論　文

不適格表現を活用する漸進的な英日話し言葉翻訳手法

非会員　松原茂樹　(名古屋大学)
非会員　浅井悟　(名古屋大学)
非会員　外山勝彦　(中京大学)
正員　稲垣康善　(名古屋大学)

Incremental English-Japanese Spoken Language Translation Utilizing Ill-formed Expressions

Shigeki Matsubara, Non-member, Satoru Asai, Non-member (Nagoya University), Katsuhiko Toyama, Non-member (Chukyo University), Yasuyoshi Inagaki, Member (Nagoya University)

This paper proposes a system for incremental English-Japanese spoken language translation, the system behaving like simultaneous interpretation. Since spontaneous speech appears continuously, each stage of a machine translation system for spoken languages should work incrementally. However, machine translation systems which have been proposed so far can not achieve high degree of incrementality because of the difference in word-order between the source language and the target one. In this paper, the system utilizing some ill-formed expressions characterizing Japanese spoken language such as repetitions, inversions, ellipses, repairs and hesitations, we have succeeded in the incremental transfer from English to Japanese. An experimental system on the basis of a chart parsing framework has been implemented. To evaluate the system we have made an experiment with 278 English dialogue sentences. 228 of them are translated correctly, providing a success rate of 82.0%, and 47msec on the average was taken to process one word. These results show our technique to be useful for spoken language translation with acceptable accuracy and high real-time nature.

キーワード：機械翻訳，話し言葉翻訳，漸進的解釈，文法的不適格性，言い直し，同時通訳

1. まえがき

機械翻訳を用いた多言語間コミュニケーションの実現のために，効率的な対話翻訳システムの開発が望まれている。対話翻訳システムは，異なる言語を使用するユーザの間で，各々が自分の言語を使って対話を遂行することを可能にするものであり，その処理対象はユーザによって生成された話し言葉である。話し言葉は書き言葉と異なり，出現が時間的に連続的であるため，翻訳処理もまた実時間で行われることが要求される。ところが，これまでに提案された機械翻訳手法の多くは書き言葉を対象とするものであり，原言語文から目標言語文への文単位での変換処理が実現されている。そのような翻訳手法を話し言葉に適用すると，原言語文が完全に入力された後でないとシステムは翻訳処理を開始することができず，結果として対話の結束性が大きく損なわれることになる。これは，従来の機械翻訳手法の多くが，話し言葉翻訳システムとして適当ではないことを意味し，高度実時間対話翻訳システムの実現のためには，目標言語文の生成が原言語文の入力とほぼ同時進行的に実行される漸進的な翻訳処理手法が必要となる[14]。それに対して，英日同時通訳モデルにおける漸進的な文生成手法[9]がKitanoによって，また，独英機械翻訳におけるHead Swiching処理のためのチャートを用いた方法[1]がAmtrupによって提案されている。これらはいずれも，翻訳の対象として話し言葉を想定しており，できる限り早い段階で翻訳結果を出力することを目的としている。

一方，そのような機械翻訳を実現するための解決すべき問題の一つとして，原言語と目標言語との間の語の生起順序に関する違いに対する処理が挙げられる。例えば，英語を日本語に漸進的に変換する場合，

（1） 英語では，動詞が文の比較的早い段階で出現するのに対して，日本語では動詞が最後に現れる。このため，英語の動詞が入力されたときに，対応する日本語を即座に生成することが難しくなる。

（2） 英語における任意格表現の生起順序は，日本語の

場合と逆行するという傾向がある。よって，すべての任意格が入力された後でないとそれらの翻訳処理を実行できない可能性がある。

などの問題が生じる。しかし，Kitano[9]やAmtrup[1]では，両言語間の語の生起順序の違いに関する問題には立ち入っておらず，原言語文の入力と目標言語文の生成とが同期する翻訳処理を実現するものではない。

それに対して本論文では，原言語文の入力に対してほぼ同時進行的に目標言語文を生成することが可能な手法として，日本語話し言葉の不適格表現を活用する漸進的な英日機械翻訳手法を提案する。話し言葉は一般に，書き言葉には見られないさまざまな不適格表現が頻繁に出現するとして特徴付けられるが[3]，人間はその高度な発話理解能力により，実際の対話に現れる不適格な発話を容易に理解することができる。そこで，日本語話し言葉の不適格性をむしろ積極的に容認し，原文の意味内容を表現する日本語文の形態を通常よりも豊富にすれば，入力された英語文に似た語の生起順序をもつ日本語翻訳文の生成が可能になると考えられる。すなわち，そのような日本語文を漸進的な翻訳結果として用い，それを入力と同時進行的に生成することが本研究の基本的な考え方である[12]。したがって，本手法は，高度な解析・変換処理により翻訳結果の品質の高さを追求するものではなく，原文の意味内容をユーザが正しく理解できさえすればよいという観点のもとで，実時間での翻訳処理を実現する。本手法に基づく翻訳システムは，いわゆる同時通訳のように振る舞うため，ユーザの待ち時間が大幅に短縮される効率的な対話翻訳システムの実現が期待できる。

本論文では，話し言葉に対する翻訳結果として，さまざまな不適格表現のうち特に繰り返し，語順の逆転，省略，言い誤り，言い直し，言い淀みを含む発話を生成することを容認することにより，英語文の入力とほぼ同期的に進行する日本語文出力を実現できることを示す。また本手法では，途中までの入力に対する漸進的な解析結果を逐次表現するために，チャートに基づく解析手法[8]を導入する。さらに，本論文で提案した翻訳手法の有効性を評価するために実験システムを実装し，ATR対話データベース[4]を用いた翻訳実験を行った。その結果，受理可能な翻訳精度をもつ効率的な話し言葉翻訳の実現のために本手法が有効であることを確認した。

本論文の構成は以下の通りである。まず2章で，漸進的な英日翻訳を実現する際に生じる語の生起順序に関する問題について述べる。3章では，日本語会話文に現れる不適格表現とそれらを活用した翻訳手法について例を用いて説明する。4章では，チャートを用いた漸進的な翻訳方法を示す。5章では，プロトタイプシステムによる翻訳実験とその評価について述べる。6章で不適格表現のうち言い直しを活用することについての考察を与える。

2. 語の生起順序と漸進的な翻訳処理

話し言葉を漸進的に翻訳するとは，それが入力される順序に従ってできる限り語単位で翻訳結果を作り上げ，それを即座に出力することをいう。しかし，英語と日本語の場合，語の生起順序が互いに大きく異なるため，英語を日本語に漸進的に翻訳することは一般には困難である。そのことを具体的に説明するために，次の英語文(2.1)に対して，その日本語翻訳文をできる限り早い段階で生成することを試みる。

(2.1) Ken could meet her in the park near the school yesterday.

この文に対する典型的な日本語翻訳文は

(2.2) 健は，昨日，学校の近くの公園で彼女に会えた。

である。これらの文の間の各構成要素の対応関係を考慮した上で，(2.2)をできる限り早い段階で生成すると，その出力のタイミングは図1のようになる。図1における上から下への順序は，入力ならびに出力の時間的な順序を示している。「昨日」が(2.2)において比較的早い段階で生成される必要があるのに対して，それに対応する英語表現“yesterday”は(2.1)の最後に現れる。このため，(2.1)の入力が終了するまで「昨日」以降の発話を生成することができない。その結果，

- 英語話者の発話の開始からその翻訳結果の出力終了までの時間が長くなり，対話の効率が低下する。
- 「健は」が発話されてから「昨日」が発話されるまでのユーザの待ち時間が大きいため，対話の結束性が損なわれる可能性がある。

などの問題が生じる。また，(2.1)の英語の述部“could meet”は比較的早い段階で出現するにも関わらず，対応する日本語の動詞「会えた」は文の最後に現れるため，

- 翻訳の即時性が損なわれ，英語話者による表情や身振りなどのノンバーバルな情報と同期をとることが難しくなる。

という問題も起こる。これらは，上記の例に特有な問題ではなく，英語文と日本語文との間の多くの場合において発生する問題である。ここで重要なことは，(2.2)を(2.1)に対する翻訳文と定める限り，これらの問題を解消することは原理的に不可能であるということである。実際，Kitanoによる翻訳文生成方法[9]も(2.2)のような標準的な翻訳文を出力するため，翻訳のタイミングは図1と同様になる。すなわち，英語文における主語の部分に対しては即座に翻訳結果を出力できるが，それ以外の部分については英語文全体が入力されるまで翻訳することができない。また，Amtrupによる手法[1]は独英機械翻訳のためのものではあるが，やはり同様の問題を有している。したがって，英語文の入力に対して，(2.2)と同じ意味内容をもち，かつ，同時進行的に作り上げることが可能な日本語翻訳文が存在するならば，それを翻訳結果として採用することにより漸進的な翻訳処理を実現することが考えら

(2.1) の入力	(2.2) の出力
Ken	健
could	は
meet	
her	
in	
the	
park	
near	
the	
school	
yesterday	昨日，学校の近くの公園で彼女に会えた

図 1　日本語翻訳文 (2.2) の出力のタイミング
Fig. 1.　Timing of outputs of the Japanese translation (2.2).

れる。それに対して本論文では，不適格表現を活用することにより，そのような処理を可能にする翻訳手法を提案する。

3.　不適格表現を活用した日本語翻訳文の生成

本章では，不適格表現である繰り返し，語順の逆転，省略，言い誤り，言い直し，言い淀みについて説明し，本論文で提案する話し言葉翻訳手法が不適格表現をどのように活用するのかについて例を用いて説明する。また，実際の日本語通訳文においてもそれらの不適格表現が頻出することを示す。

〈3•1〉　不適格表現とその活用方法

〈3•1•1〉　繰り返し　繰り返しは，発話した内容に関してより詳細な情報を追加する際に現れる言語現象である。

本手法では，英語の動詞が入力されると即座にそれに対応する日本語動詞を出力する。日本語の場合，動詞は文の最後に現れるため，そこで一旦翻訳文を閉じる。それ以降の入力に対する日本語は第 2 文において表現する。翻訳結果の品質を確保するために，第 2 文の最後に動詞を繰り返し生成する。例えば，次の英語文

(3.1) I live in a university dormitory.

に対して

(3.2) 私は 住んでいます。大学の寮に住んでいます。

を生成する。動詞 “live” の入力に対してすぐに日本語「住んでいます」を生成できるため，翻訳の即時性を満たすことができる。一方，(3.1) に対する標準的な翻訳例は

(3.3) 私は, 大学の寮に住んでいます。

である。(3.2) は「住んでいます」が繰り返し生成されていると言う点で (3.3) とは異なるものの，(3.1) の意味内容を正しく表している。また，翻訳システムのユーザからみた場合，(3.2) の第 2 文は，第 1 文の「住んでいます」に対して場所に関する情報を追加した発話とみなすことができ，自然な発話であるといえる。

〈3•1•2〉　語順の逆転　日本語の語順は，英語の語順と大きく異なる。しかし，日本語における構成要素の生起順序は比較的自由である。これは，日本語翻訳文をもとの英語文の入力順に従って作り上げたとしても，それがその翻訳結果として受理できる可能性があることを示唆する。例えば,

(3.4) How much is this tour?

に対する標準的な日本語翻訳文は「このツアーはいくらですか」であるが，ほとんどの日本人はシステムが出力する

(3.5) いくらですか。このツアーは。

の意味内容を容易に理解することができる。本研究では，そのような倒置文や標準的な語順から逸脱したやや違和感のあるような文も翻訳結果として認める。これにより，英語文の入力と同期した日本語翻訳文の出力を行える。

〈3•1•3〉　省　　略　日本語では，さまざまな表現が実に頻繁に省略される [17]。また，聞き手が容易に補うことが可能な表現については，むしろそれを省略した方がより自然な会話文になることが多い。本手法においても省略を積極的に活用する。例えば，(3.1) において動詞が入力された時点では，場所に関する情報は得られていないため，(3.2) のようにそれが省略された第 1 文をまず生成する。一方，第 2 文では主格が省略されているが，これは第 1 文から容易に補えることができるとともに，省略により，より自然な発話になっている。

〈3•1•4〉　言い誤り・言い直し　機械翻訳を困難にする原因の一つとして，入力文の曖昧性の問題が挙げられる。特に，漸進的な翻訳処理では，入力文を漸進的に解釈する必要があるが，その途中の段階は文全体の入力が未了であるため，一般に正しい解釈の候補は複数存在する。ところが漸進的な翻訳システムは，翻訳の即時性を維持するために，その中から一つの解釈を選び出し，それに対する翻訳結果を生成することが要求される。よって，それ以降の解析の進展にともない，選択された候補が除去され，結果として，生成された翻訳結果が言い誤りとなる可能性がある。例えば,

(3.6) And are these tours expensive?

の場合，“these” は指示代名詞と指示形容詞という二つの異なる品詞をもっており，“tours” が入力される段階までに，その曖昧性を解消することはできない。そこで，指示代名詞としての解釈を選択すると，それまでの日本語翻訳文は,

(3.7) それとこれらはツアー,

となり，次の “expensive” の入力に対して翻訳を続けても，意味の通らない翻訳結果となる。

この問題を解消するため，本研究では言い直しを活用する。本手法では，選択した解釈の候補が削除されると他の解釈候補を選択し，それに対する翻訳結果を生成する。その結果，(3.6) に対して

(3.8) それとこれらはツアー，これらのツアーは
高いですか？

のような翻訳結果を作り上げることができ，入力の曖昧さに対する頑健性を備えた翻訳処理を行える。

表1　不適格表現の出現頻度
Table 1. Appearance frequency of ill-formed expressions.

不適格性	頻度 (回数)
繰り返し	24
語順の逆転	26
省略	70
言い誤り	45
言い直し	49
言い淀み	393

〈3•1•5〉 言い淀み　言い淀みは，話し言葉において実に頻繁に現れるが，その出現箇所についてはいくつかの傾向がある。その一つとして，言い直しの前に出現するということがいえる[10]。本手法においても，選択されている解釈が削除された際に言い淀みを活用する。例えば，(3.6) に対する翻訳文の生成において，システムが修正した翻訳結果「これらを」を言い直す前に，言い淀み「えーと」を出力する。

(3.9) それとこれらはツアー，えーと，
　これらのツアーは高いですか？

言い淀みを用いることにより，ユーザに言い直しの発生を認識させるという効果があり，より自然な発話になると考えられる。

〈3•2〉 不適格表現の出現頻度　本論文で活用する不適格性が，実際の日本語通訳文においても頻繁に現れることを示すために，ATR 対話コーパスの中のバイリンガル対話のうち 4 対話を取り出し，英語話者によるすべての発話 278 文に対する英日通訳者による通訳文に対して上で述べた不適格表現の出現頻度を調査した。その結果を表1に示す。この表から，我々の翻訳手法で活用する不適格表現が実際の通訳文でも高頻度で現れることがわかる[10]。この点からも本手法が話し言葉翻訳のための方法として妥当であることが示される[11]。

4. 漸進的な解析・変換処理

本章では，3章で述べた不適格表現の活用方法に基づいて，漸進的な話し言葉翻訳を実現するための手法について述べる。本論文では，チャート法[8] と呼ばれる枠組を用いてこれを実現する。その理由は，

- チャート法では通常，自然言語文を左から右へ語単位で処理するため，語が入力される度に逐次解析処理を行うことができる。
- チャート法では，途中までに入力された部分的な表現に対する統語構造を，項と呼ばれるデータ構造として表現することができる。

ためである。以下ではまず，漸進的な翻訳システムの概要を示し，次に，翻訳の流れを例を用いて説明する。なお，チャート法の説明については文献[16] に詳しいのでこちらを参照されたい。

〈4•1〉 システムの概要　漸進的な話し言葉翻訳システムは，漸進的な解析，変換という二つのモジュール，ならびに文を表す統語範疇をもつ統語構造 (以下では，これを単に文構造と呼ぶ) のリストから構成される。システムの基本構成を図 2に示す。 英語単語が入力される度に次の (1)–(2) を順に実行することにより，漸進的な翻訳処理を実現する。

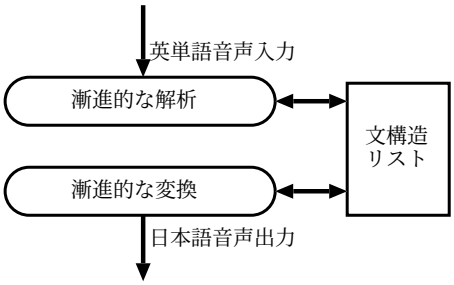

図 2　漸進的な英日話し言葉翻訳システムの構成
Fig. 2. Configuration of the incremental English-Japanese spoken language translation system.

（1） **漸進的な解析処理**　入力された単語とその統語範疇に対して，ボトムアップに文法規則を適用することにより，文構造リストの中のある文構造について，その中の最左未決定項と置き換え可能な項を可能な限り作成する。また，文構造リストの中の任意の文構造について，その最左未決定項を作成した項で置き換えた文構造のリストを新たな文構造リストとする。

（2） **漸進的な変換処理**　漸進的な解析処理によって作成された文構造リストの中の一つの文構造に対して，トップダウンに変換規則を適用することにより翻訳結果を作り上げる。なお，話し言葉翻訳においては，一旦翻訳された箇所はもはや変換される必要がないため，選択された文構造の中の日本語に変換された部分を変換済みとして印を付け，その結果得られた文構造を含むリストを新たな文構造リストとする。

漸進的な解析処理のための手続きは，到達可能性をもつ双方向チャート解析法[16] に従う。ただし，新たな単語が入力される度にそれまでの入力に対する新たな文構造を求めるため，文構造を構成する未決定項を未決定項を含む別の構造で置き換えるという操作を導入しており，通常の双方向チャート解析手法とはこの点が異なっている。一般に，解析途中で得られる文構造は複数存在するが，本手法では可能な解析結果をすべて残し，それらをリスト形式で記録する。文構造が新たに一つも得られないときは，原言語文の解析失敗であり，同時に翻訳失敗でもある。

一方，漸進的な変換処理は，リストで表現された複数の可能な文構造の中から先頭にあるものを変換の対象として選び出し，その構造にのみ変換規則を適用し，翻訳結果を作り出す。前章で説明した不適格表現の活用方法は，変換

入力英語発話	出力日本語発話
Ken	健
could	は
meet	会えた。
her	彼女に
in	
the	
park	公園で
near	
the	
school	えーと，学校の近くの公園で
yesterday	昨日，会えた。

図 3　日本語翻訳文 (4.1) の出力のタイミング
Fig. 3.　Timing of outputs of the Japanese translation (4.1).

規則において記述される。また，解析の進行にともない，選択されていた文構造が削除されたとき，新たにリストの先頭に置かれた文構造を変換処理の対象とする。その際，まず言い淀みを生成し，その上で変換処理を実行することにより自然な言い直しを実現する。なお，生成可能な言い淀み表現は一般に多数存在するが，本手法ではそれらの出現頻度に従ってランダムに生成する。

〈4・2〉 漸進的な翻訳の流れ　本論文で提案する翻訳手法に従うと，例えば英語文 (2.1) に対して次の日本語翻訳文

(4.1) 健は会えた。彼女に公園で，えーと，
学校の近くの公園で会えた。

を漸進的に生成することができる。(4.1) は，

- 「会えた」が繰り返されている。
- 「公園で」と「彼女に」の語順が逆転している。
- 「健は会えた」において誰に会ったのかが省略されている。
- 「公園で」は言い誤りであるが，「学校の近くの公園で」と言い直している。
- 「えーと」という言い淀み表現が現れる。

という点で標準的な翻訳文である (2.2) とは異なっているが，(2.1) の意味内容を正しく表現している。(4.1) の出力のタイミングを図 3に示す。 この図から，

- 英語話者の発話の開始からその翻訳結果の出力の終了までの時間が短くなり，翻訳を介した対話を効率的に進めることができる。
- ユーザの待ち時間が短縮されるため，対話の結束を保つことができる。
- 動詞を含むほとんどの英語単語に対して，対応する日本語が即座に生成されるため，翻訳の即時性を満たすことができる。

ことがわかる。ここでは，肯定文に対する翻訳例を示したが，前章で示した不適格表現を効果的に活用することにより，疑問文や否定文，さらには複文や重文なども適切に処理することができる。

表 2　英語会話 278 文の翻訳正解率
Table 2.　Translation results of 278 sentences.

タイプ	文数	割合
(A) 正しい翻訳 (言い直しを含まない)	96	34.5%
(B) 正しい翻訳 (言い直しを含む)	132	47.5%
(C) 不自然な翻訳	33	11.9%
(D) 誤った翻訳	16	5.7%
(E) 解析の失敗	1	0.4%

表 3　翻訳誤りの原因
Table 3.　Causes of incorrect translations.

原因	文数	割合
言い直しの出現の過多	33	11.9%
構造的曖昧性	8	2.9%
語彙的曖昧性	3	1.1%
慣用句	3	1.1%
不適格性	1	0.4%
音便変化	1	0.4%
解析失敗	1	0.4%

5. 翻 訳 実 験

前章までに示した漸進的な英日機械翻訳手法の実現可能性とその有効性を確認するために，実験システムを作成した。英語語彙 476 語，文法 204 規則の規模で GNU Common Lisp 2.2 を用いて実現した。また，それぞれの文法規則に対応して変換規則を作成した。

作成したシステムを用いて翻訳精度に関する実験，及び翻訳処理時間に関する実験を行った。実験の対象として ATR 対話データベース [4] の 4 対話に現れるすべての英語会話 278 文を用いた。これは 3章で述べた通訳文の調査のために用いた対話と同じものである。ただし，本研究は話し言葉の翻訳の実時間性を追求することを目的としているため，英語会話文に現れる言い淀み及び言い誤り表現についてはあらかじめ削除し，文法的に正しい入力を翻訳の対象とした。対話領域は旅行の問合せであり，対話文の平均の語の長さは 6.8 語であった。

〈5・1〉 翻訳正解率の実験　作成した実験システムの話し言葉に対する翻訳精度を評価するために，翻訳正解率を調べた。翻訳に用いた英語会話文をその翻訳結果の了解性に従って分類した。表 2に示すように，(A) または (B) に分類された 228 文 (翻訳結果の一部を付録に示す) が正解であり，82.0% の翻訳正解率を得た。これにより，本論文で提案した方法が話し言葉に対する翻訳手法として利用可能であることを確認することができた。

翻訳誤りの原因を表 3に示す。 翻訳誤りの多くは，言い直しをあまりに多く生成したために，不自然な翻訳結果が得られたことによるものである。これについては，次章で詳述する。慣用句については，翻訳を構成的に作り上げることができないため，それを漸進的に翻訳することは困難であるが，頻度が少なく全体の精度にはそれほど影響しない。それ以外の原因はすべて，通常の文単位での機械翻訳でも問題となるものである。よって，それらの問題に対し

表 4　一単語に対する翻訳処理時間
Table 4.　Processing time for one word.

項目	平均時間
翻訳処理時間	47 msec
解析処理時間	44 msec
変換処理時間	1 msec

てこれまで提案された解決法が本手法においても有効であると思われる。

〈5・2〉 翻訳処理時間の実験　実験システムの実時間性を評価するために，翻訳処理時間を測定した。時間計測は，SparcCenter 1000(SuperSPARC 50MHz) 上で，端末として XMINT CSL を用いて行った。入力された英語一単語に対する翻訳処理時間，解析処理時間，及び変換処理時間の平均を表 4に示す。翻訳処理時間 47 msec は，一単語の発話時間より少なく，本手法が効率的な話し言葉翻訳にとって有効であることがわかる。

実験の結果から，翻訳処理に要する時間のほとんどを解析処理時間が占めていることがわかる。これは，変換処理が複数の文構造からただ一つを選択し変換を実行するのに対して，解析処理では，形成されているすべての文構造に対して，入力された英単語をもとに新たな文構造を形成する必要があるためである。ここで述べた翻訳実験では，文法規則がそれほど多くなかったため翻訳処理時間の問題は生じなかったが，今後，大規模な話し言葉翻訳システムの実現に際しては，漸進的な解析のための効率的な手法を開発する必要がある。

6. 言い直しの活用

本論文では，生成した翻訳結果が言い誤りとなった場合，誤った部分を言い直すことにより正しい翻訳結果を作り上げるという手法を用いた。日本語話し言葉の解析における言い直しを処理する手法として，これまでいくつかの研究が行われているもの[3][15]，話し言葉の翻訳のために言い直しを活用して日本語翻訳文を作り上げる研究はこれまでなされていない。本章では，前章で示したプロトタイプシステムによる翻訳実験の結果をもとに，話し言葉の翻訳において言い直しを活用することについて考察する。

表 2に示すように、言い直しを含んだ正しい翻訳結果が得られた英語文は 132 文であり，全体の 47.5% を占めている。これは，言い直しを活用することにより，34.5% から 82.0% に翻訳精度が著しく向上したことを意味しており，漸進的な翻訳手法において言い直しの活用が有効であることがわかった。

一方，表 2の (C) に分類された英語文はすべて，その翻訳結果があまりに多くの言い直しを含むために不自然とみなされたものである。(B) 及び (C) に分類された英語会話文について，言い直しの回数と翻訳精度との関係を図 4に示す。この図から，言い直しの回数が増えるほど翻訳結果が不自然になりやすいという傾向がわかる。本手法では，

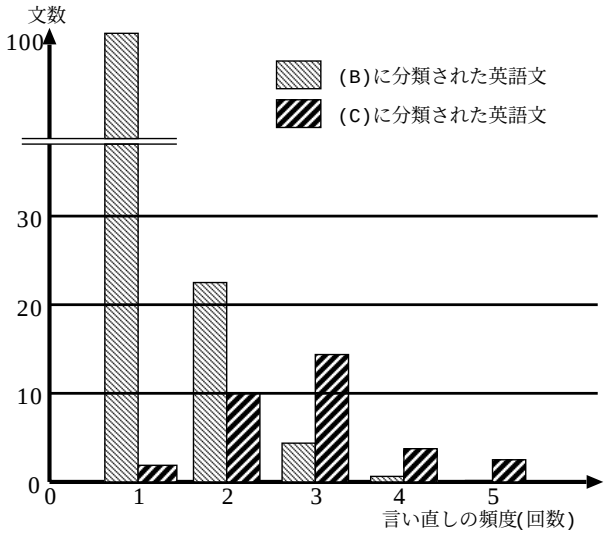

図 4　言い直しの頻度と翻訳精度の関係
Fig. 4.　Frequency of repairs in correct/unnatural translations.

選択された候補が，解析の途中で誤りであると判明したときに言い直しが生じる。図 5が示すように，一般に入力文が長くなると言い直しの回数が増加するが，これは長い文ほど多くの文構造が生成され，選択を誤る可能性が高くなるためである。より精度の高い翻訳処理を実行するために言い直しの回数を減らす必要があるが，そのために，

（1） 解析の途中で逐次生成される文構造の数をできる限り少なくする。
（2） 複数の解析候補の中から最も確からしい候補を選択するための手法を導入する。
（3） 候補の選択をできる限り語単位で実行するという制約を緩め，翻訳処理を遅延することにより，より正確な選択を行えるようにする。

などが考えられる。(1) に対しては，意味的な情報の活用により可能な候補を絞り込むことが有効であり，漸進的な曖昧性解消法[5] を利用することができる。また (2) に対しては，対訳用例を参照したり，統計的な情報を活用することが考えられる。(3) については，処理をどれくらい遅延すればよいのかを決める必要があるが，翻訳の即時性と翻訳の精度との間のトレードオフがあり，大規模な実験による検証が求められる。

7. むすび

本論文では，不適格表現を含む発話を生成する漸進的な英日機械翻訳手法を提案した。互いに語順の異なる英語文から日本語文への翻訳において，翻訳結果として繰り返し，語順の逆転，省略，言い誤り，言い直し，言い淀みを含んだ発話を容認することにより，意味内容の通じる程度の品質をもつ翻訳文を漸進的に作り出すことができる。実験システムを用いた翻訳実験により，本手法の有効性を検証した。漸進的な翻訳では，文の解析途中で生成される膨

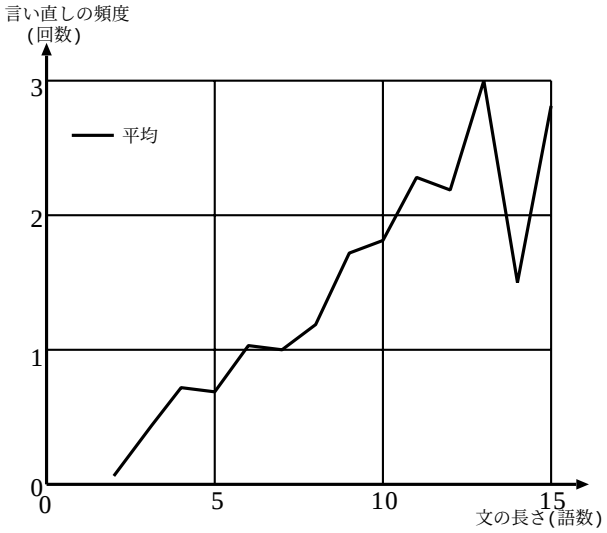

図5　入力文の長さと言い直しの回数の関係
Fig. 5.　Relation between source sentences length and frequency of repairs in translations.

大な曖昧性のために，誤った翻訳結果を生成する可能性が必然的に高くなるが，言い直しを効果的に活用することにより，ある程度の翻訳精度を維持できることがわかった。また，ユーザによる実際の発話時間に対して無視できる程度の処理時間で翻訳結果を得ることが確認できた。したがって本手法は，実時間話し言葉翻訳システムとして利用可能であり，漸進的な日英翻訳システムが開発されるならば，将来的には同時通訳を用いた英語日本語間翻訳対話の実現も期待できる[6]。

本論文では，文法的に正しい英語会話文のみを翻訳の対象とした。しかし，実際の会話文には，本論文で活用したような不適格性が頻繁に現れる。本研究で導入した双方向チャート解析法は，一般化範疇文法[2]による解析手法とともに，漸進的な統語解析に適していることが知られているが[6]，不適格性を含んだ自然言語文を漸進的に解析できるかどうかについては明らかではない。一方，チャート法を用いた文単位での非文解析手法が提案されており[7][13]，これらをもとに漸進的な不適格文解析手法を構築できる可能性がある。そのような手法を用いた不適格文に対する漸進的な翻訳処理を実現することは，今後の課題である。

謝辞　第一著者は日本学術振興会特別研究員制度の補助を受けている。また，本研究の一部は，文部省科学研究費補助金(特別研究員奨励費)の援助のもとに行われた。

(平成9年4月28日受付，同9年8月14日再受付)

文　献

(1) Amtrup J.W.: Chart-based Incremental Transfer in Machine Translation, *Proc. of 6th Int. Conf. of Theoretical and Methodological Issues in Machine Translation*, 188–195 (1995).
(2) Briscoe, E.J.: *Modelling Human Speech Comprehension: A Computational Approach*, Ellis Horwood Limited (1987).
(3) 伝 康晴：「統一モデルに基づく話し言葉の解析」，自然言語処理，4(1):23–40 (1997).
(4) 江原，井ノ上，幸山，長谷川，庄山，森元：「ATR対話データベースの内容」，テクニカルレポート TR-I-0186，ATR自動翻訳電話研究所(1990).
(5) Haddock, N.J.: Computational Models of Incremental Semantic Interpretation, *Language and Cognitive Process*, 4(3/4):337–368 (1989).
(6) Inagaki, Y. and Matsubara, S.: Models for Incremental Interpretation of Natural Language, *Proc. of 2nd Symposium on Natural Language Processing*, 51–60 (1995).
(7) 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，36(10):2343–2352 (1995).
(8) Kay, M.: Algorithm Schemata and Data Structures in Syntactic Processing, *Technical Report CSL-80-12*, Xerox PARC (1980).
(9) Kitano, H.: Incremental Sentence Production with a Parallel Marker-Passing Algorithm, *Proc. of 13th Int. Conf. on Computational Linguistics*, 217–222 (1990).
(10) 松原 茂樹，稲垣 康善：「同時通訳により生成された日本語会話文の調査・分析」，電子情報通信学会情報・システムソサイエティ大会，55 (1996).
(11) 松原 茂樹，稲垣 康善：「漸進的な英日機械翻訳により生成された日本語翻訳文の評価」，電気関係学会東海支部連合大会，622 (1996).
(12) Matsubara, S. and Inagaki, Y.: Utilizing Extra-Grammatical Phenomena in Incremental English-Japanese Machine Translation, *Proc. of 7th Int. Conf. on Theoretical and Methodological Issues in Machine Translation*, 31–38 (1997).
(13) Mellish, C.S.: Some Chart-Based Techniques for Parsing Ill-Formed Input, *Proc. of 27th Annual Meeting of Association for Computational Linguistics*, 102–109 (1989).
(14) Menzel, W.: Parsing of Spoken Language under Time Constraints, *Proc. of 11th European Conf. on Artificial Intelligence*, 560–564 (1994).
(15) Sagawa, Y., Ohnishi, N. and Sugie, N.: A Parser Coping with Self-Repaired Japanese Utterances and Large Corpus-based Evaluation, *Proc. of 15th Int. Conf. on Computational Linguistics*, 593–597 (1994).
(16) 田中 穂積：「自然言語解析の基礎」，産業図書(1988).
(17) 竹沢 寿幸，田代 敏久，森元 逞：「音声言語データベースを用いた自然発話の言語現象の調査」，人工知能学会 言語・音声理解と対話処理研究会資料，SIG-SLUD-9503，13–20 (1995).

付　録

翻訳結果の例

- Hi, yes, I am interested in booking a tour to Hong-Kong, please.
  もしもし，はい，興味があります。香港，えーと，香港へのツアーを予約することに興味があります。
- And what does that include?
  それと何は，あっ，何をそれは含みますか？
- I think it was 3days.
  思います，それが3日間であったと。
- I am looking forward to hearing from you.
  楽しみにしています。話しを聞くこと，あっ，えーと，あなたから話しを聞くことを楽しみにしています。
- I would like to stay longer if possible.
  滞在したいです。より長くもし可能なら滞在したいです。
- Does that price include meals?
  それは，えーと，その値段は含みますか？ 食事を含みますか？
- How much would this trip cost?
  いくらこの旅行は費用がかかるだろうか？
- Actually 6 or 7 would be best for me?
  実際，6時または7時は最もよい，あのー，私にとって最もよいものであるだろう。
- I am in Japan for a short visit with my husband.

日本に少しの間います。えー，私の夫といます。

- Well, I am kind of interested in seeing some Japanese temples and Buddas and things like that.
  えーと，ちょっと興味があります。いくつかの日本の寺と仏陀，それとそのような類のものを見ることにちょっと興味があります。
- I like taking showers.
  好きです。シャワーをすることが好きです。
- I am staying at the Marunouchi Hotel.
  泊まっています。えーと，丸の内ホテルに泊まっています。
- Leaving on the 24th and leaving on the 31st?
  24 日に出発と 31 日に出発？
- I don't think that's possible.
  思いません。それが可能であると。
- Sure, the name is Lenelle Degenhardt.
  ええ，名前はレネーレ デーゲンハルドです。
- I don't have a phone.
  もっていません。電話をもっていません。
- I'm a university student here in Japan and I am interested in booking a tour to India.
  大学生，えー，ここ日本の大学生です。それと興味があります。ツアー，あ，インドへのツアーを予約することに興味があります。

松 原 茂 樹（非会員） 1970 年生。1993 年 3 月 名古屋工業大学電気情報工学科卒業。1995 年 3 月 名古屋大学大学院工学研究科情報工学専攻博士前期課程修了。現在，同後期課程在学中。1996 年 4 月 より日本学術振興会特別研究員。自然言語理解，計算言語学 に関する研究に興味を持つ。電子情報通信学会，情報処理学会，人工知能学会，言語処理学会各会員。

浅 井 悟（非会員） 1974 年生。1997 年 3 月 名古屋大学工学部情報工学科卒業。現在，同研究生。自然言語処理，機械翻訳 に関心を持つ。

外 山 勝 彦（非会員） 1961 年生。1984 年 3 月 名古屋大学工学部電気学科卒業。1989 年 3 月 同大大学院工学研究科情報工学専攻博士後期課程満了。同年名古屋大学工学部情報工学科助手。1990 年 中京大学情報科学部情報科学科講師。1993 年 同助教授。工学博士。知識表現，非単調推論，自然言語理解 に関する研究に従事。電子情報通信学会，情報処理学会，人工知能学会，言語処理学会，日本認知科学会各会員。

稲 垣 康 善（正員） 1939 年生。1962 年 3 月 名古屋大学工学部電子工学科卒業。1967 年 3 月 名古屋大学大学院博士課程修了。同大助教授，三重大学教授を経て，1981 年より名古屋大学工学部教授。1997 年より同大学院工学研究科長。工学博士。オートマトン言語理論，ソフトウェア基礎論，代数的仕様記述法，人工知能，自然言語理解 に関する研究に従事。電子情報通信学会，情報処理学会，人工知能学会，日本ソフトウェア科学会，言語処理学会，IEEE， ACM， EATCS 各会員。