分散協調方式を用いた対話文の機械翻訳
Spoken Language Machine Translation Using Cooperative Distributed Method

任　福継
Fuji Ren
広島市立大学情報科学部
Faculty of Information Science, Hiroshima City University

1．はじめに

自然言語対話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，分散協調機械翻訳（CDMT: Cooperative Distributed Machine Translation）方式を提案する．これは既に開発している日中機械翻訳システム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，実際なシステム構造と若干異なるモデルを用いる．

分散協調方式は，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．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CDMT手法をSWKJC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．

2．分散協調機械翻訳の構成

分散協調機械翻訳の概念及び構成を図3に示す．

2．1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．必要なときは，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．

2．2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1】
A: 夕べ何を食べた？
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連接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1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，分散協調方式は「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．

２．３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２】
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．

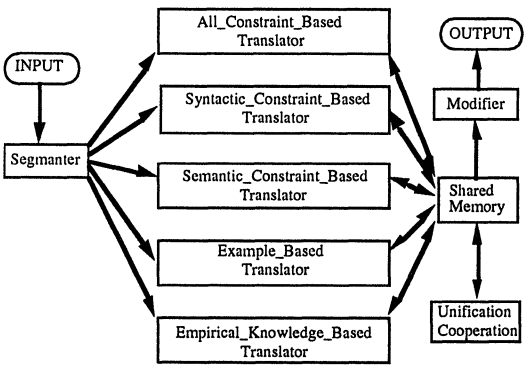

Figure 1 Concept of Cooperative Distributed MT

３．日中機械翻訳への応用例

分散協調型機械翻訳では，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，CDMT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．

３．１統語的にも意味的にも適格な場合

【例３】
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MARI',+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の部分を抜いたもの，統語的制約主導翻訳プロセスはSemの部分を抜いたものを用いる．
例３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.

3．2　統語的に不適格な場合

【例4】
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．

3．3　意味的に不適格な場合

【例5】
A. 人民は祖国を愛する．
B. 祖国は人民を愛する．

例文5A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}.....}

3．4　必須格の省略の例

【例6】
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．

これは必須格省略の例である．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，CDMT手法では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，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．

3．5　複数個所に省略した例

【例11】
A:　今晩何を食べる？
B:　私刺身食べたい．

この例のBでは2ヵ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．CPMT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).}
```

4．実験と考察

本論文で述べた分散協調型機械翻訳方式を，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．原システムとの比較評価のため，原システム（表中の「SWKJC-CF」で示す）と本文で提案した方式を組み込んだシステム（表中の「SWKJC+CF」で示す）の翻訳結果をそれぞれ表1に示す．

Table 1 The Rate of Correct Translation

	SWKJC-CF	SWKJC+CF
spoken	46.7%	81.5%
textbook	79.2%	86.3%

実験結果，格助詞「を」を省略した頻度は高く，302箇所で出現したが，CDMT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．実験結果から，CDMT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る．

5．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．

キーボートを入力の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，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分散協調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CDMT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も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，規則の自動或いは半自動的な学習機能の開発c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各位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 Discource, 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.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).

4.任，范，宮永，栃内：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).

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