

TR-I-199

対話翻訳システムのための文脈処理機構について

A Context Processing Mechanism
for a Dialogue Translation System

工藤育男

Ikuo KUDO

1991.3.8

概要

対話文を翻訳するためには、省略を補完したり、照応を同定する機構が必要になる。この処理機構は、規模の拡張に耐えられること、自然言語処理部とのマッピングを計ること、入力される文脈が乱れた場合にも耐えられることが要求される。この問題を解決するために、結束性(Cohesion)に基づいた文脈解析手法を提案する。まず、省略や照応を分析し、本当に文脈処理の必要な現象と必要のない現象とを区別する。それぞれの現象に対処するために、文脈処理機構と助述表現による省略補完の機構の二段階の機構をつくる。文脈処理機構には、入力文が、文脈上のどの文に対する発話であるのかを言語的知識(結束性に関する知識)により捕らえる。この結束性に関する知識を生成するために、形式的な定義を与え、言語データベースを利用して、多量に生成する。この知識を省略の補完や指示語照応に応用した結果、フォーカスの省略の補完には、オープンテストで66.0%、クローズドテストで76.6%、また、文脈指示の指示語の同定には、62.5%の結果を得た。また、文脈に依存しない省略には、助述表現を利用して補完を行い、95.5%の結果を得た。言語的知識のみを手掛かりとして、どこまで処理が可能なのか(効果および限界点)を示す。

©ATR Interpreting Telephony Research Laboratories

©ATR 自動翻訳電話研究所

目 次

1. はじめに	p.1
2. 対話における文脈的現象	p.2
2.1 文脈的現象	p.2
2.2 文脈的現象の機能と結束性について	p.3
2.3 対話における文脈的ロバスト性	p.4
3. 結束性に基づく文脈処理	p.6
3.1 処理の概要	p.6
3.2 結束性に基づく文脈処理	p.9
3.2.1 結束性に関する知識	p.9
3.2.2 結束性に関する知識の解釈	p.12
3.2.3 結束性に関する知識の生成方法	p.13
3.3 文脈処理の障害になりうる現象のとり扱い方	p.15
3.4 助述表現による省略の補完	p.17
4 評価実験	p.19
4.1 生成された知識に対する評価	p.19
4.1.1 生成された知識に対する適合率	p.19
4.1.2 生成された知識に対する被服率	p.20
4.2 処理機構の評価	p.20
4.2.1 処理機構に関する評価実験の方法	p.20
4.2.2 文脈処理機構に関する評価結果	p.21
4.2.2.1 フォーカスの省略の補完に関する評価結果	p.21
4.2.2.2 指示語の照応先の同定に関する評価結果	p.22
4.2.3 助述表現による省略の補完の評価結果	p.23
4.3 文脈処理部の文脈的ロバスト性の評価	p.24
4.3.1 文脈的ロバスト性に関する実験方法	p.24
4.3.2 文脈的ロバスト性に関する評価結果	p.24
4.4 考察	p.26
5. むすび	p.28
付録	p.31

1. はじめに

自動通訳システムを開発するためには、従来の書き言葉の機械翻訳システムとは違って、対話文固有の言語現象への対応を考える必要がある。対話文には、文脈と密接に絡んだ省略や照応が含まれており、対話文を翻訳するためには、文脈処理を行い、省略を補完したり、照応関係を同定する必要がある。

従来の文脈処理の研究[1,2]には、言語学と認知科学からの二つのアプローチがある。言語学の研究に、文脈省略に関する研究[3]がある。これは、日本語の対話文に頻繁に出てくるフォーカスの省略について説明を与えた研究である。最近では、Grosz等の理論[4~6]を発展させ、フォーカスの省略について扱おうとされようとしている。ただ、これらの理論は、まだ、計算機での検証は報告されておらず、すぐに、実際のシステムに応用することは期待できない。

一方、認知科学的なアプローチとして、一貫性の研究(Coherence)[7]やプランの研究[8]がある。これらの基本的な考え方は、一貫性に関する規則やプランに関する規則を用いて、談話構造を導出し、その導出過程で、ドメインに依存した知識を用いて、省略の補完等に応用するものである。しかし、これらの手法は、知識の生成を人手に頼っているため、あるトピックスに関する小規模のシステムでしか試されていないのが現状である。また、談話構造を導出するための計算量の問題やきれいな談話構造が得られないような入力にどう対処するか(文脈的なロバスト性の問題)という課題が残されており、そのままの形で翻訳システムに結合するのは難しいと思われる。

そこで、ここでは、談話構造といった発話の履歴を広域に構造化することはやめ、省略などへの応用することに重きを置いた局所的な(Local)範囲の文脈処理を行う。文脈解析手法としては、結束性(Cohesion)に基づいた解析[9,10]を提案する。すなわち、文脈的な現象を言語的な共起現象として捉え、この共起関係を言語的知識(結束性に関する知識)として定義する。この知識を使って、入力文が、文脈上のどの文に対する発話であるのか(結束性)を捕らえる。この結束性を捉えるための知識は、共起関係として形式的に定義できるので、言語データベース[11~14]を利用し、大量に抽出できる。それゆえ、対話翻訳システムの文脈処理部として使うことができる。また、この文脈解析手法は、結束性という観点から解

析しているので、一貫性のない文章にも、解析不能にならずに処理が行えるという特色がある。

2章では、実際の会話文における省略や照応などの文脈的現象を分析し、本当に文脈処理の必要な現象とその必要のない現象とを区別する。また、文脈処理の障害となる現象についても整理する。3章では、結束性に基づいた文脈処理機構について述べる。文脈処理部と解析部とのマッピングの問題を検討し、文脈処理部で用いるための知識を定義し、省略の補完や照応の同定に応用する。次に、言語データベースを利用し、知識を効率的に大量に生成する。4章では、この知識を用いて、言語的知識のみを手掛かりとして、どこまで文脈処理が可能なのかを実験により検証する。

2. 対話における文脈的現象

2.1 文脈的現象

省略や照応などの文脈的現象を分析し、本当に文脈処理の必要な現象とその必要のない現象とを区別する。

(1)省略: 日本語の省略には、文脈に依存する省略と依存しない省略がある。文脈に依存する省略として、フォーカスの省略がある。表1の例1における「持っていないのですが」の目的格の省略が良い例である。このようなフォーカスの省

表1 文脈的現象の例

Table 1 Examples of context-dependent phenomena.

種類	数	例文
文脈依存の省略	51	(例1)「クレジットカードの名前を教えてください」 「すみません。 <u>[〇ヲ]</u> 持っていないのですが」
文末の冗長性による省略	225	(例2)「 <u>[〇ヲ],[〇ニ]</u> クレジットカードの名前を教えてください」
文脈指示の照応	11	(例3)「講演をご希望なら、論文を送って下さい」 「 <u>それは</u> 、いつまでに、送ればよいですか」
現場指示の照応	7	(例4)「もしもし、 <u>そちら</u> 会議事務局ですか」 「はい、そうです」

調査対象:キーボード会話,12会話(222発話,384文)

略の場合には、補完されるべきものが、前の文脈により変わる。それに対し、日本語には、文脈に依存しない省略がある。表1の例2の「教えてくださいか」の主語や第二目的語の省略である。日本語の文末の表現「て教えてくださいか」という表現に、「誰々が」「誰々に」という情報が含まれているため、明示的に述べる必要がないのである。このような省略は、文末の表現により、補完されるべきものが決まる。省略されるべきものを陽に表現すると、文末の表現と冗長になるという意味で、「文末の冗長性による省略」と呼ぶことにする。頻度は、文末の冗長性による省略の方が高く、384文中225例、フォーカスの省略が51例あった。

(2) 指示照応: 指示語には、文脈指示と現場指示[15]の二種類の指示語がある。文脈指示とは、例3の「それ」のように、照応すべき実体が文脈中に存在するものをいい、現場指示とは、例4の「そちら」のように、照応されるべき実体が文中に表現されてないものをいう。

従って、これらの現象を扱うための処理機構としては、文脈に依存した現象とそうでない現象を区別して扱える機構を備える必要がある。

2.2 文脈的現象の機能と結束性について

フォーカスの省略や文脈指示の指示語の機能について考える。同じ内容を伝える場合でも、表2の例5から例7に示すように、(1)同一語の繰り返し、(2)フォーカスの省略、(3)文脈指示の指示語で表現する方法がある。これらの表現は、表現は異なるが、二つの文を結び付けるという言語的な機能としては同じ役割をしている。ここでは、上記(1)、(2)、(3)の表現で結び付けられた二つの文のことを、結束性がある文と呼ぶことにする。従って、フォーカスの省略や文脈指示がある文

表2 文脈的現象の機能

Table 2 The functions of context-dependent phenomena.

現象名	例文
(1)同一語の繰り返し	(例5) 「パンフレットを送りましょうか」 「もう、[パンフレットは]持っています」
(2)フォーカスの省略	(例6) 「パンフレットを送りましょうか」 「もう、[Ø]持っています」
(3)文脈指示の指示語	(例7) 「パンフレットを送りましょうか」 「[それは], もう持っています」

は、結束性が成り立っているわけである。

ここでは、どの文とどの文が結び付くかという条件を、共起現象として捉える。この共起条件を結束性に関する知識(3.2)と呼び、予め、システムに与えておく。この知識と文脈との整合を計ること(図1)により、文脈を解析する。その結果として、フォーカスの省略の補完や指示語の照応先の同定を得るものである。結束性がある文を文脈中から見つけだし、フォーカスの省略を補完したり、文脈指示語の照応先を同定することを、結束性を捉えるということにする。

会話文の中で、フォーカスの省略や文脈指示の指示語がどういう場合に起こるかということ、一つは、相手文脈[15](例えば、質問文と答えの文)、もう一つは、自己文脈[17](例えば、複文や同一発話中の文間)で起こる。しかも、文脈的現象の9割以上が一発話前の相手文脈に含まれている。従って、文脈処理といっても、むやみに履歴を保持し探索する必要はなく、ここでは、一発話前の相手文脈中に含まれる現象を取り扱うことにする。

例5)の文脈	送る<_,パンフレット,_>持つ<_,パンフレット>
例6)の文脈	送る<_,パンフレット,_>持つ<_,∅>
例7)の文脈	送る<_,パンフレット,_>持つ<_,それ,>
結束性に関する知識 (local cohesive knowledge)	送る<_,パンフレット,_>持つ<_,X> if{(X=パンフレット)or(X=∅)or(X=それ)} then(X=パンフレット).

図1 文脈と結束性に関する知識について
Fig. 1 Some contexts and local cohesive knowledge.

2.3 対話における文脈的ロバスト性

文脈処理を行うためには、質問文と答えの文の関係(結束性)を正しく捉える必要がある。この結束性を捉えるための障害となる現象を整理する(表3)。まず、二つ以上の質問がされる場合には、質問の順に答える(クロス)場合(例8)と、後の質問から先に答える(ネスト)場合(例9)のように、質問と答えの順番が変わる場合がある。また、並列名詞句(例10)や繰り返し(例11)により、質問と答えの対応関係が一对一になるとは限らない。また、文脈に依存した省略が、伝搬することがある(例12)。

これらの現象に対処し、正確に質問文と答えの文を捉えるためには、次の要件を満たす必要がある。(1)複数の文が一発話中に存在する場合には、順序関係が変わることがあるので、順序関係にかかわらず、可能性のある組合せ全てに対して結束性を調べる。(2)並列名詞句が存在すると、質問と答えの対応関係が一对多の関係になることがあるので、並列名詞句は展開し、結束性を調べる。(3)省略が伝搬していく場合があるので、解釈で得られた結果を記録する。(4)結束性を捉えるための条件を、結束性のある文だけを捉え、結束性のない文は捉えないように厳しく記述する。

表3 文脈の乱れの例
Table 3 Examples of complicated contexts.

原因	数	例
二重質問 (クロス)	2	(例8)「ホテルの名前を教えてください。京都駅から近いですか」 「[Øハ]京都ホテルと都ホテルです。どれも駅の近くです」
二重質問 (ネスト)	1	(例9)「会議の聴講をしたいのですが、どのような手続きをすればよい のですか? 参加料は、いくらですか?」 「登録料は100ドルです。登録申込み書をお送り致します。」
並列名詞句	24	(例10)「御名前と連絡先の電話番号を御聞きしたいのですが。」 「はい、会社名は、ATRです。電話番号は、06-211-5111で す。名前は、木田裕司と申します。」
繰り返し	7	(例11)「学生でもこの会議に参加できますか?」 「勿論、[Øニ]参加できます。誰でも[Øニ]参加できます。」
省略の伝搬	3	(例12)「参加料はいくらですか」 「[Øハ]100ドルです」 「いつまでに[Øヲ]支払えばよいですか」

調査対象:キーボード会話,12会話(222発話、384文)

3 結束性に基づく文脈処理機構

本章では、省略の補完や指示語の照応先の同定を行う機構を提案する。3.1で処理の概要についてのべ、3.2で文脈処理機構について述べる。3.3では、文脈処理での障害になりうる現象の処理について説明し、3.4で助述表現による省略の補完の機構について述べる。

3.1 処理の概要

処理の概要を簡単に説明する。2章で検討したように、省略や指示語には、文脈に依存した現象と依存しない現象があるので、図2に示すように二段階の処理を行う。文脈処理部では、フォーカスの省略の補完、文脈指示語の照応を行う。文脈に依存しない省略の補完は、次の助述表現[16]による省略の補完機構で行う。以下に、簡単な例について説明する。ここでは、処理の概要を簡単に説明する。

- (例13) 「申込書を郵送してください。」 ①
 「はい、[Øヲ] 送ります。」 ②

いま、②の文を処理する場合を考える。②の文には、文脈に依存した省略[Øヲ]と依存しない省略[Øハ, Øニ]がある。

(1)結束性に基づく文脈処理を行い、文脈に依存した省略を補う。中間表現は、文脈処理するには情報量が多いので、骨格となる格関係のみの構造(スケルトン)[9,10]にする。①と②の文は③と④のスケルトンになる。

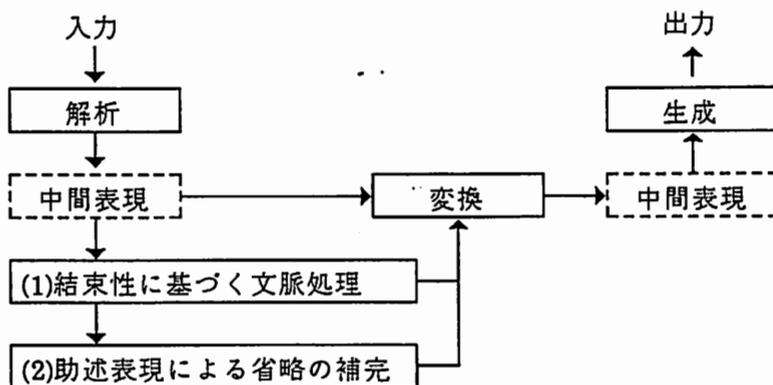


図2 対話翻訳システムにおける文脈処理機構の処理の流れ

Fig.2 A flow of a context processing mechanism for a dialogue-translation system.

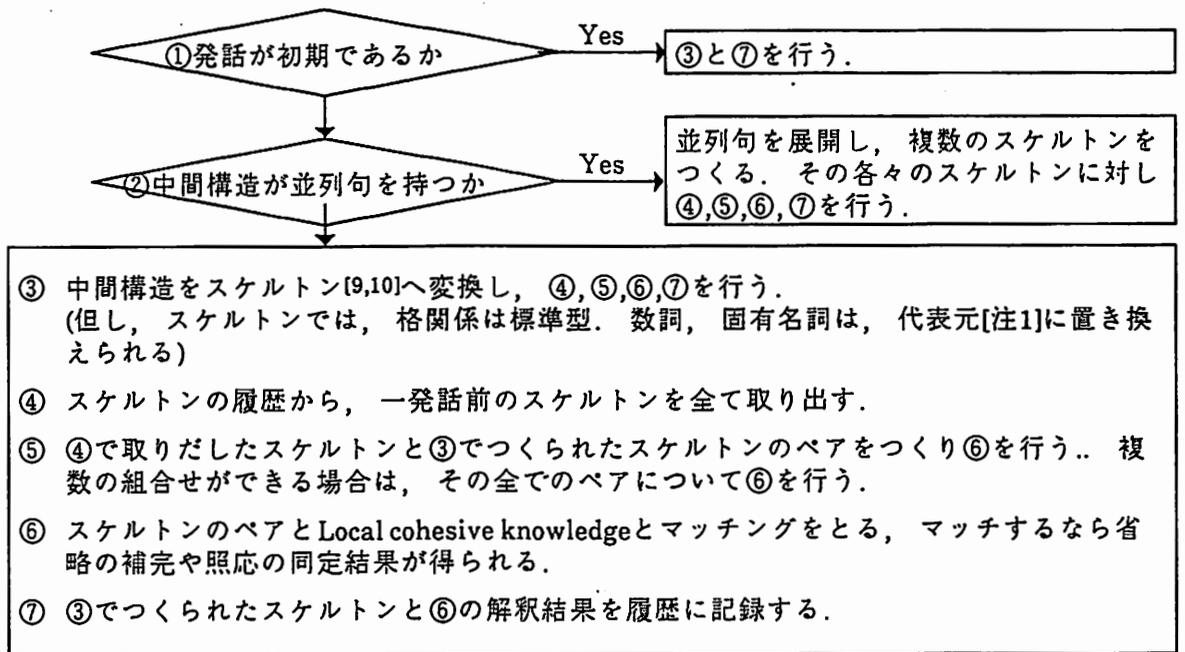
郵送する<_, 申込書, _> ③

送る<_, ∅, _> ④

文脈処理を行うために、履歴上のスケルトンのペアをつくる。③と④のスケルトンをペアにし、⑤ができる。

郵送する<_, 申込書, _>送る<_, ∅, _> ⑤

このペアと結束性に関する知識("local cohesive knowledge")とユニフィケーションさせる。もし、この知識とマッチするなら、そのペアは、結束性があり、その結果として、省略、照応の解釈が得られる。この例では、⑥の文脈処理をするための知識とマッチし、省略(X=申込書)が補完される。



スケルトンの履歴				Local cohesive knowledge	
発話 ID	述語 ID	スケルトン	解釈結果	結束性に関する制約	省略の補完 照応の同定
U ₁	f ₁	v ₁ <_, X>		v ₁ <_, X>, v ₂ <_, Y>	
U ₂	f ₉	v ₂ <_, Y>		v ₁ <_, X>, v ₃ <_, Z>	
U ₂	f ₁₀	v ₃ <_, Z>		:	
U _n	f _m	v ₄ <_, W>		v _n <_, Z>, v _m <_, W>	

図3 結束性に基づく文脈処理方法

Fig.3 A context processing mechanism based on "Local cohesive knowledge".

郵送する<_,申込書,_>送る<_,X,_>

if{(X=申込書) or (X=それ) or (X=∅)} then (X=申込書). ⑥

もし、マッチする結束性に関する知識がない場合には、そのペアは、結束性がないと判断する。得られた結果は、履歴の結果として記録する(図3)。

(2)文脈に依存しない省略は、各述語の助述表現[16]を利用し補完する。例13では、「送ります」という表現から「私ハ」「あなたニ」が推測される。データベースから大量に助述表現を抽出し、補完すべき省略といっしょに表にまとめておく。この表を引くことにより、省略を補完するものである。効率的な表を作るために、助述タイプ表と省略補完表の二つに分ける(図4)。助述タイプ表には、省略補完表を引くためのパラメータ(助述タイプ)が、省略補完表には、省略の解釈が書かれている。

- ① 必須格中の未補完の省略箇所を検知し、未補完の箇所があれば、以下を実行し省略を補完する。未補完の省略箇所がなければ、処理終了。
- ② 助述表現を中間構造から取り出す。
- ③ ④と⑤を行う時の表を決める。すなわち、例外処理の分類を行う。②の表現が、格関係を換える「もらう、やる、くれる」などの例外表現を含んでいる場合は、それぞれの例外表現に用意された助述タイプ表と省略補完表を用いて、④⑤⑥を行う。それ以外の場合は、通常の助述タイプ表と省略補完表を用いて、④⑤⑥を行う。
- ④ 助述タイプ表を引き、②助述表現の助述タイプを求める。
- ⑤ 動詞、動詞タイプ(解析用の語彙項目より得られる)、助述タイプから、省略補完表を引き、省略解釈値を求める。すなわち、未補完の補完箇所に対する、解釈値を得る。
- ⑥ ⑤で得た解釈結果をスケルトンの履歴に登録する。

助述タイプ表

助述表現	助述タイプ

省略補完表

動詞	動詞 タイプ	助述 タイプ	省略解釈値		
			主格	目的格	第二目的格

図4 助述表現による省略の補完

Fig.4 An ellipsis resolution mechanism by referring Japanese verbal phrase.

3.2 結束性に基づく文脈処理

ここでは、まず、結束性に関する知識について定義し、次に、フォーカスの省略の補完や文脈指示の指示語の同定へ応用する。次に、結束性の定義を利用して、言語データベースより、知識を効率よく生成方法について述べる。

3.2.1 結束性に関する知識

結束性に基づく文脈処理とは、どの文と文が結び付くかという条件を、共起関係として言語的知識(制約)として規定しておき、その知識と入力された文脈との整合を計ることにより、文脈を解析するものである。

この言語的知識(制約)として規定される共起条件について考える。例14は意味が通じるが、例15は通じない。このことは、文と文がどんな組合せでも、結び付きうるものではなく、結びつくためには、何らかの条件があることを示している。

(例14) 結束性のある場合(媒介となる名詞句が同一名詞):

「パンフレットを送りましょうか」

「もう、[Øヲ]持っています」

(例15) 結束性のない場合:

「パンフレットを送りましょうか」

「早く [Øヲ]食べたいです。」

フォーカスの省略について観察してみると、例14のような単純な場合の他に、受け手の文の発話者が文脈上の文のどの部分をフォーカスにするかによって、次のような現象が存在する。例16では、「論文の発表」の「論文」がフォーカスの省略として省略されている。文脈上の文の名詞句部分全体「AのB」が、

(例16) 媒介となる名詞句が「AのB」とA:

「論文の発表をしたいのですが」

「それでは、10日までに[Øヲ]送って下さい」

(例17) 媒介となる名詞句が複合語と複合語の一部:

「申込み期限を教えてください」

「もう、[Øハ]締め切りました」[Ø=申込み]

フォーカスになるのではなく、修飾語の「A」部分がフォーカスになっている。ここでは、名詞句「AのB」を「フォーカスの先行詞」、フォーカスとして省略される「A」の部分を「フォーカス部分」、この両者を併せて、「媒介となる名詞句」と呼ぶことにする。また、例17のように、複合語の一部が、フォーカスとして省略される場合もある。「申込み期限」の「申込み」の部分がフォーカスとして省略されている。従って、「媒介となる名詞句」により、フォーカスの省略を分類すると、次の三種類がある。

- (a) 同一名詞の場合 (例14)、
- (b) 名詞句「AのB」と修飾語「A」 (例16)、
- (c) 複合語と複合語の一部 (例17)。

では、どの部分がフォーカスとして省略されるのかは、何によって、決まるのだろうか。例えば、例18と例19は、同じ質問文でありながら、それを受ける文によって、フォーカス部分が変わっている。このことは、フォーカスが、ただ単に、単独の文の作用によって決まるのではなく、文脈上の文とそれを受ける文の共起によって決まると考えられる。

(例18) 媒介となる名詞句が同一名詞:

「クレジットカードの名前を教えてください」

「[Øハ]、VISAです」

(例19) 媒介となる名詞句が「AのB」とA:

「クレジットカードの名前を教えてください」

「[Øハ]、持っていないのですが」

ここでは、フォーカスの省略などの結束性に関する現象を、共起現象として捉える立場をとることにする。この共起条件には、少なくとも、次の条件が必要である。

- (1)フォーカスの先行詞(X)が、文脈上の文の格要素になること、
- (2)この先行詞(X)に対し、媒介となる名詞句を構成するフォーカス部分(Y)が存在すること、
- (3)このフォーカス部分(Y)が受ける文の格要素になること。

この三つの条件を、一つの式として定式化する。

verb1<X1, X2, X3>, verb2<Y1, Y2, Y3>

但し、verb1は文脈上の文の述語、verb2はそれを受ける文の述語を表し、格要素の配列を<X1, X2, X3>で表し、標準形での主格、目的格、第二目的格に対応するものとする。このとき、各要素X2とY2が媒介となる名詞句を構成する場合、二つの文が結束性をもつための条件は、

verb1<_, X, _>, verb2<_, Y, _>

と表す。媒介となる名詞句のみ表記し、残りの格要素は、制約とはならないので、“_”で記す。例えば、例14、例16、例17のそれぞれが、結束性をもつための条件は、以下のようなになる。

例14 送る<_,パンフレット, _>持つ<_,パンフレット>

例16 する<_,(論文)発表>送る<_,論文, _>

但し、()は、修飾語を意味する。

例17 教える<_,申込み期限, _>締め切る<_,申込み>

この条件を、結束性に関する知識(制約)と呼ぶことにする。

(例20) ですを含む結束性のある場合:

「登録料は、いくらですか」

「[Øハ]100ドルです」

但し、「です」を述語とする場合には、先の三つの条件だけでは弱く、制約を付加する必要がある。例えば、例20では、「いくら」、および、[金額](注1)の条件を追加している。

例20 です<登録料,いくら>です<登録料,[金額]>

(注1) []は、カテゴリーに置き換えてからのマッピングを意味する。

[金額]=[数字]+[金額的単位]. [金額的単位] ∈ {円,ドル, etc}.

ここでは、[]の中に記述される記号を代表元と呼ぶことにする。意味カテゴリーと呼ばない理由は、代表元の方は、統語構造が異なる名詞句も一つのカテゴリーとしてまとめているからである。例えば、

[名前]=[氏名]+[名字], [氏名], [名字].

[住所]=[地名]+[地名単位]+[地名]+[地名単位]+[地名]+[数字]+[記号]+[数字]+[記号]+[数字], [地名]+[地名単位]+[数字]+[記号]+[数字], [地名]+[地名単位], etc}.

[住所]には、「大阪,市,中央,区,城見,4,-,1,-,5」や「精華,町,2,-,1」や「東京,都」など統語構造が異なるものも含む。

3.2.2 結束性に関する知識の解釈

結束性に関する知識を、フォーカスの省略の補完や文脈指示語の照応先の同定に応用する。そのために、結束性に関する知識の一部を多段階のユニフィケーション操作に置き換える。例えば、

結束性に関する制約

送る<_,参加料,_>,送金する<_,参加料,_>.

Local cohesive knowledge

送る<_,参加料,_>,送金する<_,X,_>

if {(X=参加料) or (X=それ) or (X=∅)} then (X=参加料). ①

この機構により、フォーカスの省略の補完や文脈指示語の照応を行う。結束性に関する制約と解釈機構を併せて、Local cohesive knowledgeと呼ぶ。①式は、Xの値が、(1)値(参加料)、(2)代名詞(それ)、(3)空リスト(∅)、の3段階のユニフィケーションを行い、このいずれかが満足した場合、その値Xを(参加料)と解釈するものである。

(1)フォーカスの省略の補完

(例21) 「参加料を20日までに送ってください」

「はい、[Øヲ] 送金します」

この入力は、次のスケルトンのペアになる。

送る<X1,参加料,Z1>,送金する<Ø,Ø,Ø> ②

この②と①がマッチし、省略は(Ø=参加料)が補完される。

(2)文脈指示語の照応

(例22) 「参加料を今日中に送ってください」

「それは、今朝、送金したんですけど」

この入力は、次のスケルトンのペアになる。

送る<X1,参加料,Z1>,送金する<Ø,それ,Ø> ③

この③と①がマッチし、代名詞の照応先が(それ=参加料)同定される。

上の例では、媒介となる名詞句が同一名詞の場合について示したが、他の場合も同様である。

3.2.3 結束性に関する知識の生成方法

この文脈処理では、結束性に関する知識の生成が重要になる。ここでは、結束性の定義に基づいて、言語データベースから効率的に知識を生成する。そのために、動詞と格要素を規定した表(表4)、名詞句「AのB」を列挙した表(表5)、複合語を成分に分解した表(表6)を作成する。それぞれの集合を、集合S、集合N1、集合N2とすると、結束性に関する知識は、以下のようにして生成される。

(1)同一名詞: 集合Sから共通する名詞を格要素に持つ二つの動詞を全て抽出する。

(2)「AのB」とA: 集合N1から二つの名詞を選び、この二つの名詞を格要素として持つ二つの動詞を、集合Sから全て抽出する。

(3)複合語と複合語の一部: 集合N2から二つの名詞を選び、この二つの名詞を格要素として持つ二つの動詞を、集合Sから全て抽出する。

言語データベースの142会話(1528発話、4365文、103954異なり単語)から、結束性に関する知識を抽出した。このデータには、動詞350語(助動詞「です」を含む)、名詞1425語が含まれている。集合Sは、9112個、集合N1は、107個、集合N2は、1024個であった。但し、固有名詞、代名詞、数詞については、代表元に置き換え、要素は一個としてカウントした。その結果、表7に示すような知識が得られた。

表4 動詞と名詞の格関係の例

Table 4 Examples of case relationship.

教える<_,参加料,_> 送る<_,参加料,_> 送る<_,アブストラクト,_> 送る<_,料金,_> 払う<_,登録料金,_> 参加する<_,会議>	行く<_,会議> 行く<_,観光> 行く<_,発表> 持つ<_,アブストラクト> お願いする<_,用意> 予約する<_,ホテル> 泊まる<_,ホテル>	募集する<_,発表> 予定する<_,市内観光> する<_,講演発表> 募集する<_,講演> 始まる<会議> ある<希望>
--	---	---

表5 AのB知識

Table 5 The table of noun phrases "A of B".

A	B
ホテル	用意、希望、部屋、方、
会議	発表、場所、委員、方、記事、期間、公告、パンフレット、参加
宿泊	ホテル、方、事
公告	料金、カメラ・レディ
先生	方、来日

表6 複合語の例

Table 6 Examples of compound noun.

複合語の一部	複合語
料金	登録料金、一律料金、一般料金、掲載料金、宿泊料金、追加料金、通常料金、割引料金、予約料金、税込み料金、
観光	市内観光、観光案内、観光名所
講演	講演発表、講演者、特別講演
発表	ビデオ発表、演題発表、御発表、各発表、研究発表、講演発表、論文発表、
申込み	ホテル予約申込み、事前申込み、掲載申込み、講演申込み、御申込み、参加申込み、展示申込み、予約申込み

表7 結束性に関する知識の生成結果

Table 7 Local cohesive knowledge generated from a corpus.

媒介となる名詞句	個数	生成された知識の例と その知識で捉えうる結束性のある文の例
同一名詞	41077	教える<_,参加料,_>送る<_,参加料,_> 「参加料について教えてください」 「[Øハ],為替で学会宛に21日までにお送りください」 参加する<_,会議>行う<_,会議> 「会議に参加してくれますか」 「[Øハ],10日に行うのですね」 持つ<_,アブストラクト>送る<_,アブストラクト,_> 「アブストラクトを持っていますか」 「[Øヲ],送ってください」
「AのB」 とA	32741	お願いする<_,ホテルノ用意>予約する<_,ホテル> 「ホテルの用意をお願いします」 「[Øヲ]予約しておきます」 募集する<_,会議ノ発表>始まる<会議> 「まだ,会議の発表を募集していますか」 「えっ?もう,[Øハ]始まっていますよ」 ある<ホテルノ希望>泊まる<_,ホテル> 「ホテルのご希望はありますか」 「今週の土曜日に,[Øニ],泊まりたいのですが」
複合語と その一部	37255	払う<_,登録料金,_>送る<_,料金,_> 「登録料金をドルで払って頂けますか」 「すぐに,[料金ヲ],送ります」 予定する<_,市内観光>行う<_,観光> 「市内観光を,予定していますか」 「20日に,[観光ヲ],予定しています」 する<_,講演発表>募集する<_,講演> 「講演発表をしたいのですが」 「まだ,[講演ヲ],募集しています」

3.3 文脈処理の障害になりうる現象のとり扱い方

ここでは、2.3で述べた処理の障害になりうる現象のとり扱いについて説明する。

(例) 「お名前とあて先を教えてください」

「木田裕司と申します。大阪市東区城見23です。」

(1)並列句の場合は、入力された並列句を展開する。どのように展開すべきかは、並列句の文法規則に記述しておく。例えば、単純並列句の「お名前とあて先を」場合には、「お名前」と「あて先」の二つに展開し、次のスケルトンを作る。

- 教える<_,名前,_> ①
 教える<_,あて先,_> ②

(2)複文、ならびに、一発話中に複数の文がある場合には、可能な組み合わせ全てとマッチさせる。例えば、先の例では、展開された二つのスケルトン(①と②)と入力された二つの文のスケルトン(③と④)を組合せて、四組のスケルトンペア[①と③,②と③,①と④,②と④]をつくる。このペアを⑤や⑥の知識とマッチさせ、四組のうち二組がマッチし、「名前」と「あて先」を補う。

- 申す<_,[名前]> ③
 です<_,[住所]> ④
 教える<_,名前,_>申す<名前,[名前]> ⑤
 教える<_,あて先,_>です<あて先,[住所]> ⑥

もし、③や④の知識が抽出できずに補完できなかった場合は、助述表現による補完の機構により、デフォルト的な値がうまる。

(3)フォーカスの省略は伝搬する場合があるので、補完された格要素、および、同定された指示語は、補ってから文脈処理のための知識とマッチさせる。例えば、表3の例12の場合には、「参加料はいくらですか」と「[Øハ] 100ドルです」知識により「参加料」が補完される。この補完された「参加料」を使って、「[参加料は] 100ドルです」と「[Øヲ] いつまでに支払えばよいのですか」を⑧とマッチさせる。

- です<参加料,いくら>です<参加料,[金額]> ⑦
 です<参加料,_>支払う<_,参加料,_> ⑧

3.4 助述表現による省略の補完

ここでは、各述語の助述表現を利用し、文脈に依存しない省略に対処する。「お名前をお聞きしたいのですが」といったら、動詞に後接している助述表現「したいのですが」という表現から、省略されているものが「私ハ、あなたニ」であることが推測される。このような助述表現を、データベースから大量に抽出し、補完すべき省略といっしょに表にまとめておく。この表を引くことにより、省略を補完する。

表は、二つの表からなっている(図4)。

(1)どの格にどの語を補うべきかは、動詞の種類により決まるので、動詞の格ごとに、省略補完表(表8)を設定する。各動詞の持つ必須格の数を、動詞タイプと呼ぶ。構文解析用の辞書に埋め込まれている。

(2)助述表現により、動詞に補完すべきか値が変わるので、これをパラメータ化し、助述タイプと呼ぶ。例を表9に示した。

例えば、「お電話番号を教えてくださいか」の場合、表9より「て頂けますか」の助述タイプ=2、「教える」の動詞タイプ=3であるので、表8より、主格に「あなた」、第二目的格に「私」が補完される。

(3)例外として、「もらう」「やる」「くれる」など特定の表現により、格の数、および、補うべき語が変わる。別表を設定し対処する。

表8 省略補完表
Table 8 Default table of Japanese ellipsis.

動詞	動詞 タイプ	助述 タイプ	省略解釈		
			主格	目的格	第二目的格
ある,ござる, 込む,空く, 迫る, 違う, 出来上がる, 終わる, 始まる, できる, 当たる, 遅れる,	1	*	それ		
*	1	1	私		
		2	あなた		
です,含む, 取れる	2	*	それ	それ	
お待ちする,伺う	2	*	私	あなた	
届く	2	1	それ	私	
		2	それ	あなた	
*	2	1	私	それ	
		2	あなた	それ	
*	3	1	私	それ	あなた
		2	あなた	それ	私

* 任意の動詞を表す

表9 助述表現と助述タイプ
Table 9 The relationship between the form and its type.

	助述タイプ	
	1	2
させる	させて頂きたいのですが、させて	
せる	せて頂きたい、せたい、せします、	せ願えますか、せて頂けますか、
れる	れるのでしょうか、れます、	れるのでしょうか、れるのか、
られる	られるのでしょうか、られません、	
たい	たいんですが、たいんです、	たいんですか、
頂く	て頂きたいのです、	て頂けますか、
願える		願います
おる	ております	ておられる、でおられる、
いらっ しゃる		でいらっしゃいますか、
下さる		てくださいますか、ください
いたす	致します、致しますが、	
ます	ましようか、ます	
その他	ことはできます、 ことはできますか、 なければなりませんか、 してもいいですか	か、 のですか、 なければなりません、 してもいいです

4 評価実験

3章で提案した文脈処理機構の評価実験を行う。4.1では、(3.2.3)で生成された知識について評価を行い、4.2で、この知識を使い、省略の補完や指示語の照応先の同定に関する処理能力について調べる。4.3では、2.3で述べた文脈処理の障害になる現象に対する評価実験(文脈的ロバスト性に関する実験)を行う。4.4で、この処理機構全体に関する評価結果を整理し、今後の拡張の可能性について考察する。

4.1 生成された知識に対する評価

(3.2.3)で生成された知識を評価する。適合率と被服率(再現率)について調べる。

4.1.1 生成された知識に対する適合率

(3.2.3)で生成された知識のうちの何%に、結束性があるか調べる。すなわち、生成された知識の適合率を調べる。

ここでは、生成された知識のうち100個を、ランダムサンプリングし、その知識に対して、結束性のある文がつかれるかどうかを評価した。評価の判断は、男女3人による主観評価によって行った。評価の基準には、次の三段階を設けた。(1)現実世界と照らし合わせて、二つの文が使用される状況がすぐに思い浮かぶ。(2)二つの文が使用される状況が、すぐには思い浮かばないが、言語表現として言えそうである。(3)言語表現としても言えそうにない。

その結果を表9に示した。(1)と(2)を合わせたものを目的とする知識、(3)をノイズと考えると、各知識の適合率は、92%、82.2%、98.2%であった。生成された知識の大多数について、結束性を持つような二つの文が作れることが分かった。また、ノイズとなる知識が生成された理由は、言語データベースに翻訳調の表現が含まれていたために生じたものである。

表9 結束性に関する知識の適合率と被服率

Table 9 The precision and coverage of local cohesive knowledge.

媒介となる名詞句	主観評価				被服率の評価		
	(1)	(2)	(3)	適合率	A	B	B/A
同一名詞	79.0%	13.0%	8.0%	92.0%	47個	34個	72.3%
「AのB」とA	79.0%	3.2%	17.8%	82.2%	3個	2個	66.0%
複合語とその一部	84.7%	13.5%	1.8%	98.2%	1個	1個	100%

4.1.2 生成された知識に対する被服率

(3.2.3)で生成された知識で、あるデータを処理するのに必要な知識の何%をカバーできるかを調べた。すなわち、生成された知識の被服率を調べた。

評価の対象としては、(3.2.3)で学習に使用した学習データ(142会話)とは異なる評価用データ(12会話、384文)を使用した。この会話を処理するのに必要な知識を手で抽出し(表9のA欄)、そのうち生成された知識に何個(表9のB欄)含まれているか調べた。

その結果を表9に示した。知各識の被服率は、72.3%、66.0%、100%であった。

4.2 処理機構に関する評価

処理機構が、処理能力について評価した。(4.2.1)で実験方法について述べ、(4.2.2)で文脈処理機構の評価結果 (4.2.3)で、助述表現による省略の補完機構の結果をのべる。

4.2.1 処理機構に関する評価実験の方法

評価実験は、オープン・テストとクローズド・テストを行った。(1)オープン・テストは、評価データと学習データを分け、学習させていない評価データに対し

どのくらい有効なのかを計った。次に、(2)評価データも学習させたクローズド・テストを行い、処理する対象が限定できる場合、この手法がどの程度、有効なのかを調べた。

文脈処理の実験では、文脈が意味を持つので、評価文を任意に抽出することができない。そこで、会話全体を評価データとして使用した。評価データには、(4.1.2)で使用した会話を使用した。

評価は、対話翻訳システム DialTalk [12,19] で行った。システムの環境は、SUN-4, UNIX 4.0.1, Quintus Prolog (version 2.4) である。翻訳システムは、Lexical-functional Grammar (LFG)[20] に基づいて構文解析し、動詞と名詞の共起情報、動詞と助詞との共起情報を利用し、関数名を決定する。いわゆる、格解析を行う。辞書のエントリーは全コーパスから抽出し(活用語も含む3700エントリー)、文法規則は、評価データに関する193規則(多進木)のみ加えた。尚、解析の途中で失敗した場合は、解析結果を人手でつくり、その段階から評価を行った。

4.2.2 文脈処理機構に関する評価結果

(4.2.2.1)でフォーカスの省略の補完の結果、(4.2.2.2)で文脈指示語の照応先の同定の結果について述べる。

4.2.2.1 フォーカスの省略の補完に関する評価結果

評価データには、フォーカスの省略が51箇所あった(表10)。そのうち、オープン・テストで31箇所、クローズド・テストで36箇所の省略の補完に成功した。

失敗の原因には、知識の不足による失敗とアルゴリズムが原因の失敗がある。知識の不足による失敗には、(1)知識が抽出に失敗した場合、(2)知識を抽出するための定義がなかった場合の2種類があった。アルゴリズムが原因で失敗した場合には、(3)同一発話内にフォーカスの先行詞が存在する場合(例2-1)、(4)一発話以上前の発話にフォーカスの先行詞がある場合(例2-2)があった。

処理の対象ではなかった(3)と(4)を除くと、処理の成功率は、オープン・テストで66.0%、クローズド・テストで76.6%であった。また、文脈処理すべきでない省略に対して、処理が誤って働くことはなかった。

4.2.2.2 指示語の照応先の同定に関する評価結果

評価データには、文脈指示の指示語が11箇所あり、そのうち、5箇所に成功し、6箇所失敗した(表11)。失敗の原因には、次の3種類があった。(1)同一発話内にある語を照応する場合(表11の例2)。この問題には、アルゴリズムを拡張し対処する必要がある。(2)指示範囲を誤まった場合。(例3)の「それヲ」は前の文全体を指す。この問題については、多段階のユニフィケーション機構の部分を、動詞の種類と指示範囲の関係を捉えられるように修正する必要がある。(3)任意格にある指示語(例4)。この方法では、任意格まで扱ってないので対処できない。

処理の対象でなかった(1)を除くと、処理の成功率は、オープン・テストで62.5%であった。今回は、クローズド・テストによって改善するみこみがなかつ

表10 文脈依存の省略補完に関する結果
Table 10 The results of resolution of contextual ellipsis.

	オープン	クローズド	例
文脈的省略の補完成功	合計 31	合計 36	(例1)「学生でも、この会議に参加できますか」 「誰でも[Øニ]参加できます」 (例2)「登録料はいくらですか」 「[Øハ]100ドルです」
文脈的省略の補完失敗	合計 20	合計 15	(1-2)「会議は何語で運営されるのですか」 「[Øハ]英語と日本語です」 (2-1)「娘は会議には参加しないのですが、 [Øハ]観光だけ参加出来ますでしょうか」 (2-2)「8月に国際会議が開かれると聞いたのですが、」 「申し訳ありませんが、9月で締切らせて戴きました」 「そうですか。それでは、聴衆として [Øニ]参加させて戴きたく思います が、....」
(1)知識の問題			
(1-1) 抽出方法	14	9	
(1-2) 定義の漏れ	2	2	
(2)アルゴリズムの問題			
(2-1) 同一発話内に解がある	3	3	
(2-2) 一発話以上前の発話に解がある	1	1	

たので、この評価は行わなかった。尚、評価データには、現場指示の指示語があったが、誤って処理することはなかった。

表11 指示語の同定に関する結果
Table 11 The results of identification of anaphoras.

	結果	例
文脈指示の指示語の同定に成功	5	(例1)「そちらは会議事務局ですか」 「はい、 <u>そうです</u> 」
文脈指示の指示語の同定に失敗	合計 6	(例2)「会議の始まる1か月前までにキャンセルされる場合は、 <u>全額お返しいたしますが、それを越えると</u> 、20%のキャンセル料を申し受けます」
(1) アルゴリズムの問題、同一発話内に解がある場合	3	(例3)「会議の多くは、日本語で行われます」 「 <u>それを</u> 聞いて、安心しました」
(2) 指示範囲の誤り	1	(例4)「支払いの締切りは、6月30日です」
(3) 任意格の場合	2	「 <u>それまでに</u> 、お支払いします」
現場指示の指示語	7	(例5)「 <u>そちら</u> は会議事務局ですか」

4.2.3 助述表現による省略の補完の評価結果

3.4で述べた助述表現による省略の補完機構について調べた。処理すべき箇所が245箇所あり、そのうち234箇所について成功した(表12)。非常に簡単な機構でも、95.5%の省略の補完ができた。ただし、ここでいう成功とは、人称格が合っていれば成功とし、単複の差は区別しなかった。その理由は、翻訳システムをつくる上では、一般名詞の単複の差も扱っていないのに、代名詞だけの単複の差を論じて、意味を持たないからである。

失敗の原因には、以下のものがある。(1)「ダ文」の主語の補完: 「ダ文」の主語の補完には、文末からの情報だけでは不十分で、「ダ文」の補語の情報を使う必要がある(表12の例1~4)。(2)一つの助述表現に複数の解釈の可能性が存在する場合: 「ことができますか」には、許可と可能の二つの意味があり、助述表現の情報だけでは補完すべき省略を、「私ハ」なのか「あなたハ」なのか決定できない。現在のところ、どちらか一方が誤る結果になるが、複数の解釈を持つ

助述表現は限定できるので、この情報を利用し、対処できるかもしれない。(3)助述表現のテーブルへの登録漏れ: 助述表現が学習データから抽出できなかった場合である。(4)同一発話内の発話に解が存在する場合: 文脈処理の失敗である。自己文脈まで文脈処理を拡張する必要がある。

この助述表現による補完の機構は、文脈処理で失敗した20件のうち17件の補完に成功している。処理機構全体としては、文脈に依存した現象とそうでない現象を区別して扱う二段階の機構が機能している。

表12 文末処理のテスト結果

Table 12 The results of resolution of non-contextual ellipsis.

成功						失敗
SUBJ			OBJ2		OBJ	合計
私	あなた	それ	私	あなた	それ	
84	24	37	10	33	46	
小計145			小計43		小計46	
合計 234						11

失敗の原因	数	例
(1) だ文の主語, 「それ」以外の語を補う場合	4	(例1) 竹中様ですね。(×それ→○あなた) (例2) どちらから、お越しでしょうか。(×それ→○あなた) (例3) 東京からです。(×それ→私) (例4) 会議への参加をご希望ですか。(×それ→○あなた)
(2) 一つの助述表現に複数の解釈が存在する場合	3	(例5) 明日までに、登録料を支払うことができますか (×私ハあなたニ→○あなたハ、私ニ) (例6) バスに乗りますと(×私→○あなた)、30分から40分で行けます。 (例7) 30分から40分で行けます。(×私→○あなた)
(3) 助述表現のテーブルへの登録漏れ	1	(例8) 振り込ませて頂きます。 (×あなたハ、私ニ→○私ハ、あなたニ)
(4) 同一発話内に解がある場合	3	(例9) 娘は会議には参加しないのですが、(×私→○娘ハ) 観光だけ参加できますでしょうか

4.3 文脈処理部の文脈的ロバスト性の評価結果

2.3で述べた文脈処理の障害になる現象に関して、この処理機構がどの程度対処できるのかを調べた。

4.3.1 文脈的ロバスト性に関する実験方法

評価データから、文脈処理の障害になる現象を選び出し(表12)、処理システムにかけた。但し、この実験では、処理に必要な知識は、予め全て与えておき、アルゴリズムおよび処理機構上の問題点を探った。

4.3.2 文脈的ロバスト性に関する評価結果

(1) 答えの文の順序関係が変わる場合(クロス)には、省略の補完に成功した(表12)。ネストは、評価データ中に1件あったが、これには文脈処理を要する現象は含まれていなかった。

(2) 並列名詞句は、評価データに24件あり、このうち、省略の補完を必要としたものが5件あった。補完に成功したものが3件、失敗したものが2件(表12の例3や例4)あった。

(3) 文が繰り返す場合、評価データ中に1件あったが、補完に成功した。

(4) 省略が伝搬する場合も、補完に成功した。

今回の実験では、(1)一発話中に複数の文が存在する場合の取扱い、(3)文が繰り返す場合、(4)省略が伝搬する場合への対応は機能した。しかし、(2)並列名詞句の取扱いについては、十分ではなかった。並列句は分解してマッピングすればよいという単純な場合だけでなく、表12の例3や例4のように自然言語の答え方についての詳しい調査が必要である。

表13 文脈的ロバスト性のテスト結果
Table 13 The results of contextual robustness.

種類	A	B	C	D	例
二重質問 (クロス)	2	2	2	0	(例1)「ホテルの名前を教えてください。京都駅から近いですか」 「[Øハ]京都ホテルと都ホテルです。どれも近くです」
(ネスト)	1	0	0	0	
並列名詞句	24	5	3	2	(例2)「名前とご住所をお願いします」 「[Øハ]木村次郎です。[Øハ]大阪市東区城見2-1-4です」 (例3)「カードの名前と番号をお願いします」 「[Øハ]アメリカンエクスプレスの333-1222です」 (例4)「お名前とご住所をお願いします」 「[Øハ]竹中悦子、[Øハ]愛媛県松山市南与護町1-3、娘は真理子です」
繰り返し	7	1	1	0	(例5)「学生でもこの会議に参加できますか?」 「勿論、[Øニ]参加できます。誰でも[Øニ]参加できます。」
省略の伝搬	3	3	3	0	(例6)「参加料は、幾ら位ですか。」 「[Øハ]アメリカドルで、100ドルです。」 「[Øハ]聴講だけでですか。」

(注意) A=評価データ中に現れる現象数,
B=Aの中で文脈処理(省略や照応)を要するテストデータ数,
C=成功した数,
D=失敗した数

4.4 考察

実験結果を整理し、処理の拡張について検討する。

(1)処理アルゴリズム上の問題: 今回の実験では、処理の対象を直前の相手文脈だけに限定したが、もう少し一般化するためには、処理範囲を、自己文脈、および、一発話以前の発話まで拡張する必要がある。ただし、範囲を拡張する場合でも、人間の記憶を考えれば、無限に発話の履歴を探索する必要はなく、範囲を限

定してよいと考えられる。今回の評価データでは、二発話前まで捜せば解は存在していた。

(2) 結束性に関する知識の問題:

(a) 知識の定義方法の問題: 制約の記述としては、結束性のある文だけを捉え、結束性のない文は捉えないように厳しく記述されており、文脈とは関係ない現象を誤って処理することはなかった。ただ、多段階のユニフィケーションの機構を修正する必要がある(4.2.2.2)。

(b) 定義の漏れの問題: (3.2.1)の3つの定義以外にも、定義が存在することが明確になった(表10)。新たな定義を見つけてゆく必要がある。

(c) 言語データベース: クローズド・テストを行っても知識を完全に抽出することはできなかった。この原因の一つは、省略は省略のままデータベース化されているためである。もし、クローズド・テストで100%を期待するのなら、省略を補った言語データベースを構築する必要がある。

(d) 知識の一般化の問題: 知識を一般化するためには、抽出しきれない知識への対応を計る必要がある。このためには、抽象化したカテゴリーを導入することや新たなモデルを導入し知識を生成することを検討する必要がある。

(e) 知識の過剰生成の問題: (4.1.1)で述べたように生成された知識のうち、1割くらいはノイズとなるものが含まれていた。しかし、このノイズが処理に悪い影響を及ぼすことはなかった。

(3) 処理機構上の問題: 構造の違うマッピングの問題については、構文解析部で使用する文法規則が分かれば、文脈処理部の入力構造が決まり、ソフトウェアとして対処可能である。しかし、並列句のように言語現象として処理しづらい現象が残っており、これらの対応を計る必要がある。

(4) インプリメント上の問題: 処理速度は平均0.127秒で問題はない。記憶容量については、3000語程度の規模で8.1MBも要するので改良する必要がある。

5. むすび

対話翻訳システム構築のための省略補完方法、指示語の同定手法について述べた。文脈に依存した現象とそうでない現象を区別し、二段階で処理を行う機構を提案した。

文脈的現象を共起現象として捉える、結束性に基づく文脈処理手法を提案した。この共起関係を生成するために、形式的定義を行い、言語データベースを利用して、効率よく大量に抽出した。抽出した知識の主観評価を行い、知識の定義、および、生成方法の妥当性を検証した。

この知識を処理に応用し、省略の補完や照応関係の同定を行った結果、フォーカスの省略の補完には、オープンテストで66.0%、クローズドテストで76.6%、また、文脈指示の指示語の同定には、62.5%の結果を得た。今後の課題として、言語データベースの利用技術の向上、処理の成功率の改善(知識の抽出やアルゴリズムの拡張)や並列句などの文脈的に処理しづらい現象への対応の問題が残されている。

文脈に依存しない省略には、助述表現を利用して補完を行い、95.5%の結果を得た。「ダ文」の主語の省略の補完や一つの助述表現に対して複数の解釈が可能な場合への対応が、今後の課題である。

最後に、この研究は、言語データベースがあったからこそできた研究である。今後、言語データベースの重要性が認識され、手工業的な知識処理の時代から抜け出し、工学としての言語処理の時代が確立されていくことを願うものである。

謝辞

この研究に際しまして、貴重な助言をいただいたATR自動翻訳電話研究所 榎松明社長、江原暉将主幹研究員、鈴木雅実主任研究員、友清睦子研究員、ならびに、討論に参加していただいた研究員の方々に感謝いたします。また、言語データベースを利用させていただいた森元暹室長、小倉健太郎主任研究員(現在NTT復帰)、井上直己研究員、橋本一男研究員(現在オージー情報システム株式会社復帰)、幸山秀雄研究員、篠崎直子研究員(現在退社)に感謝いたします。

参考文献

- (1)石崎, 井佐原「文脈処理技術」情報処理学会誌, Vol.27, No8, 1986.
- (2)情報処理:「大特集:自然言語理解」Vol30, No10, (1989).
- (3)久野すすむ「日本文法研究」大修館書店(1973).
- (4) Barbara J. Grosz, Aravind K. Joshi & Scott Weinstein : “Providing A Unified Account of Definite Noun Phrases in Discourse”, Proceedings of the 21st Annual Meeting of Association of Computational Linguistics, Cambridge, Mass, pp.44-50, (1983).
- (5) Barbara J. Grosz & Candance L. Sidner : “Discourse structure and the Proper treatment of Interruptions”, Proceedings of Ninth International Joint Conference on Artificial Intelligence, pp.832-839 (1985).
- (6) Megumi Kameyama : “A Property-sharing Constraint in Centering”, Proceedings of the 24th Annual Meeting of Association of Computational Linguistics, pp.200-206, (1986).
- (7) Hobbs, Jerry.R. : “Coherence and Co-reference”, Cognitive Science, 3(1), PP.67-90.
- (8) 飯田, 有田「4階層プラン認識モデルを使った対話の理解」情報処理学会論文誌Vol.31, No.6, (1990, 1) pp.810-821.
- (9) 工藤育男:「文と文の結束性を捕らえるための知識」情報処理学会自然言語研究会資料76-7, (1990,3,9).
- (10) Ikuo Kudo : “Local Cohesive Knowledge for a Dialogue-machine Translation System”, Proceedings of the 13th International Conference on Computational Linguistics, Helsinki, August, Vol. 3, pp.391-393, (1990).
- (11) 小倉, 橋本, 森元「言語データベース統合管理システム」情報処理学会自然言語処理研究会69-4, (1988.12.6).
- (12) 橋本, 小倉, 森元「フレーム表現による検索機能を有する言語データベース管理システム」 Proceedings of Advanced Database Symposium' 89 , Kyoto Research Park, Kyoto, Japan, December 7-8, (1989).
- (13) 井上, 小倉, 森元「係り受け意味関係の問題点とその考察」電子情報通信学会, 研究会資料NLC87-25, (1987).
- (14) Terumasa Ehara, Kentaro Ogura and Tsuyoshi Morimoto, “ATR Dialogue Database” International Conference on Spoken Language Processing, Koube (1990).

- (15) 水谷静夫編:林四郎, 荻野綱男, 田中幸子, 樺島忠夫, 朝倉日本語新講座, 5.運用 I 『林四郎著「一章, 代名詞がさすもの, その指し方」』1983年, 朝倉
- (16) 首藤, 榎原, 吉田「日本語の機械処理のための文節構造モデル」電子通信学会論文誌, 79/12 Vol.J62-D No.12 pp.872-879.
- (17) Kudo, I. & Nomura, H. 'Lexical-functional Transfer: A Transfer Framework in a Machine Translation System Based on LFG', Proceedings of 11th International Conference on Computational Linguistics, Bonn, August, pp.112-114 (1986).
- (18) Kaplan, R.M.& Bresnan, J. 'Lexical-Functional Grammar: A Formal System for Grammatical Representation' In:Bresnan, J. (ed) 'The Mental Representation of Grammatical Relations', The MIT Press, Cambridge, Massachusetts, pp.173-281 (1982).

(2)媒介となる名詞句が「AのB」と「A」の例

ある<登録の用紙>致す<__登録>
 ある<通訳の電話>教える<__通訳,__>
 ある<会議の予定>聴講する<__会議>
 あう<会議の方>持つ<__会議>
 募集する<__会議の原稿>何う<__会議の原稿,__>
 です<__方(ほう)の予約>引き受ける<__方(ほう)>
 です<__宿泊の予定>なる<__宿泊>
 です<__払い戻しの名前>する<__払い戻し>
 選ぶ<__通訳者の方>要る<通訳者>
 話す<__先生の方>替く<先生,__>
 払う<__電話の方>持つ<__電話>
 引き受ける<__名前の方>言う<__名前,__>
 引き受ける<__会議の依頼,__>運営する<__会議>
 いらっしゃる<__会社の方>開く<会社>
 いる<登録の方>締め切る<__登録>
 言う<__ホテルの部屋,__>着く<__ホテル>
 限る<__時間のこと>取れる<__時間>
 考える<__座長の件>する<__座長>
 聞く<__予定の方,__>です<__予定>
 待つ<__通訳の方>居る<通訳>
 見える<時間の方>外す<__時間>
 見る<__会議のアナウンスメント>教える<__会議,__>
 持つ<事務局の方,__>来る<事務局>
 持つ<__会議のカード>締め切る<__会議>
 設ける<__会議の会場>開く<__会議>
 なる<__会議のセッション>始める<会議>
 なる<__通訳の料金>持つ<__通訳>
 なる<__会議の料金>締め切る<__会議>
 なさる<__ホテルの支払い>話す<__ホテル,__>
 行う<登録の方,__>致す<__登録>
 行う<__通訳の会議>済む<通訳>
 行う<__公告のカメラレディコピー,__>です<__公告>
 送る<__会議の番号,__>教える<__会議,__>
 送る<__登録の申込み,__>取り消す<__登録>

(3)媒介となる名詞句が複合語とその一部の場合

ある<支払い方法>かかえる<__方法>
 ある<企画会議>送る<会議,__>
 ある<小会議>する<__会議>
 扱う<__貴会議>扱う<__会議>
 合わせる<__応答時間>かかる<時間>
 募集する<__会議参加>お願いする<__参加>
 です<__コンピュータ会社>聞く<会社,__>
 です<__ビデオ発表>募集する<__発表>
 です<__昼食時間>居る<時間>
 です<__登録カウンター>設ける<__カウンター>
 です<__講演予定>教える<__予定,__>
 含む<__関係資料>ある<資料>
 含む<__関連資料>届く<資料,__>
 始まる<交歓会議>近づく<会議>
 話す<__企画会議,__>開く<会議>
 話す<__コンピュータ会議,__>なる<__会議>
 はずす<__講演時間>持つ<__時間>
 開く<__公式会議>開く<__会議>
 開く<__企画会議>締め切る<__会議>
 居る<科学者>なる<__者>
 致す<__会議参加>締め切る<参加>
 返す<__事前申込み,__>お願いする<__申込み>
 開催する<__関連会議>終わる<会議>
 かかる<ウエルカムレセプション参加,__>
 募集する<__参加>
 かかる<税込み料金,__>なる<__料金>
 抱える<__長距離電話,__>変わる<__電話>
 変わる<__国際電話,__>する<__電話>
 聞く<__学会会議,__>なる<__会議>
 聞く<__公式会議,__>何う<__公式会議,__>
 待つ<__不参加>お願いする<__参加>
 もらう<セカウンドアナウンスメント会議,__>
 始まる<会議>
 持つ<__特別講演>致す<__講演>