

日中会話文機械翻訳における不適格な表現の処理

任 福継

広島市立大学情報科学部

自然言語による日常会話には、省略を始めとして多種多様な不適格性が数多く出現する。頑健な自然言語処理システムであればこのような不適格な発話から話者の意図した意味が推定できる。一方、従来の機械翻訳システムは、多くの場合、文法的な文のみを理解し翻訳するように作られており、非文法的な文は扱えない。本論文では、機械翻訳という視点から自然言語文の省略について検討を行い、固執省略と臨界省略という新しい概念並びに複数の翻訳プロセスをもつ協調融合型機械翻訳 (CFMT) 手法を提案する。CFMT手法では、全情報制約翻訳プロセス、意味的制約主導の翻訳プロセス、統語的制約主導の翻訳プロセス、などの独立した翻訳プロセスがある。著者らはCFMT手法をSWKJCの下に翻訳実験を行った。その結果、CFMT手法による高品質で効率的な会話文機械翻訳の実現可能性を確認することができた。

Handling Technique of Ill-Formed Sentences in Japanese-Chinese Machine Translation

Fuji Ren

Faculty of Information Science, Hiroshima City University

E-mail: ren@its.hiroshima-cu.ac.jp

In natural dialogues, speakers make many kinds of ill-formed sentences such as ellipsis. A robust natural language processing system must be able to allow such ill-formed sentences and derive correct interpretations from the ill-formed sentences. But existing machine translation systems reject utterances including ellipses and ill-formed expressions. In this paper, a new concept, Persisting Ellipsis and Critical Ellipsis in machine translation, is put forward, and a method for handling ill-formed sentences, called Cooperative Fusion Machine Translation (CFMT), is presented. The CFMT advocates the ideas: (1) All-Constraint based translation process, (2) Syntactic-Constraint based translation process, and (3) Semantic-Constraint based translation process. An experimental system based on this method has been constructed and some results are given in this paper. The results show that CFMT is a promising technique for high-quality and efficient spoken language machine translation.

1. はじめに

言語を用いた自然な対話には、誤りを始めとする多種多様な不適格性が数多く現われる[1,2]. しかし、人間の聞き手は不適格な発話でも、そこから話者の意図した意味を推測することが可能である。また、人間の通訳者はこのような対話における発話者の意味を正しく相手に翻訳することができる。

一方、従来の対話システムを始めとする自然言語システムは不適格性に対して非常に弱く、人間のような柔軟性はない。特に、今までの書き言葉の機械翻訳は、多くの場合、文法的な文のみを理解し翻訳するように作られており、非文法的な文(不適格文)は扱えない。

キーボードを入力した媒体としている現在の機械翻訳システムにおいては、不適格文は少ないが、音声発話においては不適格文が非常に多い。将来、音声による入出力を備えた自然言語システムおよび自動電話翻訳システムを実用化する際には、不適格文は大きな障害になると考えられる[3].

従来の不適格文に関する研究の多くは話者の意図理解や適格文への還元に着目しているにすぎず、目的言語への翻訳を意識した研究は少ない。そのため、これらの解析では翻訳システムにそのまま流用できない場合が多いと考えられる。

本論文では、複数の翻訳プロセスからなる協調融合型機械翻訳(CFMT: Cooperative Fusion Machine Translation)手法を提案する。この手法において、独立した翻訳プロセス(あるいは翻訳エンジン)は並列に働き、それぞれのプロセスは入力された原文の適当な部分に対して解析結果を出力する。その部分結果を共有メモリに書き出す。各プロセスの出力は融合されてよりよい結果をシステム全体の出力とされる。2つ以上のプロセスが部分結果を作れば、解析プロセスとは独立に結果の統合が試みられ、より適切な解析が行われる。ただ1つのプロセスが成功した時には、そのプロセスの結果が使われる。途中でいくつかの処理が失敗しても、他のプロセスの中間結果を使い、解析を進める。この原理は分散協調処理による自然言語解析手法と類似しているが、本手法ではそれをさらに機械翻訳に拡張している。CFMT手法は、(1)実例に基づく翻訳プロセス、(2)経験知識に基づく翻訳プロセス、(3)全情報制約主導翻訳プロセス、(4)統語的制約主導翻訳プロセス、(5)意味的制約主導翻訳プロセスからなる。本論文では(1)と(2)は議論せず、(3)か

ら(5)に主眼をおいて述べる。また、本手法の有効性については、実際のシステムを用いた翻訳実験の結果から述べる。この実験はCFMT手法をSWKJC[14,15]に組み込んで行われた。その結果、この手法は格助詞のない文や意味的に不適格な文や必須格のない文を中国語に翻訳することに極めて有効であることが分かった。

本論文の構成は以下のとおりである。第2章で機械翻訳における省略の分類について述べ、第3章でSWKJCの概要について述べる。第4章で協調融合型機械翻訳方法を提案し、第5章でその応用例を提示する。そして、第6章で実験および考察について述べ、最後に、第7章でまとめと今後の課題を述べる。

2. 機械翻訳における入力文中の省略

ふつう自然言語では、聞き手に容易に判断できる要素は、文章上表現しない場合が多い。特に会話文では省略はさらに多い。このような省略現象にはさまざまなものがあるが、特に機械翻訳システムで問題となるものとして以下の2点が挙げられる。

(1) 原言語文では陽に示されていない要素が、目的言語では必須要素となる場合がある。

(2) 原言語文において省略された要素が補完されなければ、現在の技術では解析できない場合がある。

上記の(1)は目的言語の性質によりその範疇が異なる。

2. 1 省略補完方法および問題点

省略の処理は、一般に通用する制約を用いて、その補充対象を認識する処理である。そのような制約として、構文的制約や意味的制約が用いられている。しかし、これらの制約の満たす複数の補充対象を一つに絞りきることが不可能な場合もある。これに対処するには、文章あるいは対話を通じた制約が必要となり、この制約として、語用論的制約が必要となる。しかしながら、現時点では語用論的制約の実現は極めて困難である。

以下では幾つかの省略処理方法とその問題点を示し、機械翻訳という視点から省略に関する検討を述べる。

・ 構文的及び意味的制約に基づく処理

省略処理に、文の主題や複文的な特徴あるいは名詞と動詞の意味的共起性等を基礎とする処理がある。構文的制約及び意味的制約に基づく省略処理は比較的単純であるが、現象論的であり、論理的な検証方法が曖昧である。したがって、場の状況が動的に変化する対話においては省略補充処理には問題がある。

- ・ 待遇表現及び視点関係に基づく処理

構文的制約は連続した対話や文章における省略補充には不十分であり、文脈情報を利用する必要がある。例えば、対話の待遇表現や受給表現の使用に関する制約を用いた処理がある。このような言語表現の使用に関する制約は、言語使用の社会的側面をとらえており、対話の登場人物を指示する省略表現について有効であるが、対話の登場人物とその人物に属する対象以外の対象を指示する表現を扱うのは困難である。

- ・ 知識あるいは常識に基づく処理

基本的な考え方は、対話参加者の知識および一般的な常識を用いて導き、その導かれた知識が互いに矛盾してはならないという制約を用いて省略の補充候補を絞り込むというものである。前章にも述べたように、対話は動的に場の状況が変化するので、あらかじめ知識をどこまで与えればよいかを決定するのは困難である。

2. 2 機械翻訳における省略の分類

実用的な機械翻訳システムには、(1) 高い翻訳品質、(2) 速い応答、及び(3) 頑健な解析能力が要求されている。我々はこのような機械翻訳システムの要求に対し、実用化日中機械翻訳システムSWKJCを開発している[15,16]。しかしながら、このシステムでは省略に対処する機能が弱く、その改善が使用者から強く望まれている。

一方、自然言語の省略処理は極めて困難であり、いままで提案した手法はそのままSWKJCに流用できない場合が多い。本論文では、機械翻訳の視点から省略を以下の二種類に分類して定義する。

- ・ 固執省略 (Persistent Ellipsis)

原言語文のある省略において、陽に補充しなくても、ある方法でそのまま目的言語に翻訳すれば、

目的言語側でその意味を還元できるものを固執省略という。これを省略遷移とも呼ぶ。

- ・ 臨界省略 (Critical Ellipsis)

機械翻訳における原言語文のある省略において、正しく補充しなければ解析できず、正しい目的言語表現が得られないものを臨界省略という。

例えば[文献13の例]、

【例4】

NTTは新型交換機を導入した。(Φsubj1)
自己診断機能を搭載、(Φsubj2) 200システムを設置する予定だ。

ここに、括弧の部分は省略されたもので、Φの後に付いた記号は省略された格要素である。日英機械翻訳システムにおいて、このような省略は本論文の定義によれば臨界省略であり、用言意味属性を用いたゼロ代名詞照応解析方法によってこのような省略を補充できたが[13]、日中機械翻訳システムにおいては、これは固執省略であり、陽に補充する必要がないと考えられる。まず、陽に補充した訳文と補充しない訳文の比較を見よう。

例4で、Φsubj1は新型交換機であり、Φsubj2はNTTであると考えられる。

【訳文4-1】 (Φsubj1とΦsubj2を補充する)

NTT購買了一种新式交换机, 这种新式交换机具有自己诊断功能, NTT准备安置200台系统。

【訳文4-2】 (Φsubj1とΦsubj2を補充しない)

NTT購買了一种新式交换机, 具有自己诊断功能, 准备安置200台系统。

訳文4-2では、原文において省略された情報は目的言語側でも補充でき、すなわち、このような省略は遷移的に、原言語側で陽に補充しなくても翻訳できる。勿論、原言語側で省略の補充処理は常に容易ではなく大きな手がかりが必要であり、場合により多くの処理時間がかかる。

本論文では、上述した特徴に着目し、新しい日中機械翻訳システムにおける省略処理手法を提案する。以下、第3章でこの手法をインプリメント

する日中機械翻訳システムSWKJCの概要を述べる。

3. SWKJCの概要

SWKJCは翻訳規則に基づいて実用化される日中機械翻訳システムである[14,15]。翻訳規則はコーパスから半自動的に抽出されている。

近年、実例に基づく翻訳、コーパスに基づく翻訳は盛んに行われており、コーパスから翻訳知識や翻訳規則を自動獲得するための研究が多くの研究者によりなされている。しかしながら、コーパスの規模の拡大につれて、類似例文の探索や翻訳規則の獲得が困難になり膨大な処理時間が必要となる。これは実用化機械翻訳の開発に対して大きな問題となる。SWKJCは実例に基づく翻訳およびコーパスに基づく翻訳手法の利点を吸収しながら、システムの処理時間の増加を抑える。

SWKJCは、従来の手法と同じようにコーパスから翻訳規則を抽出するが、コーパス自身は翻訳システムから分離し、コーパスから抽出した翻訳規則のみを翻訳システムに組み込んでいることを特徴とする。我々はこれにより実用的な機械翻訳システムの要求を満足する試みを行っている。

実はSWKJCは規則に基づく方法、実例に基づく方法、及び統計に基づく方法を融合した翻訳システムとも言える。これは翻訳規則を獲得するため、実例を利用し、さらに、翻訳規則の適応条件を記述する際、統計的情報を使っているからである。

4. 協調融合方式

自然言語文における省略及びその他の非文法的な現象を頑健に解析して、効率の高い機械翻訳システムを実現するため、協調融合機械翻訳方式を提案する。これはSWKJCにおける

Robust_Processorに組み込まれる予定であるが、本論文では、議論の便のため、実際のシステム構造と若干異なるモデルを用いる。

協調融合方式は、全情報制約主導の翻訳プロセス、統語的制約主導の翻訳プロセス及び意味的制約主導の翻訳プロセスからなる。ここで全情報制約主導の翻訳プロセスは統語的制約、意味的制約、文脈的制約及び語用的制約による翻訳メカニズムであるが、現段階では主に統語的制約、意味的制約、時には文脈的な制約によるものである。

各翻訳プロセスは言語のある側面を重視するが、

理論的に翻訳システムは各翻訳プロセスを並列、非同期に行い、必要に応じて途中の解析結果を共通メモリに書き込み、システムの解とする。各部分の結果は、可能なら、照合及び調整される。各翻訳プロセスは、精度が異なるが、単独で何らかの結果を出す。ここでの照合は、各処理結果を統合することにより、より適当なものにすることを意味し、調整は異なるプロセスによる解析を相互に補完することを意味する。1つのプロセスしか成功しなかった場合は、このプロセスの結果を全体の結果とする。幾つかのプロセスが成功した場合は多数決により全体の結果を決める。多数決により決められない場合はあらかじめ決めたプロセスの優先順位により全体の結果を決める。その概念を図3に示す。

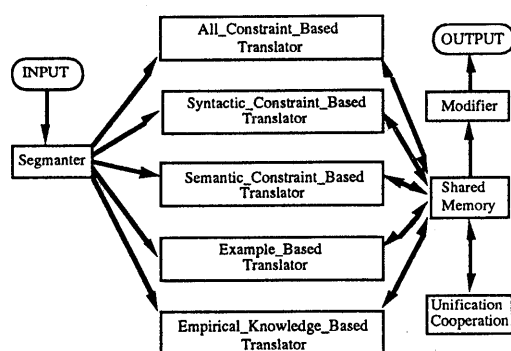


Figure 3 Concept of Cooperative Fusion Translation

・全情報制約主導の翻訳プロセス

入力した原文に対し、統語的制約、意味的制約、文脈的制約及び語用的制約により翻訳を行う。必要ときは、慣用表現、比喩、隠喩なども考えるが、完全な全情報制約主導の翻訳を実現するにはまだ極めて困難である。

・意味的制約主導の翻訳プロセス

入力文が統語的に不適格な場合は、意味的制約により解析する。特に、日常会話では必要な助詞が省略される場合が多い。例えば、

【例5】

A: タベ何を食べた？

B: 私は刺身食べた。

Bでは名詞「刺身」と動詞「食べる」(「食べた」

の原形)」の間に格助詞「を」を省略したので、統語解析では失敗する。これは<名詞 動詞>という日本語文法の接続規則がないからである。

意味的制約主導の翻訳プロセスでは語順や格助詞を無視し、各単語の意味属性関係により解析する。例5のBについて、「私 刺身 食べた」のみ解析する。各要素の意味属性および意味制約により、次の2つの結果が得られる。

(B1) 主格(私, 食べる), 目的格(刺身, 食べる)

(B2) 主格(刺身, 食べる), 目的格(私, 食べる)

B1は「私は刺身を食べる」を意味し、B2は「刺身は私を食べる」を意味する。基本的には、意味制約ではこのような曖昧性は解消されない(勿論、B2の曖昧性は意味属性の細分類及び常識を利用すれば解消されるが、一般的な解消方法はない。例えば、「荒木, 殴る, 鈴木」)。しかしながら、協調融合方式は「協調メカニズム」として上述した「照合」と「調整」を持っている。Bに対して統語的制約主導の解析では、「刺身」と「食べる」のマーチは失敗したが、「私」と「食べる」のマーチは成功したので、「私」が「食べる」の主格であることを得て、これを共有メモリに書き込む。これにより、Bの曖昧性が解消され、B1をシステムの結果として共有メモリに書き込み、その結果、「我吃生魚片」という中国語訳文が生成される。

・統語的制約主導の翻訳プロセス

ある入力文に対し、意味的制約が満たされないとき、自分では解析を終えることができない。このとき、統語的制約主導により翻訳を行う。

例えば、

【例6】

CSKは多言語機械翻訳理論を研究している。

「研究する」の主格の意味属性は「人間」という下位属性値が必要であるが、CSKは「組織, 団体」なので、意味的制約が満たされない。この場合には、統語制約主導により翻訳を行う。統語制約主導では、「CSK」は名詞であり、「は」という格助詞があるので、「研究する」の主格になることを決める。

5 協調融合型日中機械翻訳の応用例

協調融合型機械翻訳では、固執省略に対し、陽に補充しなくてもそのまま各翻訳プロセスに適応する。臨界省略に対しても必ずしも陽に補充する必要はない。多くの場合には、CFMT手法で自動的に対応する目的言語表現を生成できる。なお、原言語文では陽に示されていない要素が目的言語では必須要素となる場合の処理は別の機会に述べる。次に幾つかの応用例を示す。

5. 1 統語的にも意味的にも適格な場合

【例7】

桃太郎がマリを愛する。

この解析に用いられる辞書情報の例を次に示す。

```
Word_342 {1.Word_J {愛する, TV, NC}
          2.Number {3}
          3. {Syn ('愛AI', V1, {JPP < J1 が J2 を
$>}),
            Simp {J1 {N*}, Ind->x,
                  J2 {N*}, Ind->y}},
            Sem { Agent(x): x {HUM/ANI},
                  Patient(y):
y {HUM/ANI/ORG}},
            CHP {<x '愛AI' y>
              }
              .....}
```

```
Word_662 {1.Word_J {桃太郎, N5, #}
          2.Number {1}
          3. {Syn ('桃太郎
YAOTAILANG', +K+T),
            Simp {Ind->x},
            Sem { HUM, MAN},
            CHP {桃太郎}
              }
```

```
Word_662 {1.Word_J {マリ, N5, #}
          2.Number {1}
          3. {Syn ('馬麗MART, +K+T),
            Simp {Ind->x},
            Sem { HUM, WOMAN},
            CHP {馬麗}
              }
```

これは全情報制約主導の翻訳プロセスで用いられるものであるが、意味的制約主導翻訳プロセス

はSynとSimpの部分抜いたもの、統語的制約主導翻訳プロセスはSemの部分抜いたものを用いる。

例7に対し、全情報制約主導と統語的制約主導翻訳による結果を【結果1】に、意味的制約主導翻訳による結果を【結果2】に示す。

【結果1】

```
Fault{Center_W{愛する},Cand{1},
      1. {Agent{桃太郎},
          Patient{マリ}.....}
```

【結果2】

```
Fault{Center_W{愛する},Cand{2},
      1. {Agent{桃太郎},
          Patient{マリ}.....
      2. {Agent{マリ},
          Patient{桃太郎}.....}
```

結果2には曖昧性があるが、協調メカニズムにより結果1を全体の結果として共有メモリに書き込む。これにより、次の中国語文を得る。

【中国語訳文】 桃太郎愛馬麗。

5. 2 統語的に不適格な場合

【例8】

桃太郎がマリ愛する。

例8では名詞「マリ」と動詞「愛する」の間に必須な格助詞「を」を省略したので、全情報制約主導と統語的制約主導翻訳では失敗したが、意味的制約主導翻訳では成功し上記の【結果2】が得られた。

しかし、【結果2】には曖昧性がある。全情報制約主導と統語的制約主導翻訳は全体として失敗したが、部分的な結果が得られる。即ち、名詞「マリ」と動詞「愛する」はマーチしなかったが、名詞「桃太郎」と動詞「愛する」とのマーチは成功し、「桃太郎」が「愛する」のAgentである途中結果を生成する。この途中結果を利用して、【結果2】の曖昧性を解消することができる。さらに、この結果を全体の結果として共有メモリに書き込む。最後に、例7と同様な中国語訳文が得られる。

5. 3 意味的に不適格な場合

【例9】

- A. 人民は祖国を愛する。
- B. 祖国は人民を愛する。

例文9Aは問題がないが、Bについては祖国の意味属性がHUM(人類)もANI(動物)も持っていないので、全情報制約主導と意味的制約主導翻訳は失敗する。実はこれは比喩の例である。しかし、統語的制約主導翻訳は成功する。その結果を次の【結果3】に示す。

【結果3】

```
Fault{Center_W{愛する},Cand{1},
      1. {Agent{祖国},
          Patient{人民}.....}
```

5. 4 固執省略の例

【例10】

- 受付1：研究会で発表される方ですね。
- 来客2：はい。
- 受付3：懇親会に出席なさいますか。
- 来客4：何時からですか。
- 受付5：5時からです。
- 来客6：出席します。

これは固執省略の例である。次に来客6を例としてその翻訳過程を述べる。
この解析に用いられる辞書情報の例を次に示す。

```
Word_122{1.Word_J{出席する, IV, NC}
          2.Number{1}
          3.{Syn{'出席CHUXI,V3,(JPP<[J1
が*][J2に]>)},
          Simp{J1{N*},Ind->x,
                J2{N*,Ind->y}},
          Sem{ Agent(x): x{HUM},
                Object(y): y{THI/ORG}},
          CHP{<[x]'出席CHUXI' [y]>}
          }
          .....}
```

この例に対し、上記の方法をそのまま流用すると、翻訳プロセスは3つとも失敗する。しかしながら、CFMT手法では固執省略のために、デフォルト構造を用意している。上記のWord_122中の[]はこれを表わす。これにより「出席します」を翻訳することができる。このように、固執省略につ

いて、わざわざ省略補充しないで直接訳文を生成することは翻訳システムの効率を高めていると考えられる。

5. 5 複数個所に省略した例

【例 1 1】

- A: 今晚何を食べる？
B: 私刺身食べたい。

この例のBでは2カ所の格助詞が省略されているので、意味的制約主導翻訳プロセスのみ成功する。その結果を【結果4】に示すが、その曖昧性は解消されていない。これは統語的制約主導翻訳で利用できる情報がないためである。

【結果 4】

Fault{Center_W{食べる},Cand{2},Sen{たい},
1. {Agent{私},
Object{刺身}.....
2. {Agent{刺身},
Object{私}.....}

本例では、意味属性の細分類により曖昧性が解消できるが、一般的な解消方法は存在しないと考えられる。CFMT手法ではあらかじめデフォルト規則を用意している。例えば、次の規則は「主格と目的格に曖昧性がある時は、もっと文頭に位置する要素を主格とする」を意味している。

Rule_1{(Agent, Object){W1, W2},
Len(W1)>Len(W2): Agent(W2) &
Object(W1);
Agent(W1) &
Object(W2).}

6. 実験と考察

本論文で述べた協調融合型機械翻訳方式を、我々が開発しているSWKJCの下に翻訳実験を行い、本方式の有効性を確認した。原理的には並列処理すべき各翻訳プロセスは、現時点では直列に行われた。また、本実験は実例に基づく翻訳プロセスおよび経験知識に基づく翻訳プロセスを除いて行われていた。

以下に示すような2回の実験を行った。1回目の実験では、テキスト（日本語教科書）と情報処理関連文献から無作為に600文を抽出して実験対

象文とした。2回目の実験では、会話文80組延べ1240文を実験対象文とした。

結果、格助詞「を」を省略した頻度は高く、302箇所で出現したが、CFMT手法を用いると96.4%の正解率を得た。ここでの正解とは格助詞の補充が正しい（勿論陽に補充しなかったが）ことと、訳文が意味的に正しいことである。なお、評価は文単位とした。格助詞「と」および「に」の処理はまだ充分とはいえない。例えば、

【例 1 2】

- A: これは何ですか？
B: 鰻です。
C: 猫{Φ[は, 1]}食べるの？
D: ええ、犬{Φ[と, 1]}猫{Φ[は, 2]}{Φ[を, 3]}食べるよ。

【訳文（例 1 2）】

- A: 这是什么？
B: 是鰻魚。
C: 猫吃么？
D: 這個，狗吃猫。

ここで、Dの翻訳は間違っており、正しくは「這個，狗和猫都吃」である。この例では、猫と犬の意味分類属性はANI（動物類）であり、下位属性でも同じくEAT（食肉類）であるが、意味的制約主導の翻訳プロセスにおいて、他動詞「食べる」の特徴により、目的語要素の補充が優先したのである。特に「猫」が必ず「食べる」の目的語要素にならない理由はありえないからである。

実験結果から、CFMT手法は会話文に対し非常に有効であることが分かった。また、一般的なテキストに対しても7%の改善ができた。今後、全体システムの解析能力を高めるため、翻訳規則の充実が必要があり、また、臨界省略に関する補充方法はさらに検討する必要がある。

7 おわりに

従来の自然言語システムは不適格性に対して非常に弱く、人間のような柔軟性はない。特に、今までの書き言葉の機械翻訳は、多くの場合、文法的な文のみを理解し翻訳するように作られており、非文法的な文は扱えない。

キーボードを入力媒体としている現在の機械翻訳システムにおいては、不適格文は少ないが、音声発話においては不適格文は非常に多い。将来、

音声による入出力を備えた自然言語システムおよび自動電話翻訳システムを実用化する際には、不適格文は大きな障害になると考えられる。

本論文では、機械翻訳という視点から自然言語文の省略について検討を行い、固執省略と臨界省略という新しい概念を提案した。これにより、機械翻訳において高効な省略処理が実現できると考えられる。さらに、省略を含む不適格文の機械翻訳について、複数翻訳プロセスをもつ協調融合型機械翻訳手法を提案した。この手法において、独立した翻訳プロセスが並列に働き、それぞれのプロセスが入力した原文の適当な部分に対して解析結果を出力する。この部分結果を共有メモリに書き出す。各プロセスの出力はそのままシステム全体の出力とされ、2つ以上のプロセスが部分結果を作れば、解析プロセスとは独立に結果の統合が試みられ、より適切な解析が行われる。ただ1つのプロセスが成功した時には、そのプロセスの結果が使われる。いくつかの処理が途中で失敗しても、他のプロセスの中間結果を使い、解析を進める。この手法を用いることで、頑健な機械翻訳システムの実現が可能であると考えられる。

本論文で提案したCFMT方法をSWKJCの下に翻訳実験を行った。会話文に対し81.5%の正しい翻訳が得られ、特に格助詞省略に対しては88.7%の正解率を得た。テキストなどの対象文に対しては7%の改善が得られた。したがって、本手法は会話文における省略に対して有効であるといえる。

今後、SWKJCの翻訳規則の充実、規則の自動的または半自動的な学習機能の開発、臨界省略に関する補充方法の詳細な評価などを予定している。

謝辞

日ごろ有益な御討論、御助言を頂く研究室各位に感謝致します。また、実験システムの構築および本論文の作成を進めるにあたり種々御助言ご協力を頂いた大連理工大学自然言語処理研究室簡幼良教授、北海道大学工学研究科柝内香次教授に感謝の意を表します。

なお、本研究の一部は文部省科学研究費（課題番号07780344）により行われた。

参考文献

1. Reilly, R. G.: Types of Communication Failure in Dialogue, Communication Failure in Dialogue and

Discourse, pp.99-120, Elsevier Science Publishers, Amsterdam(1987).

2. 松本裕治：頑健な自然言語処理へのアプローチ、情報処理, Vol. 33, No.7, pp. 757-767 (1992).

3. Sagawa, Y., Ohnishi, N. and Sugie, N.: Techniques to Recover User's Utterance Failure in a Plan-Driven Dialogue System, Proc. Natural Language Processing Pacific Rim Symposium, pp. 131-138 (1991).

4. 加藤恒昭：自然言語インタフェースにおける省略の扱い、情報処理学会論文誌, Vol. 34, No.9, pp.1899-1908 (1994).

5. 久野：談話の文法、大修館(1978).

6. 石崎、井佐原：日本語文の複雑さの定性的・定量的特徴抽出、情報処理学会自然言語処理研究会報告, NL. 67-6(1988).

7. 竹本、田口、中川：質疑応答システムにおける日本語文の理解と生成、情報処理学会自然言語処理研究会報告, NL. 74-13(1988).

8. Fuji, R.: SWKJC Machine Translation System Based on Translation Rules Acquired from Corpora, Technique Report of Hiroshima City University, HCU-IS-95-036, (1996).

9. 任、宮永、柝内：日中常用文型機械翻訳システム、電子情報通信学会論文誌, DII, Vol. J74, No.8, pp. 1060-1069(1991).

10. 山本、小林、中川：音声対話文における助詞落ち・倒置の分析と解析手法、情報処理学会論文誌, Vol. 33, No.11, pp.1322-1330 (1992).

11. Takebayashi, Y., Tsuboi, H., Kenazawa, H., Sadamoto, Y., Hashimoto, H., and Sinichi, H.: A Real-Time speech Dialogue System Using Spontaneous Speech Understanding, IEICE Transactions on Information and Systems, Vol. E76-D, No.11, PP.112-119(1993).

12. 柏岡、高野、平井、北橋：対話参加者の知識状態を用いた省略語の補充、情報処理学会論文誌, Vol.33, No.10, PP.1203-1210(1992).

13. 中岩、池原：日英翻訳システムにおける用言意味属性を用いたゼロ代名詞照応解析、情報処理学会論文誌, Vol. 34, No.8, pp.1705-1715 (1993).

14. 任、范、宮永、柝内：家族モデルを用いた文の分解に基づく日中機械翻訳システム、情報処理学会論文誌, Vol.32, No.10, PP.1249-1259(1991).

15. Fuji, R., Lixin, F.: Persisting Structural Ambiguity and Its Application in Japanese-Chinese Machine Translation, Proc. International Conference of Recent Advances in Natural Language Processing, pp.301-307 (1995).