話し言葉における文法的不適格文に対する漸進的翻訳手法

3J－10

加藤 芳秀　松原 茂樹　浅井 悟　外山 勝彦　稲垣 康善

名古屋大学大学院工学研究科

1　はじめに

計算機による効率的な対話翻訳の実現のために，原言語の入力に対して同時進行的に目標言語を出力する漸進的翻訳システムが必要である [2]．著者らはこれまでに，漸進的な解析・変換に基づく英日話し言葉翻訳手法を提案している [1][4]．この手法では，英単語が入力されるごとにチャート法を用いて統語構造を構築し，この統語構造に変換規則を適用することにより漸進的に日本語翻訳文を生成する．

ところで，一般に話し言葉には，間投詞などの挿入や語句の脱落，置換など，文法的に不適格な現象が多数出現するが，これまでに著者らが作成したシステム [1] では，あらかじめ間投詞などを削除した英語文を入力する，あるいは，脱落や置換を許容する文法規則を用いるなどの方法で文法的不適格性に対処していた．しかし，このような対処法は実際の話し言葉翻訳において現実的ではない．また，Mellish[5] や加藤 [3] はチャート法に基づく不適格文解析手法を提案しているが，これらは一文全体が入力された時点で誤り修正を実行するものであり，漸進的な翻訳には適さない．

本稿では，文法的不適格文に対して，解析に失敗した時点で即座に誤り修正を行うことにより漸進的に統語構造を構築し，この統語構造に変換規則を適用することにより漸進的に翻訳結果を作成する手法を提案する．文中の適格な部分に対しては正しく統語構造を構築できるため，不適格文であっても，その意味内容を翻訳結果としてある程度再現することが可能である．翻訳実験の結果，従来の手法で 48.9% であった翻訳正解率が 60.8% に向上しており，本手法の有効性を確認した．

2　漸進的な英日話し言葉翻訳システム

漸進的な英日話し言葉翻訳システムは，日本語話し言葉に頻繁に現れる，繰り返し，語順の逆転，省略等の不適格表現を積極的に活用することにより，語順の異なる言語間において入力に対する同時進行的な出力を実現している [4]．例えば，英語文

(2.1) He met her in the park yesterday.

を翻訳する場合，入力 “He”, “met”, “her”, “in the park”, “yesterday”, に対してそれぞれ「彼は」，「会った」，「彼女に」，「公園で」，「昨日」を同期的に出力し，最後に「投げた」を再度出力する．得られた翻訳文

(2.2) 彼は会った．彼女に公園で昨日会った．

は，動詞の繰り返しや必須格の省略，語順の逆転など文法的には不適格な表現を含んでいるものの，日本語ユーザはその意味内容を正しく理解できる．

Incremental Translation of Grammatically Ill-formed Sentences
Yoshihide Kato, Shigeki Matsubara, Satoru Asai,
Katsuhiko Toyama and Yasuyoshi Inagaki (Nagoya University)

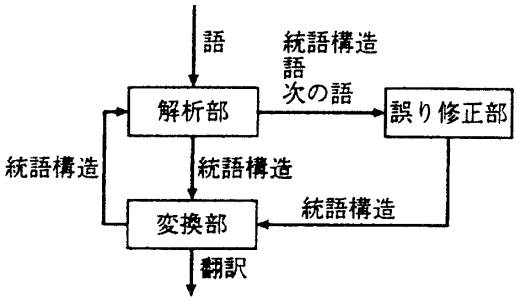

図 1: 不適格文に対する漸進的翻訳手法の構成

このようにシステムは，翻訳結果の作成に文法的不適格表現を活用するものの，逆にそのような表現を含む入力に対して適切に対処することができない．すなわち，余分な語が挿入されている英語文に対しては，あらかじめその語を人手により削除した文を入力としている．例えば，間投詞 “ah” が挿入された英語文

(2.3) Ah, that sounds interesting.

に対しては，“ah” を削除した英語文

(2.4) That sounds interesting.

を入力として用いる．また，必要な語句の脱落や他の語句への置換を含む英語文に対しては，それを許容する文法規則を用いて解析を実行する．例えば，英語文

(2.5) By train is best.

は，本来の主語 “going by train” の “going” が脱落した文であると考えられるが，システムは “by train” を名詞句と定めることが可能な文法規則を用いている．しかし，このような対処法は話し言葉翻訳において現実的ではない．より実用的な漸進的話し言葉翻訳システムの構築のため，余分な語句の削除，脱落した語句の補完，置換された語句の復元などの処理をできる限り早い段階で実行し，変換処理を実行する手法が必要である．

3　不適格文に対する漸進的翻訳手法

本稿で提案する不適格文に対する漸進的翻訳手法の構成を図 1に示す．解析部と変換部，及び誤り修正部から構成されており，語が入力されるたびに解析処理ならびに変換処理を実行することにより，漸進的に翻訳結果を作成する．

3.1　不適格文に対する漸進的解析手法

解析部は英単語が入力されるごとにチャート法を用いて統語構造を構築する．統語構造の構築に失敗するとさらに次の語を読み込み，現在までに構築されている統語構造，解析失敗した語，及び新たに入力された語を誤り修正部に渡す．本手法で扱う誤りの種類は，余分な語の挿入 (挿入誤り)，必要な語の脱落 (脱落誤り)，別の語への置換 (置換誤り) であり，修正を行った結果，得られた統語構造に対して変換処理を行う．それぞれの誤りに対する修正方法は以下の通りである．

表 1: 97 文に対する間違った誤り修正の割合

誤り修正の種類	間違って修正された文の割合
脱落誤り修正	16.5%
挿入誤り修正	17.5%
置換誤り修正	61.9%

表 2: 278 文に対する翻訳正解率

誤り修正		文数	割合 (%)
不要	翻訳成功	136	48.9
	翻訳失敗	45	16.2
必要		97	34.9

挿入誤り修正 解析失敗した語を削除

脱落誤り修正 解析失敗した語の直前に適切な語を挿入

置換誤り修正 解析失敗した語を別の語で置換

例えば, 英語文 (2.3) に対して, 入力 "ah" で解析失敗するが, 挿入誤り修正によりこれを削除し, (2.4) に対する統語構造を構築する. また, 英語文 (2.5) の解析では, 入力 "by" に対して解析失敗するが, 脱落誤り修正により "by" の前に, 例えば動名詞を挿入する.

しかし, 英語文 (2.5) に対して挿入誤り修正を行い,

(3.1) Train is best.

とすることも考えられ, どの誤り修正を優先するかが問題となる. 一般には, 間違った修正を行う可能性が低い誤り修正手法を優先することが望ましい. そこで, 文法的に不適格な 97 文に対して各誤り修正が間違って修正する文の割合を調べたところ, 表 1に示すような結果が得られた. このことから, 本手法では脱落誤り修正, 挿入誤り修正, 置換誤り修正の順に誤り修正処理を行うこととし, ある誤り修正に成功したらそれ以外の誤り修正を試みないこととした. すなわち, 英語文 (2.5) に対しては, まず脱落誤り修正を行い, その修正は成功するので他の修正は行わない.

3.2 不適格文に対する漸進的変換手法

変換部では, 解析部が統語構造を構築するたびに, それに変換規則を適用することにより翻訳結果を作成する. しかし, 脱落誤り修正, 置換誤り修正においてはそれぞれ, 挿入する語, 置き換える語のカテゴリだけしか求めることができないため, これらの語に対しては変換処理を実行しない. 例えば, 英語文 (2.5) に対する翻訳結果は

(3.2) 電車でが一番いいです

となる. しかし, 日本語ユーザはこの意味内容を正しく理解することができる.

4 評価

本手法の有効性を調べるため, Common Lisp を用いてプロトタイプシステムを作成し翻訳実験を行った. 実験には, ATR 対話データベースの旅行申し込み対話のうち, 英語話者による発話 278 文を用いた. なお, 文法は規則数 94, 終端記号数 40, 非終端記号数 38, 辞書は登録単語数 391 である.

表 3: 97 文に対する誤り修正の成功率

誤り修正の結果	文数	割合 (%)
誤り修正成功	88	90.7
誤り修正失敗	9	9.3

表 4: 誤り修正された英語文 88 文に対する翻訳正解率

項目	文数	割合 (%)
翻訳成功	33	37.5
翻訳失敗	55	62.5

実験結果を表 2に示す. 181 文は適格文であり, 誤り修正を行うことなく解析処理に成功し変換部に入力された. 一方, 残りの 97 文は誤り処理部が処理を行った. その結果, 表 3 に示すように, 278 文全体の 31.7% に相当する 88 文について新たに統語構造を構築でき, その構造は変換部に入力された. 翻訳結果を表 4に示す. 誤り修正を行なわない場合, 翻訳成功率は 48.9% であったが, 誤り修正を行うことにより 278 文全体の 11.9% に相当する 33 文について新たに翻訳に成功し, 翻訳正解率は 60.8% に向上した. 以上の実験により, 本稿で述べた不適格文翻訳手法が翻訳精度の向上に有効であることが分かった. しかし, 誤り修正に成功した 88 文の 62.5% に相当する 55 文が翻訳に失敗している. その主な理由として, 88 文のうち 39 文が間違って修正されていることが挙げられる. より精度の高い不適格文翻訳の実現のため, 今後はより正確な誤り修正手法を考案する必要がある.

5 おわりに

本稿では, 不適格文の入力に対してできる限り早い段階で, 語の削除, 語の挿入, 語の置換といった誤り修正を行うことにより, 不適格文の意味内容をある程度再現可能な漸進的翻訳手法を提案した. 翻訳実験を行った結果, 本手法の有効性を確認した.

参考文献

[1] 浅井 悟, 松原 茂樹, 稲垣 康善, 外山 勝彦: 言い直しを用いた漸進的な英日翻訳手法, 人工知能学会第 11 回全国大会, 360-363 (1997).

[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 恒昭 : 一般化弧を用いた A* 探索による非文の解析, 情報処理学会論文誌, Vol.36, No.10, 2343-2352 (1995).

[4] Matsubara, S. et al.: Incremental Spoken Language Translation Utilizing Grammatically Ill-formed Expressions, *Proc. of 3rd Conference of Pacific Association for Computational Linguistics* (1997).

[5] Mellish, C.S.: Some Chart-Based Techniques for Parsing Ill-Formed Input, *Proc. of 27th Conference of Association for Computational Linguistics*, 102-109 (1989).