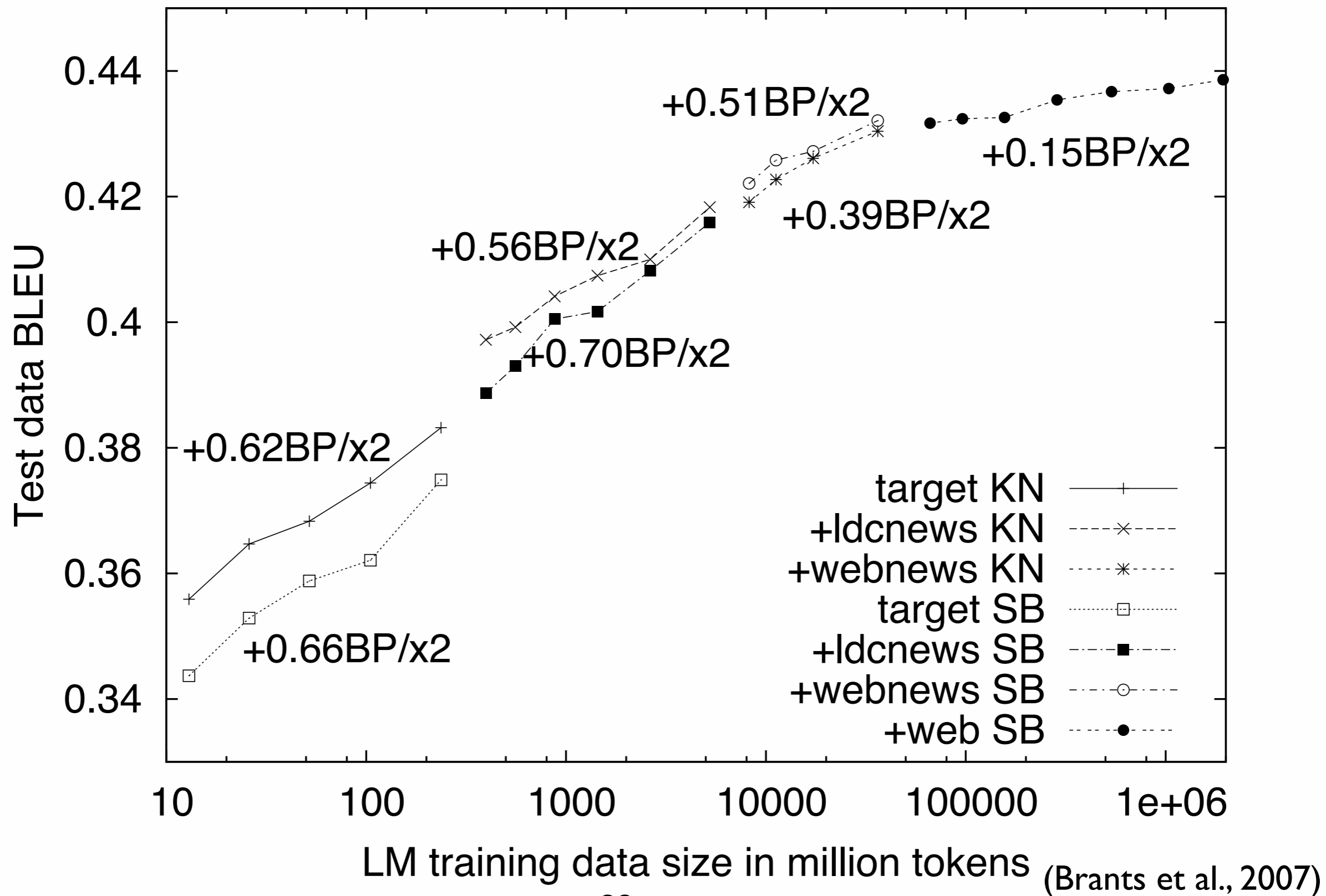
$$p(\text{I do not know}) = p(\text{I})p(\text{do}|\text{I})p(\text{not}|\text{do})p(\text{know}|\text{not})$$

- 学習: 目的言語の言語データから最尤推定

$$p(\text{know}|\text{not}) = \frac{\text{count}(\text{not know})}{\text{count}(\text{not})}$$

- + smoothing (Good-Turing, Witten-Bell, Kneser-Ney etc.)

# Bigger LM, Better MT



# 研究課題

- “There is no data like more data” (Mercer, 1985)
- 現在のトレンド:
  - 大規模データに対応したツールキット (irstlm, kenlm, expgram)
  - よりコンパクトなデータ構造 (Watanabe et al., 2009; Heafield, 2011; Yasuhara et al., 2013)
  - (Recurrent) Neural Network LM (Schwenk, 2007; Tomáš et al., 2010)

単語アライメント



49.	咕 咗 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 咗 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 the famous red barbecued pork</i>	



# 単語アライメント

$f$  = わかりません

$e$  = I do not know

$$Pr(f|e) = ??$$



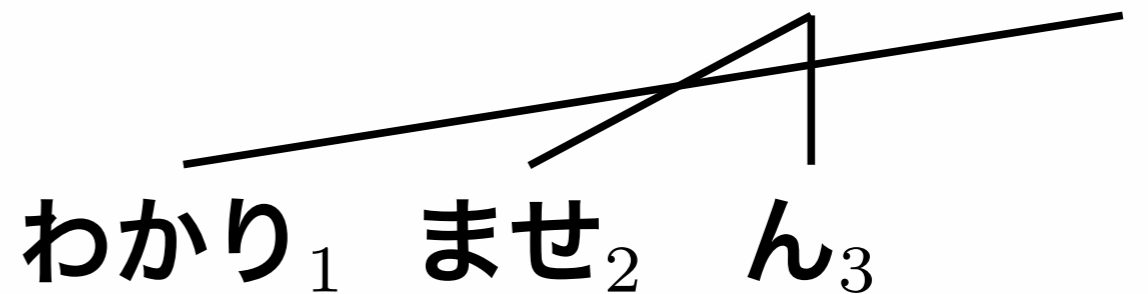
- 「単語アライメント」に基づく翻訳モデル
- IBM Model (Brown et al., 1993):
  - どのように $P(f|e)$ を表現するか?: 生成モデル
  - どのように $P(f|e)$ を推定するか?: EMアルゴリズム
  - $f$ と $e$ の単語アライメントは?: ビタビアライメント

# アライメントによる表現

$$Pr(f|e) = \sum_a Pr(f, a|e)$$

know <sub>4</sub>	■		
not <sub>3</sub>		■	■
do <sub>2</sub>		■	■
I <sub>1</sub>			

NULL<sub>0</sub>    I<sub>1</sub>    do<sub>2</sub>    not<sub>3</sub>    know<sub>4</sub>



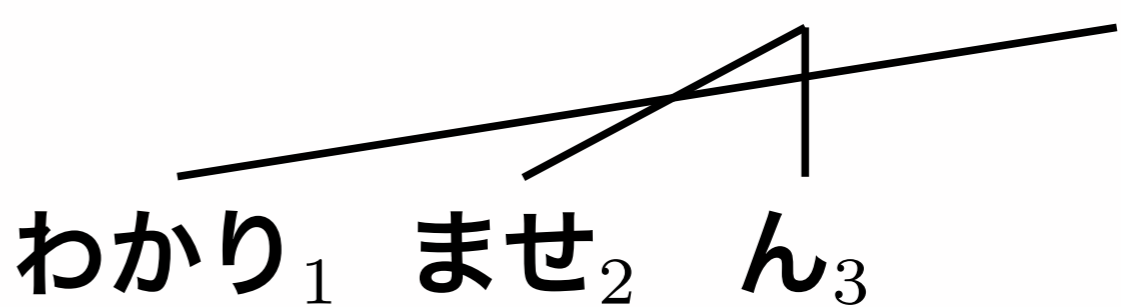
$a = \{ \quad 4 \quad 3 \quad 3 \quad \}$

わかり<sub>1</sub>    ませ<sub>2</sub>    ん<sub>3</sub>

- 単語アライメント  $a$  を導入:  $a$  に基づき、 $e$  から  $f$  を生成
- 1-to-many、NULL により削除を直接表現

# 生成モデル: Model 1



1.  $f$ の単語数を決定  $\text{NULL}_0$      $I_1$      $\text{do}_2$      $\text{not}_3$      $\text{know}_4$
2.  $a$ を決定  

3. 単語の置換     $a = \{ \quad 4 \quad \quad 3 \quad \quad 3 \quad \quad \}$

- $f$ の順番で生成
- Model 1、2、3、HMMによるモデル化

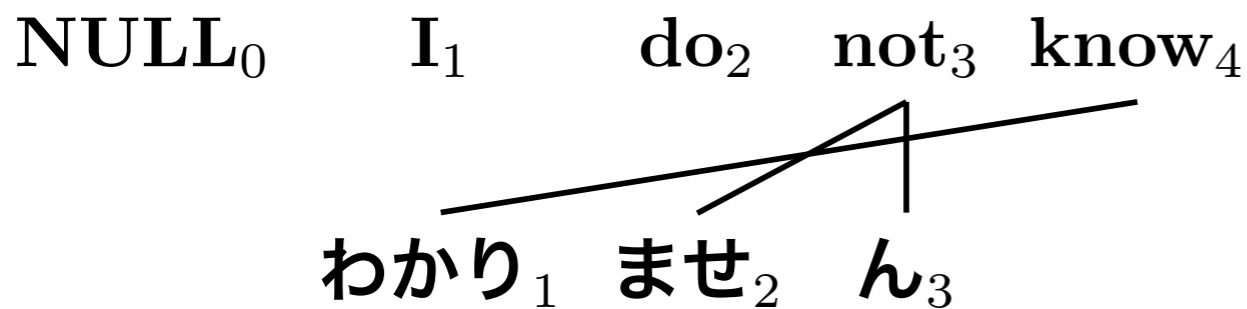


# 例: Model 1

$$\begin{aligned}
 Pr(f|e) &= \sum_a Pr(f, a|e) \\
 &= \sum_a Pr(f|a, e) Pr(a|e) && \text{2. uniformに} \\
 & && \text{アライメント} \\
 &= Pr(m|e) \sum_a Pr(f|a, m, e) Pr(a|m, e) \\
 &\approx \epsilon \sum_a \prod_{j=1}^m t(f_j|e_{a_j}) \frac{1}{(l+1)^m} \text{ s.t. } \forall e : \sum_f t(f|e) = 1
 \end{aligned}$$

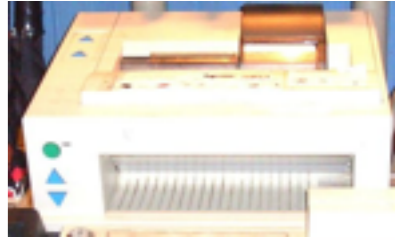
1. 単語数を決定

3. 単語単位に置換



$$\begin{aligned}
 &\epsilon \times t(\text{わかり}_1 | \text{know}_4) \\
 &\times t(\text{ませ}_2 | \text{not}_3) \\
 &\times t(\text{ん}_3 | \text{not}_3) \\
 &\times \frac{1}{5^3}
 \end{aligned}$$

# 生成モデル: Model 4



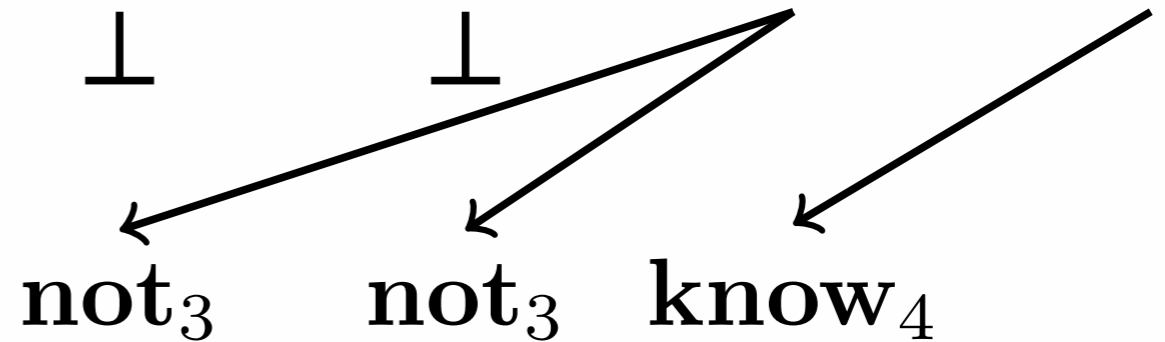
1. 単語数を決定

2. 置換

3. 並び替え

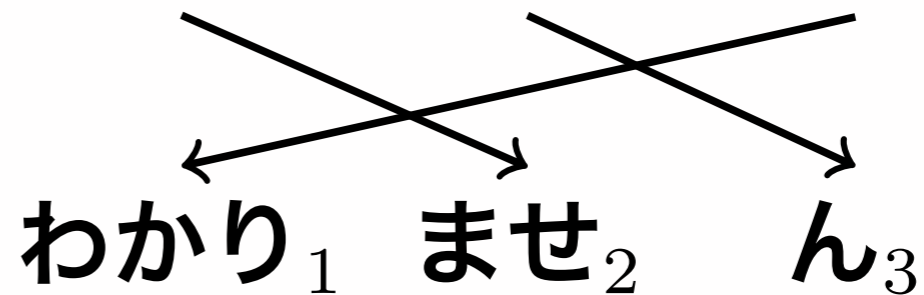
NULL<sub>0</sub>    I<sub>1</sub>    do<sub>2</sub>    not<sub>3</sub>    know<sub>4</sub>

⊥            ⊥            ⊥



↓            ↓            ↓

ませ            ん            わかり



$a = \{ \quad 4 \quad 3 \quad 3 \quad \}$

- eの順番で生成(Model 3、4、5): 複雑な生成過程

# モデルの学習

- 対訳データ  $D: (f, e)$  の集合  $D = \langle \mathcal{F}, \mathcal{E} \rangle$
- データの尤度:  $\prod_{\langle f, e \rangle \in D} Pr(\mathbf{f} | \mathbf{e})$
- データの対数尤度を最大化するパラメータ  $\Theta$  を学習:  $\hat{\theta} = \operatorname{argmax}_{\theta} \sum_{\langle f, e \rangle \in D} \log P_{\theta}(\mathbf{f} | \mathbf{e})$

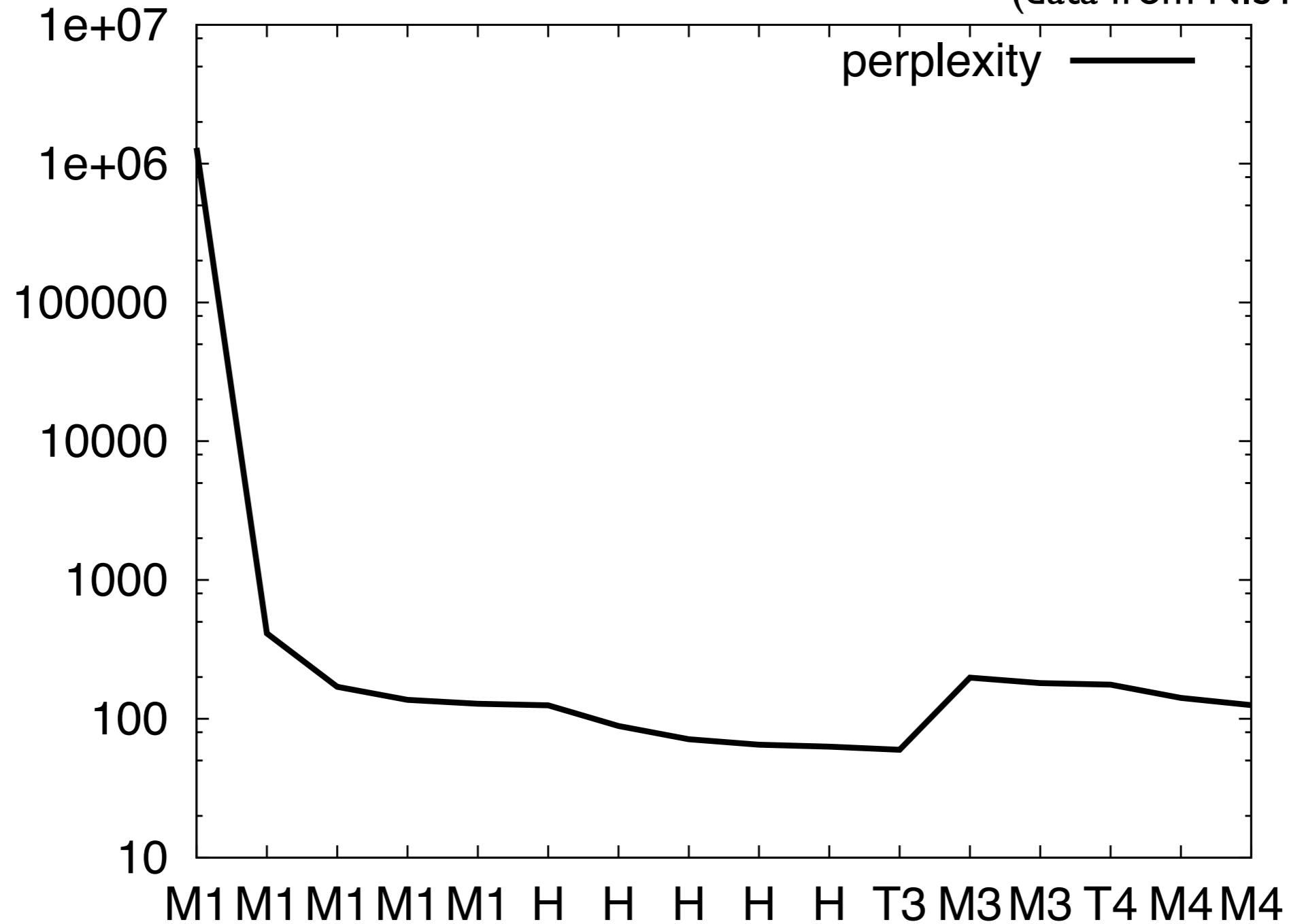
Model 1 の場合、  $\Theta = t(\mathbf{f} | \mathbf{e})$

# EMアルゴリズム

- Eステップ  $q^{(t+1)}(a|f, e) = P_{\theta^{(t)}}(a|f, e)$
- Mステップ  $q^{(t+1)}(a|f, e) = \frac{P_{\theta^{(t)}}(f, a|e)}{\sum_{a'} P_{\theta^{(t)}}(f, a'|e)}$
- $\theta^{(t+1)} = \arg \max_{\theta} \sum_{\langle f, e \rangle} \sum_a q^{(t+1)}(a|f, e) \log P_{\theta}(f, a|e)$
- $\theta^{(t+1)} = \arg \max_{\theta} \mathbb{E}_{q^{(t+1)}} [\log P_{\theta}(f, a|e)]$
- 期待値を計算(q)、qを頻度とみなして  $\Theta$ を更新、繰り返し
- 最適化は困難: Model I 以外は凸関数でない
- 簡単なModel Iから学習、徐々に複雑なモデルを推定

# 对数尤度

(data from NIST08: zh-en)



perplexity:  $\exp(-\text{log-likelihood} / \# \text{ of words in } F)$

# ビタビアライメント

$$Pr(\mathbf{f}|\mathbf{e}) = \sum_a Pr(\mathbf{f}, \mathbf{a}|\mathbf{e})$$

- $\mathbf{f}$ と $\mathbf{e}$ が与えられた時、確率を計算
- 「最も良い」アライメントは?

$$\hat{\mathbf{a}} = \arg \max_a Pr(\mathbf{f}, \mathbf{a}|\mathbf{e})$$

- $\Sigma$ の代わりに、 $\arg\max$



# 単語アライメント例

NULL<sub>0</sub> Bolivia<sub>1</sub> Holds<sub>2</sub> Presidential<sub>3</sub> and<sub>4</sub> Parliament<sub>5</sub> Elections<sub>6</sub>

玻利维亚<sub>1</sub> 举行<sub>2</sub> 总统<sub>3</sub> 与<sub>4</sub> 国会<sub>5</sub> 选举<sub>6</sub>

NULL<sub>0</sub> The<sub>1</sub> polling<sub>2</sub> stations<sub>3</sub> closed<sub>4</sub> at<sub>5</sub> 4<sub>6</sub> p.m.<sub>7</sub> ( <sub>8</sub> 4<sub>9</sub> a.m.<sub>10</sub> on<sub>11</sub> )

投票所<sub>1</sub> 于<sub>2</sub> 下午<sub>3</sub> 四时<sub>4</sub> ( <sub>5</sub> 台北<sub>6</sub> 时间<sub>7</sub> 七月<sub>8</sub> 一日<sub>9</sub> 清晨<sub>10</sub> 四时<sub>11</sub> )

NULL<sub>0</sub> The<sub>1</sub> Bolivian<sub>2</sub> President<sub>3</sub> Quiroga<sub>4</sub> ,<sub>5</sub> whose<sub>6</sub> term<sub>7</sub> soon<sub>8</sub> expires<sub>9</sub>

任期<sub>1</sub> 即将<sub>2</sub> 届满<sub>3</sub> 的<sub>4</sub> 玻利维亚<sub>5</sub> 总统<sub>6</sub> 吉洛卡<sub>7</sub> 说<sub>8</sub> :<sub>9</sub>

NULL<sub>0</sub> The<sub>1</sub> Organization<sub>2</sub> of<sub>3</sub> American<sub>4</sub> States<sub>5</sub> has<sub>6</sub> sent<sub>7</sub> 50<sub>8</sub> observers<sub>9</sub>

美洲<sub>1</sub> 国家<sub>2</sub> 组织<sub>3</sub> 派遣<sub>4</sub> 五十<sub>5</sub> 名<sub>6</sub> 观察员<sub>7</sub> 监督<sub>8</sub> 玻利维亚<sub>9</sub>

NULL<sub>0</sub> Government<sub>1</sub> officials<sub>2</sub> said<sub>3</sub> that<sub>4</sub> the<sub>5</sub> border<sub>6</sub> has<sub>7</sub> been<sub>8</sub> closed<sub>9</sub> so<sub>10</sub>

政府<sub>1</sub> 官员<sub>2</sub> 说<sub>3</sub> ,<sub>4</sub> 边界<sub>5</sub> 已经<sub>6</sub> 关闭<sub>7</sub> 以<sub>8</sub> 防止<sub>9</sub> 选举<sub>10</sub>

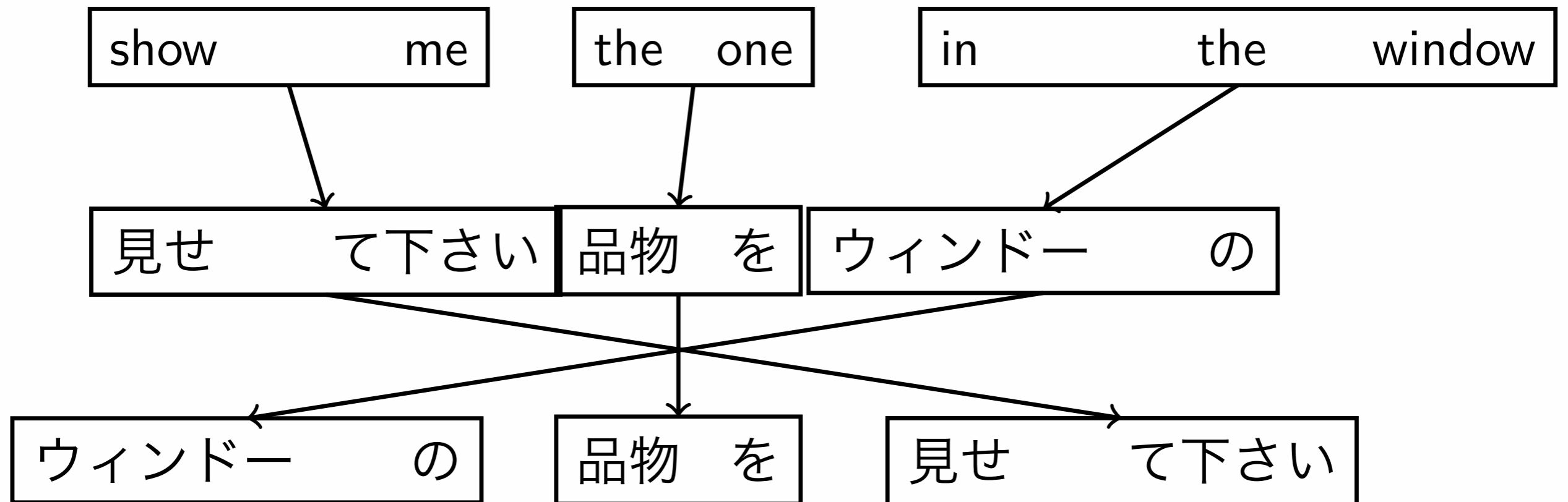
(data from NIST08: zh-en)

# 研究課題

- one-to-manyの呪縛: ヒューリスティック、事後確率による閾値(Liang et al., 2006)、事後確率正則化(Ganchev et al., 2010)、many-to-manyモデル(Fraser and Marcu, 2007)
- スムージング:  $L_0$ 正則化(Vaswani et al., 2012)、ナイーブベイズ
- 句単位のアライメント (Marcu and Wong, 2002; DeNero et al., 2008; Blunsom et al., 2009; Neubig et al., 2011)
- ツール: Giza++, PostCat, cicada, pialign

句に基づく機械翻訳

# 句に基づくと機械翻訳



- fを句へと分解 + 各句を翻訳 + 並び替え

# 句に基づくモデル

句単位の変換

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

並び替え

言語モデル

- $\Phi$ :  $(f, e)$ の句単位の分割、 $\alpha$ : 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

# 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 $h(e, d, f)$ 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)
- $h$ の例:  $h(f, d, e) = \begin{pmatrix} \log p_d(f, \alpha | \phi) \\ \log p_\phi(\phi | e) \\ \log p_{lm}(e) \end{pmatrix}$
- $w$ : 各素性の重み付け

# Questions

$$\hat{e} = \arg \max_e \mathbf{w}^\top \cdot \mathbf{h}(e, d, f)$$

- 学習: 句とパラメータをどのように学習するか(d and h)?
- デコード(探索): どのようにして最適な翻訳を見つけるか(argmax)?
- チューニング(最適化): どのようにして重み付けをするか(w)?
- 評価: 最適な翻訳は良い翻訳か( $\hat{e}$ )?



# 学習



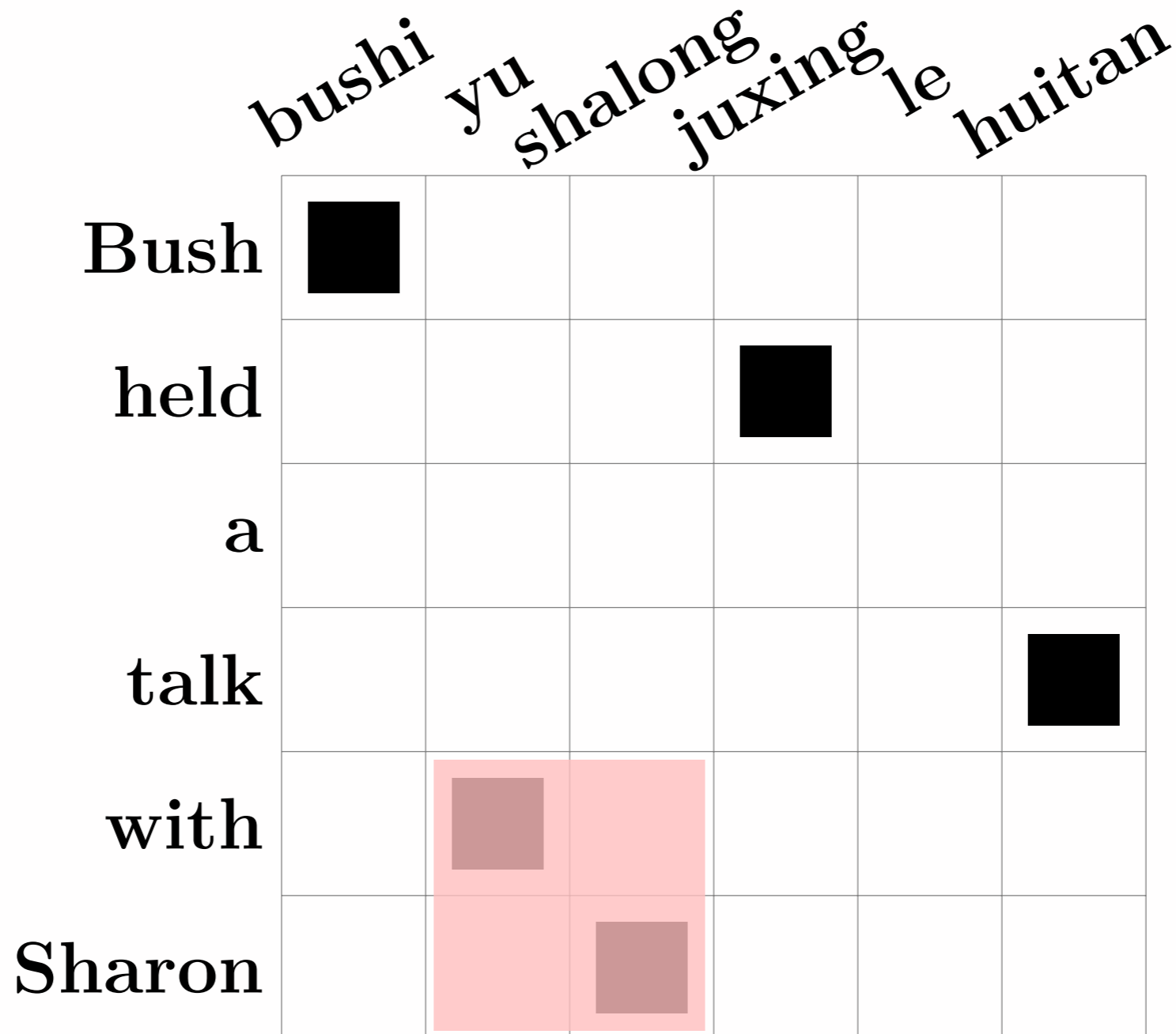
- $D = \langle \mathcal{F}, \mathcal{E} \rangle$  からフレーズペア  $\Phi$  を学習
- 標準的なヒューリスティックな手法  
(Koehn et al., 2003)
  1. 単語アライメントの計算
  2. フレーズペアの抽出
  3. フレーズペアのスコアリング

# 単語アライメント

	<i>bushi</i>	<i>yu</i>	<i>shalong</i>	<i>juxing</i>	<i>le</i>	<i>huitan</i>
Bush	■					
held				■		
a						
talk						■
with		■				
Sharon			■			

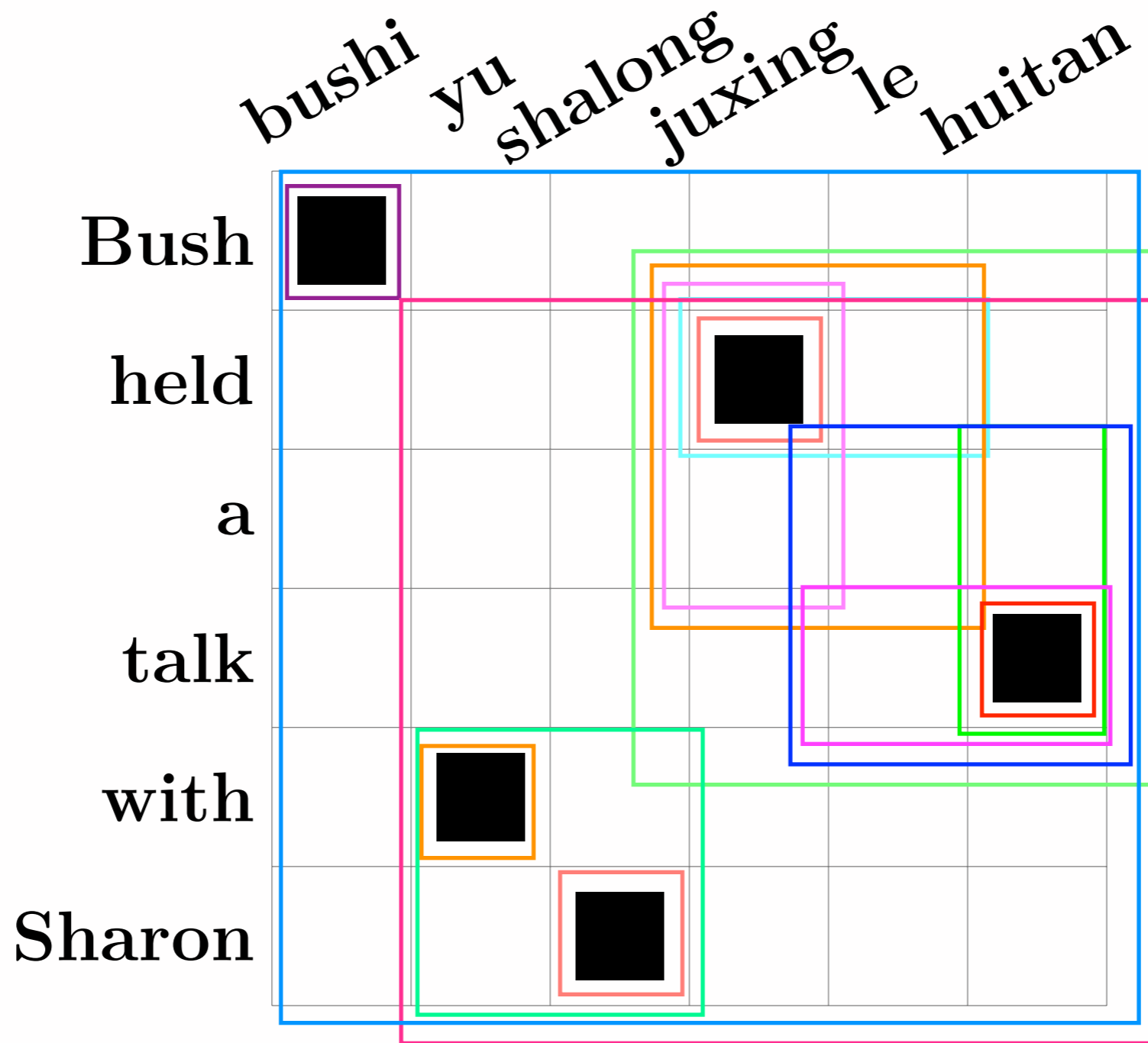
(Example from Huang and Chiang, 2007)

# フレーズペアの抽出



- 一貫した句(単語アライメントが閉じている句)を抽出

# 網羅的に抽出



# 句に対応した素性

$$\log p_{\phi}(\bar{f}|\bar{e}) = \log \frac{\text{count}(\bar{e}, \bar{f})}{\sum_{\bar{f}'} \text{count}(\bar{e}, \bar{f}')}$$

$$\log p_{\phi}(\bar{e}|\bar{f}) = \log \frac{\text{count}(\bar{e}, \bar{f})}{\sum_{\bar{e}'} \text{count}(\bar{e}', \bar{f})}$$

- データから全ての句を抽出、頻度に基づく最尤推定 + その他の素性

不 熟 悉 ||| 'm not familiar ||| -1.4859937213 -7.2301988107 -0.3036824138 -3.0311892056

不 熟 悉 ||| do n't know ||| -1.2064088591 -5.3571402084 -3.4402617349 -6.8870595804

不 熟 悉 ||| i 'm not familiar ||| -2.522085653 -9.1804032749 -1.06784063 -3.0311892056

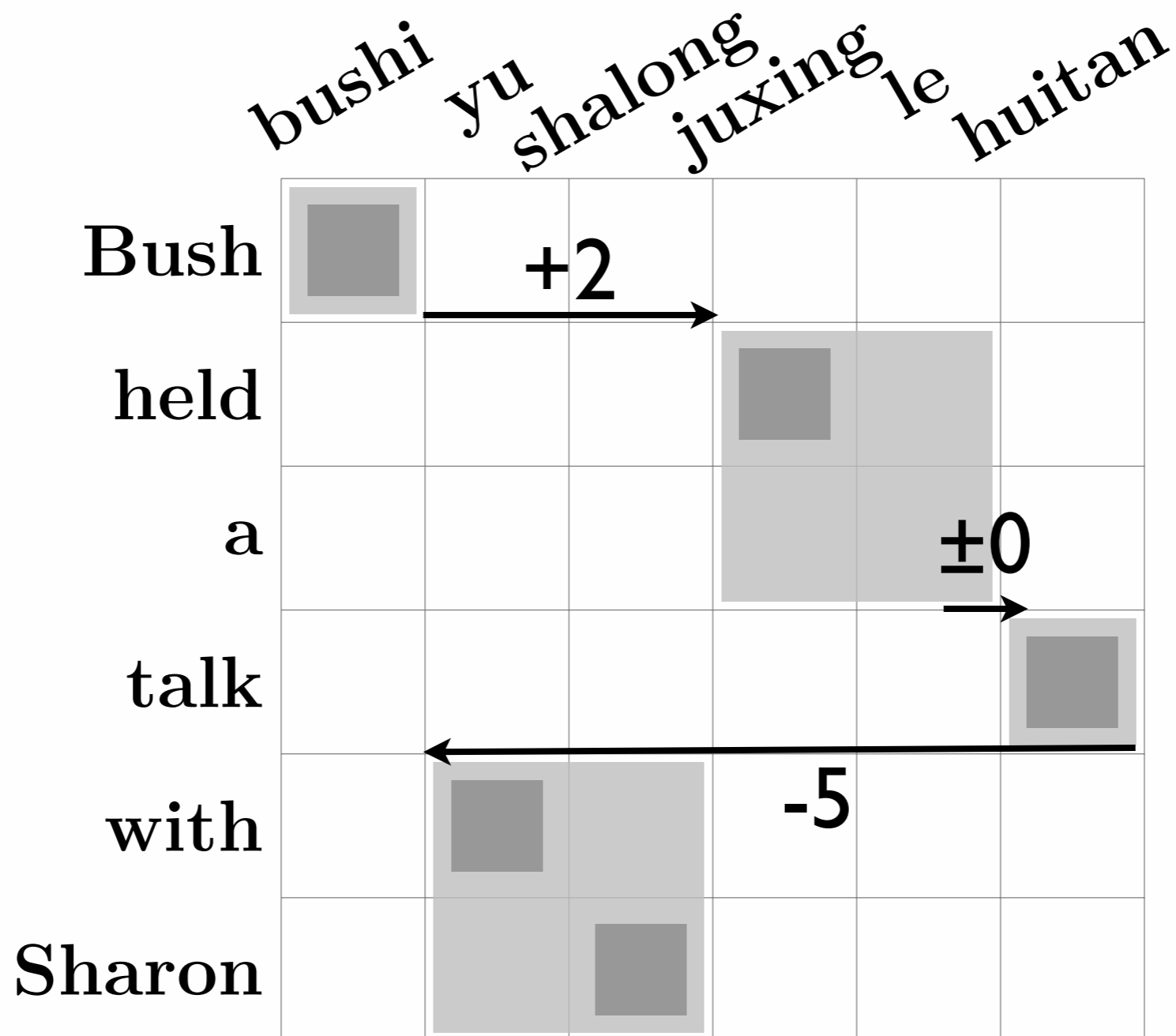
不 熟 悉 ||| it will be great ||| -2.522085653 -20.871716142 0.0 -11.4593095552

不 熟 悉 ||| not accustomed ||| -2.522085653 -5.5628513514 -0.6931471806 -2.2177906617

不 熟 悉 ||| not accustomed to ||| -2.522085653 -8.5631752395 0.0 -2.2177906617

不 熟 悉 ||| not familiar ||| -1.8754584881 -3.4150084505 -0.4212134651 -2.4210642434

# 並び替え素性



- 距離に基づく素性 (注意: 確率モデルではない)

$$\log p_d(f, \alpha | \phi) = -|+2| - |0| - |-5| = -7$$



# 研究課題

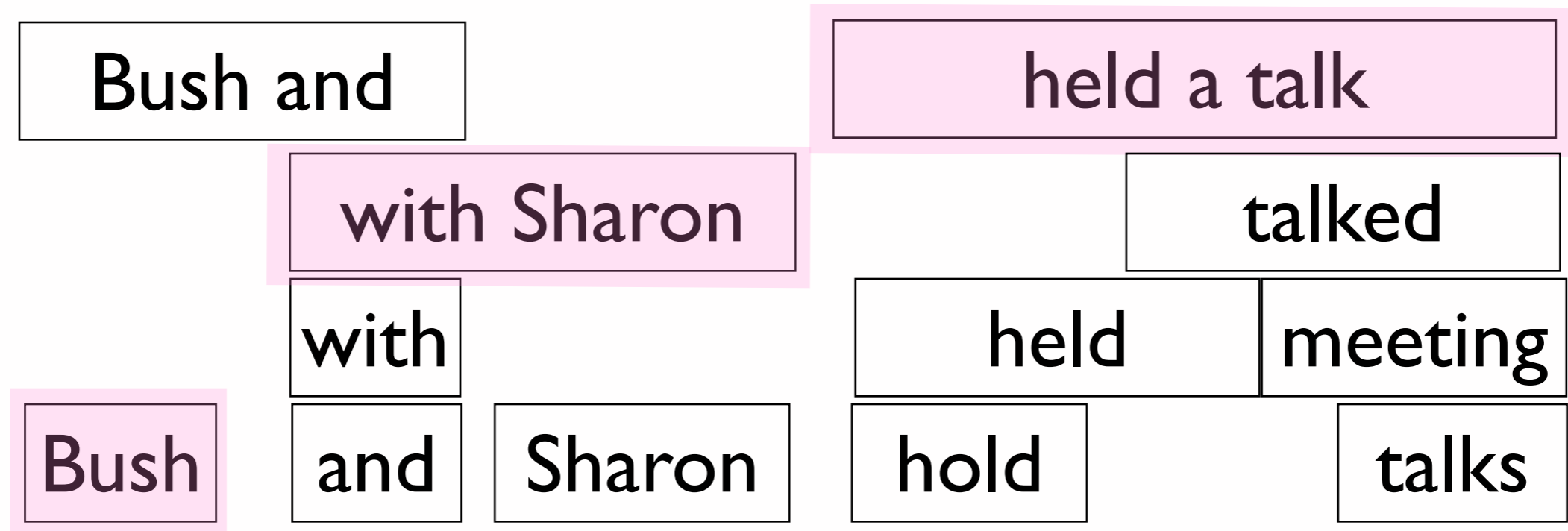
- ヒューリスティックなモデル?
- 句を全て列挙してペアを求める問題は、#P困難
- ノンパラメトリック・ベイズ法による推定  
(DeNero et al., 2008; Blunsom et al., 2009; Neubig et al., 2011)
- 並び替え素性? 非常に重要とわかってきた
- 大量の素性(Cherry, 2013)
- 文全体を考慮したモデル(Goto et al., 2013)

# Questions

$$\hat{e} = \arg \max_e \mathbf{w}^\top \cdot \mathbf{h}(e, d, f)$$

- 学習: 句とパラメータをどのように学習するか(d and h)?
- デコード(探索): どのようにして最適な翻訳を見つけるか(argmax)?
- チューニング(最適化): どのようにして重み付けをするか(w)?
- 評価: 最適な翻訳は良い翻訳か( $\hat{e}$ )?

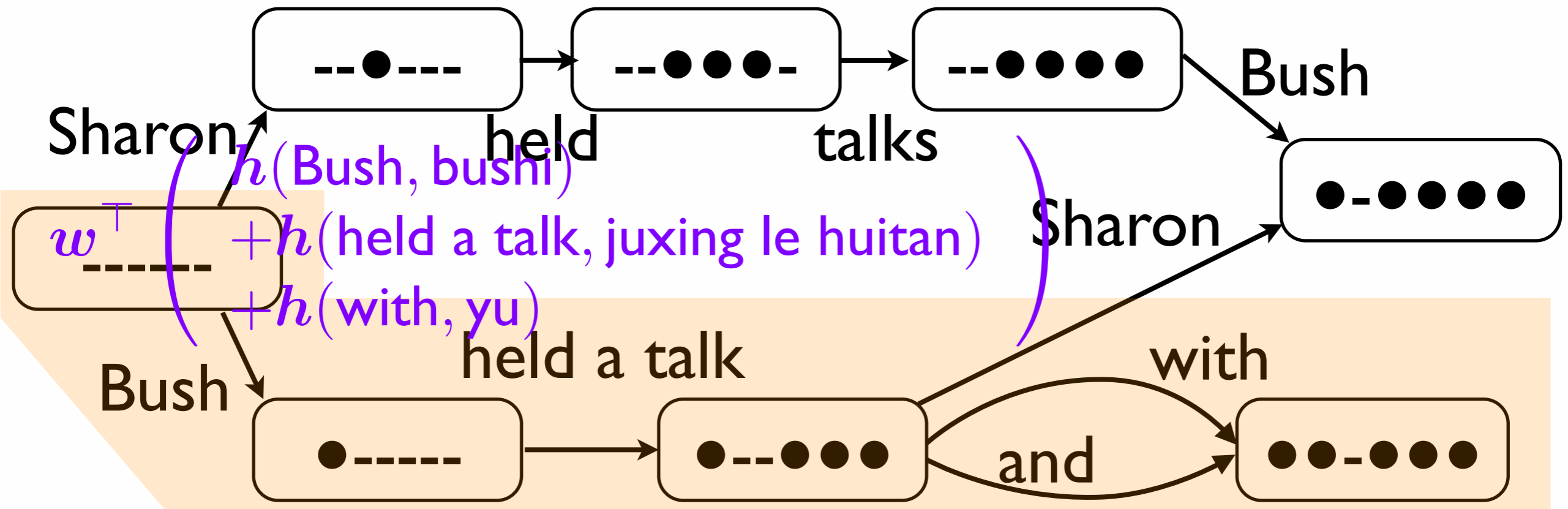
# フレーズペアの列挙



bushi yu shalong juxing le huitan

- 入力文fに対し、原言語側がマッチする句を列挙
- 最もよい、フレーズペアの選択 + 並び替え

# フレーズベースな探索空間



bushi yu shalong juxing le huitan

- ノード: 翻訳された原言語の単語位置を表す bit-vector
- エッジ: left-to-right に組み合わされる目的言語側の句
- 探索空間:  $O(2^n)$ 、時間:  $O(2^n n^2)$  (Why?)

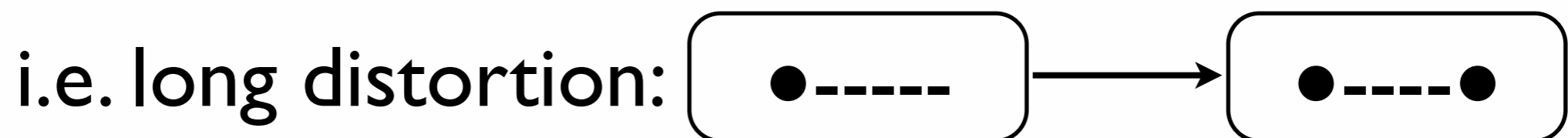
# 巡回セールスマン問題

- NP-hard problem: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)

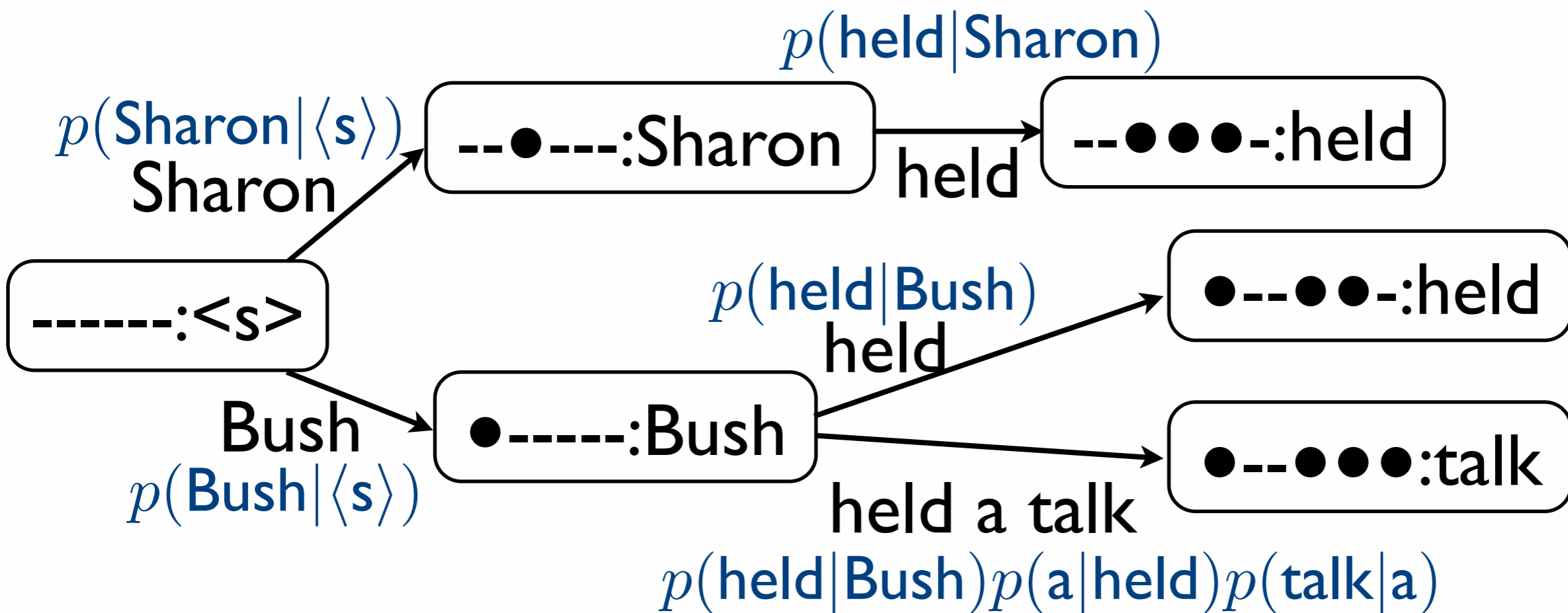
- 原言語の各単語 = 都市
- 動的計画法(DP)による解:
  - State: 訪れた都市 (bit-vector)
  - 探索空間:  $O(2^n)$



- 探索空間を小さくするため、並び替えに制約

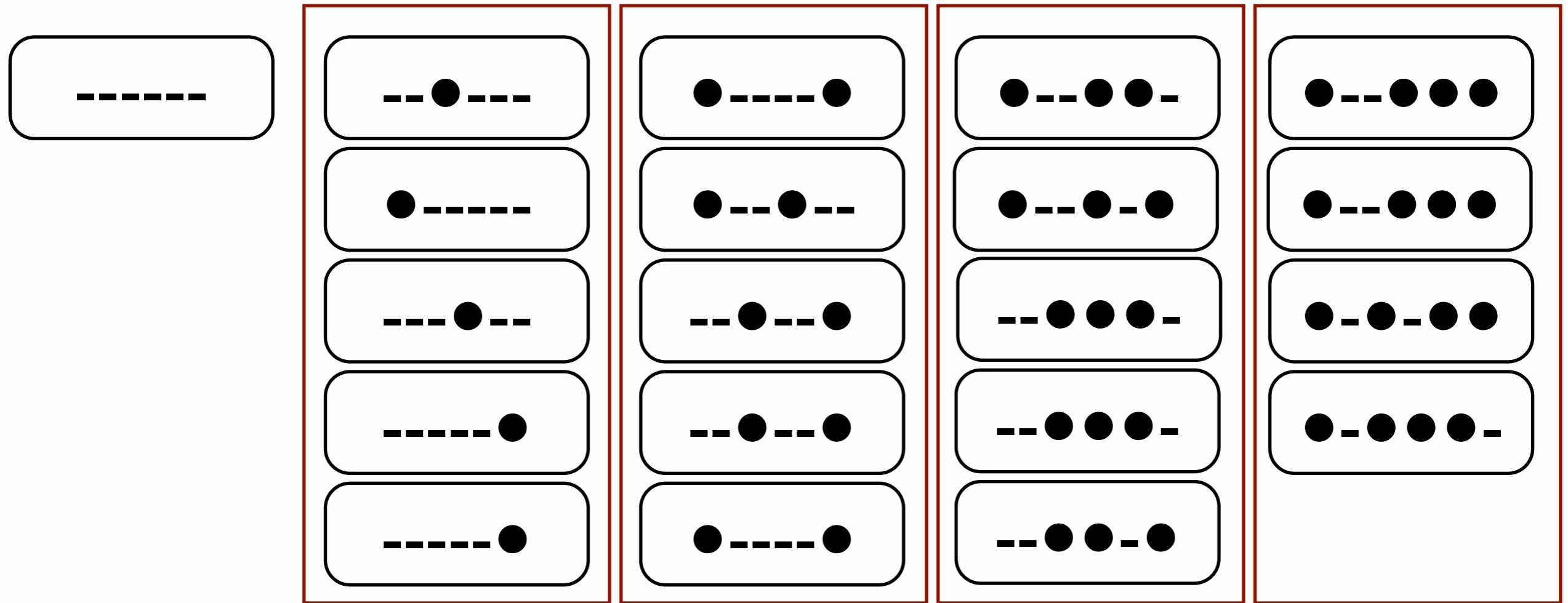


# 局所的でない素性



- フレーズに閉じていない素性: bigram 言語モデル
- 「将来のスコアの計算」のために、1単語保持
- m-gram LM: 探索空間:  $O(2^n V^{m-1})$ , 時間:  $O(2^n V^{m-1} n^2)$

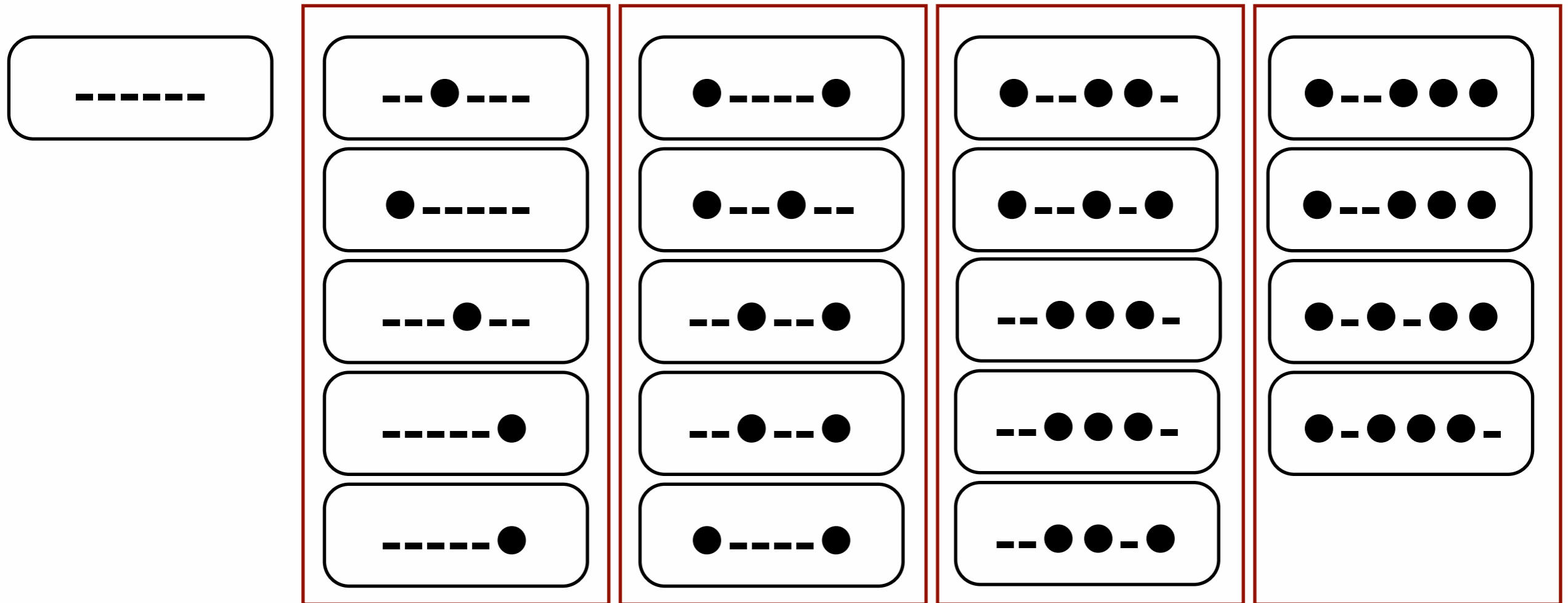
# フレーズベースなデコーディング



- 探索空間を「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



# プルーニング



- 同じグループの仮説内部でプルーニング
- 数あるいはスコアによるプルーニング
- $O(2^n)$ の項を $O(nb)$ へ縮小

# 研究課題

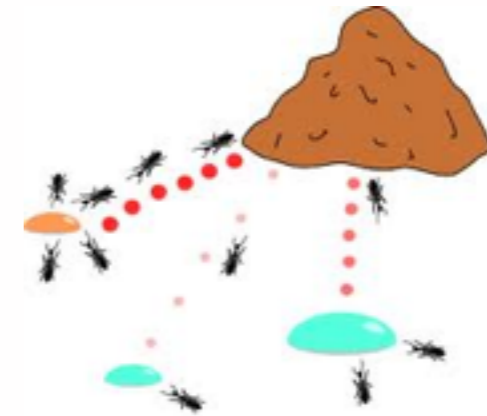
- より複雑な素性を入れたデコーディング
- graph-based依存構造解析に基づいた並び替え制約  
(Cherry et al., 2012)
- 統語論的な言語モデルの統合(Schwartz et al., 2011)
- 探索エラーのないデコーディング
- ラグランジアン緩和法(Chang and Collins, 2011)
- ビーム幅の最適化(Rush et al., 2013)

# Questions

$$\hat{e} = \arg \max_e \mathbf{w}^\top \cdot \mathbf{h}(e, d, f)$$

- 学習: 句とパラメータをどのように学習するか(d and h)?
- デコード(探索): どのようにして最適な翻訳を見つけるか(argmax)?
- チューニング (最適化): どのようにして重み付けをするか(w)?
- 評価: 最適な翻訳は良い翻訳か( $\hat{e}$ )?

# 最適化



$$\begin{aligned}\hat{\boldsymbol{w}} &= \arg \min_{\boldsymbol{w} \in \mathcal{W}} \mathbb{E}_{Pr(F, E)} [\ell(F, E; \boldsymbol{w})] \\ &= \arg \min_{\boldsymbol{w} \in \mathcal{W}} \ell(F, E; \boldsymbol{w}) + \lambda \Omega(\boldsymbol{w})\end{aligned}$$

- ある損失関数  $\ell$  を仮定し、対訳データ  $(F, E)$  に対するリスクを最小化□□
- 真の分布は未知なため、正則化 ( $\Omega$ ) された経験リスクを最小化

# MERT

$$\hat{\boldsymbol{w}} = \arg \min_{\boldsymbol{w}} \ell \left( \left\{ \arg \max_e \boldsymbol{w}^\top \cdot \boldsymbol{h}(\boldsymbol{e}, \boldsymbol{f}_s) \right\}_{s=1}^S, \{\boldsymbol{e}_s\}_{s=1}^S \right)$$

- MERT (Minimum Error Rate Training) (Och, 2003)
- 統計的機械翻訳では標準(でも他のNLPなタスクでは使われない)
- $\ell$  (.)に対して、様々なエラー関数を使用可能
- $\boldsymbol{w}$ を更新するたびに、argmaxを計算:n-best近似

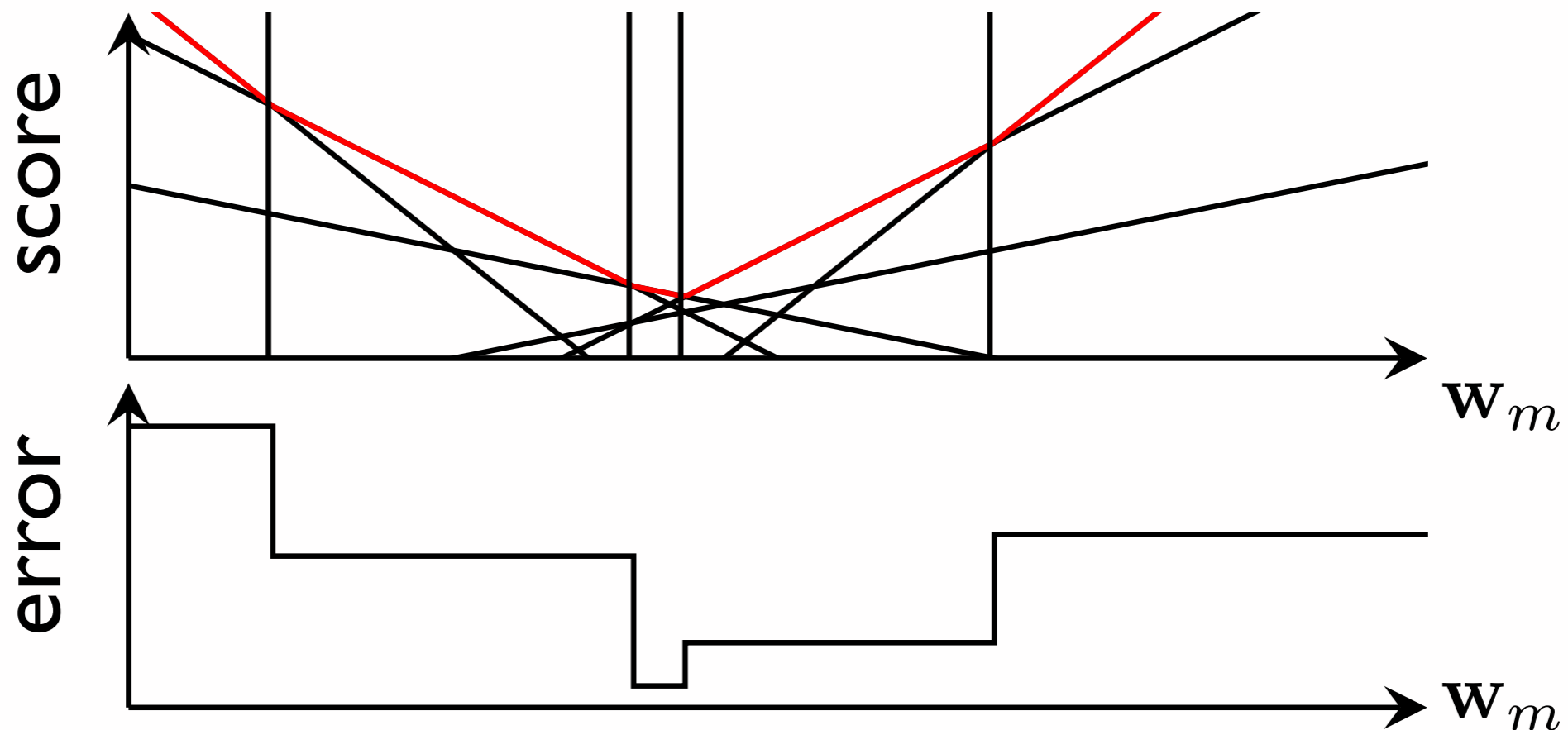
# n-best 結合による近似

```
1: procedure MERT( $\{(e_s, f_s)\}_{s=1}^S$ )
2:   for  $n = 1 \dots N$  do
3:     Decode and generate nbest list using  $w$ 
4:     Merge nbest list
5:     for  $k = 1 \dots K$  do
6:       for each parameter  $m = 1 \dots M$  do
7:         Solve one dimensional optimization
8:       end for
9:       update  $w$ 
10:    end for
11:  end for
12: end procedure
```

- 現在の $w$ でn-bestを生成、結合(N回)
- M次元( $M =$  素性の数)の各次元に対して、最適化、 $w$ を更新(K回)

# Line Searchの効率化

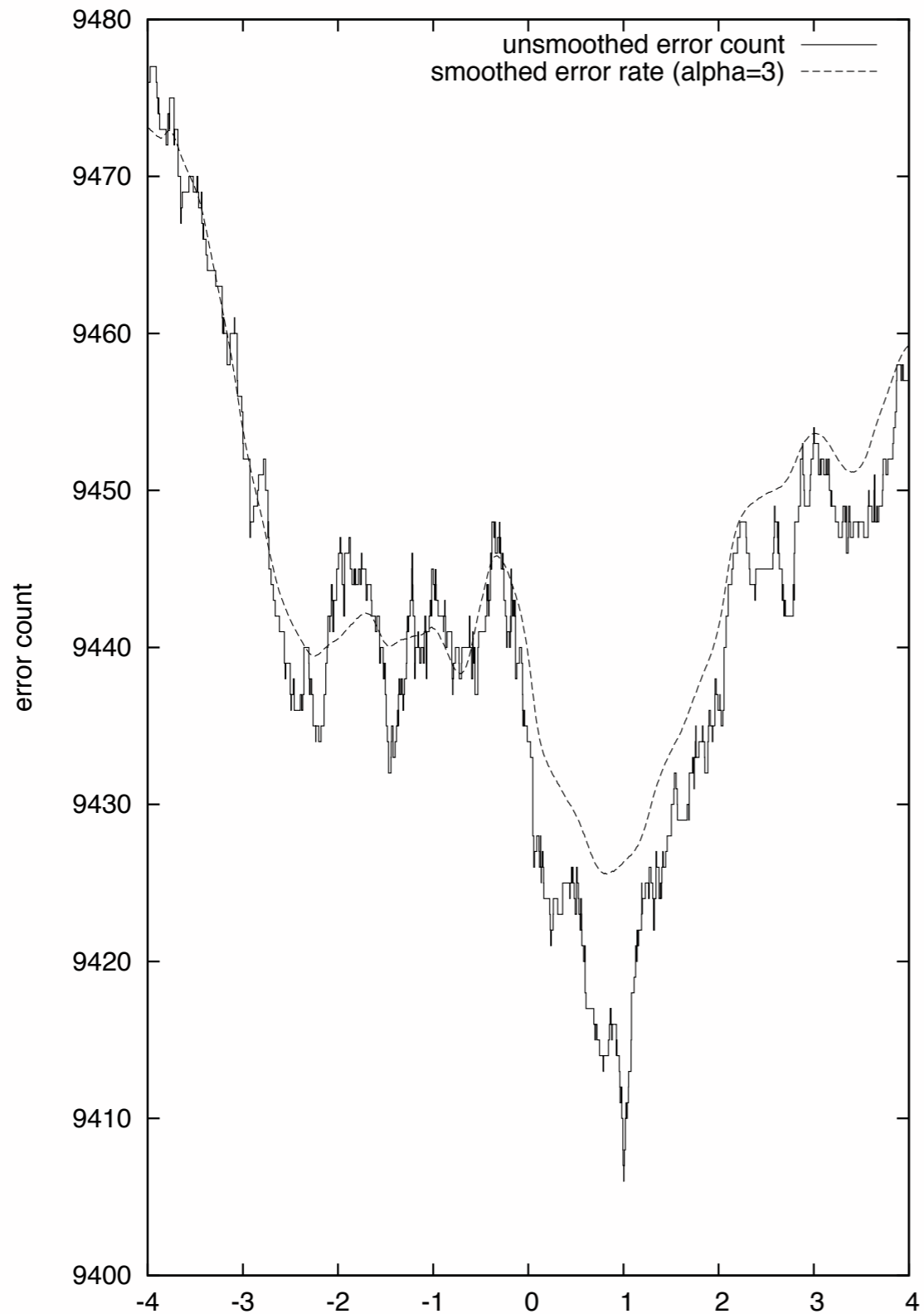
$$\hat{e} = \operatorname{argmax}_e \underbrace{w_m^\top \cdot h_m(e, f_s)}_{\text{slope}} + \underbrace{w_{m-}^\top \cdot h_{m-}(e, f_s)}_{\text{constant}}$$



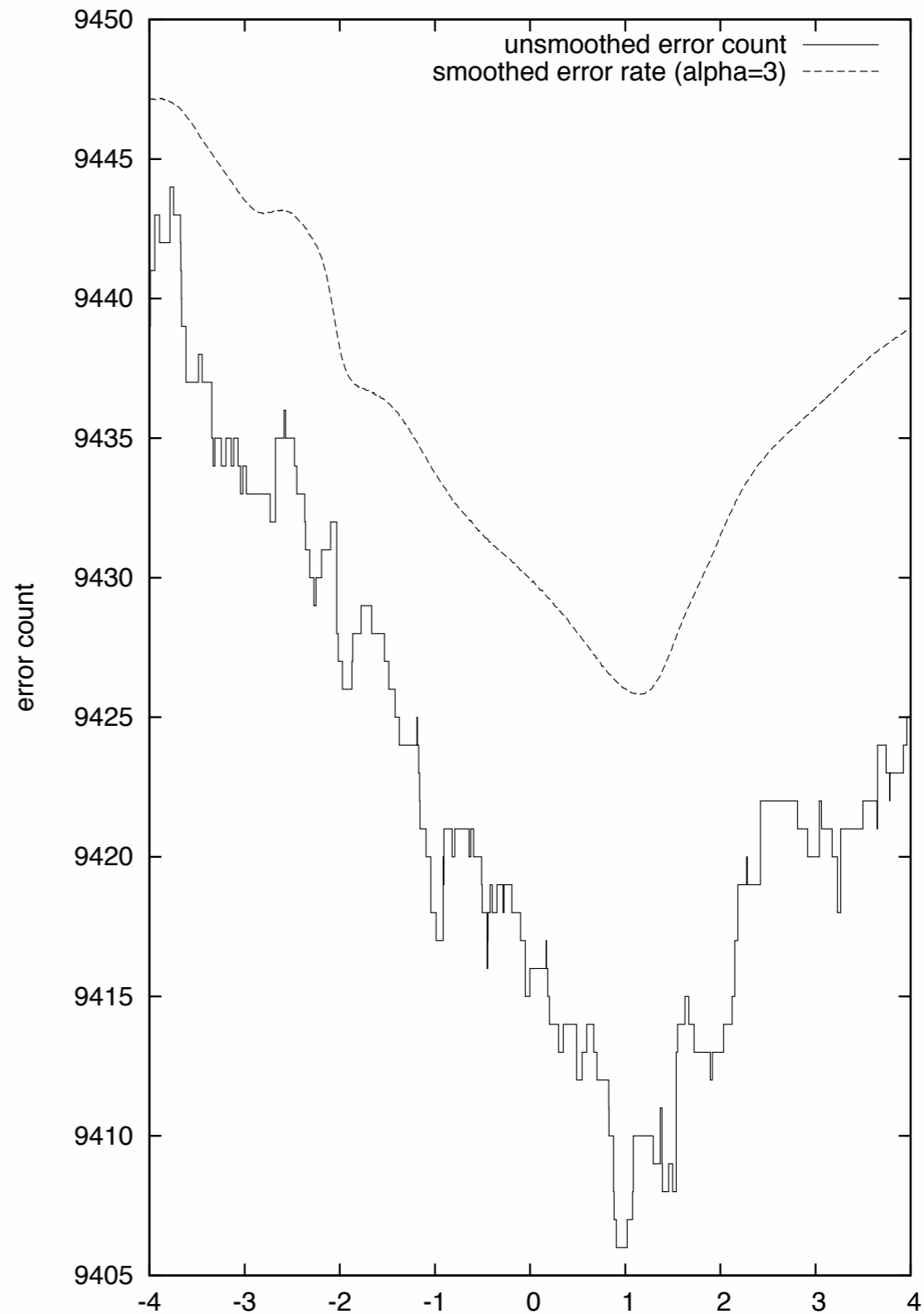
- 一つの次元を選択した場合、その仮説を「線」としてみなせる
- 「線」の集合から、凸包(convex hull)を計算



# エラー曲線



63



(Och, 2003)

# MERTの現実

- ランダムな初期値 (Macherey et al., 2008; Moore and Quirk, 2008)、方向 (Macherey et al., 2008)
- スムージング (Cer et al., 2008)、Regularization (Hayashi et al., 2009 ; Galley et al., 2013)
- Forest/LatticeからのMERT (Macherey et al., 2008; Kumar et al., 2009)
- 凸包を計算、その後最適化 (Galley and Quirk, 2011)
- 期待BLEUから「方向」を決定、その方向で線分探索 (Galley et al., 2013)
- 最低3回MERT、平均BLEUを報告しなさい (Clark et al., 2011)  
(そんなアホな)

# 研究課題

- MERT以外の手法、目的関数が主流に
  - PRO(Pair-wise Ranking Optimization) (Hopkins and May, 2011)、MIRA(Cherry and Foster, 2012)
  - 期待BLEU、xBLEU (Rosti et al., 2010)
- オンライン学習(Watanabe et al, 2007; Chiang et al., 2008)
  - 素性選択 (Levenberg et al., 2012)
  - ミニバッチ+勾配の調整(Watanabe, 2012)
  - Violation-fixing パーセプトロン(Yu et al., 2013)

# Questions

$$\hat{e} = \arg \max_e \mathbf{w}^\top \cdot \mathbf{h}(e, d, f)$$

- 学習: 句とパラメータをどのように学習するか(d and h)?
- デコード(探索): どのようにして最適な翻訳を見つけるか(argmax)?
- チューニング (最適化): どのようにして重み付けをするか(w)?
- 評価: 最適な翻訳は良い翻訳か( $\hat{e}$ )?

# 評價: ngram適合率

Well , I 'd like to stay five nights beginning October twenty-fifth to thirty .

- I 'd like to stay there for five nights , from October twenty fifth to the thirtieth .
- I want to stay for five nights , from October twenty fifth to the thirtieth .
- I 'd like to stay for five nights , from October twenty fifth to the thirtieth .
- I would like to reserve a room for five nights , from October twenty fifth to the thirtieth .

# 評價: ngram適合率

Well , I 'd like stay five nights beginning  
October twenty-fifth to thirty .

$$p_1 = \frac{11}{15}$$

- I 'd like to stay there for five nights , from October twenty fifth to the thirtieth .
- I want to stay for five nights , from October twenty fifth to the thirtieth .
- I 'd like to stay for five nights , from October twenty fifth to the thirtieth .
- I would like to reserve a room for five nights , from October twenty fifth to the thirtieth .

# 評價: ngram適合率

Well , I 'd like to stay five nights beginning October twenty-fifth to thirty .

$$p_1 = \frac{11}{15} \quad p_2 = \frac{5}{14}$$

- I 'd like to stay there for five nights , from October twenty fifth to the thirtieth .
- I want to stay for five nights , from October twenty fifth to the thirtieth .
- I 'd like to stay for five nights , from October twenty fifth to the thirtieth .
- I would like to reserve a room for five nights , from October twenty fifth to the thirtieth .



# 評價: ngram適合率

Well , I 'd like to stay five nights beginning October twenty-fifth to thirty .

$$p_1 = \frac{11}{15} \quad p_2 = \frac{5}{14} \quad p_3 = \frac{3}{13}$$

- I 'd like to stay there for five nights , from October twenty fifth to the thirtieth .
- I want to stay for five nights , from October twenty fifth to the thirtieth .
- I 'd like to stay for five nights , from October twenty fifth to the thirtieth .
- I would like to reserve a room for five nights , from October twenty fifth to the thirtieth .

# 評價: ngram適合率

Well , I 'd like to stay five nights beginning October twenty-fifth to thirty .

$$p_1 = \frac{11}{15} \quad p_2 = \frac{5}{14} \quad p_3 = \frac{3}{13} \quad p_4 = \frac{2}{12}$$

- I 'd like to stay there for five nights , from October twenty fifth to the thirtieth .
- I want to stay for five nights , from October twenty fifth to the thirtieth .
- I 'd like to stay for five nights , from October twenty fifth to the thirtieth .
- I would like to reserve a room for five nights , from October twenty fifth to the thirtieth .

# 評価: BLEU

$$\exp \left( \sum_{n=1}^4 \frac{1}{4} \log p_n \right) \cdot \min \left\{ \exp \left( 1 - \frac{r}{c} \right), 1 \right\}$$



- 適合率の対数線形結合 (Papineni et al., 2002)
- 短すぎる翻訳に対する簡潔ペナルティ
- $r$  = 参照訳の長さ,  $c$  = 翻訳候補の長さ
- 文単位でなく、文章単位の評価尺度

# 研究課題

- BLEUは10年以上も標準的な評価尺度
- 計算が簡単なため、BLEUを直接最適化システムを実現可能 (例: MERT)
- 文単位の評価?
- 他の尺度: WER、PER、TER、METEOR、RIBES等

# Questions

$$\hat{e} = \arg \max_e \mathbf{w}^\top \cdot \mathbf{h}(e, d, f)$$

- 学習: 句とパラメータをどのように学習するか(d and h)?
- デコード(探索): どのようにして最適な翻訳を見つけるか(argmax)?
- チューニング (最適化): どのようにして重み付けをするか(w)?
- 評価: 最適な翻訳は良い翻訳か( $\hat{e}$ )?

# Answered?

- 文法のないモデル(でも結構頑健)
- 高速なデコーディング
- $w$ の最適化はなぜ必要?
- “Phrase-based MT is wrong. It is against my linguistic intuition.” (Mercer, 2013)
- 統語論的な知識の導入: 木構造に基づく機械翻訳