

# 漸進的な構文解析における出力タイミング決定の一手法

加藤 芳秀†

松原 茂樹†

外山 勝彦†

稲垣 康善†

†名古屋大学大学院工学研究科

†名古屋大学言語文化部

## 1 はじめに

実時間対話処理システムを実現するためには、自然言語文をその出現順序に従って順次解釈する枠組、すなわち、漸進的解釈手法が不可欠となる [2]。相手の発話を漸進的に解釈することにより、発話途中での割り込みの生成や即時的な応答が可能となり、自然で円滑な対話の実現が期待できる。

漸進的に意味解析や文脈解析を行うには、入力途中の段階でそれまでの入力に対する構文構造を作成する必要がある [1]。これを実現する手法として、漸進的チャート解析が提案されている [4]。この手法は、語が入力されるごとにそれまでの入力に対する構文構造を随時作成する。しかし、作成された構文構造の中には、それ以降の入力に対して解析を続行できない構造 (正しくない構造) が含まれる可能性がある。そのような構造をもとに、意味解析や文脈解析を実行すると、システムは相手の発話に対して適切に応答することはできない。これを解決する方法として、語単位で構文解析結果を出力するという制約を緩和することが考えられる。

そこで本稿では、漸進的な構文解析において、構文構造が正しいと判明するまで出力を遅延することとして、出力タイミングを動的に決定する手法を提案する。本手法では、構文構造間の包含関係を用いることにより、入力途中の段階で随時、構文構造の正しさを判別する。これにより、漸進性を保持しつつ、正しい構文構造のみを出力することが可能となる。

## 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 [4] は、従来の上昇型チャート解析 [3] に対して、活性弧に文法規則を適用する操作、及び活性弧の項の最左未決定項を別の活性弧の項で置き換える操作を導入した枠組である。ここで項とは、構文構造を表現するデータ構造である。 $i$  番目の語  $w_i$  が入力されたときに用いられる手続きを以下に示す。

- 1) 辞書引き 語  $w_i$  の範疇が  $X$  ならば、項  $[w_i]_X$  をもつ不活性弧をチャートの節点  $i-1$  と節点  $i$  の間に追加する。
- 2) 文法規則の適用 チャートの節点  $i-1$  と節点  $i$  を結ぶ項  $[\dots]_X$  をもつ弧に対して、文法規則  $A \rightarrow XY \dots Z$  が存在するならば、項  $[[\dots]_X [?]_Y \dots [?]_Z]_A$  をもつ弧をチャートの節点  $i-1$  と節点  $i$  の間に追加する。
- 3) 項の置き換え チャートの節点  $0$  と節点  $i-1$  を結ぶ活性弧の項  $\sigma$  の最左未決定項を  $[?]_X$  とする。このとき、チャートの節点  $i-1$  と節点  $i$  を結ぶ弧の項  $\tau$  の範疇が  $X$  ならば、 $\sigma$  の最左未決定項を  $\tau$  で置き換えた項をもつ弧をチャートの節点  $0$  と節点  $i$  の間に追加する。

文法	辞書
s $\rightarrow$ np vp \$	pron $\rightarrow$ I / her
np $\rightarrow$ pron	det $\rightarrow$ her / the
np $\rightarrow$ det n	n $\rightarrow$ parent / book
np $\rightarrow$ det n pp	p $\rightarrow$ with
pp $\rightarrow$ p np	vt $\rightarrow$ saw
vp $\rightarrow$ vt np	vi $\rightarrow$ saw
vp $\rightarrow$ vi	\$ $\rightarrow$ .

図 1: 構文解析のための文法と辞書

この手法では、語が入力されるごとにそれまでの入力に対する構文構造を漸進的に作成することが可能である。このため、入力途中の段階での構文構造を用いた処理、例えば、意味解析や文脈解析などを実行するための枠組として適している。

例えば、図 1 に示す文法と辞書を用いるとき、英語文

$$I \text{ saw her parent with the book.} \quad (2.1)$$

における語 “saw” が入力された段階では、“I saw” に対する項

$$[[[I]_{pron}]_{np} [[saw]_{vi}]_{vp} [?]_{\$}]_s \quad (2.2)$$

$$[[[I]_{pron}]_{np} [[saw]_{vt} [?]_{np}]_{vp} [?]_{\$}]_s \quad (2.3)$$

をそれぞれ作成することができ、この段階で “saw” の主語は “I” であるなどといった文の構成要素間の関係を捉えることができる [1]。

しかし、漸進的チャート解析において作成された項の中には、それ以降の入力に対して解析を続行できないものが含まれている可能性がある。例えば、英語文 (2.1) において、語 “saw” が入力された段階で作成された項 (2.2) は、“saw” 以降の入力に対して解析を続行できず、それに対して意味解析を実行しても正しい解析結果を得ることはできない。この問題を解決する方法として、語単位での項の出力という制約を緩和し、正しい項の特定に十分な情報が得られるまで出力を遅延することが考えられる。このとき、出力をどの程度遅延するかを考慮する必要があるが、本稿では、項が正しいと判明するまで出力を遅延する方法を採用する。

## 3 漸進的構文解析における出力タイミング決定

本稿で提案する手法は、漸進的チャート解析に対して出力タイミングを制御する枠組を導入している。語が入力されるごとに、解析処理により項を作成し、その正しさを判別し、正しければ即座にこれを出力する。なお、入力文全体に対して、項は一つのみ存在すると仮定する。

### 3.1 項の正しさの判別

漸進的チャート解析において正しい項のみを出力するためには、入力途中のできるだけ早い段階で、項が正しいかど

表 1: “I saw her parent with the book.” に対する漸進的な解析過程と出力タイミング

入力語	チャート			出力された 構文構造
	word	#	loc term	
I	1	0-0	$[?]_s$	#1
	2	0-1	$[[[I]_{pron}]_{np}[?]_{vp}]_s$	#2
saw	3	0-2	$[[[I]_{pron}]_{np}[[saw]_{vi}]_{vp}]_s$	
	4	0-2	$[[[I]_{pron}]_{np}[[saw]_{vt}]_{np}]_s$	
her	5	0-3	$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{pron}]_{np}]_{vp}]_s$	#4
	6	0-3	$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[?]_n]_{np}]_{vp}]_s$	
	7	0-3	$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[?]_n[?]_{pp}]_{np}]_{vp}]_s$	
parent	8	0-4	$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[parent]_n]_{np}]_{vp}]_s$	
	9	0-4	$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[parent]_n[?]_{pp}]_{np}]_{vp}]_s$	
with	10	0-5	$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[parent]_n[[with]_p[?]_{np}]_{pp}]_{vp}]_s$	#10
the	11	0-6	$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[parent]_n[[with]_p[[the]_{det}[?]_n]_{np}]_{pp}]_{vp}]_s$	#12
book	12	0-7	$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[parent]_n[[with]_p[[the]_{det}[book]_n]_{np}]_{pp}]_{vp}]_s$	#12
.	13	0-8	$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[parent]_n[[with]_p[[the]_{det}[book]_n]_{np}]_{pp}]_{vp}[\cdot]_s]$	#13

うかを判別する必要がある。そこで本手法では、語が入力されるごとに項の正しさを判別する。

まず、項の間の包含関係と項の正しさを定義する。項  $\sigma$  に対して項の置き換えの手続きを有限回実行することにより項  $\tau$  が得られるとき、 $\sigma$  は  $\tau$  を包含するといひ、 $\sigma \triangleright \tau$  と書く。これは、 $\tau$  が  $\sigma$  を具体化した項であることを意味する。例えば、項  $[[[I]_{pron}]_{np}[?]_{vp}]_s$  は項 (2.3) を包含している。次に、 $n$  語からなる入力文  $w_1 \dots w_n$  に対する項を  $\tau$  とするとき、 $\sigma \triangleright \tau$  ならば、 $w_1 \dots w_n$  に対して  $\sigma$  は正しいという。例えば、英語文 (2.1) に対して項 (2.3) は正しい。

次に  $i$  番目の語が入力されたときの項の正しさの判別法を以下に示す。

項の正しさの判別 チャートの節点 0 と節点  $i$  を結ぶ弧の項全体からなる集合を  $\mathbf{T}$  とする。また、チャートの節点 0 と節点  $j (j \leq i)$  を結ぶ弧の項を  $\sigma$  とする。このとき、任意の  $\tau \in \mathbf{T}$  に対して、 $\sigma \triangleright \tau$  であるならば、 $\sigma$  は正しい項である。

これにより入力途中の段階で項の正しさを判別することができる。例えば、英語文 (2.1) における語 “her” が入力された段階では、“I saw her” に対する項

$$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{pron}]_{np}]_{vp}]_s \quad (3.4)$$

$$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[?]_n]_{np}]_{vp}]_s \quad (3.5)$$

$$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[?]_n[?]_{pp}]_{np}]_{vp}]_s \quad (3.6)$$

は、項 (2.3) の最左未決定項をそれぞれ  $[[her]_{pron}]_{np}$ ,  $[[her]_{det}[?]_n]_{np}$ ,  $[[her]_{det}[?]_n[?]_{pp}]_{np}$  で置き換えることにより作成される。すなわち、項 (2.3) は (3.4) ~ (3.6) を包含するので、正しい項であるとわかる。実際、項 (2.3) は正しい。

上記の方法により、入力途中の段階で随時、項の正しさを判別し、これを即座に出力することにより、漸進的チャート解析において正しい項のみを出力することができる。

#### 4 解析例

英語文 (2.1) に対する漸進的な解析処理の過程と項の出力タイミングを表 1 に示す。各行はチャートの弧に対応してお

り、# の欄は弧の作成順、loc の欄は弧の始点と終点の対を表す。

語 “saw” が入力された時点では、正しい項を決定できないため、出力しない。語 “her” が入力された時点では、“I saw her” に対する正しい項を判別できない。しかし、3 節で述べたように、“I saw” に対する項 #4 は “I saw her” に対する項 #5, #6, #7 を包含する、すなわち正しい項であるので、#4 を出力する。語 “with”, “the”, “book”, “.” が入力された時点では、それぞれ項 #10, #11, #12, #13 が作成され、これらを出力する。このように、項が正しいとわかった段階で即座にこれを出力することができる。

#### 5 おわりに

本稿では、漸進的な構文解析において入力途中の段階で随時、構文構造の正しさを判別し、これを出力する手法を提案した。これにより、文法的に曖昧でない入力文を漸進的に、かつ正しく解釈することができる。本手法を曖昧な入力文に適用できるようにすることは今後の課題である。

#### 参考文献

- [1] 秋葉 友良, 田中 穂積: 拡張部分木を用いた漸進的構文解析, 情報処理学会第 45 回全国大会 (3), pp.175-176 (1992).
- [2] Inagaki, Y. and Matsubara, S.: Models for Incremental Interpretation of Natural Language, *Proc. of 2nd Symposium on Natural Language Processing*, 51-60 (1995).
- [3] Kay, M.: Algorithm Schemata and Data Structures in Syntactic Processing, *Technical Report CSL-80-12*, Xerox PARC (1980).
- [4] Matsubara, S., Asai, S., Toyama, K. and Inagaki, Y.: Chart-based Parsing and Transfer in Incremental Spoken Language Translation, *Proc. of 4th Natural Language Processing Pacific Rim Symposium*, pp.521-524 (1997).