

文法の種類

- 定義
 - 終端記号(terminal symbol, category) a, b, \dots
 - 非終端記号(nonterminal ...) A, B, \dots
 - 記号列(string) α, β, \dots
- 正規文法(regular grammar)
 - $A \rightarrow a, \quad A \rightarrow aB$
- 文脈自由文法(context-free grammar)
 - $A \rightarrow \beta$
- 文脈依存文法(context-sensitive grammar)
 - $\alpha \rightarrow \beta$
 - 1型: $|\alpha| \leq |\beta|$
 - 0型: 制限なし

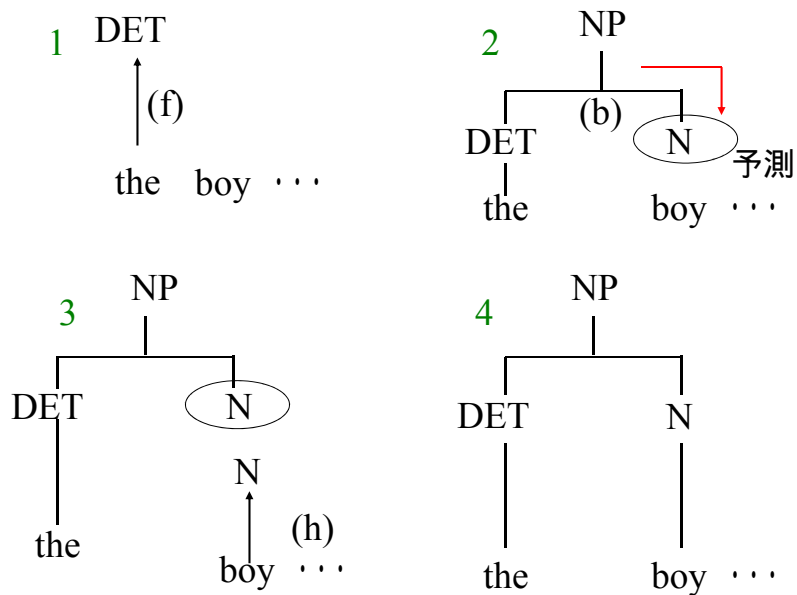
Top-Down Parsing

1. S the boy hits the dog with a rod.
2. NP VP the boy hits the dog with a rod.
3. DET N VP **the** boy hits the dog with a rod.
4. N VP **boy** hits the dog with a rod.
5. VP **hits the dog with a rod.**
6. V NP **hits** the dog with a rod.
7. NP the dog with a rod.
8. DET N **the** dog with a rod.
9. N **dog** with a rod.

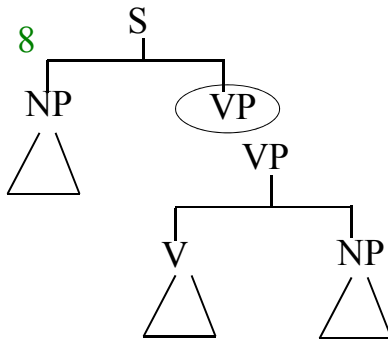
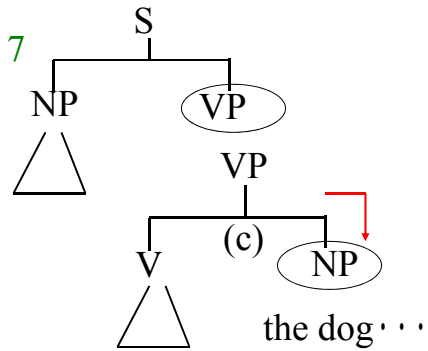
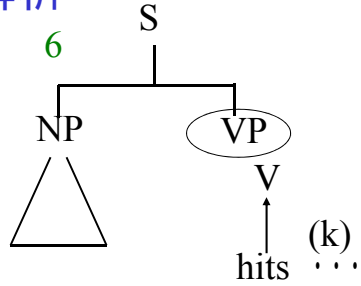
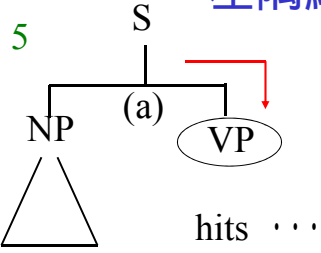
Top-Down Parsing

- | | |
|-------------|---|
| 1. S | the boy hits the dog with a rod. |
| 2. NP VP | the boy hits the dog with a rod. |
| 3. DET N VP | the boy hits the dog with a rod. |
| 4. N VP | boy hits the dog with a rod. |
| 5. VP | hits the dog with a rod. |
| 10. VP PP | hits the dog with a rod |
| 11. | |
| 12. PP | with a rod. |
| 13. PREP NP | with a rod. |
| 14. DET N | a rod. |

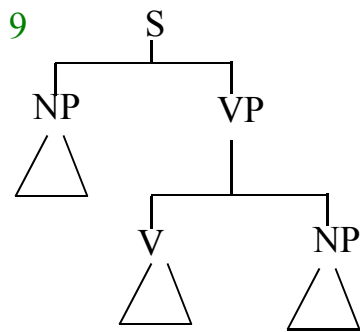
左隅統語解析



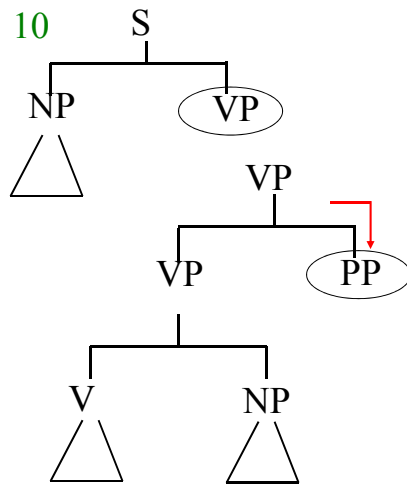
左隅統語解析



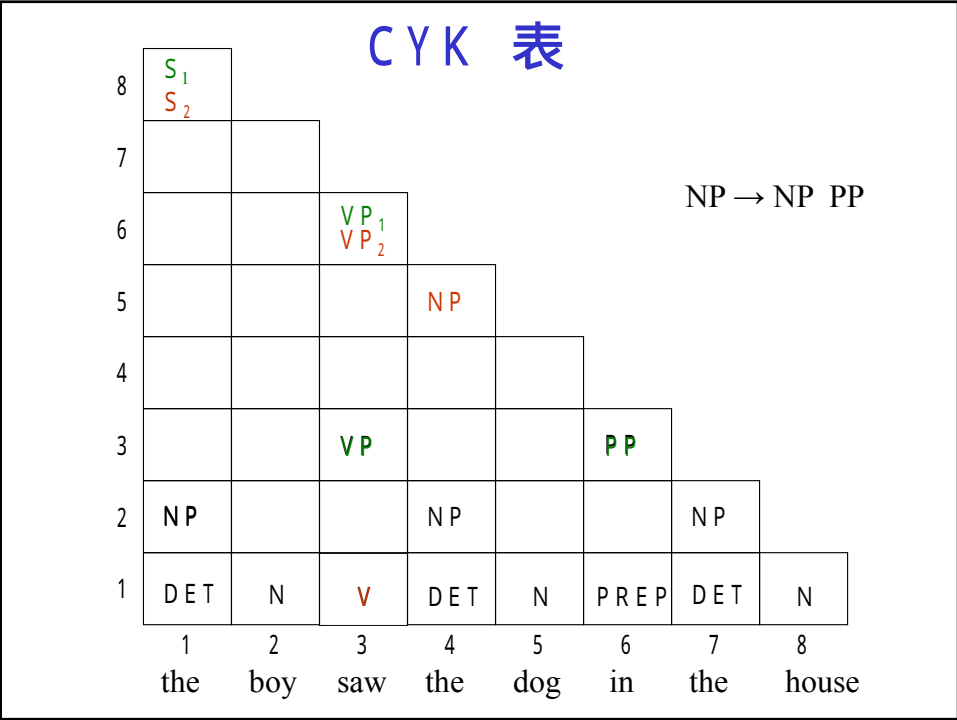
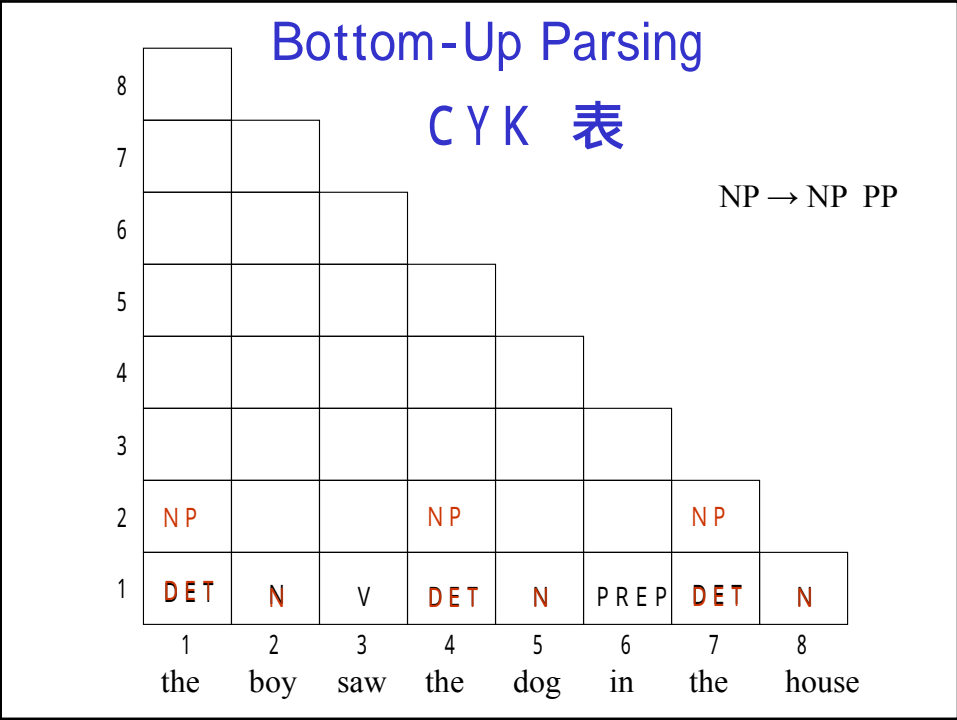
左隅統語解析



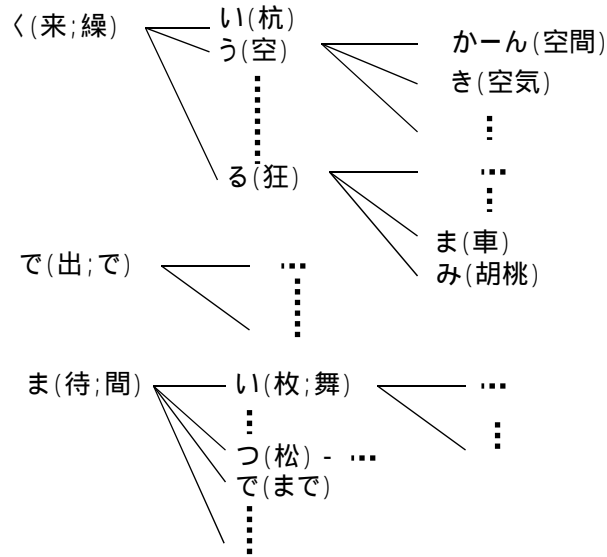
sentence が完成するが、
文が残っている



8へ後戻りし、解析を続ける



形態素解析のための辞書



形態素辞書に基づく CYK 表の途中結果

3		話 歯無し					
2	今	花 鼻	梨 無し	志賀		或	
1	医 位	間 真	は 歯	名 菜	四 死	が 我	有 亜
	1	2	3	4	5	6	7
	い	ま	は	な	し	が	あ
							る

形態素解析結果とCYK表

8	今Hが有る 今は梨が有る	
7	今Hが有 今は梨が有	H=[話 歯無し]
6	今Hが 今は梨が	G=[名 菜]
5	今H 今は梨	F=[花 鼻]
4	今はG 今F	
3	今は 今歯	いまはなしがある

コーパスに基づく統語解析

(統語解析を規定する)

確率文脈自由文法に基づく統計モデル

$$P(\alpha | A) = \frac{C(A \rightarrow \alpha)}{\sum_i C(A \rightarrow \alpha_i)}$$

依存関係に基づく統計モデル

He wrote a letter

$$P(w_i \rightarrow w_j) = \frac{C(w_i, w_j, \text{依存})}{C(w_i, w_j, \text{依存}) + C(w_i, w_j, \text{非依存})}$$

$$\text{Score}(D) = \prod_{w_i \rightarrow w_j \in D} P(w_i \rightarrow w_j)$$

同音語

(1) 自立語単語

しょうひん：商品、賞品、小品

(2) 活用形

いった　：行った、言った、入った

(3) 文節

ひとで　：人手、人で、火とで

(4) 接辞付き

しんぶんや：新分野、新聞屋

(5) べたがき

きょうはきものの：今日は着物の、今日履物の

文節

<文節> := <通常文節> | <数詞文節> | <固有名詞文節>

<通常文節> := (〔接頭辞〕 自立語 〔接尾辞〕)*

〔付属語〕*

<数詞文節> := 〔前置助数詞〕 数詞 〔後置助数詞〕

〔接尾語〕* 〔付属語〕*

<固有名詞文節> := 〔接頭語〕 固有名詞

〔接尾語〕* 〔付属語〕*

〔 〕は省略可、* は繰り返し可を表す。

文節単位変換(分かち書きあり)

(1) 自立語と付属語、付属語と付属語の接続行列

$C(i,j)=1$: 行 i が列 j に接続可能 (大きさは250程)

$C(i,j)=0$: 行 i が列 j に接続不可能

(2) 文節終端条件

$T(i,j)=1$: 文節終端可能

$T(i,j)=0$: 文節終端不可能

例 「おもったが」

尾もったが(尾もから後が接続不可能)

重ったが(用言の語尾変化が接続不可能)

連文節変換

(1) 文節最長一致法

ていあんしたけいかくを (失敗したら次の候補へ)

(2) 2文節最長一致法(最初の文節を決めるだけ)

けんきゅうの もくてきは 研究のも九

(3) 文節数最小法

にほんの れきしを まなぶ 日本 乗れ 岸を 学ぶ

(4) 前処理法(特徴的な部分を抽出してから解析)

ぶんしょうの にゅうりょくに (熟語の熟語に)

(5) 共通区切り探索法(n 文節最長一致法の共通区切り)

みせでは かった せいひんの

(店では買った、店で測った) 共通区切りで分割

あいまい性への対処

(1) 体言に直接動詞がつくものはX

私は/知っている (私/走っていると)

(2) 1字語名詞はX

増えてきたと/聞く (増えてきたとき/区)

(3) 漢字熟語の結合

行政改革 (行政か/威嚇)

(4) 意味情報の利用 (分類語彙表や類語辞典を参考にして辞書に付加される)

本を/読んだ (本を/呼んだ)

先生と/生徒が (宣誓と/生徒が)

第一階述語論理 (FPC)

「私は本を持つ」

$$\exists x[\text{have}(I, x) \wedge \text{book}(x)]$$

「私は本かノートを持つ」

$$\exists x[\text{have}(I, x) \wedge \text{book}(x)] \vee \exists x[\text{have}(I, x) \wedge \text{notebook}(x)]$$

「すべての女性はケーキが好きだ」

$$\forall x[\text{girl}(x) \rightarrow \exists y[\text{loves}(x, y) \wedge \text{cake}(y)]]$$

「誰もそれをできない」

$$\neg \exists x[\text{human}(x) \wedge \text{do-it}(x)]$$

「ペンギン以外の鳥は飛ぶ」

$$\forall x[\text{bird}(x) \wedge \neg \text{penguin}(x) \rightarrow \text{fly}(x)]$$

NL \rightarrow parser \rightarrow FPC \rightarrow Database

格文法

break

O

(a) "The window broke"

O: 対象格

break

John

A

the window

O

(b) "John broke the window"

A: 動作主格

break

John

A

the window

O

a hammer

I

(c) "John broke the window with a hammer"

I: 道具格

Fillmore の与えた深層格の集合

(1971年当時)

動作主格 (A): 動作を引き起こす者

経験者格 (E): 心理事象を体験する者

道具格 (I): 出来事の直接原因

対象格 (O): 移動する対象物や変化する対象物など

源泉格 (S): 対象物の移動の起点、および最初の状態

目標格 (G): 対象物の移動の終点、および最終的な状態

場所格 (L): 出来事が起こる場所

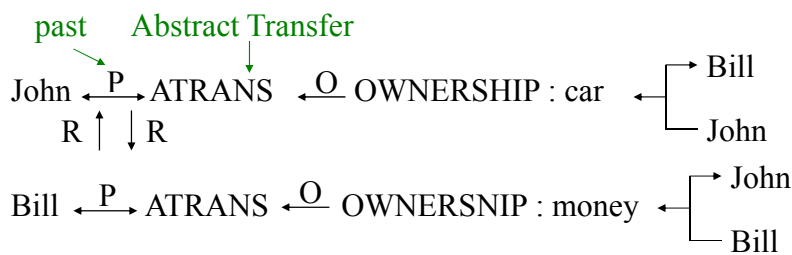
時間格 (T): 出来事が起こる時間

概念依存理論 (Conceptual Dependency Theory)

Schank, R. C. : Conceptual Information Processing,
North-Holland, Amsterdam and American Elsevier ,
New York , 1975

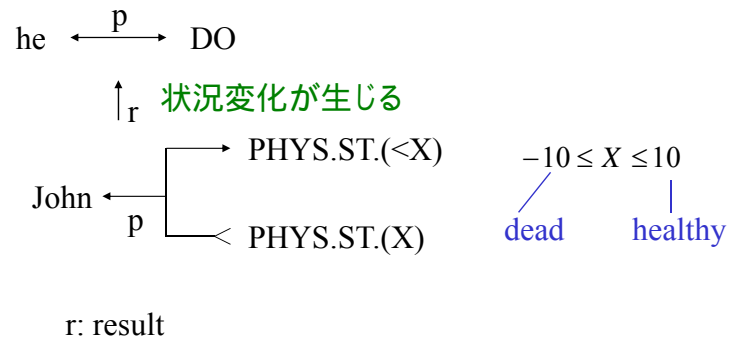
一つのまとまりを概念化 Conceptualization

“John sold his car to Bill.” に対応する概念依存構造

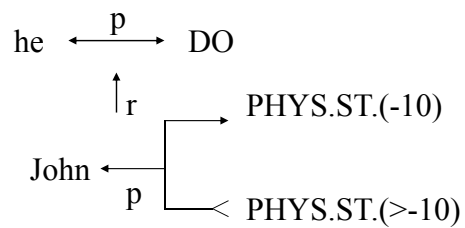


$\alpha_1 \xleftrightarrow{\quad} \alpha_2$ $\beta_1 \xleftarrow{\beta_2}$ が理由・原因 (Reason) となって
 $\beta_1 \xleftrightarrow{\quad} \beta_2$ $\alpha_1 \xleftarrow{\alpha_2}$ なる行為が引き起こされること
 P: past

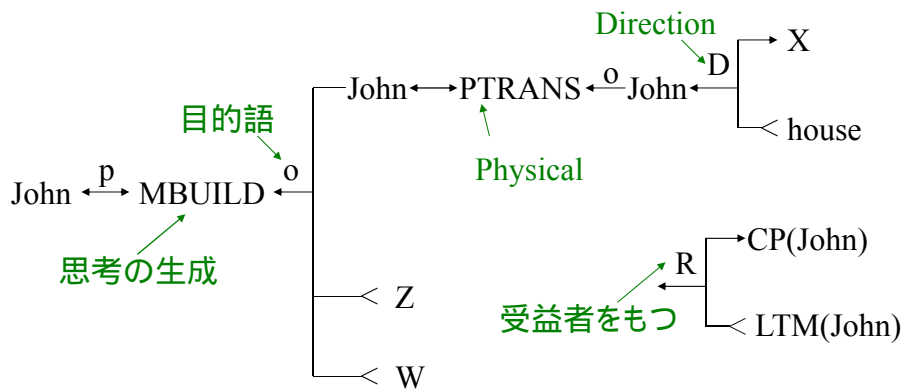
“He hurt John.” に対する概念依存構造



“He killed John.” に対する概念依存構造

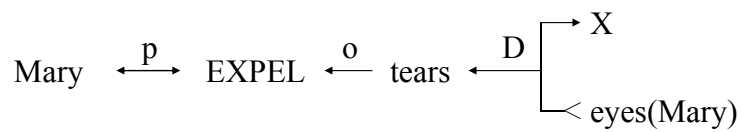


“John decided to leave the house.” の概念依存構造



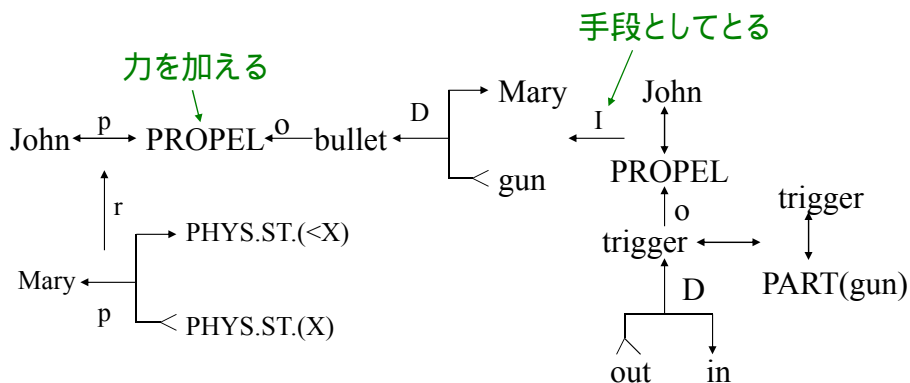
「ジョンは、ジョンがジョンを家からXに移動させる(PTRANS)という考えを、LTM(長期記憶)からCP(概念プロセッサ)に移動した。」

“Mary cried.”の概念依存構造



「メアリーはメアリーの目から涙をXに排出した。」

“John shot Mary.”の概念依存構造



「ジョンは、gunの一部である trigger に力を加えてout から in にすることによって鉄砲からメアリーに bullet を動かした。その結果メアリーの物理的な状態がXからX未満に悪化した。」

“John went to New York from Texas.” に対するMARGIEの出力

- (1) JOHN CAME TO BE IN NEW YORK.
- (2) JOHN CEASED TO BE IN TEXAS.
- (3) JOHN WANTED TO DO SOMETHING IN NEW YORK.
- (4) JOHN THOUGHT HE WOULD ENJOY BEING
IN NEW YORK.

PTRANSの概念依存構造での推論規則



上の概念依存構造から次のようなことを推論する。

- (1) Yは今Zに位置している。
JOHN CAME TO BE IN NEW YORK.
- (2) Yはもはや位置Wにはない。
JOHN CEASED TO BE IN TEXAS.
- (3) もしXが人間でC1を要求したか、XとYが同じであれば、Xは通常Zでできることをすると思われる。
JOHN WANTED TO DO SOMETHING IN NEW YORK.
- (4) (3)をすることがXを喜ばせるだろう。
JOHN THOUGHT HE WOULD ENJOY BEING IN NEW YORK.

PTRANSの概念依存構造での推論規則



上の概念依存構造から次のようなことを推論する。

- (1) Yは今Zに位置している。
- (2) Yはもはや。位置Wにはない
- (3) もしXが人間でC1を要求したか、XとYが同じであれば、Xは通常Zでできることをすると思われる。
- (4) (3)をすることがXを喜ばせるだろう。