

漸進的構文解析における構文構造の動的決定手法

加藤芳秀†

松原茂樹‡

外山勝彦†

稲垣康善†

†名古屋大学工学研究科情報工学専攻

‡名古屋大学言語文化部

{yoshihide,matu,toyama,inagaki}@inagaki.nuie.nagoya-u.ac.jp

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]。そ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 [6]。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、作成される構造は、以後の入力に対する解析に用いられることのない不適切な構造である可能性がある。そのような構造をもとに、意味解析や文脈解析を行い、発話内容を定めると、対話システムは相手の発話に対して不適切な応答をすることになる。これに対して我々は、語単位での構文構造の出力という制約を緩和し、適切な構文構造のみを出力するように出力タイミングを動的に決定する手法を提案した [4]。しかし、この手法では、曖昧でない入力文のみを処理の対象としていた。

そこで本稿では、曖昧な入力文に対しても適切な構文構造を漸進的に決定できる手法を提案する。本手法では、従来の手法に加えて、複数の構文構造に対して、それらの未決定部分の範疇を比較することにより、構文構造が適切であるかどうかを判別する。

本稿の構成は以下の通りである。次の 2 節では、従来の構文構造の動的決定手法について説明する。3 節では、曖昧な入力文に対する構文構造の動的決定手法について述べる。4 節では、解析例を示す。

2 従来の構文構造決定手法

本節では、これまでに我々が提案した構文構造決定手法 [4] について簡単に説明し、その問題点を明らかにする。この手法では、漸進的チャート解析 [6] に対して、構文構造が適切であるかどうかを判別する枠組を導入している。語が入力されるごとに、構文構造を作成し、それが適切であるかどうかを判別し、適切であれば即座にこれを出力する。

以下ではまず、漸進的チャート解析について説明し、次に、構文構造の適切さを判別する手法とその問題点について述べる。

2.1 漸進的チャート解析

漸進的チャート解析は、従来の上昇型チャート解析 [5] 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 [6]。ここで項とは、構文構造を表現するデータ構造である。通常のチャート解析については文献 [7]

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

図 1: 構文解析のための文法と辞書

を参照されたい。漸進的チャート解析では、 i 番目の語 w_i が入力されたとき、以下の手続きを順次実行する。なお、弧のラベルが項 σ であるとき、 σ をその弧の項と呼ぶ。

- 1) 辞書引き 語 w_i の範疇が X ならば、項 $[w_i]_X$ をラベルとして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 を結び、項 $[\dots]_X$ をラベルとしてもつ弧に対して、文法規則 $A \rightarrow XY \dots Z$ が存在するならば、項 $[[\dots]_X][?]_Y \dots [?]_Z]_A$ をラベルとして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 σ 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 τ の範疇が X ならば、 σ の最左未決定項を τ で置き換えた項をラベルとしてもつ弧をチャートの節点 0 と節点 i の間に追加する。

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。そのため、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 saw her aunt with the telescope. (1)

における語 “saw” が入力された段階では、“I saw” に対する項

$$[[[I]_{pron}]_{np}[[saw]_{vi}]_{vp}][?]_s]_s \quad (2)$$

$$[[[I]_{pron}]_{np}[[saw]_{vt}][?]_{np}][?]_{pp}]_{vp}][?]_s]_s \quad (3)$$

をそれぞれ作成することができ、この段階で動詞 “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]。

2.2 構文構造の動的な決定

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不適切なものが含まれている可能性がある。そのような項をもとに、意味解析や文脈解析を実行し、発話内容を定めると、システムは相手の発話に対して不適当な応答をすることになる。例えば、英語文 (1) において、語 “saw” が入力された段階で作成された項 (2) は、“saw” 以後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。

それに対して、文献 [4] の手法では、項の間の包含関係を用いて、作成された項が適切であるかどうかを判別する。これにより、漸進的な構文解析において、適切な項のみを出力することができる。

まず、項の間の包含関係を定義する。

定義 2.1 (包含関係) 項 σ, τ に対して、ある項 θ が存在し、 σ の最左未決定項を θ で置き換えることにより得られる項が τ に等しいとき、 $\sigma \triangleright \tau$ と書く。さらに、 \triangleright^* を \triangleright の反射推移閉包とする。 $\sigma \triangleright^* \tau$ であるとき、 σ は τ を包含するといふ。

σ が τ を包含するとは、 τ が σ を具体化した項であることを意味する。例えば、項

$$[[[I]_{pron}]_{np}[[?]_{vp}]]_s \quad (4)$$

は項 (2) 及び (3) を包含している。

次に、項の適切さを定義する。

定義 2.2 (項の適切さ) 入力文 $w_1 \cdots w_n$ に対して項 σ が適切であるとは、次の 1. ~ 3. を満たす項 τ が存在するときである。

1. $\sigma \triangleright^* \tau$.
2. τ は $w_1 \cdots w_n$ に対する項である。
3. τ は未決定項を含まない。

$w_1 \cdots w_n$ に対して σ が適切であるとは、項 σ に対して置き換え手続きを有限回実行することにより、入力文 $w_1 \cdots w_n$ に対する項が作成されることを意味する。例えば、項 (3) に対して置き換え手続きを実行することにより、最終的に、英語文 (1) に対する項が得られるので、文 (1) に対して項 (3) は適切である。また、項 (2) に対して置き換え手続きを実行しても、文 (1) に対する項は得られない、すなわち、文 (1) に対して項 (2) は適切でない。

i 番目の語 w_i が入力されたときの項の適切さの判別法を以下に示す。

定理 2.3 (項の適切さの判別) 入力文 $w_1 \cdots w_n$ に対して、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 σ 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、 $w_1 \cdots w_n$ に対して σ は適切な項である。

定理 2.3 により、項 σ が \mathbf{T} のすべての項を包含するとき、 σ は適切であると判定できる。 \mathbf{T} には入力文に対する項を包含する項が必ず存在するが、 σ はそのような項を必ず包含する。よって、包含関係の推移性により、 σ も入力文に対す

る項を包含することが保証され、 σ は適切であると決定できる。

例えば、英語文 (1) における語 “her” が入力された段階では、“I saw her” に対する項として、

$$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{pron}]_{np}]]_{pp}]]_s \quad (5)$$

$$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}]]_n]]_{pp}]]_s \quad (6)$$

の 2 つが作成されるが、これらは、項 (3) の最左未決定項をそれぞれ

$$[[her]_{pron}]_{np} \quad (7)$$

$$[[her]_{det}]]_n]]_n \quad (8)$$

で置き換えることにより作成される。すなわち、項 (3) は (5) 及び (6) を包含するので、英語文 (1) に対して適切な項であるとわかる。

上記の方法により、入力途中の段階で随時、項の適切さを判別し、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適切な項のみを出力することができる。しかし、ここで注意すべき点は、あらゆる適切な項を定理 2.3 の判別法でもって適切であると判定できるわけではないということである。すなわち、定理 2.3 において、項が適切であると判定できるのは、適切な項が一つだけ存在する場合のみであり、複数の適切な項が存在するときには、これらの項を適切であると判定できない。実際、項 σ, σ' が適切であるとすると、 σ は \mathbf{T} のある項 τ を包含するが、このとき σ' は τ を包含しない。逆もまた同様である。したがって、 σ, σ' のいずれも、 \mathbf{T} のすべての項を包含することはない。そのため、定理 2.3 では、適切な項が複数存在する曖昧な入力文に対して、それらの項が適切であると判定することはできない。

例えば、図 1 の文法において、英語文 (1) は曖昧でなく、前置詞句 “with the telescope” が動詞 “saw” にかかるとする構文構造のみが考えられる。しかし、文法規則

- $vp \rightarrow vt \ npl$
- $npl \rightarrow det \ n \ pp$

を加えることにより、前置詞句 “with the telescope” が名詞 “aunt” にかかるとする構文構造も考えられる。すなわち、英語文 (1) は曖昧になるが、この場合、語 “her” が入力された段階では、項 (5), (6) に加えて、項

$$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}]]_n]]_{pp}]]_{npl}]]_s \quad (9)$$

も作成される。これらの項の中で、英語文 (1) に対して適切な項は、(6) 及び (9) である。文 (1) に対しては、(A) 前置詞句 “with the telescope” が動詞 “saw” にかかるとする項、(B) 前置詞句 “with the telescope” が名詞 “aunt” にかかるとする項、の 2 つが作成されるが、(6) は項 (A) のみを、(9) は項 (B) のみを包含し、“I saw her” に対する項の中に (A), (B) を共に包含する項は存在しない。したがって、定理 2.3 の方法では、これらの項が適切であると判定できず、結局、入力が完了した時点においてもこれらの項を出力することができない。

この問題を解決するためには、適切な項が複数存在する場合においても、それらが適切であるかどうかを判別する方法が必要となる。

3 曖昧な入力文に対する構文構造の決定手法

本節では、曖昧な入力文に対しても適切な構文構造を決定できるように、2.2 節で述べた手法を拡張する。項の適切さは以後の入力に応じて定まるが、本稿ではこの点に着目し、複数の項をまとめることにより、曖昧な入力文に対する項の適切さの判別を実現する。

まず、項の未決定範疇列を定義する。

定義 3.1 (項の未決定範疇列) 項 σ に出現する未決定項を左から順に並べたものを τ_1, \dots, τ_m とする。また、 τ_k ($1 \leq k \leq m$) の範疇を X_k とする。このとき、 X_1, \dots, X_m を σ の未決定範疇列という。□

項 σ の未決定範疇列が X_1, \dots, X_m であるとは、範疇がそれぞれ X_1, \dots, X_m である構成要素がそれ以後順に入力されたとき、 σ が適切な項となることを意味する。例えば、“I saw her” に対する項 (6) の未決定範疇列は $n, pp, \$$ であるが、範疇がそれぞれ $n, pp, \$$ である構成要素、例えば、“aunt”, “with the telescope”, “.” が以後この順で入力されたとき、(6) は適切な項となる。

項 σ, σ' の未決定範疇列が一致するとき、項の適切さも一致する。なぜなら、 σ, σ' の未決定範疇列をともに X_1, \dots, X_m とすると、範疇が X_1, \dots, X_m である構成要素が順に入力されたとき、 σ, σ' はどちらも適切となり、そうでないときはどちらも適切とならないからである。例えば、項 (6), (9) の未決定範疇列はともに $n, pp, \$$ であるが、以後に $n, pp, \$$ であるような構成要素が入力されたときにはどちらも適切となり、そうでないときはどちらも適切とならない。以下では、記法 $\sigma \equiv \sigma'$ で、 σ と σ' の未決定範疇列が等しいことを表す。

i 番目の語 w_i が入力されたときの項の適切さの判別法を以下に示す。

定理 3.2 (項の適切さの判別) 入力文 $w_1 \dots w_n$ に対して、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 σ とする。このとき、任意の $\tau \in \mathbf{T}$ に対して、 $\sigma \triangleright^* \tau$ 、かつ $\tau' \equiv \tau$ である $\tau' \in \mathbf{T}$ が存在するならば、 $w_1 \dots w_n$ に対して σ は適切な項である。□

定理 3.2 により、 \mathbf{T} の任意の項 τ に対して、ある項 $\tau' \in \mathbf{T}$ が存在して、 $\sigma \triangleright^* \tau'$ 、及び $\tau \equiv \tau'$ が成り立つとき、すなわち、図 2 に示すような関係が成り立つとき、項 σ は適切であると判定できる。

例えば、英語文 (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 (5), (6) 及び (9) が作成される。(3) \triangleright^* (5), (3) \triangleright^* (6) であり、(6) \equiv (9) であるので、(3) は適切な項であるとわかる。さらに、語 “aunt” が入力された段階では、“I saw her aunt” に対して、項

$$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[aunt]_n]_{np}[pp]_{vp}[?]\$]]_s \quad (10)$$

$$[[[I]_{pron}]_{np}[[saw]_{vt}[[her]_{det}[aunt]_n][pp]_{np1}[vp]_{vp}[?]\$]]_s \quad (11)$$

が作成される。(10) \triangleright^* (10), かつ (10) \equiv (11) であるので、(10) は適切な項であるとわかる。同様に、項 (11) も適切な項であるとわかる。このように、複数の適切な項が存在するとき、これらを適切であると判定することができる。

上記の方法を用いて、入力途中の段階で項の適切さを判別し、これを即座に出力することにより、曖昧な入力文に対しても適切な項のみを出力することができる。

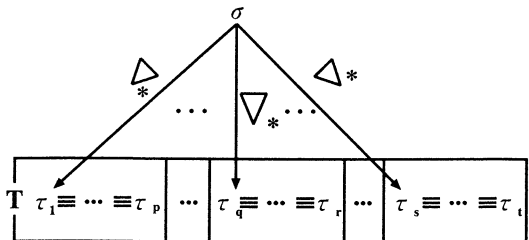


図 2: 適切な項の判別

4 解析例

英語文 (1) に対する漸進的な解析処理の過程を表 1 に示す。各行がチャートの弧に対応しており、# の欄の数は弧の作成順序、loc は弧の場所、term は弧がラベルとしてもつ項を表す。また、表 2 に項の出力タイミングを示し、図 3 に作成された項の間の包含関係 \triangleright^* 、及び同値関係 \equiv を示す。

語 “saw” が入力された時点では、適切な項を決定できないため、項を出力しない。また、語 “her” が入力された時点では、“I saw her” に対する適切な項を判別することができない。しかし、“I saw” に対する項 #4 は、“I saw her” に対する項 #6, #7 を包含し、#7 \equiv #8 である、すなわち適切な項であるので、#4 を出力する。語 “aunt” が入力された時点では、項 #9, #10 を作成する。3 節で述べたように、#9 \triangleright^* #9, かつ #9 \equiv #10 である、すなわち適切な項であるので、項 #9 を出力する。同様に、項 #10 も適切な項であると分かるので、これも出力する。語 “with” が入力された時点では項 #11 及び #12 が、“the” が入力された時点では #13 及び #14 が、“telescope” が入力された時点では #15 及び #16 が、“.” が入力された時点では #17 及び #18 が、それぞれ作成され、それらを出力する。このように、曖昧な入力文に対しても、入力途中の段階で随時、項が適切であるかどうかを判別することができる。

5 おわりに

漸進的な構文解析では一般に、語が入力されるたびに構文構造が作成されるが、その構文構造は以後の解析に用いられない不適切な構造である可能性がある。このため、語単位での出力という制約を緩和し、そのタイミングを動的に決定する方法の実現が重要な課題となっている。それに対して本稿では、漸進的な構文解析において、入力途中の段階で随時、構文構造の適切さを判別し、適切な構文構造のみを出力する手法を提案した。構文構造の未決定範疇列を比較することにより、曖昧な入力文に対しても、構文構造が適切であるかどうかを判別できることを例をもって示した。これは、漸進性よりもむしろ項の適切さを重視し、構文構造が適切であると判明するまで出力を遅延する手法であるといえる。

本手法では、構文構造が以後の解析に用いられる適切な構造であることが保証された段階で、これを出力する。このため、場合によっては入力が完了するまで構文構造を出力できず、漸進性を損なうことがある。漸進性の度合と構文構造の適切さの間にはトレードオフが存在すると考えられ、構文構造の適切さに加えて漸進性の度合も考慮し、出力タイミングを決める必要があるが、これは今後の課題である。

表 1: “I saw the aunt with the telescope” に対する漸進的な解析過程

入力語	チャート			未決定 範疇列
word	#	loc	term	
	1	0-0	[?] _s	s
I	2	0-1	[[[I] _{pron}] _{np}][?] _{vp} [_] _s	vp, \$
saw	3	0-2	[[[I] _{pron}] _{np}][saw] _{vi} [_] _s	\$
	4	0-2	[[[I] _{pron}] _{np}][saw] _{vt} [_] _{np} [_] _{pp} [_] _s	np, pp, \$
	5	0-2	[[[I] _{pron}] _{np}][saw] _{vt} [_] _{np1} [_] _s	np1, \$
her	6	0-3	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{pron} [_] _{pp} [_] _s	pp, \$
	7	0-3	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [_] _n [_] _{pp} [_] _s	n, pp, \$
	8	0-3	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [_] _n [_] _{pp} [_] _{np1} [_] _s	n, pp, \$
aunt	9	0-4	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [aunt] _n [_] _{pp} [_] _s	pp, \$
	10	0-4	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [aunt] _n [_] _{pp} [_] _{np1} [_] _s	pp, \$
with	11	0-5	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [aunt] _n [_] _{pp} [_] _{with} [_] _{np} [_] _s	np, \$
	12	0-5	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [aunt] _n [_] _{pp} [_] _{with} [_] _{np} [_] _{pp} [_] _{np1} [_] _s	np, \$
the	13	0-6	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [aunt] _n [_] _{pp} [_] _{with} [_] _p [_] _{the} [_] _{det} [_] _n [_] _{pp} [_] _s	n, \$
	14	0-6	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [aunt] _n [_] _{pp} [_] _{with} [_] _p [_] _{the} [_] _{det} [_] _n [_] _{pp} [_] _{np1} [_] _s	n, \$
telescope	15	0-7	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [aunt] _n [_] _{pp} [_] _{with} [_] _p [_] _{the} [_] _{det} [telescope] _n [_] _{pp} [_] _s	\$
	16	0-7	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [aunt] _n [_] _{pp} [_] _{with} [_] _p [_] _{the} [_] _{det} [telescope] _n [_] _{pp} [_] _{np1} [_] _s	\$
.	17	0-8	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [aunt] _n [_] _{pp} [_] _{with} [_] _p [_] _{the} [_] _{det} [telescope] _n [_] _{pp} [_] _s]	
	18	0-8	[[[I] _{pron}] _{np}][saw] _{vt} [[her] _{det} [aunt] _n [_] _{pp} [_] _{with} [_] _p [_] _{the} [_] _{det} [telescope] _n [_] _{pp} [_] _{np1} [_] _s]	

表 2: 構文構造の出力タイミング

入力語	出力された構文構造
I	#2
saw	
her	#4
aunt	#9, #10
with	#11, #12
the	#13, #14
telescope	#15, #16
.	#17, #18

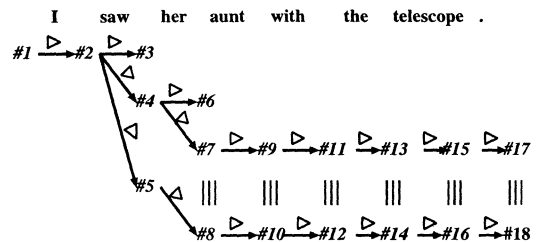


図 3: 項の間の包含関係 ▷* と同値関係 ≡

また本手法では、語が入力される毎に構文構造が少なくとも1つ必ず作成されると仮定しており、作成された構造の間の関係から構文構造の適切さを判別している。すなわち、文法的に適格な入力文のみを対象としていたが、対話には文法的に不適格な文が多数出現するため、これに対処する必要がある。著者らはこれまでに頑健な漸進的構文解析手法を提案しているが[3]、これに対して本稿で提案した手法を適用することにより、文法的不適格文に対しても、入力途中の段階で適切な構文構造のみを出力する枠組の実現が期待できる。

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] 加藤 芳秀, 松原 茂樹, 外山 勝彦, 稲垣 康善: 文法的不適格文に対する漸進的構文解析手法, 電子情報通信学会技術研究報告, NLC98-6, pp.39-46(1998).
- [4] 加藤 芳秀, 松原 茂樹, 外山 勝彦, 稲垣 康善: 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, 情報処理学会第57回全国大会(2) pp.203-204(1998).
- [5] Kay, M.: Algorithm Schemata and Data Structures in Syntactic Processing, *Technical Report CSL-80-12*, Xerox PARC (1980).
- [6] 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).
- [7] 田中 穂積: 自然言語解析の基礎, 産業図書 (1989).