

左再帰規則を含む CFG に対する漸進的構文解析

加藤 芳秀

松原 茂樹

外山 勝彦

稲垣 康善 (名古屋大学)

1 はじめに

実時間対話処理システムの実現には、自然言語文をその出現順序に従って解釈する枠組が必要である。そのための手法として、漸進的チャート解析が提案されている [2]。この手法は文脈自由文法 (CFG) に基づいて構文解析を実行するが、CFG が左再帰規則を含むとき解析が無限ループに陥るといった問題がある。本稿では、漸進的チャート解析に対して木接合文法 (Tree Adjoining Grammar, TAG) [1] における接合操作を導入する。これにより、CFG が左再帰規則を含んでいる場合でも、無限ループに陥ることなく漸進的に解析できることを示す。

2 漸進的チャート解析とその問題点

漸進的チャート解析 [2] は、上昇型解析と下降型解析を融合した構文解析手法であり、語が入力されるごとに、それまでに入力された文の断片に対する構文木を作成することができる。例えば、図 1 に示す文法と辞書を用いて、英語文 “The boy met the girl.” を解析する場合、語 “boy” が入力された段階では、文の断片 “The boy” に対する構文木として、(1) $[[the]_{dt}[boy]_{np}]_s$ を作成する。次の語 “met” が入力された段階では、まず語 “met” に対する構文木 (2) $[[met]_{v}]_{vp}$ を作成し、構文木 (1) の $[?]_{vp}$ を (2) で置き換えることにより、“The boy met” に対する構文木として (3) $[[the]_{dt}[boy]_{np}[[met]_{v}]_{vp}]_s$ を作成する。

しかし、漸進的チャート解析は下降的な側面をもった解析処理であるため、左再帰規則を含む CFG に対して解析が無限ループに陥るといった問題がある。例えば、図 1 の文法規則 (g4) は左再帰規則であり、文の断片 “the boy met” に対する構文木は、(3) に加えて次のように無限に存在する。

$$\begin{aligned} &[[[the]_{dt}[boy]_{np}[[met]_{v}]_{vp}[[met]_{v}]_{vp}]_{pp}]_s \\ &[[[the]_{dt}[boy]_{np}[[met]_{v}]_{vp}[[met]_{v}]_{vp}[[met]_{v}]_{vp}]_{pp}]_{pp}]_s \\ &\vdots \end{aligned}$$

漸進的チャート解析では、それらすべての解析木の作成を試みるため、解析処理が停止しない。

3 漸進的チャート解析への接合操作の導入

本稿で提案する手法では、まず、解析処理の無限ループを防止するために、左再帰規則を取り除いた CFG に基づき漸進的チャート解析を実行する。その結果として、左再帰規則に対応した構造を含む構文木を作成できなくなるので、左再帰規則を処理する特別な操作として接合操作を新たに導入する。

TAG では、再帰的な構造を補助木と呼ばれる木により表現する。補助木はフットノードと呼ばれる特殊な葉ノードをもち、その範囲は根ノードと同一である。接合操作は、1) 構文木のある内部ノードで上下に分割し、2) そのノードと同一の範囲を根ノードにもつ補助木の根ノードと分割された上側の木を、フットノードと下側の木を結合する操作である。例を図 2 に示す。(b) が補助木、 vp^* がフットノードである。構文木 (a) の節点 vp に (b) を接合して得られた構文木が (c) である。

- (g1) $s \rightarrow np\ vp$ (g4) $vp \rightarrow vp\ pp$ (d1) $dt \rightarrow the$ (d4) $p \rightarrow in$
- (g2) $np \rightarrow dt\ n$ (g5) $pp \rightarrow p\ np$ (d2) $n \rightarrow boy/girl/park$
- (g3) $vp \rightarrow v\ np$ (d3) $v \rightarrow met$

図 1: 文法と辞書

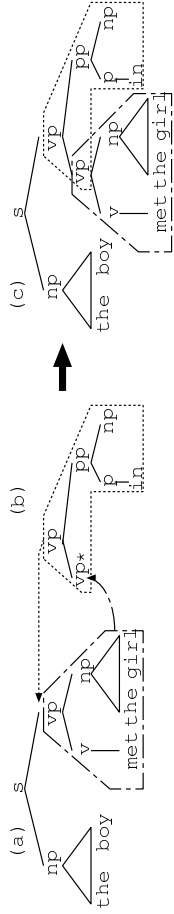


図 2: 接合操作

本手法を導入する接合操作では、次の補助木作成手続きにより左再帰規則に対応した構造を含む補助木を作成し、可能ならば、即座にそれを文の断片に対する構文木に接合する。

補助木作成 語に対する構文木 $[...]_X$ に対して、左再帰規則 $A \rightarrow AXY \dots Z$ が存在するならば、補助木 $[[*]_A[...]_X[?]_Y \dots [?]_Z]_A$ を作成する ($[*]_A$ はフットノードを表す)。

提案手法では、左再帰規則に対応した構造を含む構文木を即座に作成するのではなく、必要になった段階でそのような構文木を接合操作により作成するため、CFG が左再帰規則を含んでいる場合でも、無限ループに陥ることはない。また、語に対して補助木を作成し、それを即座に接合するため、漸進的に解析を進行させることができる。

例として、英語文 “The boy met the girl in the park.” を図 1 の文法で解析する場合を考える。文の断片 “The boy met the girl” に対する構文木は無限に存在するが、語 “girl” が入力された段階では、左再帰規則 (g4) を取り除いた文法で解析するため、図 2 の構文木 (a) のみを作成する。次の語 “in” が入力された段階で、“in” に対して構造木 $[[in]_p[?]_{np}]_{pp}$ を作成し、それに対して補助木作成手続きを適用することにより、図 2 の構文木 (b) を作成する。構文木 (a) に構文木 (b) を接合することにより、“The boy met the girl in” に対する構文木 (c) を作成する。

4 おわりに

本稿では漸進的チャート解析に TAG の接合操作を導入し、CFG が左再帰規則を含む場合でも、無限ループに陥ることなく漸進的に解析する手法を提案した。提案手法を GNU Common Lisp を用いて実装し、それが正しく動作することを確認した。

参考文献

- [1] Joshi, A.: Tree Adjoining Grammar: How Much Context-Sensitivity is required to Provide Reasonable Structural Descriptions?, In D.R. Dowty, L. Karttunen, and A. M. Zwicky (eds), *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, Cambridge University Press, New York (1985).
- [2] Matsubara, S. et al.: Chart-based Parsing and Transfer in Incremental Spoken Language Translation, *Proceedings of NLP95-97*, pp.521-524 (1997).