

階層的フレーズベース

機械翻訳

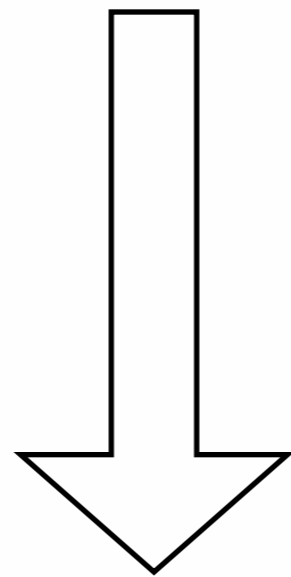
渡辺太郎

taro.watanabe at nict.go.jp



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

機械翻訳について勉強したい。



どないしましょう？

I want to study about machine translation.

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

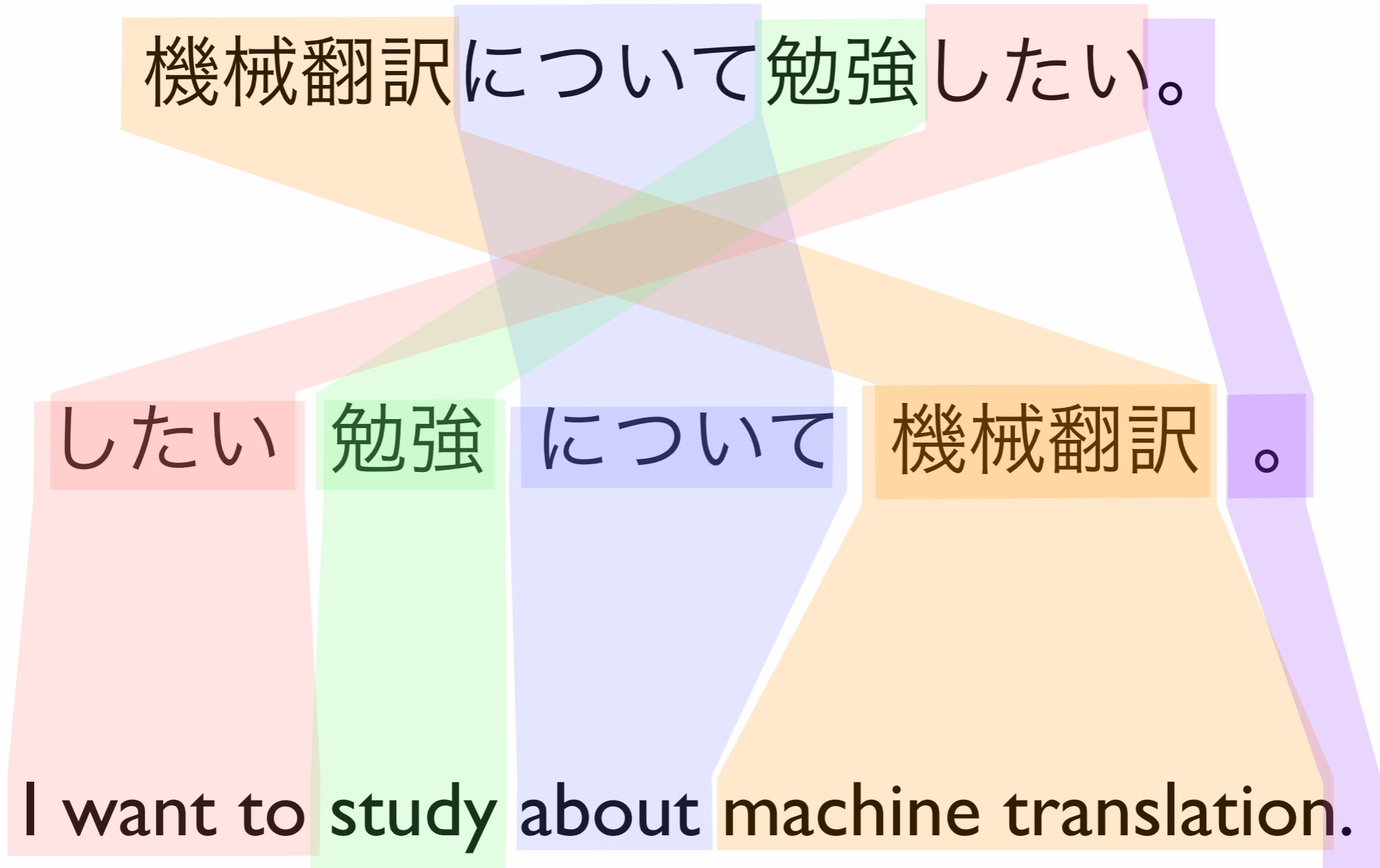
$Pr(\text{勉強} | \text{study})$ about

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

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

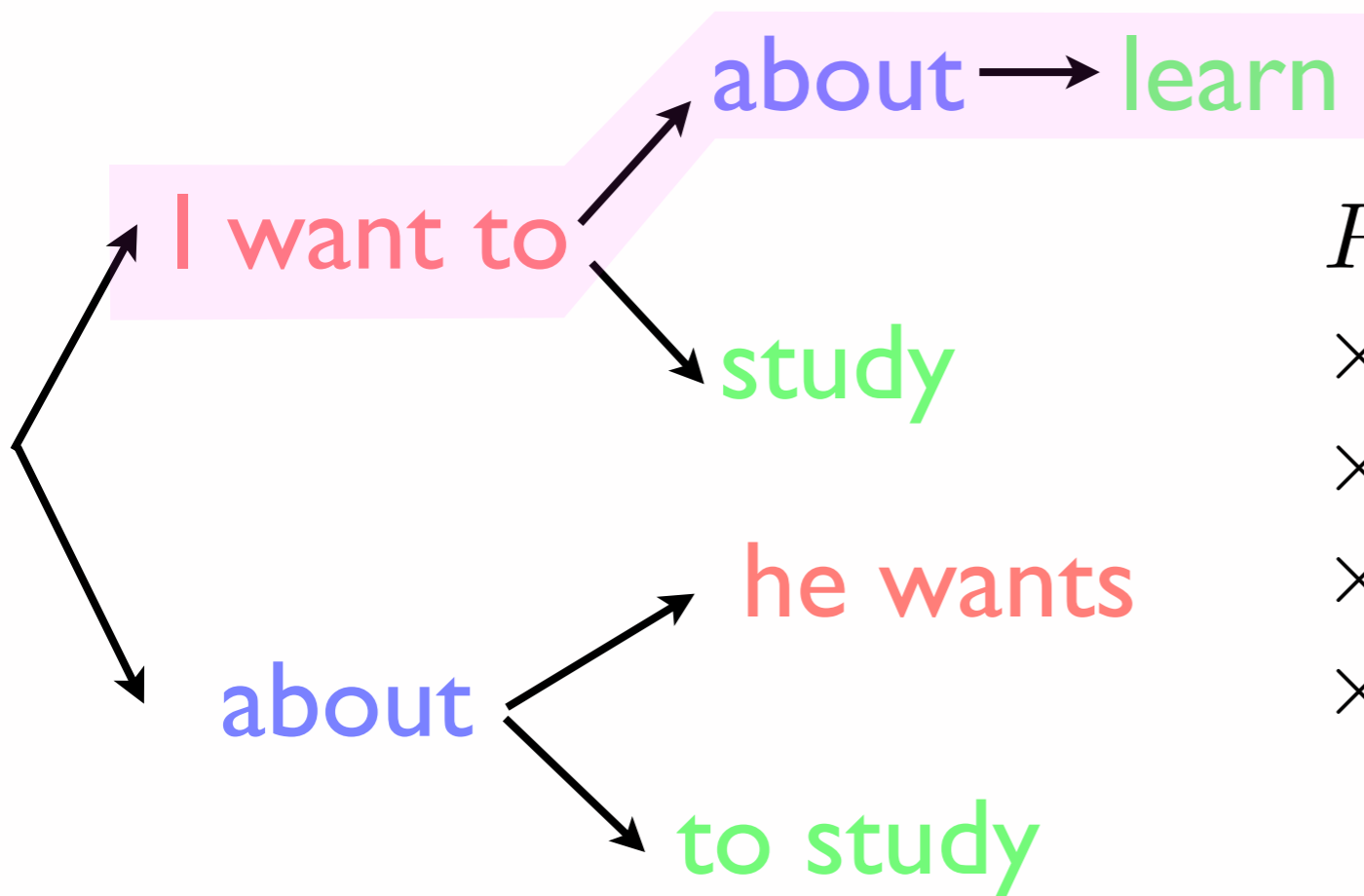
$Pr(\circ | \bullet)$

問題分割

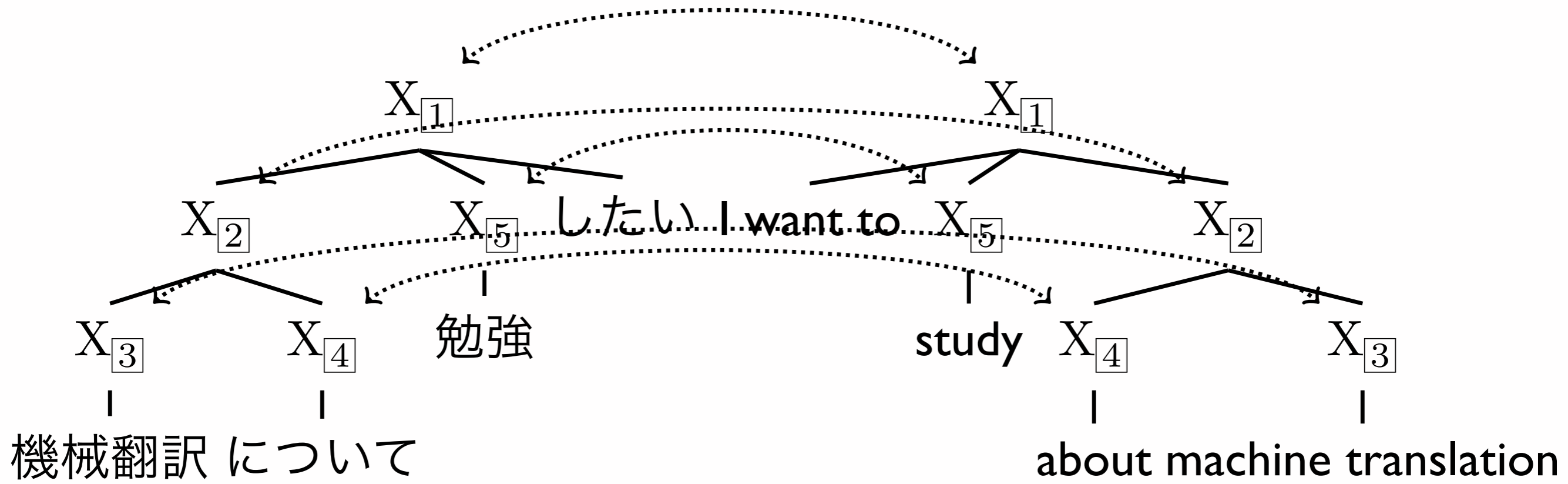


探索

機械翻訳について勉強したい。

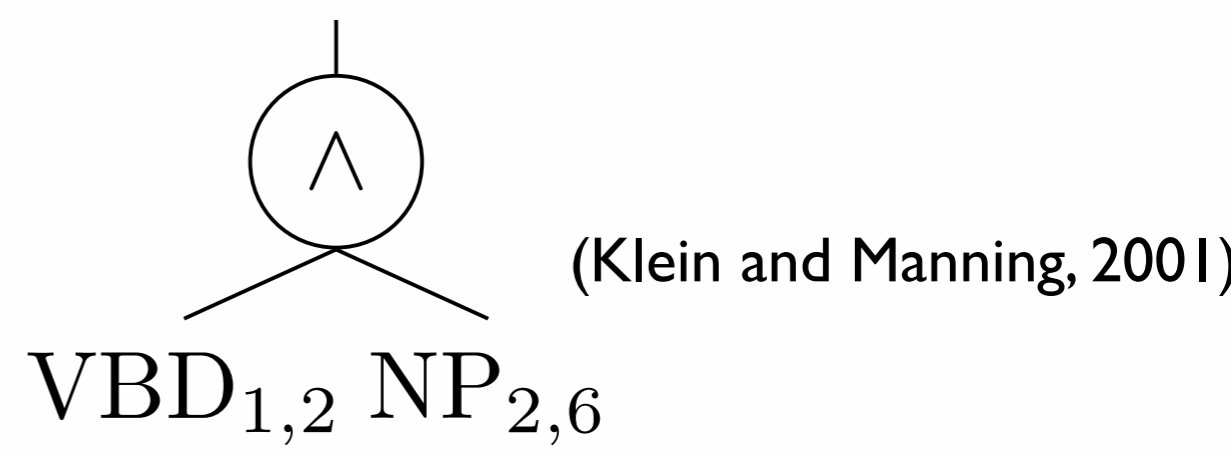
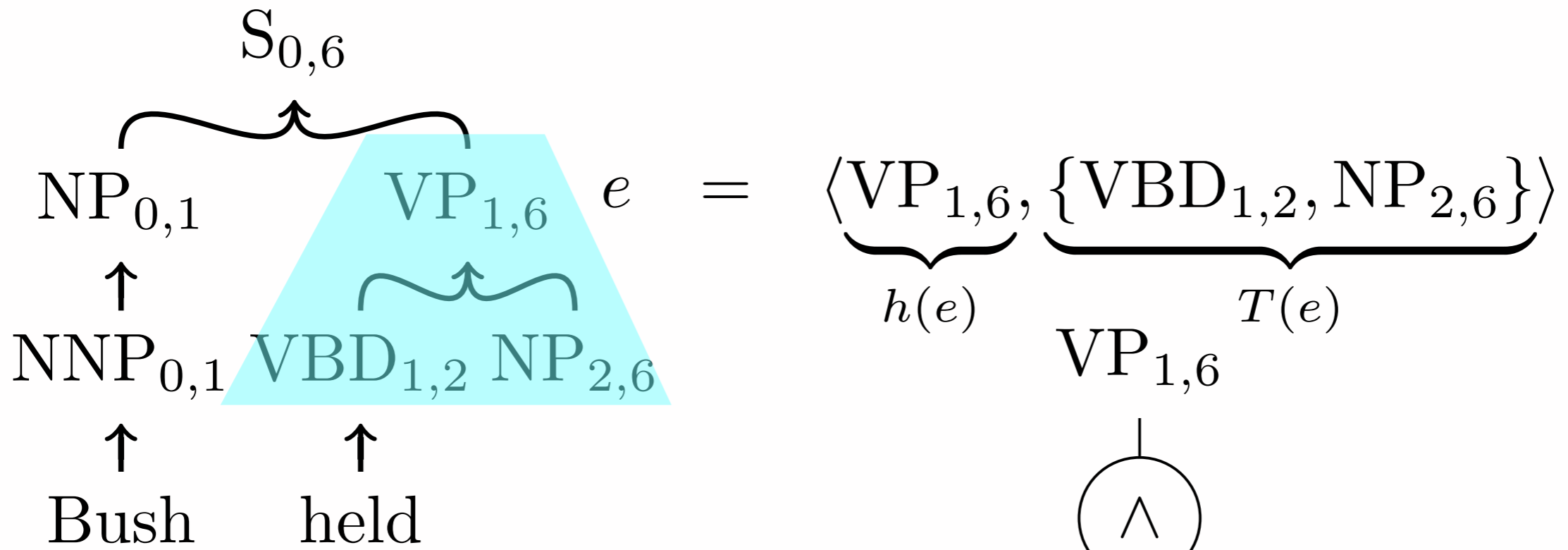


$$\begin{aligned} &Pr(\text{I want to about learn}) \\ &\times Pr(\text{したい, について, 勉強}) \\ &\times Pr(\text{したい} | \text{I want to}) \\ &\times Pr(\text{について} | \text{learn}) \\ &\times Pr(\text{勉強} | \text{learn}) \end{aligned}$$



(Chiang, 2007)

Hypergraph



- グラフの一般化:
- $h(e)$: 超辺 (hyperedge) e の head ノード、 $T(e)$: 超辺 e の tail ノード、 $arity = |T(e)|$
- 超辺 = インスタンス化されたルール
- and-or グラフとしても表記可能

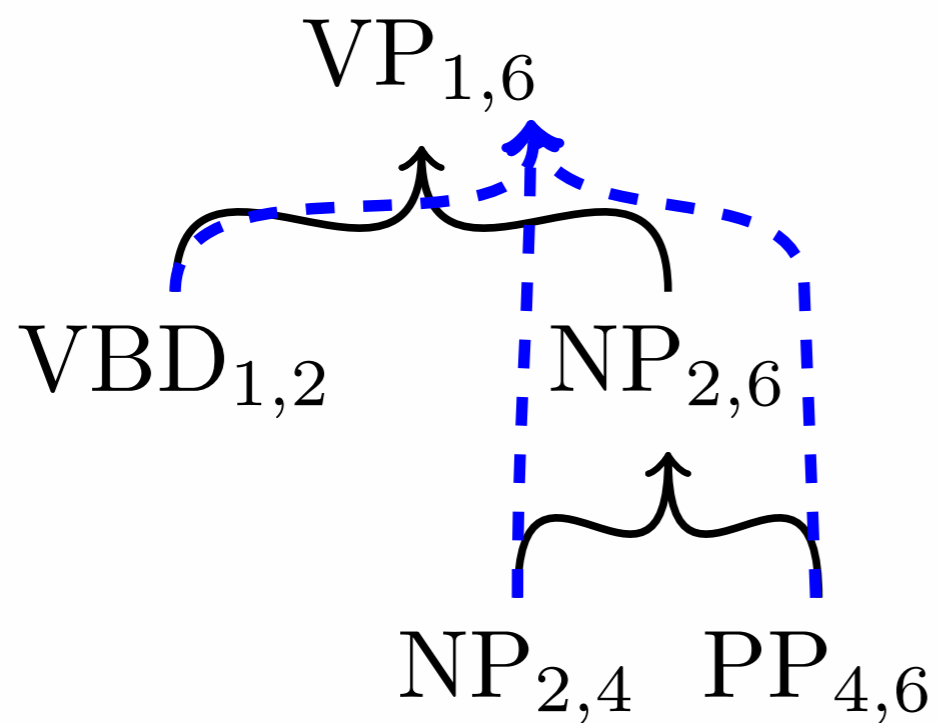
Deductive System

$$\begin{array}{c} \text{VBD}_{1,2} \quad \text{NP}_{2,6} \\ \underbrace{\hspace{10em}} \\ \text{VP}_{1,6} \end{array} \quad \frac{\overbrace{\text{VBD}_{1,2} \quad \text{NP}_{2,6}}^{\text{antecedents}}}{\underbrace{\text{VP}_{1,6}}_{\text{consequent}}} \text{VP}_{[i,j]} \rightarrow \text{VBD}_{[i,k]} \quad \text{NP}_{[k,j]}$$

(Shieber et al., 1995)

- 構文解析アルゴリズムは、演繹法(deduction system)で記述可能
- 公理(axiom)から始め、goalへたどり着くまで推論規則を適用
- 前件(antecedent)が証明されたら、その後件(consequent)が証明される
- 推論規則の導出 = 超辺

Packed Forest



$$\frac{VBD_{1,2} \frac{NP_{2,4} PP_{4,6}}{NP_{2,6}}}{VP_{1,6}}$$

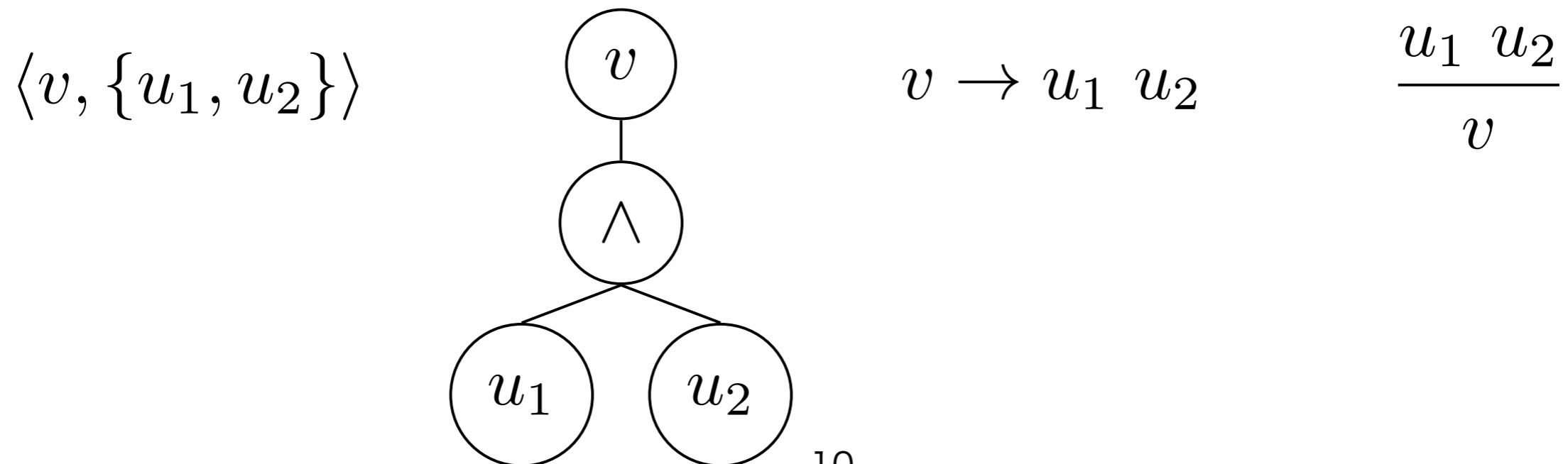
$$\frac{VBD_{1,2} NP_{2,4} PP_{4,6}}{VP_{1,6}}$$

(Klein and Manning, 2001; Huang and Chiang, 2005)

- ノードを共有することにより、複数の導出をコンパクトに表現
- 一つの導出 = 木

Summary of Formalisms

hypergraph	AND/OR graph	CFG	deductive system
vertex	OR-node	symbol	item
source-vertex	leaf OR-node	terminal	axiom
target-vertex	root OR-node	start symbol	goal item
hyperedge	AND-node	production	instantiated deduction



Weight and Semiring

VP $\xrightarrow{w_1}$ VBD NP

NP $\xrightarrow{w_2}$ NP PP

$$VP_{1,6} : w_1 \otimes c \otimes d$$

$$\frac{VBD_{1,2} : c \quad NP_{2,6} : d}{VP_{1,6} : w_1 \otimes c \otimes d} : w_1$$

$$VBD_{1,2} : c \quad NP_{2,6} : d$$

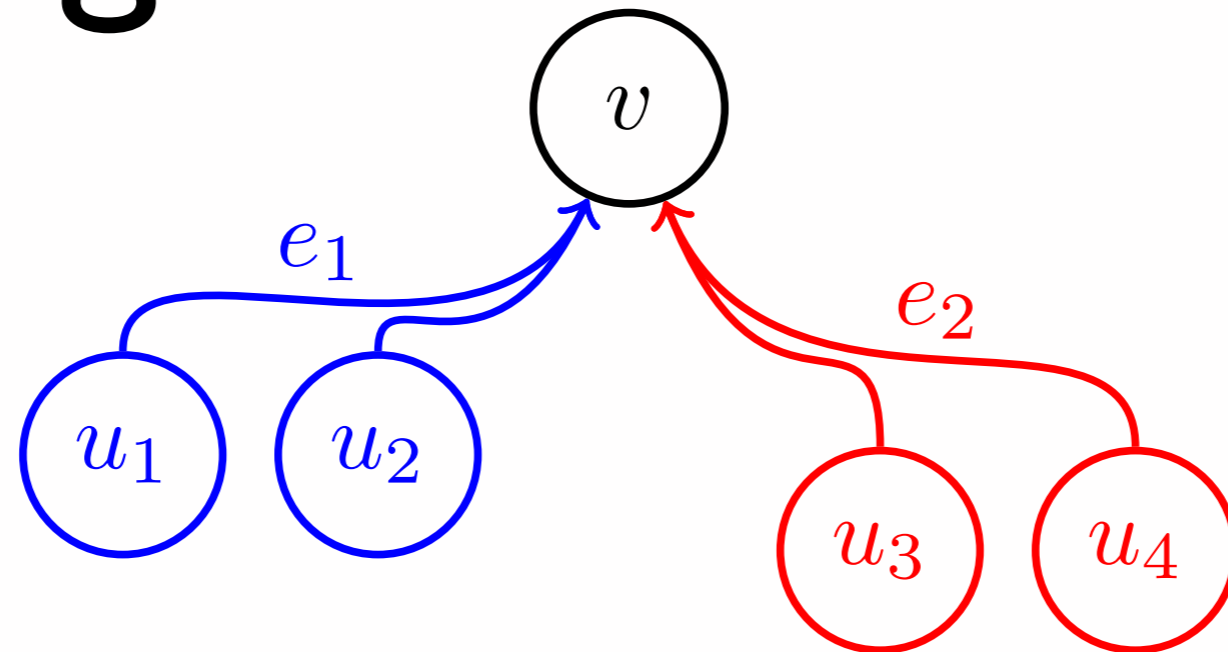
$$NP_{2,6} : w_2 \otimes a \otimes b$$

$$\frac{NP_{2,4} : a \quad PP_{4,6} : b}{NP_{2,6} : w_2 \otimes a \otimes b} : w_2$$

$$NP_{2,4} : a \quad PP_{4,6} : b$$

- WFSTのように、各超辺にweightを関連付ける
- \otimes : extension (multiplicative), \oplus : summary (additive)

Weight and Semiring



$$d(v) = (w(e_1, u_1, u_2) \otimes d(u_1) \otimes d(u_2)) \oplus (w(e_2, u_3, u_4) \otimes d(u_3) \otimes d(u_4))$$

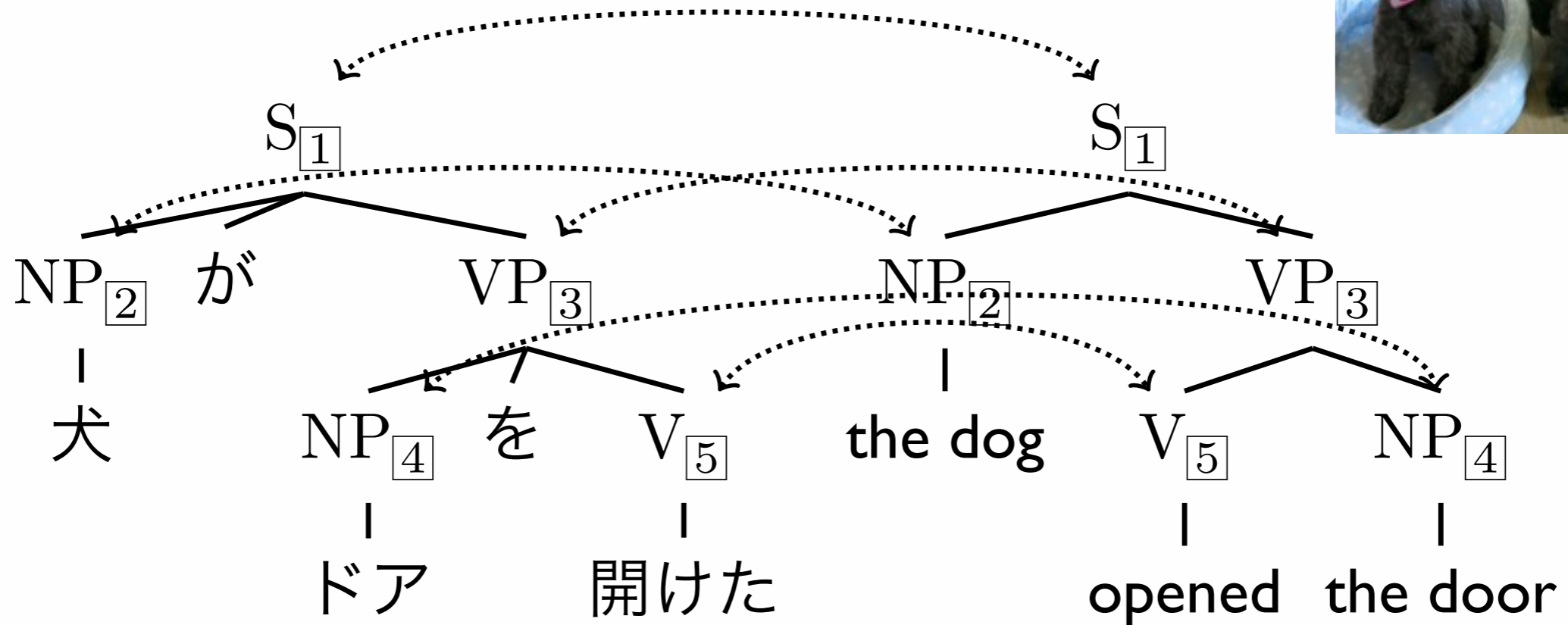
- 超辺の各weightは、その前件のノードに依存(non-monotonic)
- 一つの導出のweight = 超辺の各weightの積
- あるノードのweightは、それを含む導出のweightの和

Semirings

$$\mathbf{K} = \langle K, \oplus, \otimes, \mathbf{0}, \mathbf{1} \rangle$$

semiring	K	\oplus	\otimes	0	1
Viterbi	[0, 1]	max	\times	0	1
Real	R	+	\times	0	1
Log	R	logsumexp	+	$+\infty$	0
Tropical	R	min	+	$+\infty$	0
Expectation	$\langle P, R \rangle$	$\langle p_1 \oplus p_2, r_1 \oplus r_2 \rangle$	$\langle p_1 \otimes p_2, p_1 \otimes r_2 \oplus p_2 \otimes r_1 \rangle$	$\langle 0, 0 \rangle$	$\langle 1, 0 \rangle$

同期文脈自由文法



(Chiang, 2007)

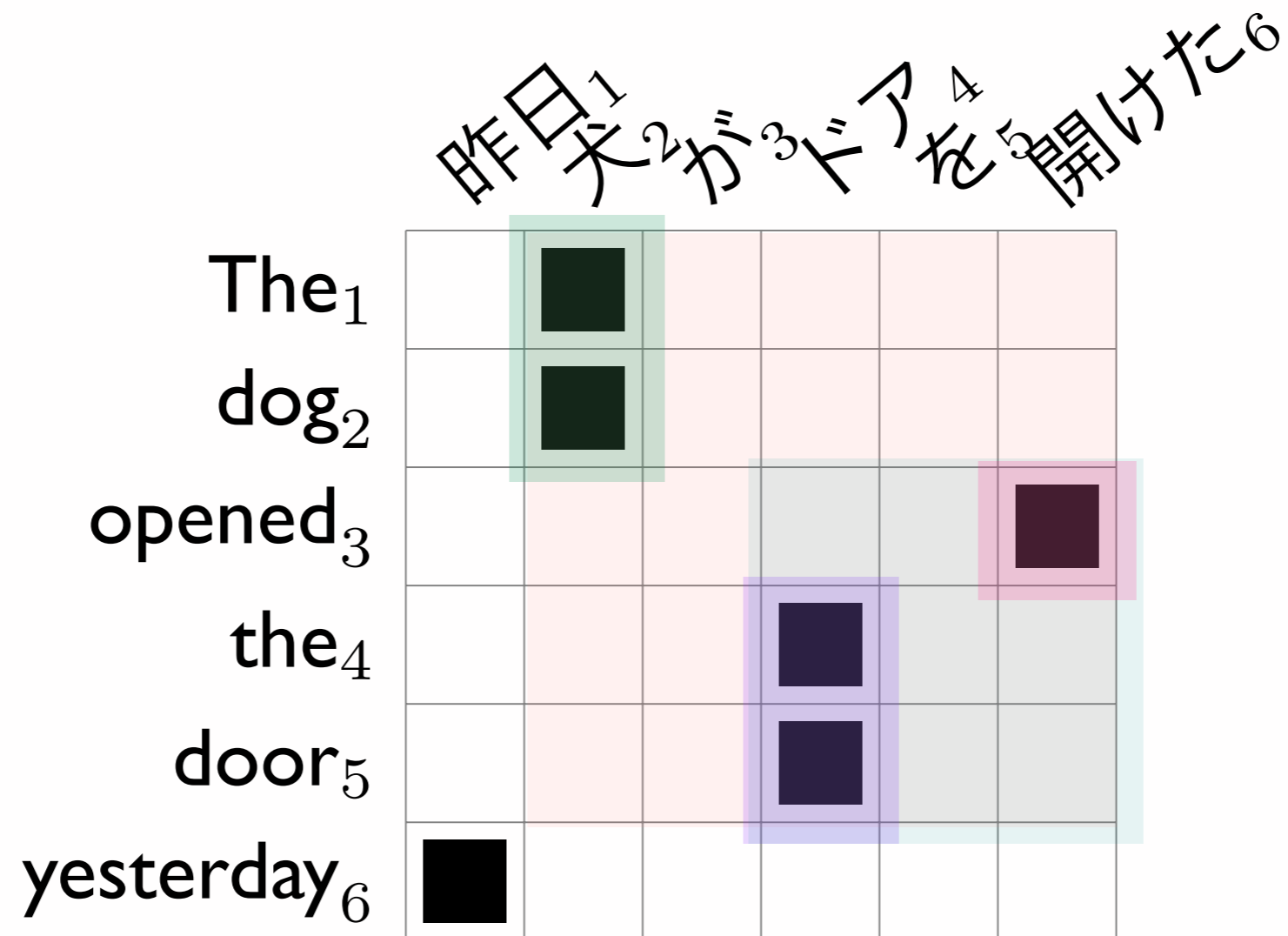
- 文脈自由文法の二言語への拡張

同期文脈自由文法

S → $\langle \text{NP}_{[1]} \text{が} \text{VP}_{[2]}, \text{NP}_{[1]} \text{VP}_{[2]} \rangle$
VP → $\langle \text{NP}_{[1]} \text{を} \text{V}_{[2]}, \text{V}_{[2]} \text{NP}_{[1]} \rangle$ V → $\langle \text{開けた, opened} \rangle | \langle \text{座った, sat} \rangle$
VP → $\langle \text{PP}_{[1]} \text{V}_{[2]}, \text{V}_{[2]} \text{PP}_{[1]} \rangle$ P → $\langle \text{上に, on} \rangle$
VP → $\langle \text{NP}_{[1]} \text{V}_{[2]}, \text{V}_{[2]} \text{NP}_{[1]} \rangle$ NP → $\langle \text{犬, the dog} \rangle | \langle \text{ドア, the door} \rangle$
PP → $\langle \text{NP}_{[1]} \text{の} \text{P}_{[2]}, \text{P}_{[2]} \text{NP}_{[1]} \rangle$ | $\langle \text{本, the book} \rangle | \langle \text{上に, the upper} \rangle$
NP → $\langle \text{NP}_{[1]} \text{の} \text{NP}_{[2]}, \text{NP}_{[2]} \text{of NP}_{[1]} \rangle$

- 原言語、目的言語側で同じ非終端記号を共有
- 非終端記号による並び替えの表現

同期ルールの抽出

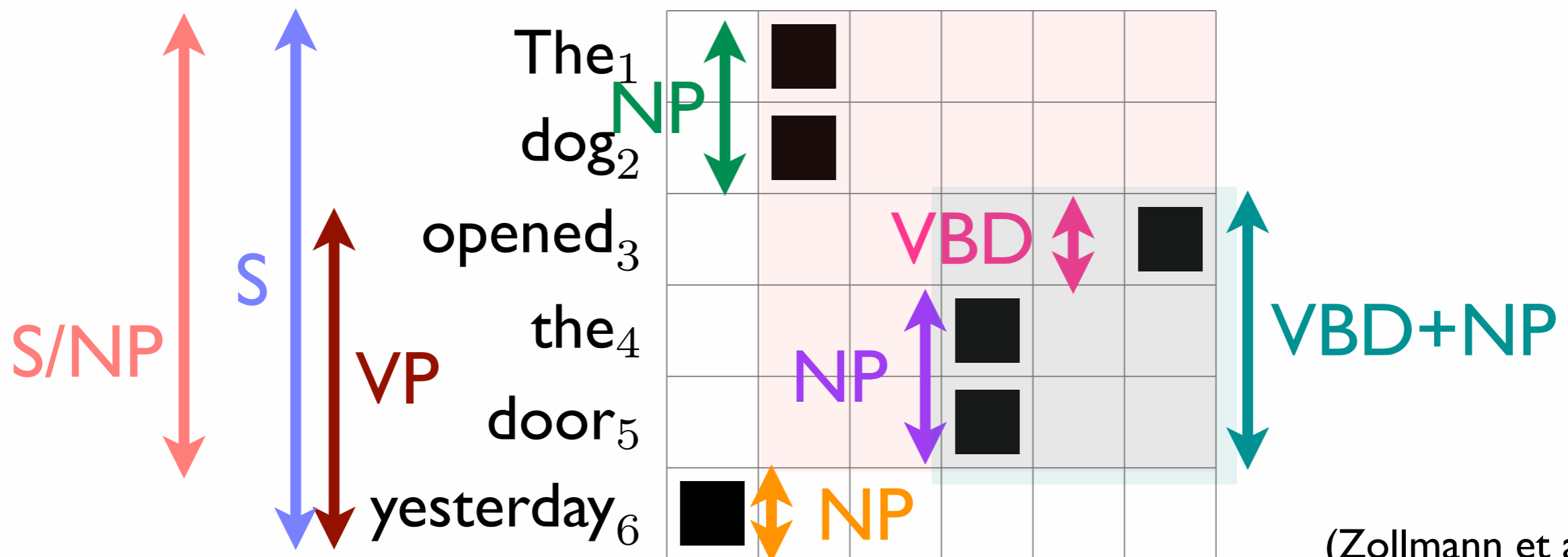


$X \rightarrow \langle X_1 \text{ が } X_2 \text{ を 開けた, } X_1 \text{ opened } X_2 \rangle$

$X \rightarrow \langle X_1 \text{ を } X_2, X_2 X_1 \rangle$

統語的な知識の導入

昨日¹犬²が³ドア⁴を⁵開けた⁶



(Zollmann et al., 2006)

S/NP → ⟨NP₁ が NP₂ を 開けた, NP₁ opened NP₂⟩

VBD+NP → ⟨NP₁ を VBD₂, VBD₂ NP₁⟩

モデル



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

$$h_{\text{gen}}^\triangleright(f, d, e) = \sum_{\langle \alpha, \beta \rangle \in d} \log p_{\text{gen}}(\alpha | \beta) \quad \text{生成モデル}$$

$$h_{\text{gen}}^\triangleleft(f, d, e) = \sum_{\langle \alpha, \beta \rangle \in d} \log p_{\text{gen}}(\beta | \alpha)$$

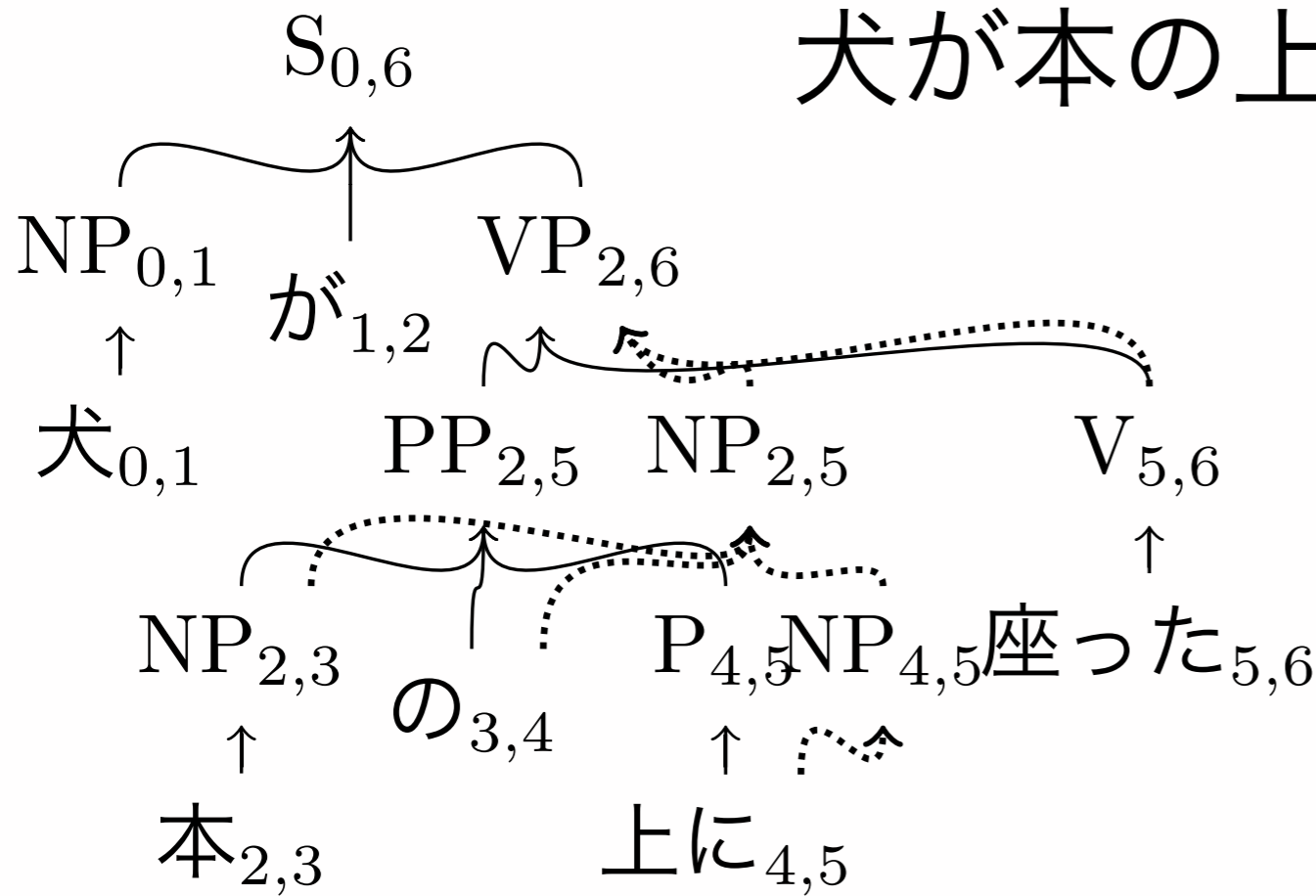
$$h_{\text{lex}}^\triangleright(f, d, e) = \sum_{\langle \alpha, \beta \rangle \in d} \log p_{\text{lex}}(\alpha | \beta) \quad \text{語彙翻訳モデル}$$

$$h_{\text{lex}}^\triangleleft(f, d, e) = \sum_{\langle \alpha, \beta \rangle \in d} \log p_{\text{lex}}(\beta | \alpha)$$

フレーズの翻訳モデル
とほぼ同じ素性

デコード

犬が本の上に座った

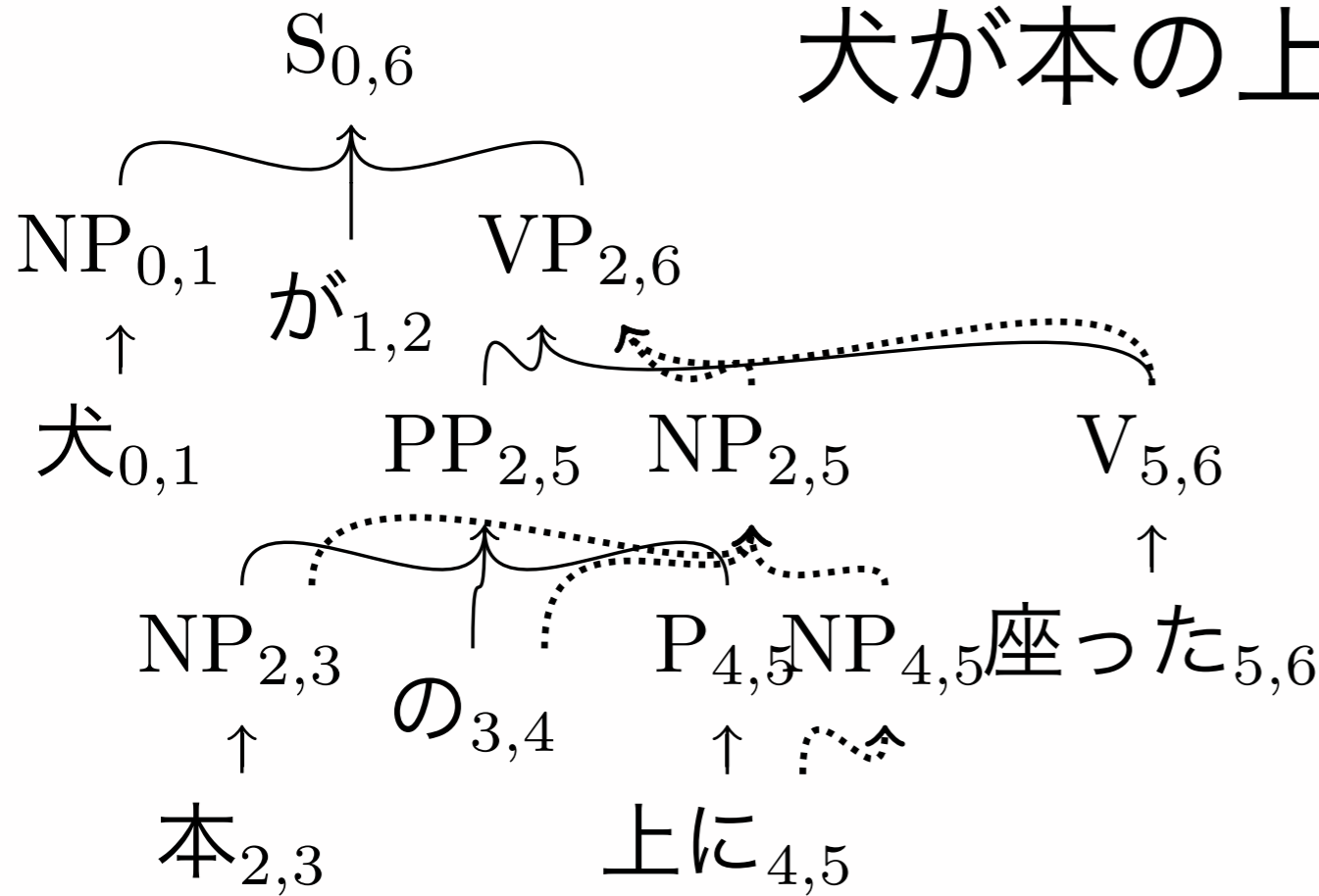


- $S \rightarrow \langle NP_{[1]} \text{が} VP_{[2]}, NP_{[1]} VP_{[2]} \rangle$
- $VP \rightarrow \langle NP_{[1]} \text{を} V_{[2]}, V_{[2]} NP_{[1]} \rangle$
- $VP \rightarrow \langle PP_{[1]} V_{[2]}, V_{[2]} PP_{[1]} \rangle$
- $VP \rightarrow \langle NP_{[1]} V_{[2]}, V_{[2]} NP_{[1]} \rangle$
- $PP \rightarrow \langle NP_{[1]} \text{の} P_{[2]}, P_{[2]} NP_{[1]} \rangle$
- $NP \rightarrow \langle NP_{[1]} \text{の} NP_{[2]}, NP_{[2]} \text{of} NP_{[1]} \rangle$

- 同期ルールの方言語側で構文解析
- 目的言語側で森を生成: 翻訳森

デコード

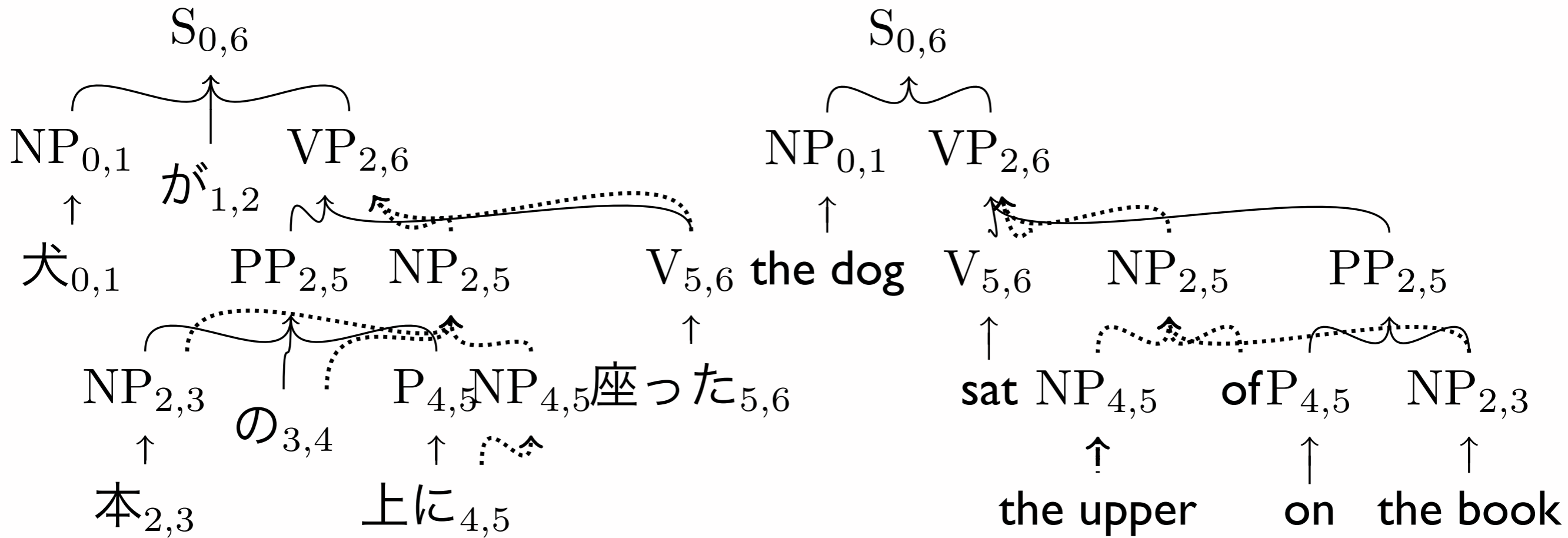
犬が本の上に座った



- S → $\langle \text{NP}_{[1]} \text{が} \text{VP}_{[2]}, \text{NP}_{[1]} \text{VP}_{[2]} \rangle$
- VP → $\langle \text{NP}_{[1]} \text{を} \text{V}_{[2]}, \text{V}_{[2]} \text{NP}_{[1]} \rangle$
- VP → $\langle \text{PP}_{[1]} \text{V}_{[2]}, \text{V}_{[2]} \text{PP}_{[1]} \rangle$
- VP → $\langle \text{NP}_{[1]} \text{V}_{[2]}, \text{V}_{[2]} \text{NP}_{[1]} \rangle$
- PP → $\langle \text{NP}_{[1]} \text{の} \text{P}_{[2]}, \text{P}_{[2]} \text{NP}_{[1]} \rangle$
- NP → $\langle \text{NP}_{[1]} \text{の} \text{NP}_{[2]}, \text{NP}_{[2]} \text{of} \text{NP}_{[1]} \rangle$

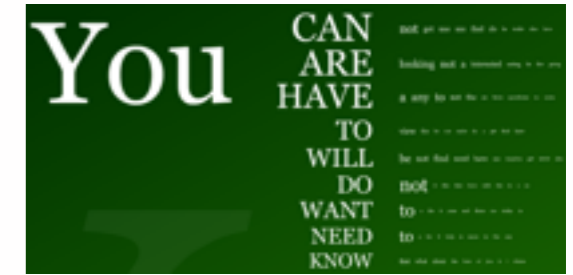
- 同期ルールの方言語側で構文解析
- 目的言語側で森を生成: 翻訳森

デコード



- 同期ルールの方言語側で構文解析
- 目的言語側で森を生成: 翻訳森

言語モデル?



$\log p_{1m}(\text{sat})$ $\log p_{1m}(\text{on})p_{1m}(\text{the}|\text{on})p_{1m}(\text{book}|\text{the})$

$V_{5,6} : u, \text{sat}$ $PP_{2,5} : v, \text{on} \star \text{the book}$

$VP_{2,6} : u \otimes v \otimes w, \text{sat} \star \text{the book}$

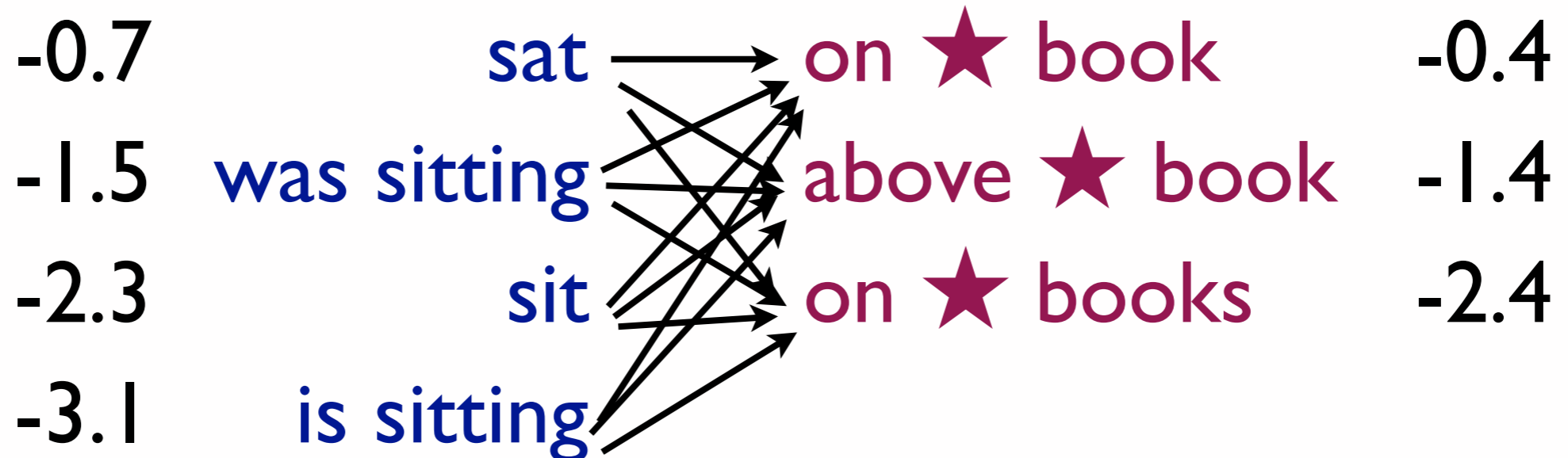
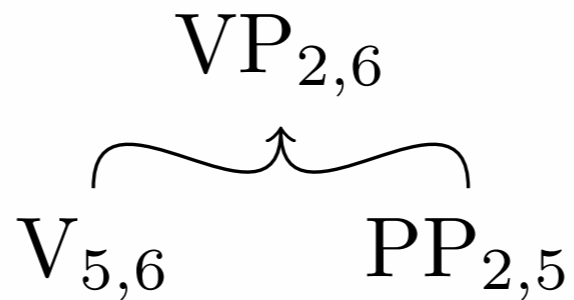
$\log p_{1m}(\text{on}|\text{sat}) - \log p_{1m}(\text{on})$

$w = \omega(VP \rightarrow \langle PP_{[1]} V_{[2]}, V_{[2]}, PP_{[1]} \rangle)$

- 「境界」で再計算: prefix と suffix を保持
- 計算量: $O(n^3 V^{2(m-1)})$: ★は?

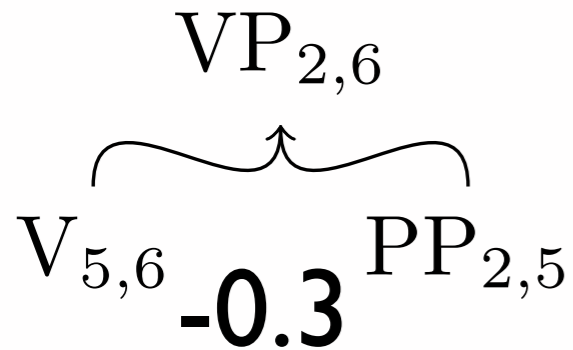
再計算

- 2.1 sat on ★ book
- 2.5 sat above ★ book
- 3.8 was sitting above ★ book



各ノードで網羅的な計算: $O(b^2)$

キューブ枝刈り



on ★ book above ★ book on ★ books
 -0.4 -1.4 -2.4

sat -0.7	-1.4 -0.7	-2.4 -0.1	-3.4 -0.7
was sitting -1.5	-2.2 -1.8	-3.2 -0.6	-4.2 -1.8
sit -2.3	-3.0 -0.9	-4.0 -0.3	-5.0 -0.9
is sitting -3.1	-3.8 -1.8	-4.8 -0.6	-5.8 -1.8

キューブ枝刈り

queue: (0,0)

k-best:

on ★ book above ★ book on ★ books

-0.4

-1.4

-2.4

sat -0.7

-2.1

was sitting -1.5

sit -2.3

is sitting -3.1

-2.1		

キューブ枝刈り

queue:

k-best: (0,0)

on ★ book above ★ book on ★ books

-0.4

-1.4

-2.4

sat -0.7

-2.1

was sitting -1.5

sit -2.3

is sitting -3.1

-2.1		

キューブ枝刈り

queue: (0,1)(1,0)

k-best: (0,0)

on ★ book above ★ book on ★ books

-0.4

-1.4

-2.4

sat -0.7

-2.1

-2.5

was sitting -1.5

-4.0

sit -2.3

is sitting -3.1

キューブ枝刈り

queue: (1,0)

k-best: (0,0)(0,1)

on ★ book above ★ book on ★ books

-0.4

-1.4

-2.4

sat -0.7

-2.1

-2.5

was sitting -1.5

-4.0

sit -2.3

is sitting -3.1

キューブ枝刈り

queue: (1,1)(1,0)(0,2)

k-best: (0,0)(0,1)

on ★ book above ★ book on ★ books

-0.4

-1.4

-2.4

sat -0.7

-2.1

-2.5

-4.1

was sitting -1.5

-4.0

-3.8

sit -2.3

is sitting -3.1

キューブ枝刈り

queue: (1,0)(0,2)

k-best: (0,0)(0,1)(1,1)

on ★ book above ★ book on ★ books

-0.4

-1.4

-2.4

sat -0.7

-2.1

-2.5

-4.1

was sitting -1.5

-4.0

-3.8

sit -2.3

is sitting -3.1

	-2.1	-2.5	-4.1
	-4.0	-3.8	

キューブ枝刈り

queue: (1,0)(0,2)(2,1)(1,2)

on ★ book above ★ book on ★ books

k-best: (0,0)(0,1)(1,1)

-0.4

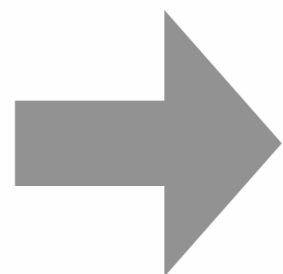
-1.4

-2.4

sat -0.7	-2.1	-2.5	-4.1
was sitting -1.5	-4.0	-3.8	-6.0
sit -2.3	-3.9	-4.3	
is sitting -3.1	探索エラー		

デコード

機械翻訳について勉強したい。



構文解析



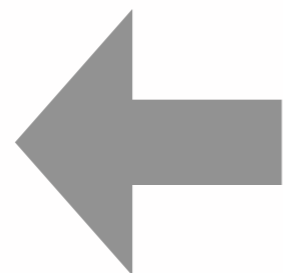
変換



再計算



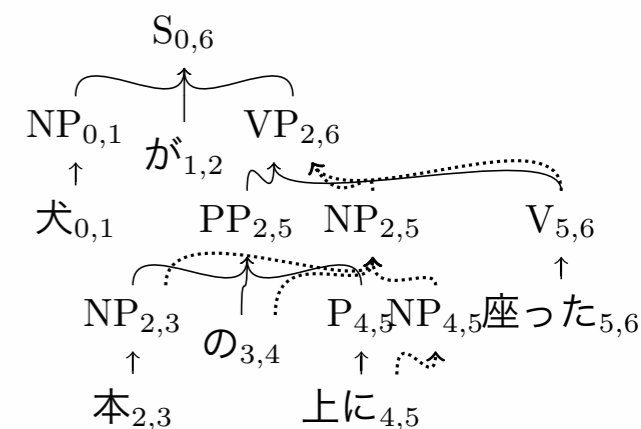
k-best



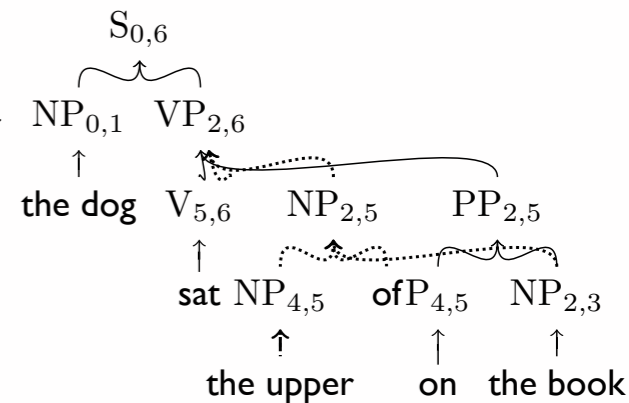
I want to study about machine translation.



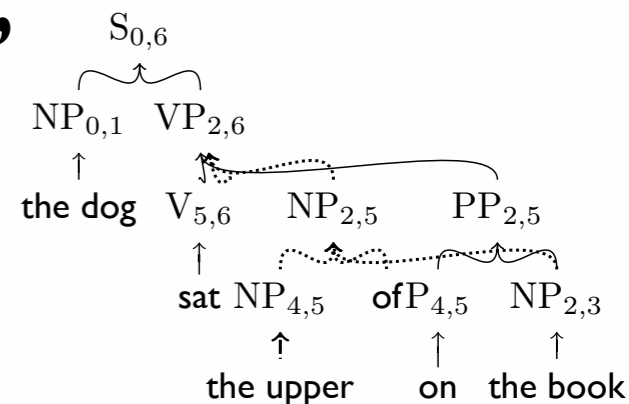
構文森



翻訳森



翻訳森'



同期文法の特徴



- 同期文脈自由文法には、標準形が存在しない
- 文脈自由文法: Chomsky、Greibach

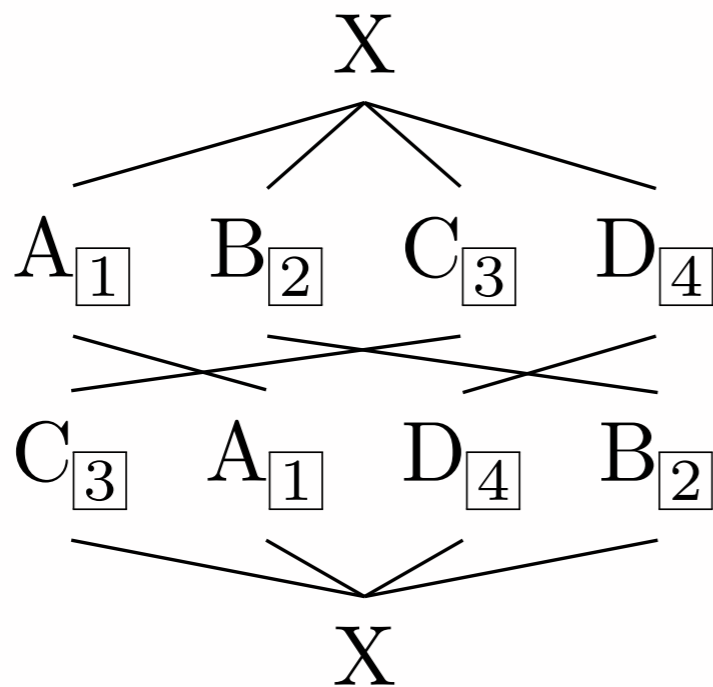
二分化可能 $X \rightarrow \langle A_{[1]} B_{[2]} C_{[3]}, C_{[3]} B_{[2]} A_{[1]} \rangle$

$$X \rightarrow \langle X'_{[1]} C_{[2]}, C_{[2]} X'_{[1]} \rangle$$

$$X' \rightarrow \langle A_{[1]} B_{[2]}, B_{[2]} A_{[1]} \rangle$$

これは? $X \rightarrow \langle A_{[1]} B_{[2]} C_{[3]} D_{[4]}, C_{[3]} A_{[1]} D_{[4]} B_{[2]} \rangle$

Inside-Out



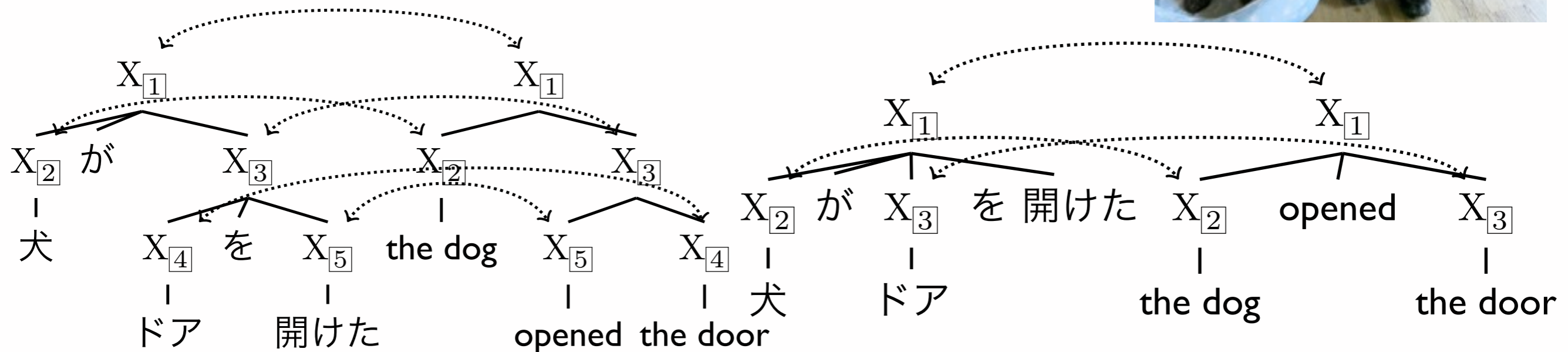
	A ₁	B ₂	C ₃	D ₄
C ₃			■	
A ₁	■			■
D ₄				
B ₂		■		

(Wu, 1997)

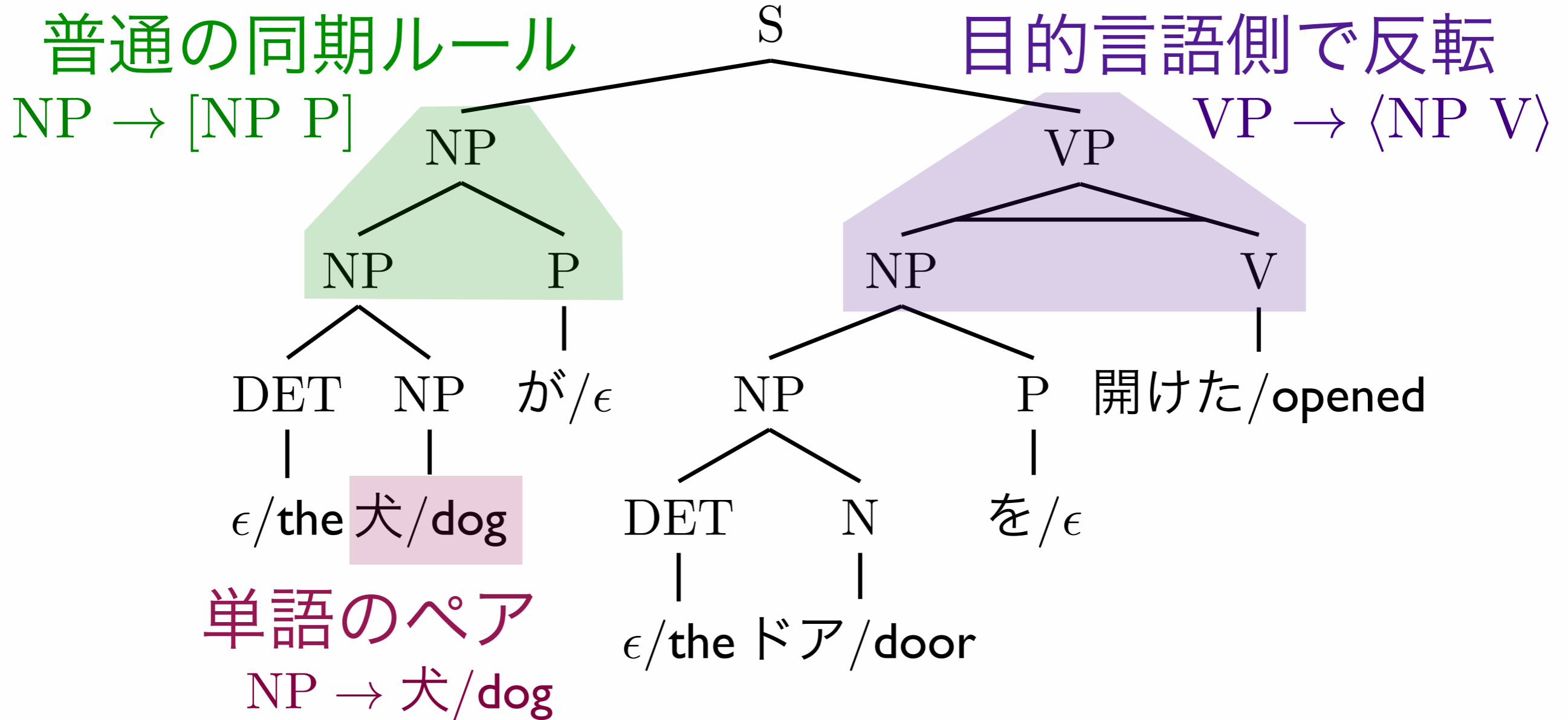
- 他にも: Greibach標準形も不可能、合成できない

二言語同期解析

- 単言語の構文解析 → 二言語の構文解析
- 例: 最もらしい導出を決定

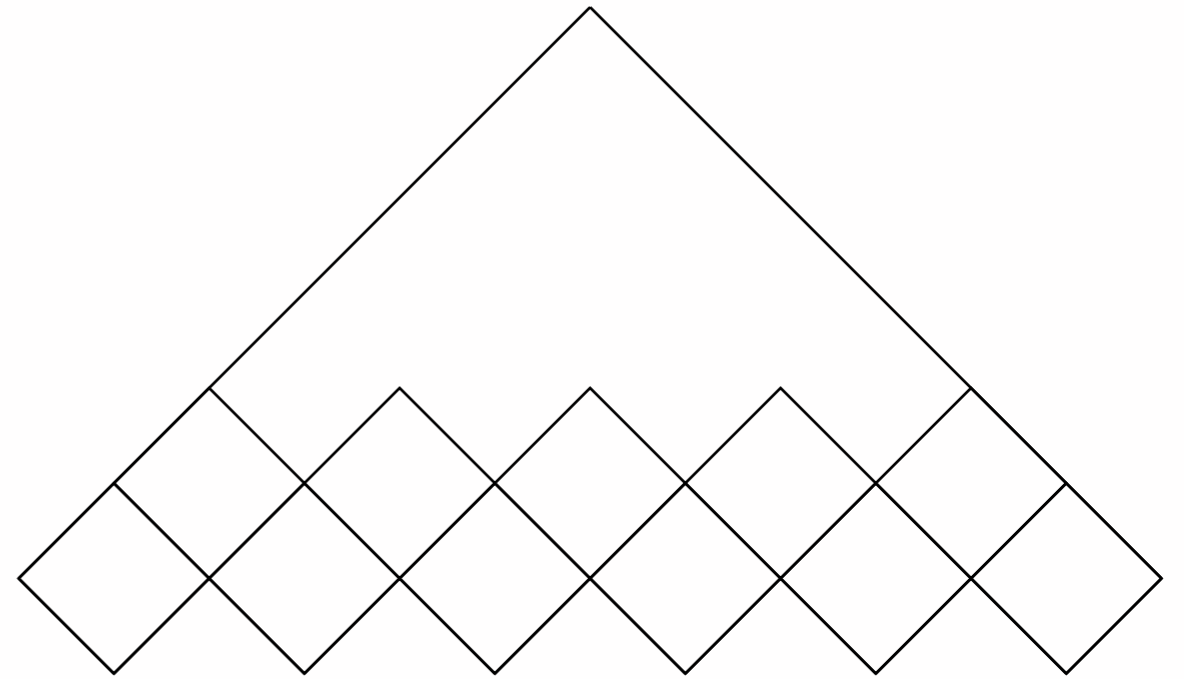
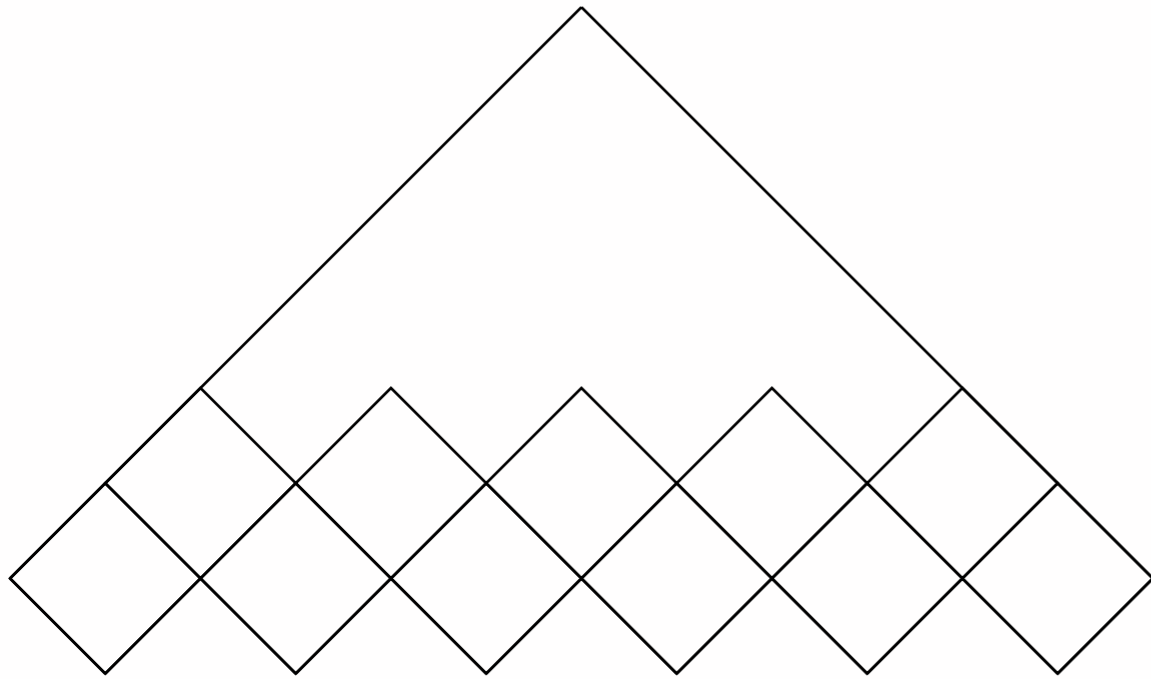


簡単な同期文法



- 反転トランスダクション文法(ITG) (Wu, 1997)

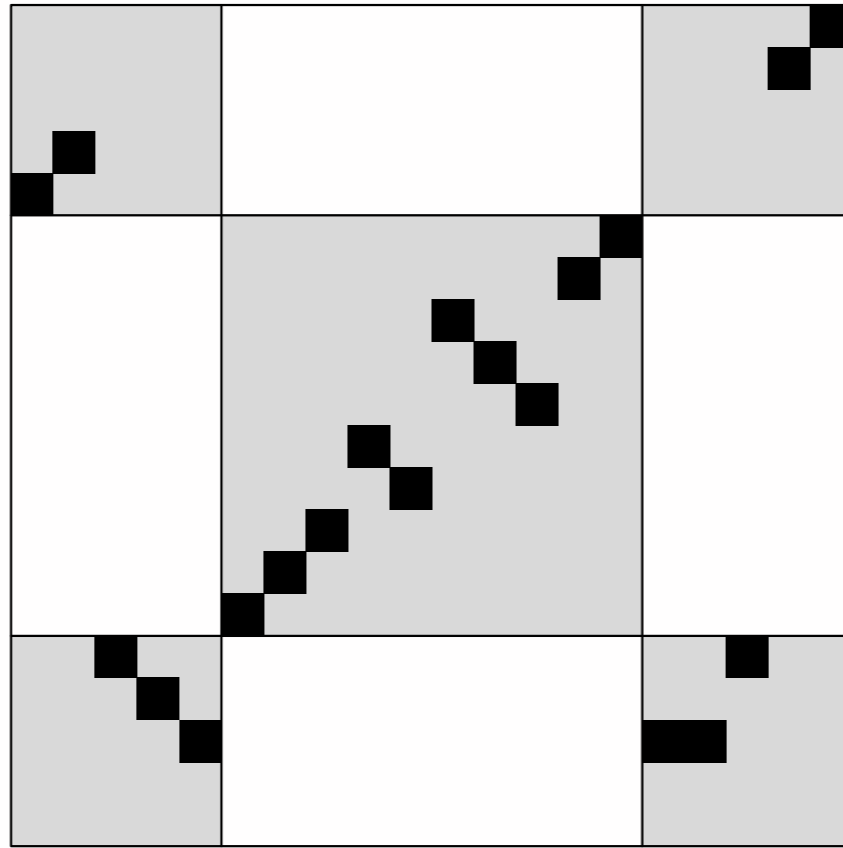
二言語解析



(Wu, 1997)

- CYKの二言語への拡張: スパンの対応を計算
- ITGでは、 $O(N^3 M^3)$: 各長さ n と m 、各位置 i と j 、各ルール $X \rightarrow YZ$ 、各分岐点 k と l

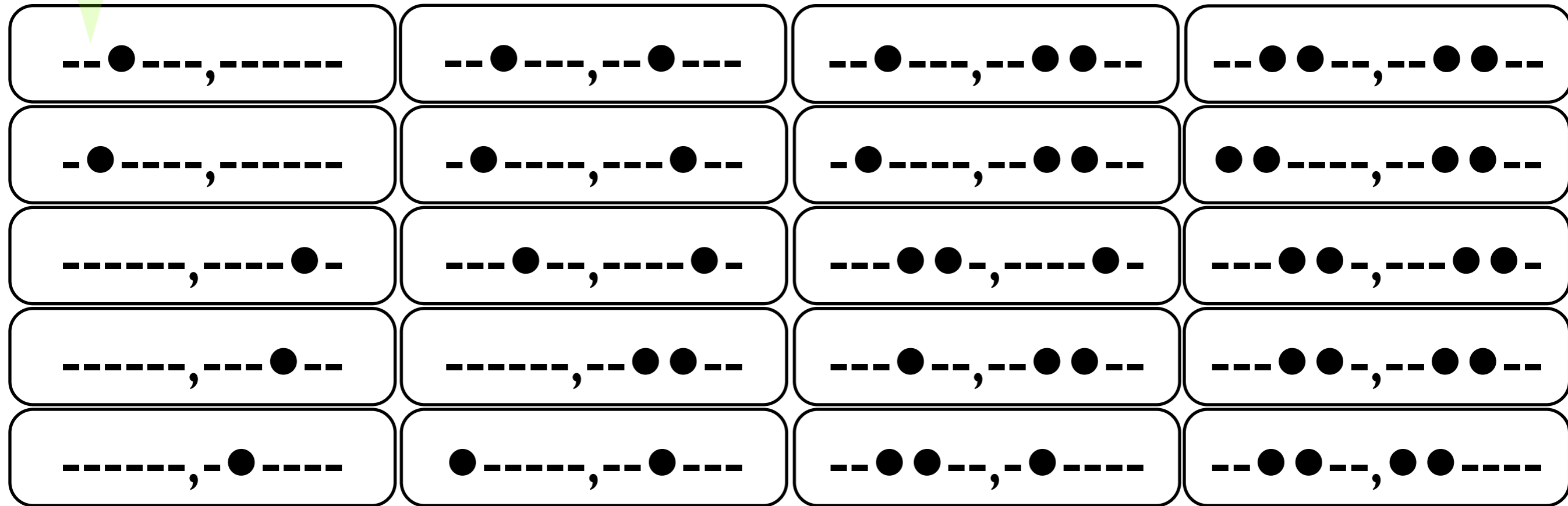
span枝刈り



- 予めスパンのペアを枝刈りすることで高速化
- $O(N^4)$ 簡単なアルゴリズム (Zhang and Gildea, 2005)
- $O(N^3)$ DPアルゴリズム (Zhang et al., 2008)

ビーム探索

● = 解析されたspan



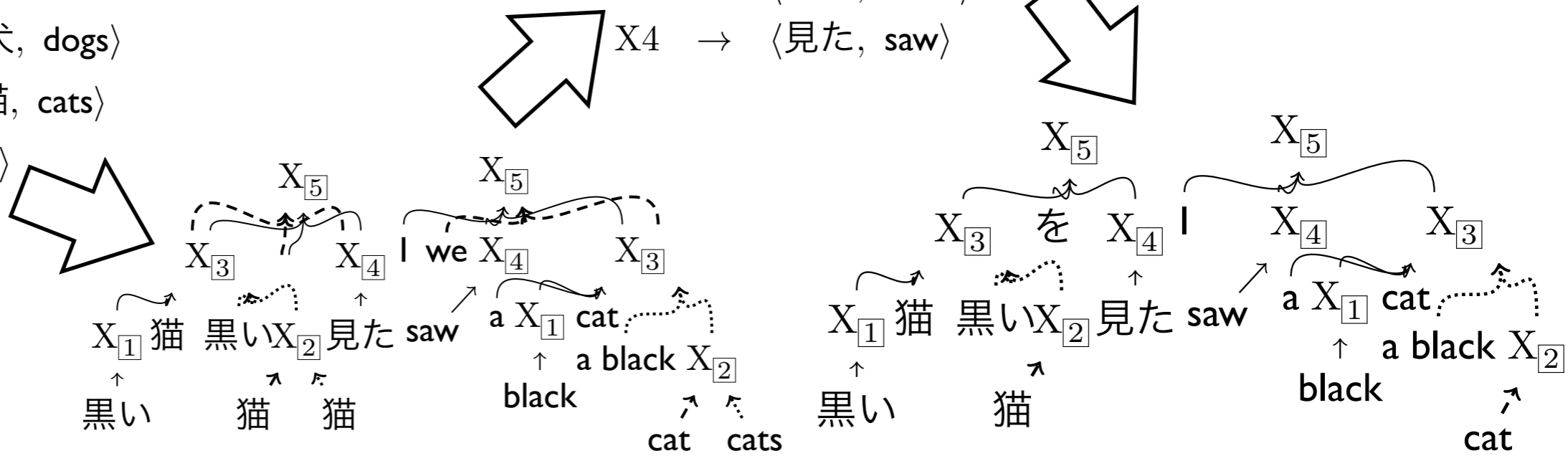
解析された単語数で同期

二段解析

黒い猫を見た \leftrightarrow I saw a black cat

- X \rightarrow $\langle X_1$ を X_2 , I X_2 X_1 \rangle
- X \rightarrow $\langle X_1$ を X_2 , we X_2 X_1 \rangle
- X \rightarrow $\langle X_1$ は X_2 , X_1 is X_2 \rangle
- X \rightarrow \langle 黒い X_1 , a black X_1 \rangle
- X \rightarrow $\langle X_1$ 猫, a X_1 cat \rangle
- X \rightarrow \langle 犬, dog \rangle | \langle 犬, dogs \rangle
- X \rightarrow \langle 猫, cat \rangle | \langle 猫, cats \rangle
- X \rightarrow \langle 黒い, black \rangle
- X \rightarrow \langle 見た, saw \rangle

- X5 \rightarrow $\langle X3_1$ を $X4_2$, I $X4_2$ $X3_1$ \rangle
- X5 \rightarrow $\langle X3_1$ を $X4_2$, we $X4_2$ $X3_1$ \rangle
- X3 \rightarrow \langle 黒い $X2_1$, a black X_1 \rangle
- X3 \rightarrow $\langle X1_1$ 猫, a $X1_1$ cat \rangle
- X2 \rightarrow \langle 猫, cat \rangle | \langle 猫, cats \rangle
- X1 \rightarrow \langle 黒い, black \rangle
- X4 \rightarrow \langle 見た, saw \rangle



原言語で構文解析、同期ルールを抽出、目的言語で構文解析
(Dyer, 2010)

まとめ

- 同期ルールによる明示的な並び替え
- 構文解析による翻訳
- 二言語構文解析
- 注意: Mosesで実装されている木構造のモデルは、同期文脈自由文法

機械翻訳システム

- Moses: <http://www.statmt.org/moses/>
- Joshua: <http://joshua-decoder.org>
- Cdec: <http://www.cdec-decoder.org>
- Cicada: http://www2.nict.go.jp/univ-com/multi_trans/cicada/

参考文献

- David Chiang. 2007. Hierarchical phrase-based translation *Computational Linguistics*, 33(2): 201-228.
- Chris Dyer. 2010. Two monolingual parses are better than one (synchronous parse). In *Human Language Technologies The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 263-266 Los Angeles, California, June.
- Liang Huang and David Chiang. 2005. Better k-best parsing. In *Proc. of IWPT'05*, pages 53-64, Vancouver, British Columbia, October.
- Liang Huang and David Chiang. 2007. Forest rescoring Faster decoding with integrated language models. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 144-151, Prague, Czech Republic June. Association for Computational Linguistics.
- Dan Klein and Christopher D. Manning. 2001. Parsing and hypergraphs. In *Proc. of IWPT-2001*, pages 123-134.
- Markus Saers, Joakim Nivre, and Dekai Wu. 2009. Learning stochastic bracketing inversion transduction grammars with a cubic time biparsing algorithm. In *Proc. of IWPT'09*, pages 29-32, Paris, France, October.

参考文献

- Stuart M. Shieber, Yves Schabes, and Fernando C. N Pereira. 1995. Principles and implementation of deductive parsing. *Journal of Logic Programming*, 24(1-2):3-36, July-August.
- Hao Zhang and Daniel Gildea. 2005. Stochastic lexicalized inversion transduction grammar for alignment. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 475-482, Ann Arbor Michigan, June.
- Hao Zhang, Chris Quirk, Robert C. Moore, and Danie Gildea. 2008. Bayesian learning of non-compositiona phrases with synchronous parsing. In *Proceedings of ACL-08 HLT*, pages 97-105, Columbus, Ohio, June.
- Andreas Zollmann and Ashish Venugopal. 2006. Syntax augmented machine translation via chart parsing. In *Proc. of StatMT '06*, pages 138-141, Morristown, NJ, USA.