

言語解析論

講師 竹内孔一

本日の内容

- 構文解析
 - 構文解析の難しさ
 - Chart Parser

構文解析 (parsing)

- 特徴
 - 入れ子構造を扱うモデル
 - マルコフモデルとの違い: 離れた関係を捉える
- 言語理論から
 - 文脈自由言語 (CFL)
- 問題点
 - 多数の書き換え規則の適用に時間がかかる
 - まちがった解析と正しい解析の区別がつかない
- 解決法の一つ
 - 確率モデルと学習を利用した 確率文脈自由文法

利用例

- 翻訳
 - google、excite, 辞書英辞郎
- 事実関係の抽出 (Web)
 - 「11時半ごろAでBがCしたことが明らかになった」
いれ子の内容が重要
 - Web上から事件, 事故, など記事情報の抽出
- 遺伝子情報の解析
 - 塩基配列 ATGC が実は自然言語と同じく入れ子構造やATGC記号列のまとまりが機能をもつ
 - 原始生物からその部分は共通として使われている
 - 構文解析を利用して生物間のまとまりを調べる

規則の適用の難しさ

- CFLで解析してみよう

NP ::= A NP | AP N | AP NP | A N

AP ::= N P | AP A | NP P

A ::= 黒い | 美しい

N ::= 目 | 人

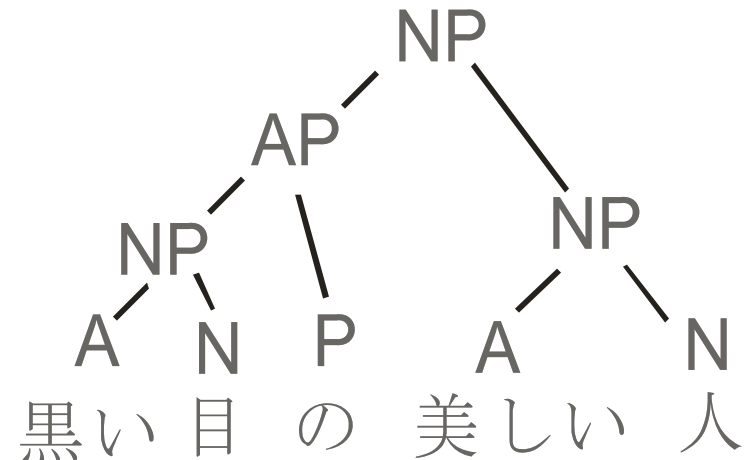
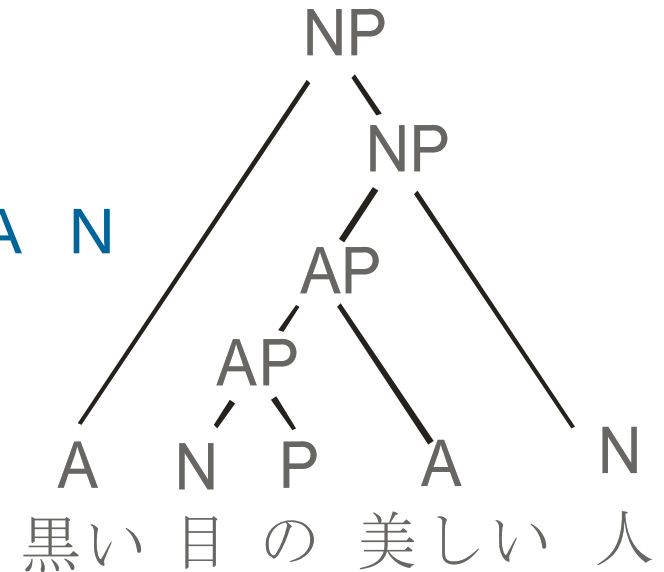
P ::= の

- トップダウン解析

- ボトムアップ解析

いくつかの冗長な解析が
おこなわれる

(どのルールをどこで使うかは自由)



top-down parsing

- トップダウンで行ってみよう

					NP ::= A NP AP N AP NP A N
NP					AP ::= N P AP A NP P
A NP					A ::= 黒い 美しい
黒い NP					N ::= 目 人
黒い	A	NP(失敗)			P ::= の
黒い	AP	NP			$\sigma =$ NP
黒い	N	P	NP		
黒い	目	の	NP		
黒い	目	の	A	N	
黒い	目	の	美しい	人	

どのルールを使うか
文脈は自由

bottom-up parsing

- ボトムアップで解析してみよう

黒い 目 の 美しい 人

A 目 の 美しい 人

A N の 美しい 人

NP の 美しい 人

NP P 美しい 人

AP 美しい 人

AP A 人

AP A N

AP NP

NP

この場合は深さ優先探索

$NP ::= A\ NP \mid AP\ N \mid AP\ NP \mid A\ N$

$AP ::= N\ P \mid AP\ A \mid NP\ P$

$A ::= \text{黒い} \mid \text{美しい}$

$N ::= \text{目} \mid \text{人}$

$P ::= \text{の}$

$\sigma = NP$

Chart Parser

- CFLの解析の難点
 - 繰り返し規則の展開 $NP ::= A NP$
 - 無限ループに陥る可能性がある
 - 多重の解が木の枝にできる
 - 組み合わせで書くと大変多い
- Chart Parser
 - CFLに基づく構文解析アルゴリズム
 - バックトラックしない
 - 無限ループに陥らない
 - 多重解釈の解を提示できる

基本原理は何？

Chart Parser

- 逐次型のアルゴリズム(概要)

1. 入力単語を一つを非終端記号(Xとする)に書き換える

以下を収束するまで繰り返し, なければ次の単語に進む

- 1つの非終端に置き換え可能なら置き換える

- $B := X Y \dots$ という規則を探し Yは空きとして記録

 - つまり右側の左隅がX

- $B := Y X[\text{空き}] Z[\text{空き}] \dots$ というルールが注目する単語よりまえまでにある場合, 埋めて更新. もし $B := Y X$ ですべてうまったら非終端 Bについてこの手続きを繰り返し止るまで行う

後で説明する

2. 最後の単語まで行き, S ができれば成功

- アイデア

- 単語の入力に応じて非終端を繰り返す

 - 無限に繰り返す必要は無い

- 単語の位置を利用して「解析済み」「未解析」を分ける

Chart Parser

埋まってない
活性弧

- 実演

– 入力文: 目の美しい

NP ::= A NP | AP N | AP NP | A N

AP ::= N P | AP A | NP P

A ::= 黒い | 美しい

N ::= 目 | 人

P ::= の

目 の 美しい

0 1 2 3

弧をはるための節点番号

小練習:

「目の美しい人」まで
解析してみる

回答
弧と適応規則

(0 1 N 目)	↓
(0 1 AP N P*)	
(1 2 P の)	
(0 2 AP N P)	
(0 2 NP AP NP*)	
(0 2 NP AP N*)	
(0 2 AP AP A*)	
(2 3 A 美しい)	
(0 3 AP AP A)	

これで終了

小練習

0 目 1 の 2 美しい 3 人 4

- ①(0 1 N 目)
- ②(0 1 AP N P*) ①より

- ③(1 2 P の)
- ④(0 2 AP N P) ② ③より
- ⑤(0 2 NP AP NP*) ④より
- ⑥(0 2 NP AP N*) ④より
- ⑦(0 2 AP AP A*) ④より

- ⑧(2 3 A 美しい)
- ⑨(2 3 NP A N*) ⑧より
- ⑩(2 3 NP A NP*) ⑧より

- ⑪(0 3 AP AP A) ⑦⑧より

- ⑫(0 3 NP AP NP*) ⑪より
- ⑬(0 3 NP AP N*) ⑪より
- ⑭(0 3 AP AP A*) ⑪より

- ⑮(3 4 N 人)
- ⑯(2 4 NP A N) ⑨⑮より
- ⑰(0 4 NP AP NP) ⑤⑯より
- ⑱(0 4 NP AP N) ⑬⑮より

よって解は⑰ と ⑱

練習14

- 「黒い目の美しい人」をChartParserで解析してそのステップを記述せよ
(規則は前と同じ)