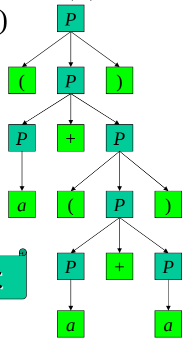
5. 文脈自由文法と言語(2):
(テキスト5.2)

5.2. 構文木

導出のプロセスは木構造
で表現されることが多い。

例) $P \Rightarrow (P) \Rightarrow (P+P)$
$\Rightarrow (a+P) \Rightarrow (a+(P))$
$\Rightarrow a+(P+P))$
$\Rightarrow (a+(a+P))$
$\Rightarrow (a+(a+a))$

構文木



1/34

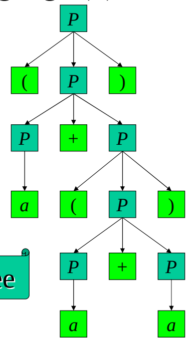
5. Context Free Language (2):
(Text 5.2)

5.2. Parse Tree

The process of derivations
are usually described by
Tree Structure.

Ex) $P \Rightarrow (P) \Rightarrow (P+P)$
$\Rightarrow (a+P) \Rightarrow (a+(P))$
$\Rightarrow a+(P+P))$
$\Rightarrow (a+(a+P))$
$\Rightarrow (a+(a+a))$

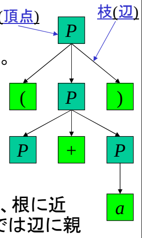
Parse Tree



2/34

5.2.* 木

節点(頂点) 枝(辺)

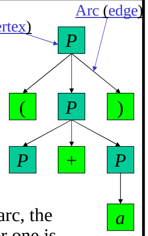


- 節点(頂点): 点。ここではラベルを持つ矩形。
- 枝(辺): 二つの点を結ぶ線。
- 木: 頂点が辺で結ばれたもので、
閉路がないもの。
- 根: 木の特定の1頂点。
(習慣的に)一番上に描く。
- 親子関係: 辺で結ばれた二つの頂点のうち、根に近いほうを親、そうでない方を子という。(ここでは辺に親から子へ方向をつける)
- 葉: 1つの辺にしかつながっていない頂点
- 先祖/子孫:
  1. 頂点 v は頂点 v の先祖かつ子孫
  2. 頂点 v の子孫の子供は v の子孫
  3. 頂点 v の先祖の親は v の先祖
- 子供は左の子と右の子は区別される。(順序つき木)

3/34

5.2.* Tree

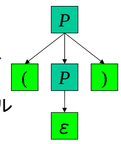
Node (vertex) Arc (edge)



- Node (vertex): Point. Rectangle with label.
- Arc (edge): Line joining two nodes.
- Tree consists of nodes and edges,
which contains no cycle.
- Root: One special node in a tree
(Usually) it placed on the top.
- Parent/children: For two nodes joined by an arc, the node closer to the root is parent, and the other one is child.
- Leaf: a node associated to exactly one arc.
- Ancestor/Successor:
  1. Node v is its ancestor and successor.
  2. A child of a successor is successor.
  3. A parent of an ancestor is ancestor.
- Left child and right child are distinguished. (So-called ordered tree.)

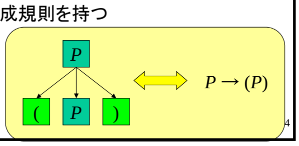
4/34

5.2.1. 構文木の構成



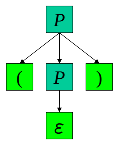
文法 $G=(V,T,P,S)$ に対して、Gの構文木とは、
以下の条件を満たす木:

1. 葉でない頂点にはV(非終端記号)がラベル
2. 葉のラベルは次のどれか一つ
   1. Tの要素(導出が終わっている頂点)
   2. Vの要素(まだ導出途中の頂点)
   3. ε (その葉が親の唯一の子供のとき)
3. 葉でない頂点のラベルがAで、子のラベルが左から
$X_1, X_2, \ldots, X_k$
なら、G は以下の生成規則を持つ
$A \rightarrow X_1X_2\ldots X_k$



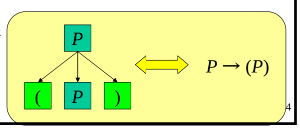
4

5.2.1. Construction of a parse tree



For a grammar $G=(V,T,P,S)$, the parse tree of G
is a tree that satisfies the following conditions:

- Inner nodes (non-leaf) have labels in V (Nonterminal)
- Leaf has one of the following labels
  - Element in T (derivation ends for the node)
  - Element in V (derivation does not end for the node)
  - ε (the leaf is the unique child of its parent)
- For each inner node of label A has children of labels
$X_1, X_2, \ldots, X_k$,
G has the production
rule $A \rightarrow X_1X_2\ldots X_k$.



4

5. 文脈自由文法と言語(2):
(テキスト5.2)

5.2.1. 構文木の構成

例)

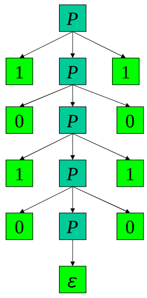
G_p＝{{P}, {0,1}, A, P}

A: $P \to \varepsilon$ | 0 | 1 | 0P0 | 1P1

10100101の導出

$P \Rightarrow 1P1 \Rightarrow 10P01 \Rightarrow 101P101$
$\Rightarrow 1010P0101 \Rightarrow 10100101$

10100101の構文木



5. Context Free Grammar (2):
(Text 5.2)

5.2.1. Construction of a parse tree

Ex)

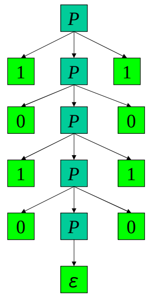
G_p＝{{P}, {0,1}, A, P}

A: $P \to \varepsilon$ | 0 | 1 | 0P0 | 1P1

The derivation of 10100101

$P \Rightarrow 1P1 \Rightarrow 10P01 \Rightarrow 101P101$
$\Rightarrow 1010P0101 \Rightarrow 10100101$

The parse tree of 10100101



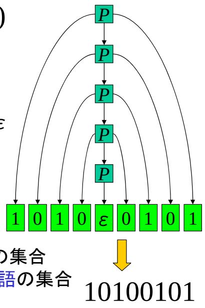
5. 文脈自由文法と言語(2):
(テキスト5.2)

5.2.2. 構文木の成果

構文木が
1. 根のラベルが出発記号
2. 葉のラベルがすべて終端記号か ε

のとき、葉のラベルを左から並べた文字列を構文木の成果と言う。

(c.f. $a\varepsilon = \varepsilon a = a$)



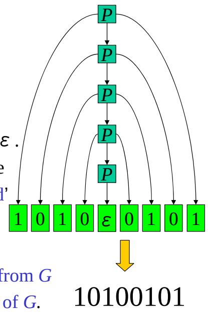
[観測] 文法 G によって導出される語の集合
＝文法 G の構文木の成果である語の集合

5. Context Free Grammar (2):
(Text 5.2)

5.2.2. Yield of a parse tree

Suppose a parse tree satisfies
1. the label of the root is start symbol
2. all labels of leaves are terminals or ε.

Then the string obtained by the sequence of labels from left to right is called ‘yield’ of the parse tree. (c.f. $a\varepsilon = \varepsilon a = a$)



[Observation] The set of words derived from G
＝The set of yields of the parse tree of G.

5. 文脈自由文法と言語(2):
(テキスト5.2)

5.2.3. 推論･導出･構文木

- 再帰的推論
  文字列(語＝終端記号列)から出発記号(非終端記号)
- 導出(最左導出と最右導出)
  出発記号(非終端記号)から文字列(語)
- 構文木

5. Context Free Grammar (2):
(Text 5.2)

5.2.3. Recursive Inference･Derivation･Parse Tree

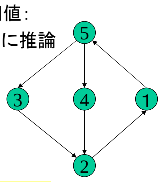
- Recursive Inference
  From a Word (=Terminal symbols) to the Start symbol (Nonterminal symbol)
- Derivations (Leftmost derivation and Rightmost derivation)
  From the Start symbol (Nonterminal symbol) to a Word (=Terminal symbols)
- Parse Tree

5. 文脈自由文法と言語(2):
(テキスト5.2)

5.2.3. 推論・導出・構文木

文法 G=(V,T,P,S) について以下はすべて同値:

1. 終端記号列 w から変数 S が再帰的に推論できる
2. $S \overset{*}{\Rightarrow} w$
3. $S \overset{*}{\underset{\text{左}}{\Rightarrow}} w$
4. $S \overset{*}{\underset{\text{右}}{\Rightarrow}} w$
5. S を根とし、w を成果とする構文木が存在。



• 3→2, 4→2 は自明
• 5→3, 5→4 は対称

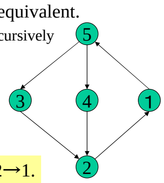
1→5, 5→3, 2→1 を示す。

5. Context Free Grammar (2):
(Text 5.2)

5.2.3. Recursive Inference・Derivation・Parse Tree

For a grammar G=(V,T,P,S), the followings are equivalent.

1. From a word w, the start symbol S can be recursively inferred.
2. $S \overset{*}{\Rightarrow} w$
3. $S \overset{*}{\underset{\text{左}}{\Rightarrow}} w$
4. $S \overset{*}{\underset{\text{右}}{\Rightarrow}} w$
5. There is a parse tree such that the label of the root is S, and its yield is w.



• 3→2, 4→2 : trivial
• 5→3, 5→4 :symmetric

We show 1→5, 5→3, 2→1.

5. 文脈自由文法と言語(2):
(テキスト5.2)

5.2.4. 再帰的推論から構文木へ(1→5)

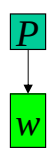
[定理] CFG G=(V,T,P,S) に対し、再帰的推論で語 w が変数 S の言語に属しているなら、S を根として、w を成果とする構文木が存在する。

[証明] w が S の言語に属していることを示す導出のステップ数に関する帰納法。

[基礎] w が S から1ステップで導出できる場合

生成規則 $S \rightarrow w$ が P に入っている。

したがって構文木が存在。



5. Context Free Grammar (2):
(Text 5.2)

5.2.4. From Recursive Inference to Parse Tree(1→5)

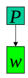
[Theorem] For a CFG G=(V,T,P,S), if w can be inferred to the start symbol S, there is a parse tree T such that T has a root with label S and yield w.

[Proof] Induction for the number of steps of the derivation to check if w is inferred to the start symbol S.

[Base] w is derived from S with one step.

P contains the rule $S \rightarrow w$.

Hence there is a parse tree (right figure).



5.2.4. 再帰的推論から構文木へ(1→5)

[定理] CFG G=(V,T,P,S) に対し、再帰的推論で語 w が変数 S の言語に属しているなら、S を根として、w を成果とする構文木が存在する。

[証明] w が S の言語に属していることを示す導出のステップ数に関する帰納法。

[帰納] w が S から n+1 ステップ(n>1)で導出できる場合

[帰納法の仮定] G において語 x が変数 B の言語に属していて、かつ x が B から n ステップ以下で導出できるなら、B を根として、x を成果とする構文木が存在。

5.2.4. From Recursive Inference to Parse Tree(1→5)

[Theorem] For a CFG G=(V,T,P,S), if w can be inferred to the start symbol S, there is a parse tree T such that T has a root with label S and yield w.

[Proof] Induction for the number of steps of the derivation to check if w is inferred to the start symbol S.

[Inductive step] Suppose that w can be derived from S with n+1 (n>1) steps.

[Inductive hypothesis] On G, if a word x is inferred to a nonterminal B, and x can be derived from B at most n steps, there is a parse tree T' such T' has the root with label B, and yield x.

5.2.4. 再帰的推論から構文木へ(1→5)

[証明] w が S の言語に属していることを示す導出のステップ数に関する帰納法。

[帰納] w が S から n+1 ステップ(n＞1)で導出できる場合

[帰納法の仮定] G において語 x が変数 B の言語に属していて、かつ x が B から n ステップ以下で導出できるなら、B を根として、x を成果とする構文木が存在。

w は S から n+1 ステップで導出できるので、P は生成規則

$S \to X_1X_2\ldots X_k$

をもち、かつ

$X_i \overset{*}{\Rightarrow} w_i$

$w = w_1w_2\ldots w_k$

を満たす文字列 $w_1, w_2, \ldots, w_k$ が存在する。

X_iからw_iの導出は高々 n ステップ ($X_i = w_i$ もありえる)

5.2.4. From Recursive Inference to Parse Tree(1→5)

[Proof] Induction for the number of steps of the derivation to check if w is inferred to the start symbol S.

[Inductive step] Suppose that w can be derived from S with n+1 (n＞1) steps.

[Inductive hypothesis] On G, if a word x is inferred to a nonterminal B, and x can be derived from B at most n steps, there is a parse tree T' such T' has the root with label B, and yield x.

Since w can be derived from S with n+1 steps, P has a production rule

$S \to X_1X_2\ldots X_k$,

and there are substrings $w_1, w_2, \ldots, w_k$ such that

$X_i \overset{*}{\Rightarrow} w_i$

$w = w_1w_2\ldots w_k$.

The derivation from X_i to w_i has at most n steps (It can be $X_i = w_i$.)

5.2.4. 再帰的推論から構文木へ(1→5)

[帰納] w が S から n+1 ステップ(n＞1)で導出できる場合

[帰納法の仮定] G において語 x が変数 B の言語に属していて、かつ x が B から n ステップ以下で導出できるなら、B を根として、x を成果とする構文木が存在。

w は S から n+1 ステップで導出できるので、P は規則

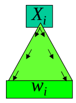
$S \to X_1X_2\ldots X_k$をもち、かつ

$X_i \overset{*}{\Rightarrow} w_i$ (nステップ以下で導出できる)

$w = w_1w_2\ldots w_k$

を満たす文字列 $w_1, w_2, \ldots, w_k$ が存在する。

G において語 w_i は変数 X_i の言語に属し、かつ n ステップ以下で導出できるので、帰納法の仮定より、X_i を根として w_i を成果とする構文木が存在する。



5.2.4. From Recursive Inference to Parse Tree(1→5)

[Inductive step] Suppose that w can be derived from S with n+1 (n＞1) steps.

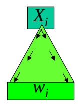
[Inductive hypothesis] On G, if a word x is inferred to nonterminal B, and x can be derived from B at most n steps, there is a parse tree T' such T' has the root with label B, and yield x.

Since w can be derived from S with n+1 steps, P has a production rule $S \to X_1X_2\ldots X_k$, and there are substrings $w_1, w_2, \ldots, w_k$ such that

$X_i \overset{*}{\Rightarrow} w_i$

$w = w_1w_2\ldots w_k$.

In G, the words w_i is inferred to symbol X_i, which can be derived at most n steps. Hence, by inductive hypothesis, there is a parse tree with the root labeled X_i that has yield w_i.



5.2.4. 再帰的推論から構文木へ(1→5)

[帰納] w が S から n+1 ステップ(n＞1)で導出できる場合

w は S から n+1 ステップで導出できるので、P は生成規則

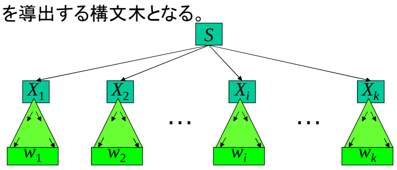
$S \to X_1X_2\ldots X_k$

をもち、かつ

$X_i \overset{*}{\Rightarrow} w_i$, $w = w_1w_2\ldots w_k$

を満たす文字列 $w_1, w_2, \ldots, w_k$ が存在する。

帰納法の仮定より、X_i を根として w_i を成果とする構文木が存在する。これらの構文木から以下の構文木を構成すると、S から w を導出する構文木となる。



5.2.4. From Recursive Inference to Parse Tree(1→5)

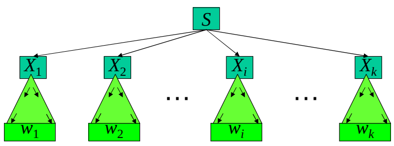
[Inductive step] Suppose that w can be derived from S with n+1 (n＞1) steps.

Since w can be derived from S with n+1 steps, P has a production rule $S \to X_1X_2\ldots X_k$, and there substrings $w_1, w_2, \ldots, w_k$ such that

$X_i \overset{*}{\Rightarrow} w_i$

$w = w_1w_2\ldots w_k$.

By inductive hypothesis, there is a parse tree with the root labeled X_i that has yield w_i. From the parse trees, we construct the following parse tree which derives w from S.



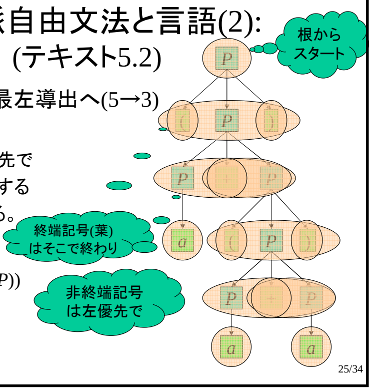
5. 文脈自由文法と言語(2):
(テキスト5.2)

5.2.5. 構文木から最左導出へ(5→3)

直感的には…

構文木を左優先で
深さ優先探索する
ことに対応する。

例) $P \Rightarrow (P) \Rightarrow (P+P)$
$\Rightarrow (a+P) \Rightarrow (a+(P))$
$\Rightarrow a+(P+P))$
$\Rightarrow (a+(a+P))$
$\Rightarrow (a+(a+a))$

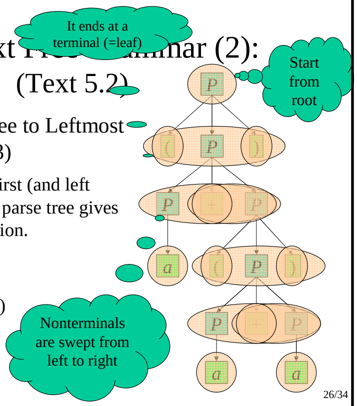


5. Context Free Grammar (2):
(Text 5.2)

5.2.5. From Parse Tree to Leftmost Derivation (5→3)

Intuitively, Depth first (and left first) search for the parse tree gives the leftmost derivation.

Ex) $P \Rightarrow (P) \Rightarrow (P+P)$
$\Rightarrow (a+P) \Rightarrow (a+(P))$
$\Rightarrow a+(P+P))$
$\Rightarrow (a+(a+P))$
$\Rightarrow (a+(a+a))$



5. 文脈自由文法と言語(2):
(テキスト5.2)

5.2.5. 構文木から最左導出へ(5→3)

[定理] CFG $G=(V,T,P,S)$ に対し、変数 S を根とし、w を成果とする構文木があれば、G の最左導出 $S \overset{*}{\underset{左}{\Rightarrow}} w$ が存在する。

[略証] 木の高さ i についての帰納法で証明する。

(木の高さ = 各葉から根までの辺の個数の最大値)

木の高さが0のときは根しかないので、これは構文木ではない。したがって意味のある木の高さの最小値は1。

[基礎] i=1のとき: $S \rightarrow w$ が規則に入っている。

これは最左導出。

5. Context Free Grammar (2):
(Text 5.2)

5.2.5. From Parse Tree to Leftmost Derivation (5→3)

[Theorem] For a CFG $G=(V,T,P,S)$, if there exists a parse tree with the root labeled by S and yield w, there exists the leftmost derivation $S \overset{*}{\underset{左}{\Rightarrow}} w$ of G.

[Proof (Sketch)] Induction for the height i of the tree.

(The height of a tree = the maximum number of edges from a leaf to the root)

The tree of height 0 consists of just its root, which is not a parse tree. Hence, the smallest parse tree has height 1.

[Base] i=1: $S \rightarrow w$ is a production rule, which is a leftmost derivation.

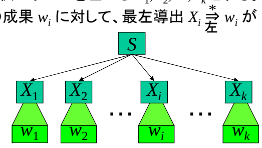
5.2.5. 構文木から最左導出へ(5→3)

[定理] CFG $G=(V,T,P,S)$ に対し、変数 S を根とし、w を成果とする構文木があれば、G の最左導出 $S \overset{*}{\underset{左}{\Rightarrow}} w$ が存在する。

[略証] 木の高さ i についての帰納法で証明する。

[帰納] $i \geqq 2$のとき:

- 根のラベルを S とし、S の子供のラベルを左から $X_1,X_2,\ldots,X_k$ とする。
- 帰納法の仮定から、各 X_i の成果 w_i に対して、最左導出 $X_i \overset{*}{\underset{左}{\Rightarrow}} w_i$ が存在する。
- $w = w_1w_2\ldots w_k$ である。



導出 $S \underset{左}{\Rightarrow} X_1X_2\ldots X_k$ から、
最左導出

$S \overset{*}{\underset{左}{\Rightarrow}} w_1w_2\ldots w_k$

が構成できることを示す。具体的には j=1,2,…,k について、

$S \overset{*}{\underset{左}{\Rightarrow}} w_1w_2\ldots w_j X_{j+1}\ldots X_k$

であることを j に関する帰納法で示す。（以下省略）

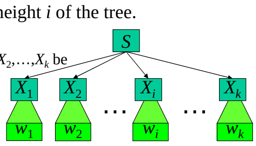
5.2.5. From Parse Tree to Leftmost Derivation (5→3)

[Theorem] For a CFG $G=(V,T,P,S)$, if there exists a parse tree with the root labeled by S and yield w, there exists the leftmost derivation $S \overset{*}{\underset{左}{\Rightarrow}} w$ of G.

[Proof (Sketch)] Induction for the height i of the tree.

[Inductive step] $i \geqq 2$:

- Let S be the label of the root, $X_1,X_2,\ldots,X_k$ be the labels of children of the root from left to right in this order.
- By the inductive hypothesis, for each yields w_i of X_i, there is the leftmost derivations $X_i \overset{*}{\underset{左}{\Rightarrow}} w_i$.
- $w = w_1w_2\ldots w_k$.



From a derivation $S \underset{左}{\Rightarrow} X_1X_2\ldots X_k$, we construct the leftmost derivations $S \overset{*}{\underset{左}{\Rightarrow}} w_1w_2\ldots w_k$.
To do this, for each j=1,2,…,k, we show $S \overset{*}{\underset{左}{\Rightarrow}} w_1w_2\ldots w_j X_{j+1}\ldots X_k$ by an induction of j, which is omitted here.

5. 文脈自由文法と言語(2):
(テキスト5.2)

5.2.6. 導出から再帰的推論へ(2→1)

[定理] CFG $G=(V,T,P,S)$ に対して、導出 $S\overset{*}{\Rightarrow}w$ があれば、w が S の言語に属することが再帰的推論によって確かめられる。

[略証] 導出 $S\overset{*}{\Rightarrow}w$ の長さに関する帰納法による。

[基礎] 長さが 1 のとき: $S\rightarrow w$ が生成規則に入っている。
したがって1ステップの再帰的推論により確認できる。

[帰納] 導出 $S\overset{*}{\Rightarrow}w$ の長さが $n+1$ とし、長さ n 以下のすべての導出が再帰的推論によって確かめられるとする。
導出は生成規則 $S\rightarrow X_1X_2\ldots X_k$ により

$$S\Rightarrow X_1X_2\ldots X_k\overset{*}{\Rightarrow}w$$

という形で表現できる。

5. Context Free Grammar (2):
(Text 5.2)

5.2.6. From Derivation to Recursive Inference (2→1)

[Theorem] For a CFG $G=(V,T,P,S)$, if there is a sequence of derivations $S\overset{*}{\Rightarrow}w$, the start symbol S can be recursively inferred from a word w.

[Proof (Sketch)] Induction for the number of the derivations $S\overset{*}{\Rightarrow}w$.

[Base] Number=1: $S\rightarrow w$ is a production rule.
Hence one step inference gives S from w.

[Inductive Step] Suppse that the number of the derivations of $S\overset{*}{\Rightarrow}w$ is $n+1$, and any derivations less than n times can be checked by a recursive inference. Then, the derivation can be represented by

$$S\Rightarrow X_1X_2\ldots X_k\overset{*}{\Rightarrow}w$$

for a production $S\rightarrow X_1X_2\ldots X_k$.

5.2.6. 導出から再帰的推論へ(2→1)

[定理] CFG $G=(V,T,P,S)$ に対して、導出 $S\overset{*}{\Rightarrow}w$ があれば、w が S の言語に属することが再帰的推論によって確かめられる。

[略証] 導出 $S\overset{*}{\Rightarrow}w$ の長さに関する帰納法による。

[帰納] 導出 $S\overset{*}{\Rightarrow}w$ の長さが $n+1$ とし、長さ n 以下のすべての導出が再帰的推論によって確かめられるとする。
導出は生成規則 $S\rightarrow X_1X_2\ldots X_k$ により

$$S\Rightarrow X_1X_2\ldots X_k\overset{*}{\Rightarrow}w$$

という形で表現できる。さらに

» $X_i\overset{*}{\Rightarrow}w_i \quad (1\leqq i\leqq k)$

» $w=w_1w_2\ldots w_k$

であり、帰納法の仮定から、すべての導出 $X_i\overset{*}{\Rightarrow}w_i$ は再帰的推論によって確かめられる。したがって $S\rightarrow X_1X_2\ldots X_k$ とこれらの推論から、w が S の言語に属することが推論によって確かめられる。

5.2.6. From Derivation to Recursive Inference (2→1)

[Theorem] For a CFG $G=(V,T,P,S)$, if there is a sequence of derivations $S\overset{*}{\Rightarrow}w$, the start symbol S can be recursively inferred from a word w.

[Proof (Sketch)] Induction for the number of the derivations $S\overset{*}{\Rightarrow}w$.

[Inductive Step] Suppse that the number of the derivations of $S\overset{*}{\Rightarrow}w$ is $n+1$, and any derivations less than n times can be checked by a recursive inference. Then, the derivation can be represented by $S\Rightarrow X_1X_2\ldots X_k\overset{*}{\Rightarrow}w$ for a production $S\rightarrow X_1X_2\ldots X_k$.

Moreover,

$X_i\overset{*}{\Rightarrow}w_i \quad (1\leqq i\leqq k)$

$w=w_1w_2\ldots w_k$

and inductive hypothesis, each derivation $X_i\overset{*}{\Rightarrow}w_i$ can be checked by a recursive inference. Thus, $S\rightarrow X_1X_2\ldots X_k$ with the inferences, w can be recursively inferred to S.