

(トップダウン方式の) 簡単な構文解析アルゴリズム

サイエンス社「自然言語処理の基礎[新改版]」には、トップダウン方式の簡単な構文解析アルゴリズムが、以下の通り載っている。

少し分りにくいところが有ったので、() で説明を追加すると共に2つの解析例を示す。

/\*\*\*\*\*本の記述に少しコメントを付けたもの\*\*\*\*\*/

step1:open(スタック)を空( $\Phi$ )にする。

step2:open に開始記号と入力文の対[S,input]を入れる。

step3:open= $\Phi$ ならば、解析失敗として終了する。

step4:open の先頭の要素を取り出す(pop)。この要素を[Tree,input]とする。

step5:Tree に未解決の非終端記号がないならば、

step5-1:input= $\Phi$ ならば、Tree を解析結果として終了する。

step5-2:input! $\Phi$ ならば、step3 へ。

step6: (Tree に未解決の非終端記号があるなら) Tree の最も左にある未解決の非終端記号 A を左辺に持つ全ての書き換え規則  $A \rightarrow \alpha$  について、

step6-1: $\alpha$  の先頭が終端記号で input の先頭に等しいならば、input の先頭を取り除く。

( $\alpha$  の先頭が終端記号で) input の先頭に等しくないならば、step6-2 と step6-3 を行わず次の書き換え規則の処理に進む(step3 へ行く)。

step6-2:Tree 中の A を、A の小文字 a と  $\alpha$  を要素とするリストで置き換えた Tree1 を作る。

step6-3:対[Tree1,input]を open の先頭に追加する(push)。

step7:step3 へ行く。

/\*\*\*\*\*\*\*/

2つの解析例:

例1:バックトラックなしの最も簡単な例

(サイエンス社「自然言語処理の基礎[新改版]」p.89を簡単化したもの)

文法:

$S \rightarrow PP VP$

$PP \rightarrow NP P$

$VP \rightarrow V$

$NP \rightarrow N$

$N \rightarrow \text{学校}$

$P \rightarrow \text{に}$

V→行く

解析する文 (input) は「学校へ行く」

/\*コメント：以下の図で —————>はスタックの底を示す\*/

解析結果は次ようになれば成功。

[s [pp [np [n 学校]] [p に]] [vp[v 行く]]]

それでは、処理過程を追っていきます。

open (スタック) | TOS を取り出した Tree, input

=====

(step1,step2 によりスタックは次の様になる)

[S,学校に行く]

—————>

(step4 でポップ、スタックは空になる) | S,学校に行く

(S は未解決非終端なので step6-2,6-3 で、S を[s PP VP]に置き換えて、プッシュ)

[s PP VP,学校に行く]

—————>

(step4 でポップ、スタックは空になる) | s PP VP,学校に行く

(PP は未解決非終端なので step6-2,6-3 により PP を[pp NP P]に書き換えてプッシュ)

[s [pp NP P] VP,学校に行く]

—————>

(step4 でポップ、スタックは空になる) | s [pp NP P] VP,学校に行く

(NP は未解決非終端なので step6-2,6-3 により NP を[np N]に書き換えてプッシュ)

[s [pp [np N] P] VP,学校に行く]

—————>

(step4 でポップ、スタックは空になる) | s [pp [np N] P] VP,学校に行く

(N->学校、で学校が終端、かつ input の先頭と同じなので step6-1,6-2,6-3 により)

(注意 : input の先頭である学校は消える)

[s [pp [np [n 学校] ]P] VP,に行く]

—————>

(step4 でポップ、スタックは空になる) | s [pp [np [n 学校] ] P] VP,に行く

(P->に、で「に」は終端、かつ input の先頭と同じなので step6-1,6-2,6-3 により)

[s [pp [np [n 学校] ] [p に]] VP,行く]

—————>

(step4 でポップ、スタックは空になる) | s [pp [np [n 学校] ] [p に]] VP,行く

(VP は未解決非終端なので step6-2,6-3 により VP を [vp V] に書き換えてプッシュ)

[s [pp [np [n 学校] ] [p に]] [vp V],行く]

—————>

(step4 でポップ、スタックは空になる) | s [pp [np [n 学校] ] [p に]] [vp V],行く

(V->行く、で「行く」は終端かつ input の先頭と同じなので step6-1,6-2,6-3 により)

s [pp [np [n 学校] ] [p に]] [vp [v 行く]]

—————>

step3,4,5 と進み、5-1 で終了。

[s [pp [np [n 学校] ] [p に]] [vp [v 行く]]]

が解析結果。

例2：バックトラックのある例

(昭晃堂「言葉と言語処理」 p.105 より。この本の説明は、データ構造は少し異なるが、アルゴリズム上と同じで分かり易い) 説明は例1より少し雑にします。

解析する文：

the japanese saw him の解析

文法：

S→NP VP

NP→DET NP1

NP→NP1

NP1→ADJ NP1

NP1→N

VP→V NP

DET→the

ADJ→japanese

N→japanese | saw | him

V→saw

open (スタック) | TOS を取り出した Tree, input

=====

[S,the japanese saw him]

→

(step4:pop) | S,the japanese saw him

(step6-2,6-3により)

[s NP VP,the japanese saw him]

→

(step4:pop) | s NP VP,the japanese saw him

(step6、6-2,6-3により、2つ push)

[s [np DET NP1] VP,the japanese saw him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop)

| s [np DET NP1] VP,the japanese saw him

(step6-1,6-2 により)

[s [np [det the] NP1] VP, japanese saw him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop)

| s [np [det the] NP1] VP, japanese saw him

(step6、 6-2,6-3 により)

[s [np [det the] [np1 ADJ NP1]] VP, japanese saw him]

[s [np [det the] [np1 N]] VP, japanese saw him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop)

| s [np [det the] [np1 ADJ NP1]] VP, japanese saw him

(step6-1,6-2,6-3 により)

[s [np [det the] [np1 [adj japanese] NP1]] VP, saw him]

[s [np [det the] [np1 N]] VP, japanese saw him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop)

| s [np [det the] [np1 [adj japanese] NP1]] VP, saw him

(step6-2,6-3 により)

[s [np [det the] [np1 [adj japanese] [np1 ADJ NP1]]] VP, saw him]

[s [np [det the] [np1 [adj japanese] [np1 N]]] VP, saw him]

[s [np [det the] [np1 N]] VP, japanese saw him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop) | s [np [det the] [np1 [adj japanese] [np1 ADJ NP1]]] VP, saw him

(step6-1 が NO なので)

[s [np [det the] [np1 [adj japanese] [np1 N]]] VP, saw him]

[s [np [det the] [np1 N]] VP, japanese saw him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop) | s [np [det the] [np1 [adj japanese] [np1 N]]] VP, saw him

(step6-1 が YES(N=saw)なので)

[s [np [det the] [np1 [adj japanese] [np1 [n saw]]]] VP, him]

[s [np [det the] [np1 N]] VP, japanese saw him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop) | s [np [det the] [np1 [adj japanese] [np1 [n saw]]]] VP, him

(step6-1 が NO(VP!=him)なので)

[s [np [det the] [np1 N]] VP, japanese saw him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop) | s [np [det the] [np1 N]] VP, japanese saw him

(step6-1 が YES(N=japanese)なので)

[s [np [det the] [np1 [n japanese]]] VP, saw him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop) | s [np [det the] [np1 [n japanese]]] VP, saw him

(step6-1,6-2,6-3 により)

[s [np [det the] [np1 [n japanese]]] [vp V NP], saw him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop) | s [np [det the] [np1 [n japanese]]] [vp V NP], saw him

(step6-1 が YES(V=saw)なので)

[s [np [det the] [np1 [n japanese]]] [vp [v saw] NP], him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop) | s [np [det the] [np1 [n japanese]]] [vp [v saw] NP], him

(step6,step6-1,6-2,6-3 より)

[s [np [det the] [np1 [n japanese]]] [vp [v saw] [np DET NP1]], him]

[s [np [det the] [np1 [n japanese]]] [vp [v saw] [np NP1]], him]

[s [np NP1] VP,the japanese saw him]

—————>

(step4:pop) | s [np [det the] [np1 [n japanese]]] [vp [v saw] [np DET NP1]], him

(step1 が NO(DET!=him)なので何もせず)

[s [np [det the] [np1 [n japanese]]] [vp [v saw] [np NP1]], him]

[s [np NP1] VP,the japanese saw him]

---

(step4:pop) | s [np [det the] [np1 [n japanese]]] [vp [v saw] [np NP1]], him

(step6-1,6-2,6-3 より)

[s [np [det the] [np1 [n japanese]]] [vp [v saw] [np [np1 ADJ NP1]]], him]

[s [np [det the] [np1 [n japanese]]] [vp [v saw] [np [np1 N]]], him]

[s [np NP1] VP,the japanese saw him]

---

(st4) | s [np [det the] [np1 [n japanese]]] [vp [v saw] [np [np1 ADJ NP1]]], him

(step6-1 が NO(ADJ!=him)なので何もせず)

[s [np [det the] [np1 [n japanese]]] [vp [v saw] [np [np1 N]]], him]

[s [np NP1] VP,the japanese saw him]

---

(step4:pop) | s [np [det the] [np1 [n japanese]]] [vp [v saw] [np [np1 N]]], him

(step6-1 が YES(N=him)なので)

[s [np [det the] [np1 [n japanese]]] [vp [v saw] [np [np1 [n him]]]], null]

[s [np NP1] VP,the japanese saw him]

---

(step5-1 で終了)

s [np [det the] [np1 [n japanese]]] [vp [v saw] [np [np1 [n him]]]]

が解析結果、

スタックに一つ残っているが、これを同様にやっているとやがて破たんして終了。