

文の差分に基づく言語獲得

神奈川大学理学部情報科学科

田中研究室

研究の背景

【言語獲得の特徴】

- 単純な1語文・2語文→複雑な文
- 僅かな差分から言語知識の獲得

例： 車(1語文) → 車 乗る(2語文)
↓
一度帰って、先月買った車に乗る(複雑な文)

差分に基づく言語獲得法

- 既知の統語規則と入力との**差分**から新たな統語規則を獲得
- カテゴリの**統合**、規則の**汎化**をくりかえして、統語規則を獲得

言語獲得系への入力

- 時刻 t における入力を $s(t)$ とする

$$s(t) = w_1 w_2 \dots w_n$$

$s(t)$: 入力

t : 時刻

w_n : 語

例: $s(0) = [\text{ミルク}]$

$s(0) = [\text{dr ink}]$

$s(1) = [\text{ミルク}][\text{飲む}]$

$s(1) = [\text{I}][\text{dr ink}]$

$s(2) = [\text{I}][\text{dr ink}][\text{mi lk}]$

言語獲得系からの出力

- 時刻 t における出力を、規則の集合 $R(t)$ 、カテゴリの集合 $C(t)$ 、変数の集合 $V(t)$ とする

$$R(t) = \{ \alpha \rightarrow \beta \mid \alpha \in V(t), \beta \in (V(t) \cup C(t))^* \}$$

$$C(t) = \{ c \mid c \in \{ w_1, w_2, \dots \} \}$$

$$V(t) = \{ S_1, S_2, \dots \}$$

c_i : カテゴリ

w_j : 語

S_k : 変数

例: $R(1) = \{ S_1 \rightarrow c_1, S_2 \rightarrow c_1 c_2 \}$

$$C(1) = \{ c_1, c_2 \}$$

$$V(1) = \{ S_1, S_2 \}$$

c_1	c_2
ミルク	飲む

$$S_1 \rightarrow c_1$$

$$S_2 \rightarrow c_1 c_2$$

$$R(1) = \{ S_1 \rightarrow c_1, S_2 \rightarrow c_2 c_1 \}$$

$$C(1) = \{ c_1, c_2 \}$$

$$V(1) = \{ S_1, S_2 \}$$

c_1	c_2
drink	I

$$S_1 \rightarrow c_1$$

$$S_2 \rightarrow c_2 c_1$$

獲得の3つの原理

- カテゴリ化
 - レベル1: 語
 - レベル2: 格
 - レベル3: 文
 - カテゴリへの差分の付加
 - カテゴリの統合
-

カテゴリのレベル

 c_j^i i : レベル j : カテゴリ番号

【レベル1 : c^1 】 統語規則を構成する最小単位となるカテゴリ

c_1^1

c_2^1

 ...

【レベル2 : c^2 】 レベル1カテゴリに w を付加したものを要素とするカテゴリ

c_1^2				
<table border="1"><tr><td>c_a^1</td></tr></table>	c_a^1	w	<table border="1"><tr><td>c_b^1</td></tr></table>	c_b^1
c_a^1				
c_b^1				
		$w \dots$		

c_2^2				
<table border="1"><tr><td>c_c^1</td></tr></table>	c_c^1	w	<table border="1"><tr><td>c_d^1</td></tr></table>	c_d^1
c_c^1				
c_d^1				
		$w \dots$		

 ...

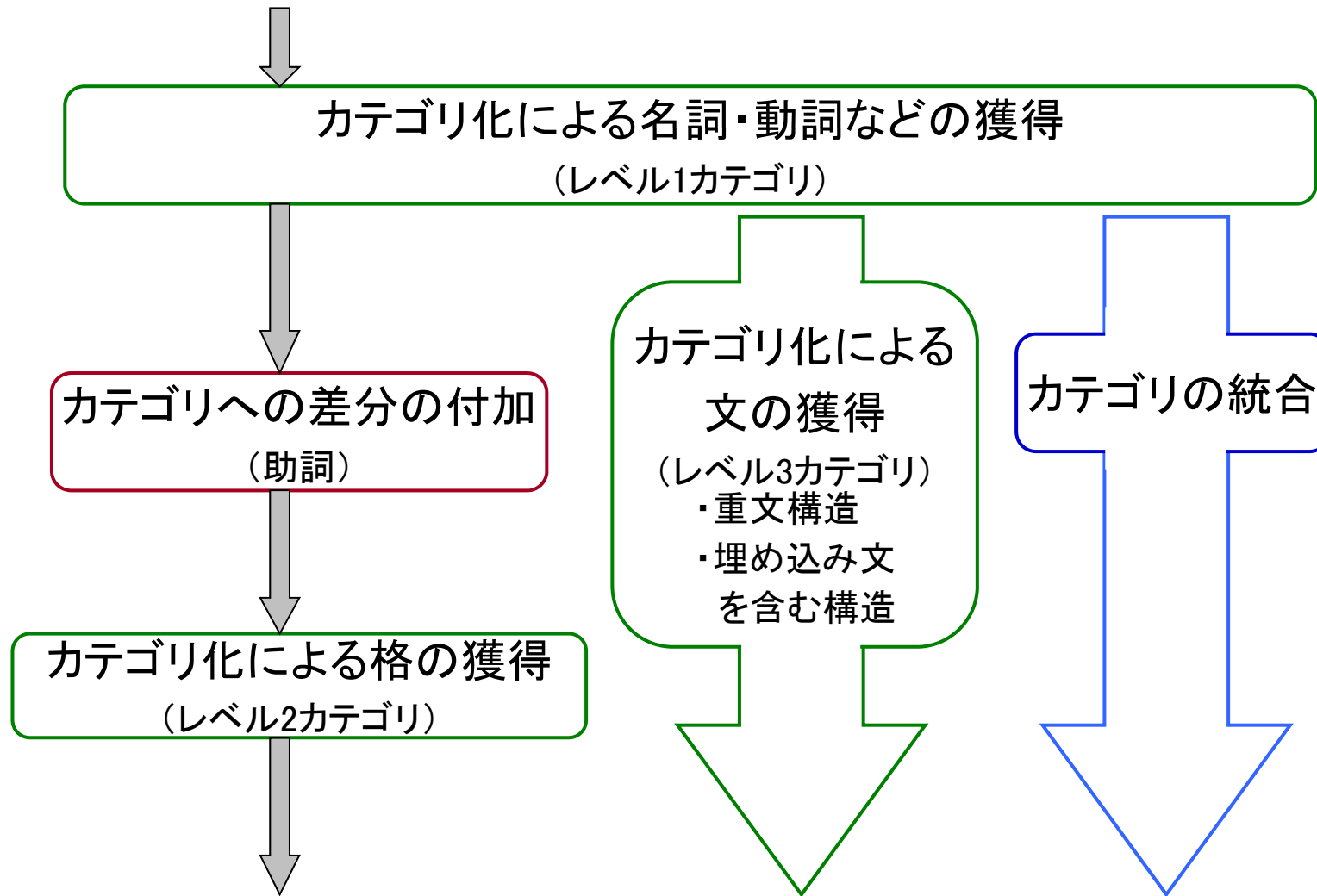
【レベル3 : c^3 】 文を要素とするカテゴリ

c_1^3		
s_a	s_b	$\dots$

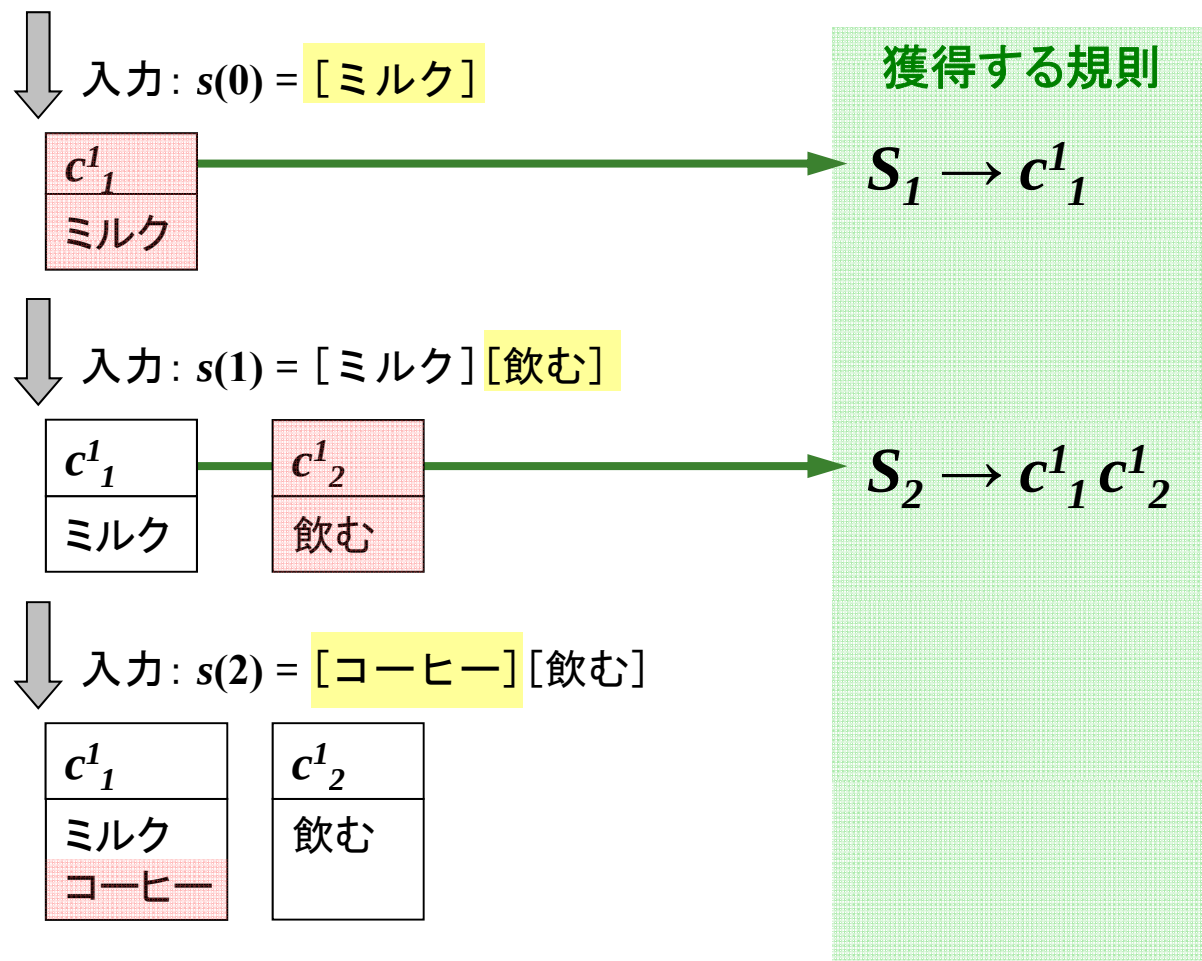
c_2^3		
s_c	s_d	$\dots$

 ...

獲得の流れ



レベル1カテゴリ c^1 (差分のカテゴリ化)

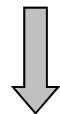


カテゴリへの差分の付加 (助詞)

- 個々のカテゴリに付随して助詞を獲得

c_1^1	c_2^1
ミルク	飲む
コーヒー	

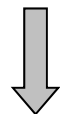
$$\begin{array}{l} \overline{S \rightarrow S_1 \mid S_2} \\ S_1 \rightarrow c_1^1 \\ S_2 \rightarrow c_1^1 c_2^1 \end{array}$$



入力 : $s(3) = [\text{コーヒー}][\text{を}]$

c_1^1	c_2^1
ミルク	飲む
コーヒー	

$$\begin{array}{l} \overline{S \rightarrow S_1 \mid S_2} \quad V_{\text{を}} \rightarrow c_1^1 \text{を} \\ S_1 \rightarrow c_1^1 \\ S_2 \rightarrow c_1^1 c_2^1 \end{array}$$



c_1^1	c_2^1
ミルク	飲む
コーヒー	

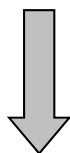
$$\begin{array}{l} \overline{S \rightarrow S_1 \mid S_2 \mid S_3} \quad V_{\text{を}} \rightarrow c_1^1 \text{を} \\ S_1 \rightarrow c_1^1 \\ S_2 \rightarrow c_1^1 c_2^1 \quad S_3 \rightarrow V_{\text{を}} c_2^1 \end{array}$$

レベル2カテゴリ c^2 (カテゴリ化による格の獲得)

- 格の可換性からレベル2カテゴリ c^2 の獲得

c^1_1	c^1_2	c^1_3	c^1_4
ミルク	飲む	僕	公園
コーヒー	遊ぶ	彼	遊園地

$$\begin{array}{lcl}
 S \rightarrow S_1 | S_2 & S_1 \rightarrow V_{\text{は}} V_{\text{を}} c^1_2 & \\
 V_{\text{を}} \rightarrow c^1_1 \text{を} & S_2 \rightarrow V_{\text{を}} V_{\text{は}} c^1_2 & \\
 V_{\text{は}} \rightarrow c^1_3 \text{は} & &
 \end{array}$$



レベル2カテゴリ c^2 に整理

c^1_1	c^1_2	c^1_3	c^1_4	c^2_1
ミルク	飲む	僕	公園	$V_{\text{を}}$
コーヒー	遊ぶ	彼	遊園地	$V_{\text{は}}$

$$\begin{array}{lcl}
 S \rightarrow S_1 & S_1 \rightarrow c^2_1 c^1_2 & \\
 V_{\text{を}} \rightarrow c^1_1 \text{を} & & \\
 V_{\text{は}} \rightarrow c^1_3 \text{は} & &
 \end{array}$$

- レベル2カテゴリ c^2 の獲得により規則を汎化

c^1_1	c^1_2	c^1_3	c^1_4	c^2_1	$S \rightarrow S_1 S_2 S_3$	$S_1 \rightarrow c^2_1 c^1_2$
ミルク	飲む	僕	公園	$V_{を}$	$V_{を} \rightarrow c^1_1を$	$S_2 \rightarrow V_{は} V_{で} c^1_2$
コーヒー	遊ぶ	彼	遊園地	$V_{は}$	$V_{は} \rightarrow c^1_3は$	$S_3 \rightarrow V_{で} V_{は} c^1_2$
					$V_{で} \rightarrow c^1_4で$	

c^1_1	c^1_2	c^1_3	c^1_4	c^2_1	$S \rightarrow S_1$	$S_1 \rightarrow c^2_1 c^1_2$
ミルク	飲む	僕	公園	$V_{を}$	$V_{を} \rightarrow c^1_1を$	
コーヒー	遊ぶ	彼	遊園地	$V_{は}$	$V_{は} \rightarrow c^1_3は$	
				$V_{で}$	$V_{で} \rightarrow c^1_4で$	

c^2_1 で3つの格を用いた表現が可能に

例えば

「彼は 公園で コーヒーを 飲む」

という文が扱えるようになる

同じ助詞に連なるカテゴリの統合

- レベル2カテゴリ内で同じ助詞に連なることを条件に統合

c^1_1	c^1_2	c^1_3	c^1_4	c^2_1
ミルク コーヒー	飲む 遊ぶ 見る	僕 彼	公園 遊園地	c^1_1 を c^1_4 に c^1_3 は c^1_3 を c^1_4 で c^1_3 と

(「 c^1_1 を」と「 c^1_3 を」で c^1_1 と c^1_3 が
同じ助詞に連なっている)



c^1_1	c^1_2		c^1_4	c^2_1
ミルク コーヒー 僕 彼	飲む 遊ぶ 見る		公園 遊園地	c^1_1 を c^1_4 に c^1_1 は c^1_1 を c^1_4 で c^1_1 と

(c^1_3 を c^1_1 に統合)

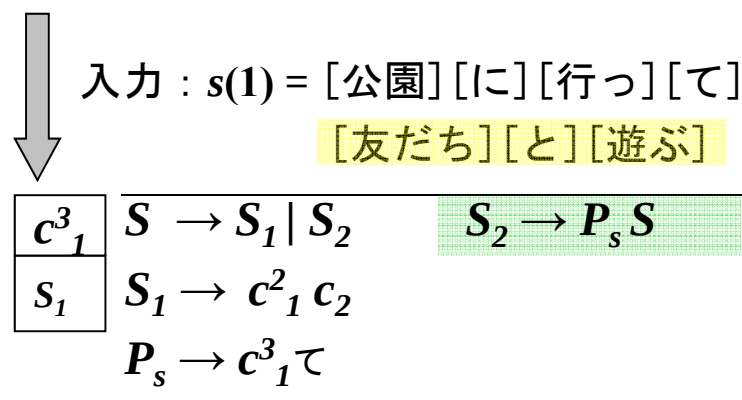
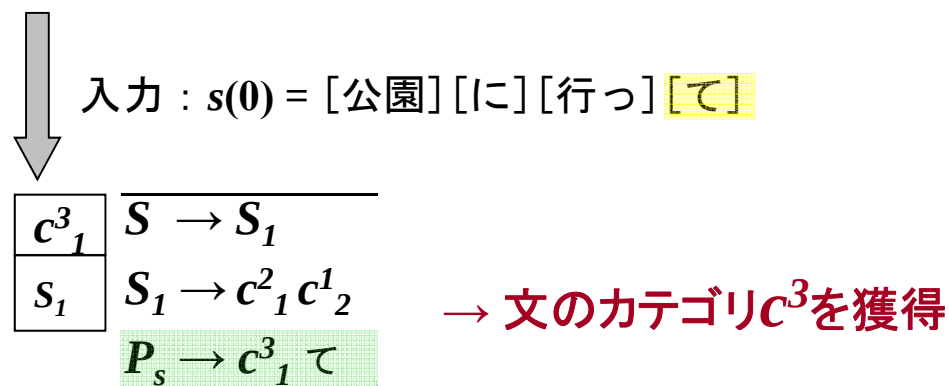
レベル3カテゴリ c^3 (重文構造)

- 重文の最初にくる文を従属節 P_s として獲得

c^1_1	c^1_2	c^1_3	c^1_4
ミルク	飲む	公園	友だち
コーヒー	見る	遊園地	
猫	遊ぶ		
僕	行く		

c^2_1
V を V は
V に V で
V と

$S \rightarrow S_1$	V を $\rightarrow c^1_1$ を
$S_1 \rightarrow c^2_1 c^1_2$	V に $\rightarrow c^1_3$ に
	V と $\rightarrow c^1_4$ と
	V は $\rightarrow c^1_1$ は
	V で $\rightarrow c^1_3$ で



レベル3カテゴリ c^3 (埋め込み文を含む構造)

- 名詞節を P_n として獲得

c^1_1	c^1_2	c^1_3	c^1_4
ミルク	飲む	公園	友だち
コーヒー	見る	遊園地	
猫	遊ぶ		
僕	行く		

c^2_1
$V_を$ $V_は$
$V_に$ $V_で$
$V_と$

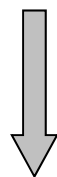
$S \rightarrow S_1$ $V_を \rightarrow c^1_1を$
 $S_1 \rightarrow c^2_1 c^1_2$ $V_に \rightarrow c^1_3に$
 $V_と \rightarrow c^1_4と$
 $V_は \rightarrow c^1_1は$
 $V_で \rightarrow c^1_3で$

↓ 入力: $s(2) = [\text{公園}][\text{で}][\text{遊ん}][\text{だ}]$

$S \rightarrow S_1 | S_3$

$S_1 \rightarrow c^2_1 c^1_2$

$S_3 \rightarrow S_1だ$



入力: $s(3) = [\text{公園}][\text{で}][\text{遊ん}][\text{だ}][\text{猫}]$

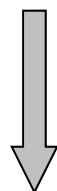
c^3_2
S_1

$S \rightarrow S_1 | S_3$

$P_n \rightarrow c^3_2だc^1_1$

$S_1 \rightarrow c^2_1 c^1_2$

$S_3 \rightarrow S_1だ$



入力: $s(4) = [\text{公園}][\text{で}][\text{遊ん}][\text{だ}][\text{猫}][\text{を}]$
 $[\text{遊園地}][\text{で}][\text{見る}]$

c^1_1	c^1_2	c^1_3	c^1_4
ミルク	飲む	公園	友だち
コーヒー	見る	遊園地	
猫	遊ぶ		
僕	行く		

c^3_2
S_1

$S \rightarrow S_1 | S_3$

$S_1 \rightarrow c^2_1 c^1_2$

$S_3 \rightarrow S_1だ$

$P_n \rightarrow c^3_2だc^1_1$

カテゴリの統合

- 獲得を進めるとカテゴリ数が増大
↳ **カテゴリの統合**で抑制

統合条件の例：

- 要素の一致が見られるカテゴリ
 - 同じ格が連なるカテゴリ
 - レベル2カテゴリ獲得後, 同じ助詞に連なるカテゴリ
-

要素の一致があるカテゴリの統合

- 単語カテゴリ内の要素の一致を条件に統合

c_1^1	c_2^1	c_3^1	c_4^1	c_5^1
ミルク	飲む	僕	公園	友だち
コーヒー	遊ぶ	彼	遊園地	彼
	行く	猫		

(c_3^1 と c_5^1 「彼」が一致)



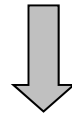
c_1^1	c_2^1	c_3^1	c_4^1	
ミルク	飲む	僕	公園	
コーヒー	遊ぶ	彼	遊園地	
	行く	猫		
		友だち		

(c_5^1 を c_3^1 に統合)

同じ格が連なるカテゴリの統合

- 同じ格が連なることを条件に統合

c^1_1	c^1_2	c^1_3	c^1_4	c^1_5
ミルク コーヒー	飲む	僕 彼猫	公園 遊園地	遊ぶ



c^1_1	c^1_2	c^1_3	c^1_4	
ミルク コーヒー	飲む 遊ぶ	僕 彼猫	公園 遊園地	

$S \rightarrow S_1 | S_2 | S_3 | S_4$
 $S_1 \rightarrow c^1_3$ は c^1_1 を c^1_2
 $S_2 \rightarrow c^1_1$ を c^1_3 は c^1_2
 $S_3 \rightarrow c^1_3$ は c^1_4 で c^1_5
 $S_4 \rightarrow c^1_4$ で c^1_3 は c^1_5

(「 c^1_3 は」が c^1_2 と c^1_5 の
両方に連なる)

$S \rightarrow S_1 | S_2 | S_3 | S_4$
 $S_1 \rightarrow c^1_3$ は c^1_1 を c^1_2
 $S_2 \rightarrow c^1_1$ を c^1_3 は c^1_2
 $S_3 \rightarrow c^1_3$ は c^1_4 で c^1_2
 $S_4 \rightarrow c^1_4$ で c^1_3 は c^1_2

(c^1_5 を c^1_2 に統合)

課題

- カテゴリ統合条件の検討
 - 日本語文法の品詞から見たカテゴリとレベルの位置づけの検討
 - 英語の言語獲得との対応
 - 例： ● レベル1: 単語
 - レベル2: [前置詞_名詞]
 - レベル3: 節、文
 - 言語獲得における意味の役割の形式化
-