

漸進的な話し言葉翻訳のためのチャート解

松原 茂樹

浅井 悟

外山 勝彦

稲垣 康善 (名古屋大学大学院)

1 はじめに

計算機を用いた効率的な英語日本語間対話翻訳の実現のため、著者は英語話者の発話とほぼ同時進行的に日本語を生成する漸進的な話し言葉翻訳システムについて研究を進めている [2]。一般に機械翻訳システムは、(1) 原言語文の解析、(2) 原言語構造から目標言語構造への変換、及び (3) 目標言語の生成、という3つのステージから構成されるが、漸進的な翻訳においてもこれらを語単位で順次実行することが考えられる。しかし、これまでに提案されている解析手法の多くは、文単位での処理を基本としており、それらを漸進的な翻訳処理に用いるのは適当ではない。本稿では、単語が入力される度にその時点までの原言語構造を生成することが可能な漸進的解析手法について述べる。

2 チャート：大域チャートと局所チャート

本研究では、解析結果の表現形式としてチャート (*chart*) [1] を導入する。チャートは、統語解析の途中経過を表現するデータ構造であり、通常ラベル付きグラフとして定義される。ここでラベルは項 (*term*) と呼ばれる統語解析木である。例えば、範疇 v の単語 “met” に対する項は、範疇が名詞句である未決定項 $[?]_{np}$ を用いて $[[met]_v [?]_{np}]_{vp}$ と表すことができる。なお、未決定項を含む項をラベルとしてもつ弧 (*edge*) を活性弧 (*active edge*)、そうでない弧を不活性弧 (*inactive edge*) という。部分的な表現に対してもその統語構造を活性弧として表現できるため、チャートを活用することは漸進的な解析処理において有効である。

チャートは、漸進的解析の途中で導出されるすべての弧を表現するが、変換処理で活用されるものはその一部である。本研究では、変換処理で用いられる弧からなるチャートを大域チャートと呼び、解析処理のみで用いられる弧からなる局所チャートと区別する。大域チャートは、単語が入力されたその時点までの統語構造を表す項をもつ弧のみからなる。

3 漸進的チャート解析手法

漸進的チャート解析は、大域チャートの節点 0 と 0 を結ぶ項 $[?]_s$ をもつ弧を初期値として、単語が入力される度に弧を大域チャートに追加する。単語の入力に対して追加すべき弧が存在しなければ解析失敗である。 k 番目の単語 w が入力されたときの解析手続きを以下に示す。なお、記法 $\alpha \rightsquigarrow \sigma$ で範疇 α が項 σ の最左未決定項の範疇に到達可能であることを表す。

【漸進的チャート解析手続き】

(手続き 1) 辞書引き w の範疇 α が大域チャートの節点 0 と $k-1$ を結ぶ活性弧の項 σ に対して $\alpha \rightsquigarrow \sigma$ なら項 $[w]_\alpha$ の不活性弧を局所チャートの節点 $k-1$ と k の間に追加する。

(手続き 2) 解析規則の適用 局所チャートの節点 $k-1$ と k を結ぶ弧の項 ϕ の範疇を β_1 とする。このとき、大域チャートの節点 0 と $k-1$ を結ぶ活性弧の項 σ に対して $\beta \rightsquigarrow \sigma$ である規則 $\beta \rightarrow \beta_1 \beta_2 \dots \beta_n$ が存在するならば、項 $[\phi [?]_{\beta_2} \dots [?]_{\beta_n}]_{\beta}$ をもつ活性弧を局所チャートの節点 $k-1$ と k の間に追加する。この操作を適用可能な限り繰り返す。

(手続き 3) 項の置き換え 大域チャートの中の節点 0 と $k-1$ を結ぶ活性弧の項を σ とし、その最左未決定項を $[?]_\gamma$ とする。このとき、局所チャートの節点 $k-1$ と k を結ぶ弧の項

表 1: 漸進

入力 <i>word</i>	局所チャート		
	#	<i>loc</i>	<i>term</i>
Ken	2	0-1	$[Ken]_{np}$
	3	0-1	$[[Ken]_{np} [?]_{vp}]_s$
met	5	1-2	$[met]_v$
	6	1-2	$[[met]_v [?]_{np}]_{vp}$
her	8	2-3	$[her]_{np}$
	9	2-3	$[her]_{da}$
baby	10	2-3	$[[her]_{da} [?]_n]_{np}$
	13	3-4	$[baby]_n$

ϕ の範疇が γ なら、 $[?]_\gamma$ を項 ϕ で置き換えて、局所に追加する。

通常の上昇型チャート解析に対して、活性弧の項の中の未決定項を別の項に置き換える操作は、文法的に適切な入力文に対しては、力に対する弧を少なくとも一つ導出する。

英語文 “Ken met her baby” に対する局所チャートの生成順序を、*loc* は弧が結ぶ節点を示す。語の入力順序に対しては #11 と #12 という複数の弧が導出される。一方を選択しそれに対してのみ変換処理を行う。

4 まとめ

漸進的な解析、及び変換処理のためのチャート解析を定義し、単語が入力される度に大域チャートに追加する。本手法は、漸進的な英日話し言葉翻訳に有効である。

参考文献

- [1] Kay, M.: Algorithm Schemata and Incremental Report CSL-80-12, Xerox PARC.
- [2] Matsubara, S. and Inagaki, Y.: Utilizing Incremental English-Japanese Machine Translation.
- [3] Matsubara, S. et al.: Incrementally Generated Partially Ill-formed Expressions, *Proceedings of the 1991 Conference on Computational Linguistics*.