

## 文章理解のための意味・文脈解析の試み

金子 和恵 八木沢 津義 藤田 稔

キヤノン(株) 情報メディア研究所  
〒 211 神奈川県川崎市幸区鹿島田 890-12  
Tel: (044)549-5111  
E-mail: kaneko@cis.canon.co.jp

あらまし 10万語の語彙辞書と、オーソドックスな解析手法の組み合わせで、どの程度の文章理解システムが実現できるかの試みを行なった。本研究では、対象とする文章を限定しないことを目標とし、a) 対象文章に依存しない意味・文脈解析を実現する。b) 対象分野についての知識や対象世界モデルを持たず、高度な推論は行なわない。検証のためのシステムとしては、比較的短めの物語文章や新聞記事などを解析し、その文章に関するユーザからの質問に対して応答文を生成する質問応答システム StoryReader を作成した。その正答率は多く見積もって5割程度であるが、語彙辞書と解析系の精度を向上させることで、ある程度汎用的な文章理解システムが可能であると考えている。

キーワード 文章理解, 意味解析, 文脈, 副詞, 省略, 指示照応, 因果関係, 時間関係

### A trial of semantic and context analysis for text understanding:

Kazue KANEKO, Tsuyoshi YAGISAWA and Minoru FUJITA

Media Technology Research Center, Canon Inc.  
890-12, Kashimada, Saiwai-ku, Kawasaki-shi, Kanagawa, 211, Japan  
Phone: +81-44-549-5111  
E-mail: kaneko@cis.canon.co.jp

**Abstract** In this paper, we tested how well a text understanding system is able to perform, using just a combination of orthodox methods and 100,000 vocabulary size dictionary. Our system is directed to understand any type of text. Thus, a) It utilizes a semantic and a context analyzers which are independent to the contents of the text to be analyzed. b) It does not have any kind of domain knowledge nor any target world models, moreover it does not utilize high-level inference. We also implemented a question-answer type evaluation system, "StoryReader". This system analyzes what is written in the given short story or the article and answers the questions about the contents. The correct answer rate is 50 % at best. The wrong answers are due to roughness of the dictionary and the analyzers. Then we suppose that the further improvement of them enables a text understanding system in general use.

key words text understanding, semantic analysis, context, adverb, ellipsis, anaphora, causal relation, temporal relation

# 1 はじめに

本研究は、10万語の語彙辞書と、オーソドックスな解析手法の組み合わせで、どの程度の文章理解システムが実現できるかの試みである。

これまで、文章理解システムでは、「文章を理解する上では、その対象とする分野の知識は欠かせない」ことを前提として、対象分野や扱える言語現象を限定したアプローチが多い([1]、[2]、[3])。しかし、それでは、応用対象分野の知識と自然言語処理が混在し、新しい分野へ応用する場合、自然言語処理部分を最初から作り直す必要性が出てくる。対象分野そのものが広範囲にわたるテキストデータの検索などへの応用可能性も限定される。

それに対し、我々は、応用対象に依存する部分と自然言語処理の部分とをできるだけ切り分け、対象分野に依存しない自然言語処理を作成し、応用可能性を広げることを目指す。

今回作成したシステムは、対象分野についての知識を持たず、対象世界モデルの作成や高度な推論を行わない。言語情報のみを用いて、明示的に記述された現象について意味・文脈モデルの作成が行なえた時に文章を理解したとする。例えば「私は寒い」という文は、「私という主体が寒いという状態である」と解析する。「窓を閉めてくれ」、「暖房を入れてくれ」という遠回しな要求だったり、「私は寂しい」という暗喩だったりする可能性があるといった深い解析は行なわない。その代わりに、対象とする分野を限定せずに書かれた文章を受け入れる。

解析処理は各処理モジュールの組み合わせで実現している。意味解析としては、個々の文に対し、ボイス、格関係、副詞、テンス・アスペクト、取り立て、ムードについての解析を行ない、用言を中心としたネットワークを作成する。文脈解析としては、意味解析で作成された複数文のネットワークについて、省略補完、指示照応、因果関係・時間関係の推定を行なう。

また、この解析処理を検証するためのシステムとして、比較的短めの物語文章や新聞記事などを解析し、その文章に関するユーザからの質問に対して応答文を生成する質問応答システム StoryReader を作成し、評価を行なった。

## 2 意味解析

### 2.1 用言の5階層モデル

意味解析では、意味解析結果として、図1のような意味ネットワークを作成する。これは、佐野の文の階層構造[4]を参考にしており、用言の階層を5階層に分けて

いる。それぞれの階層に対して、格関係解析、ボイス解析、テンス・アスペクト解析、取り立て解析、ムード解析の各モジュールを適用し、すべての階層に対して副詞解析モジュールを適用する。

この意味解析の特徴として、オーソドックスな手法<sup>1)</sup>に、これまであまり扱われていなかった副詞の利用を合わせた点がある。モジュールの適用順序は、ボイス解析、格関係解析、副詞解析、取り立て解析、ムード解析、テンス・アスペクト解析の順であるが、まず副詞解析から説明する。

彼女は、多分、母親に褒められるだろう。

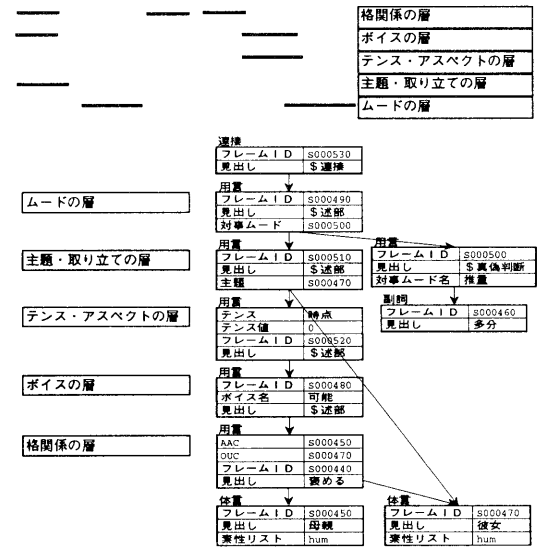


図1 例文と解析結果

### 2.2 副詞解析

ここでは、副詞とは必須格以外の連用修飾する語や句や節(任意格や従属節を含む)のことを言う。自立的に副詞をなす自立副詞のみでなく、形式副詞<sup>2)</sup>によって作られる副詞句や副詞節も同じ副詞とする。

自立副詞 太郎はとても疲れた。

形式副詞 太郎は死ぬほど疲れた。

また自立副詞は、自立語の副詞として品詞分類されるものでだけでなく、接続詞や、副詞的な働きをする名詞や数詞、形容詞や形容動詞の副詞的用法も含む。

昨日、本を三冊、素早く 読んだ。

<sup>1)</sup>名詞句や名詞述語の正確な解析には、概念辞書などの意味の細かい記述が必要であると考えて、名詞句と名詞述語の解析については保留している。

<sup>2)</sup>[5]、[6] 参照

これら、副詞として扱えるものの語義数は1万4千語義である。

このような副詞を、仁田による分類[7]を参考にして、表1のように素性付けを行ない、用言の5階層のそれぞれの階層を修飾するものとして割りふりそれぞれの階層での解析に利用する。

表1 用言の階層と修飾する副詞の素性

階層	素性	例文
格関係	動きの質	雲が <u>もくもく</u> わいている。
	動きの強さ	雨足が <u>激しく</u> なった。
	進行の速さ	水が <u>だーっと</u> 流れる。
	取りかかりの早さ	雨が <u>さっと</u> 止んだ。
	結果	洗濯物が <u>びしょびしょ</u> に濡れた。
	評価づけ	鯉が、 <u>うるさく</u> 鳴きだした。
	数量	水たまりが <u>たくさん</u> 出来た。
	程度	日射しが <u>たいそう</u> 強い。
	場所	親子づれが、 <u>公園</u> で遊んでいる。
道具	子供は、 <u>傘</u> で水たまりをかき混ぜる。	
ボイス	意図	子供は、 <u>わざと</u> 水たまりに飛び込む。
	態度	母親は <u>あわてて</u> 子供を叱る。
テンス・アスペクト	時間	子供は <u>しばらく</u> おとなしくしている。 数分後には、元気になる。
	頻度	子供は、 <u>何度も</u> 手を叩く。 手を叩く <u>度に</u> 、子供は叫ぶ。
主題・取り立て	視点	子供は <u>たいでい</u> 、いたずらっ子だ。 これに <u>関しては</u> 、言うことはない。
	目的	<u>濡らさないように</u> 、気をつけなさい。
	理由	<u>服を汚したので</u> 、着替えさせられる。
	仮定	<u>いたずらをしたら</u> 、叱られる。
	順接	母親は子供を叱り、家へ帰って行く。
	逆接	家へ帰っても、子供は遊ぶ。
対比	<u>かつ</u> 、いたずらを続ける。	
ムード	対事	<u>きつと</u> 、叱られるだろう。
	伝達	<u>どうか</u> 、叱られませんように。

### 2.3 ボイス解析

ここでは、「間接受身、強制使役、許可使役、被役、可能、不可能、自発、尊敬、意思、希望、拒否、試行、準備、経験、予定、放置、受益、授益、難易、強制推移、自然推移」の22種類をボイスとして、格の置換などの処理を行なう。

ボイスを示す助動詞が出現した場合、ボイス名をボイスの層に設定し、ボイスの主体へのリンクを張り表層格の置き換えを行なう。

また、副詞解析でボイスの層に割りふられた副詞は、ボイスの主体の態度や意図を示すものとし、副詞からボイスの主体へのリンクを張っている。

例えば、「彼は、自ら彼女に殴られた」という例では、「自ら」は主体の意図を示す副詞であるので、ボイスの

階層へ割りふって、「受身主」である「彼」の態度を規定するものとしてリンク付けする。

### 2.4 格関係解析

動詞と形容詞・形容動詞の格関係解析では、動詞・形容詞・形容動詞の独自の分類(184分類)と、個々の格に与えられる名詞の意味素性([8]、[9])を使って、格役割を示すラベルの付与と動詞の語義選択を行なっている。格関係解析後、3.1で述べる省略補完で、ラベルと意味素性の情報を再び利用する。

格役割を示すラベルを与えた後、程度や数量を示す副詞がどの格の体言の数量を規定するものであるかの解析も行なっている。これは、程度・数量の副詞を「測定・集合・まとめ・比較・多少・疑問」の6種類に細分類して格役割の種類と副詞との位置関係を調べて、組み合わせのルールとして62通りの記述を行ない、次の例のような対処を行なっている。

- 「子豚はみんな餌を食べた」のような例では、「みんな」が「集合」の副詞であり、「運動主格」である「子豚」の直後に現れている。これは「みんな」の規定対象を「運動主格」の「子豚」とする。
- 「子豚は餌をみんな食べた」では、「みんな」が「集合」の副詞であり、「非変化対象格」である「餌」の直後に現れている。これは「みんな」の規定対象を「非変化対象格」の「子豚」とする。
- 「子豚はたくさん餌を食べた」では、「たくさん」は「多少」の副詞である。格としては「運動主格」と「非変化対象格」がある。これは「たくさん」の規定対象を「非変化対象格」の「餌」とする。
- 「みんな餌を食べた」では、「みんな」が「集合」の副詞であり、すべての格の前に現われている。これは、「みんな」の規定対象を、省略されている「運動主格」とする。ここでは、「非変化対象格」の可能性もあるが、ルールの適用上、「運動主格」としている。
- 「たくさん餌を食べた」では、「たくさん」は「多少」の副詞である。これは、「たくさん」の規定対象を「非変化対象格」の「餌」とする。

### 2.5 テンス・アスペクト解析

テンス・アスペクト解析は、井上の説[10]と、森山の説[11]を参考にしている。

アスペクトの決定には、アスペクトを示す助動詞と、動詞の状態性の有無等を示す分類を用いるが、アスペクトを示す副詞がある場合は、その副詞を使ってさらに絞り込みを行なう。

例えば、助動詞「ている」の示すアスペクトとして、「進行相/結果状態相/継続相/反復相」の複数があげられる。「彼が枝を折っている」という例では「折る」が動作性や継続性を持つ動詞であり、アスペクトとして、「進行相/結果状態相/継続相/反復相」が考えられる。

このようなものについては、副詞のアスペクトとの共起を調べて以下にある例のような絞り込みを行なう。

- 「彼が枝をぼきぼき折っている」は、「ぼきぼき」という副詞が「反復相」を表すものであるから、アスペクトは「反復相」に限定される。
- 「彼が枝をすっかり折っている」は、「すっかり」という副詞が「完結相／結果状態相」を表すものであるから、アスペクトは「結果状態相」に限定される。

「彼女はその本を読んでいる」の例でも、「読む」のアスペクトとしては、「進行相／結果状態相／継続相／反復相」が考えられるが、以下にある例のような絞り込みが行なえる。

- 「彼女はメモを取りながらその本を読んでいる」は、「...ながら」という副詞が「進行相／反復相」を示すことから、アスペクトは「進行相／反復相」に限定される。
- 「彼女はここに来る度にその本を読んでいる」は、「...度に」が「頻度」すなわち「反復相」を示すことから、アスペクトは「反復相」に限定される。

テンスの決定は、用言の状態性の有無と完了を示す助動詞「た」の有無を用いて行なうが、時を示す副詞も利用する。例えば、「彼はここにいる」では、状態性の用言に「た」がついていないので「現在」を示すが、「彼は明日はここにいる」では、未来の副詞があるので「未来」を示す。

現象が起きた時と発話した時との時間関係は、発話した時を0とし、過去を-、未来を+の数値で示すテンス値で表現する。また、複文での時間の前後関係は、接続の種別を考慮して、テンス値の大小で記述する。

例えば、「本を読んでから、テレビを見た」の例では、「テレビを見た」のは発話した時よりも過去であるのでテンス値を-1とし、さらに「...でから」が「継起」を示す時の副詞であるので、「本を読む」のテンス値を「テレビを見た」のテンス値から1だけ小さい値-2としている。

## 2.6 取り立て解析

取り立て解析では、取り立ての層から、副助詞「は」のつくものへ「主題」としてのリンク付けを行なう。この「主題」の情報は、3.1の省略補完で利用する。

取り立ての層には、「視点・目的・理由・仮定・順接・逆接・対比」などの副詞をリンクする。このうち「目的」「理由」の副詞は3.3の因果関係推定で利用する。

## 2.7 ムード解析

ムードは、話し手の判断である対事ムードと、話し手の聞き手に対する態度である伝達ムード [12] の2つに大きく分け、さらに細かく分類している (表2、3参照)。

これらのムードは、助動詞と終助詞と副詞と感動詞の組合せで決定している。

表2 対事ムードの分類

大分類	小分類
真偽判断	推量、伝聞
説明	帰結、反実仮想、失敗、意外、驚嘆、回想、披瀝、既定
価値判断	当然、義務、禁止、当為、許可、推奨、不必要、不可避、過剰、相応、不足

表3 伝達ムードの分類

大分類	小分類
問いかけ	YN 質問、WH 質問、依頼、依頼の禁止、意思、提案、希望1、希望2、希望3
働きかけ	命令、依頼、禁止、依頼の禁止、勧誘
表出	意思、拒否、希望1、希望2、希望3、願望、疑念

## 3 文脈解析

文脈表現は、一般的には、意味表現とは別のレイヤであり、文と文との接続関係や指示照応関係などを示すものである。しかし、ここでは、インプリメントの関係上、複数文の意味表現ネットワークに、省略補完・指示照応・因果関係の設定・時間関係の推定の各モジュールを適用してリンク付けを行なったものを、文脈表現ネットワークとする。

### 3.1 省略補完

格関係解析の結果、主格や対象格について省略があると見なされた場合、同じ文中もしくは直前の文中に出現した体言で、省略された格の補完を行なう<sup>3</sup>。ここでは、用言の一致、主題・格の意味素性などの情報を使い、次のような条件を満たすかどうかを調べる。

**用言一致:** 前文と後文の用言が一致し、前文にある格 A と同じ格 B が後文で省略され、

- 前文と後文の用言が同じ語義である。
- 前文と後文の用言が同じ見出しである。

ex. 彼女はペンを落した。そして、ノートを落した。

**主題:** 前文の主題 A に対し、

- その意味素性が後文の (主格または対象格の) 省略格 B の意味素性と等しい。
- その意味素性が後文の (主格または対象格の) 省略格 B の意味素性に包括される。

<sup>3</sup> 主格や対象格以外の格について省略補完を行なうと弊害が大きいため、この2つに限定した。

- 後文または従属節の主格 B が省略されている。

ex. 彼女は駅へ行った。切符を買って電車に乗った。

意味素性: 前文の格 A の意味素性が

- 後文の (主格または対象格の) 省略格 B の 意味素性と等しい。
- 後文の (主格または対象格の) 省略格 B の 意味素性に包括される。

ex. 彼女はりんごを食べた。私も食べた。

主格: 前文に主格があり、後文の主格が省略されている。

ex. 彼女がクッキーをかじった。そして紅茶を飲んだ。

上記の項目の条件にあう場合、条件の厳しいもの (先に記載した項目のものの方が条件が厳しい) を採用して、B が A を指すよう、省略格の補完を行なう。

### 3.2 指示照応

照応先の推定では、「コソアド」については、金水らの指示詞の分類 [13] の中から文章中で照応関係が既定できるものをピックアップして、17 個のルールを作成した。その一部を次に示す。

対象指示: 対象を指し示す動作の後の指示詞は、指し示された対象を指す。

ex. 彼女は、私の食べているパフェを見た。  
「それ、おいしそうですね」  
彼女は、私の本を指さした。  
「その本は、おもしろそうね」

場所: 場所を示す指示詞は、それ以前に出てきた場所を示す。

ex. 彼女は、昨日の朝、公園へ行きました。  
そこから、直接、学校へ行きました。  
彼女は、昨日の朝、鎌倉へ行きました。  
その街に、家族が住んでいます。

時: 時を示す指示詞は、直前の文の時間の層を指す。

ex. 私達は山の上で焚火をしました。  
あの日は、とても幸せでした。  
きっと、彼は帰って来ます。  
その時に、来て下さい。

繰り返し: 連体修飾する指示詞は、連体修飾した体言が前の文の繰り返しのとき、前の文の体言を指す。

ex. 夏になると、蟻の行列をよく見掛けます。  
その行列は、蟻の巣から餌の在る所まで、  
ずっと続いています。

また、人称代名詞については、前の文に人の素性を示すものがあれば、それを指し示すというルールを使用している。

### 3.3 因果関係

因果関係の推定は、「理由」「目的」といった人の行動の動機付けの言明や、「原因」「結果」といった明示的な因果関係の言明があった場合にのみ行ない、仮定や条件等については時間関係として扱う。

因果関係を表現するものとしては、次のような、名詞、動詞、副詞、分裂文、接続詞などを認める。

名詞: 理由、原因、目的、結果...

ex. ストレスが、肥満の原因である。  
その仕事を完成することが生きる目的になった。

動詞: 帰する、基づく、依る、困む、起こす、生む...

ex. 当社の発展は彼の努力に依る。  
多くの失敗が、成功を生む。

副詞: から、ため(過去+ため)、ので、のに、ため、ように....

ex. 少年は空腹だったからパンを盗んだ。  
妻はやせるためダイエットを始めた。

分裂文: ... のは... からです、... のは... ためです

ex. 彼女が泣いたのは、悔しかったからだ。  
彼女が来たのは、彼に会うためだ。

接続詞: 何故なら、というのは、そのため、故に....

ex. 豚が塔の上にいるはずがない。  
何故なら、豚は空を飛ばない。  
我思う。ゆえに 我あり。

### 3.4 時間関係

文章中に記述された事柄の時間関係を推定し、その経過や前後関係を記述する。隣接する文の用言のテンス・アスペクトの層を見てテンス値のずれや時の副詞を調べ、同時であると判断した場合に、時間帯という漠然とした帯域にまとめあげる。文章の文の流れに沿って時間帯を連結し、隣合う時間帯の時間経過を表 4 のタグの中から選んで付与する。この時間帯の連鎖を時間関係の表現とする。

表 4 時間帯間の時間経過タグ

タグ	説明
=	時間が断続なく連続していることを示す。
+	時間が経過したことを示す。経過を示す時の副詞、もしくは、順接や累加を示す副詞を持つ。
-	時間を遡ったことを示す。遡及を示す時の副詞を持つ。
?	時間の流れが途切れたことを示す。過去か未来か分からないが、任意の時間を指定する副詞を持つ。

事象の開始や終了や事象間の時間の重複関係などについては、詳しい研究 ([3]、[14]) が行なわれているが、本研究では詳しい検討を行っていない。現段階では、朝の次に昼が来るといった時間に関する知識や、今が何月何日の何時であるといった状況把握などの機能を持ち合わせていないため、時間の配置を厳密に行なうことはできない。

## 4 システム

以上述べた意味・文脈解析処理を検証するためのシステム StoryReader を作成した。これは、物語を読み込み、それに関する質問を受け入れ、応答を作成するアプリケーションである。

このアプリケーションを作成するに当たっては、意味・文脈解析の他に、文脈と質問文とのマッチングをとる質問応答処理と、簡単な文生成処理をインプリメントしている。

## 4.1 構成

アプリケーションの構成を図2に示す。複数の文からなる物語文章を、一文ごとに、形態素・句関係<sup>4</sup>・意味解析処理を行ない、意味解析結果をシンボル・文脈領域に蓄積して行く。物語文章の全ての文を解析し終わったら、文脈解析を行なって文脈表現を作成する。

物語文章に対する質問と応答は一文単位で行なわれる。入力された質問文は形態素・句関係・意味解析処理の順で解析し、質問応答領域に質問意味表現を作成する。質問応答処理は、質問意味表現と文脈とのマッチングを行ない、質問文の疑問詞に相当する箇所を文脈領域から取り出し、応答意味表現を質問応答領域に作成する。応答文生成処理は、応答意味表現を意味領域や意味辞書領域に再構築し、表層格を使って応答文を生成する。

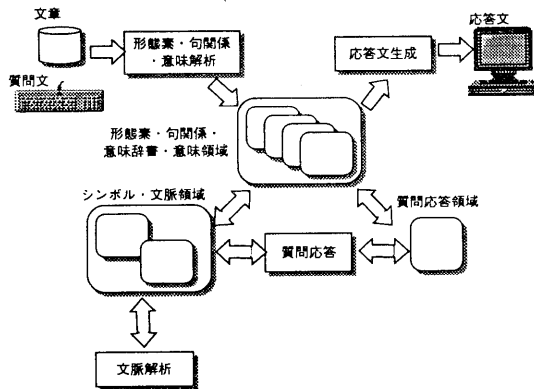


図2 アプリケーションの構成

### 4.1.1 質問応答処理

質問応答処理は、一問一答形式の簡単なもので、談話処理などは行なわない。質問応答領域に作成された質問の意味表現を使って、文脈領域と以下のようなマッチングを行ない、応答にあたる意味表現を作成する。

**WH 質問:** 疑問詞のある疑問に対しては、疑問詞に相当する箇所を取り出す。

- 疑問詞が格の時、格に相当する体言部分を取り出す。  
ex. 文章 太郎は学校へ行きました。  
質問 誰が学校へ行きましたか。  
応答 太郎。

<sup>4</sup>このシステムでは、形態素を句にまとめ上げ、係り受けを決定する処理を句関係解析と呼んでいる。

- 疑問詞が副詞の時、副詞に相当する体言部分を取り出す。  
ex. 文章 今日、太郎は学校へ行きました。  
質問 何時、太郎は学校へ行きましたか。  
応答 今日。
  - 「何をしましたか」のような、動作を示す用言に対応する疑問の場合は、文に相当する部分を取り出す。  
ex. 文章 太郎が学校へ行きました。  
質問 太郎が何をしましたか。  
応答 太郎が学校へ行った。
  - 「Aは何ですか」のような、状態を示す用言に対応する疑問の場合は、文に相当する部分を取り出す。  
ex. 文章 太郎は学生です。  
質問 太郎は何ですか。  
応答 太郎が学生だ。
  - 「どうしましたか」のような、用言に対応する疑問の場合は、文に相当する部分を取り出す。  
ex. 文章 太郎は学生です。  
太郎は学校へ行きました。  
質問 太郎がどうしましたか。  
応答 太郎が学生だ。  
太郎が学校へ行った。
- YN 質問:** 質問文の文に相当する部分を応答とする。  
ex. 文章 太郎は学生ではない。  
質問 太郎は学生ですか。  
応答 太郎が学生でない。

**条件を限定した質問:** 条件部分のマッチングをした上で、WH 質問やYN 質問とのマッチングを行なう。

- 「...した時、...か」のように時に限定した疑問の場合は、時の部分にあたる文とマッチングする文を捜し、その時間帯のスコープにはいる文を限定し質問とのマッチングを行なう。  
ex. 文章 太郎はりんごを取りました。  
次郎はみかんを取りました。  
質問 太郎がりんごを取った時、  
次郎は何をしましたか。  
応答 次郎はみかんを取った。
- 「どんな目的で...か」のように、理由や目的に対する質問は、理由・目的のリンク付けされた箇所を取り出す。  
ex. 文章 彼女は痩せるためダイエットを始めた。  
質問 彼女がダイエットを始めた目的は何ですか。  
応答 彼女が痩せる。

マッチングは、質問文の意味表現のネットワークが、文脈中にある意味表現のネットワークのサブセットならば成功とする。

質問文に含まれる副詞や格については、文脈中に同じ記述があるかどうかをチェックする。文脈中の副詞や格に指示詞が含まれる場合は、その指示対象と質問文中のものとのマッチングを行なう。

用言の層の内容であるボイスやテンス・アスペクトやムードの表現については、マッチングの条件を緩くするため敢えてチェックしない。ただし、否定表現については、WH 質問の場合はチェックして、否定と肯定の違いがあるものはマッチングに成功しないものとする。YN 質問の場合は、否定と肯定の違いがあってもマッチングに成功し、応答文中で肯定もしくは否定を表現する。

意味フレームのマッチングは「語彙」「字面」「意味分類」の3段階で行ない、「語彙」レベルでのマッチングには3、「字面」レベルでのマッチングには2、「意味分類」レベルでのマッチングには1の確信度を与える。

#### 4.1.2 応答文生成

質問応答領域に生成された応答の意味表現ネットワークを使って、語の表記や動詞の活用情報などから応答文を生成する。この文生成は、体言に格助詞をつけて動詞の前に並べ過去や現在のテンスをつける程度の便宜的なものであり、ボイスやムードなどは扱わない。

### 4.2 実行例

例題として使用した文章<sup>5</sup>を図3に、アプリケーションの実行例を図4に示す。なお図4の左側に付与されている矢印は、質問応答のログの順番を示し、図3の左側に付与されている矢印は、その番号の質問に対する応答を作成するために使われた文章を示す。

ユーザは、ウィンドウの上部にある質問文入力部に質問をタイプ入力する。システムは応答を作成し、質問と応答のログをその下に表示する。質問文は茶色、応答のログは青色で表示する。応答は、質問文とのマッチングの度合により、濃いものから、薄いものへ3段階で表示される。

おうむの留守番

1),2),6) □ 森の中に孔雀のお父さんとお母さんと子供達が住んでいました。  
 3) □ 孔雀のお父さんは、食べ物を捜しに、毎日、遠くまで行きました。  
 ある日、暗くなっても、お父さんは帰ってきませんでした。  
 8) □ お母さんは、お父さんを捜すことにしました。  
 4),8) □ 留守番をお隣のおうむに頼んで行きました。  
 7) □ おうむは鳴き真似が上手です。  
 6) □ 孔雀の子供達に、動物の鳴き真似を色々聞かせました。  
 □ 子供達は、小さくちばしをばくばくさせて、喜びました。  
 おうむが猿の鳴き真似をしていたとき、あちらから、一匹の豹が来ました。  
 豹は孔雀の子供達の方へ向かってきます。  
 その時、ライオンが「ウォー、ウォー」と吠える声がありました。  
 豹は、立ち止まって、辺りを見回しました。  
 だけど、ライオンは、どこにも見えません。  
 豹は、また、孔雀たちの方へ、向かってきました。  
 そうすると、また、「ウォー、ウォー、ウォー」と、ライオンの声がありました。  
 豹はあわてて逃げて行きました。  
 孔雀のお父さんとお母さんが、一緒に帰って来ました。  
 自分達のうちのそばまで来たとき、「ウォー、ウォー」と、ライオンの声がありました。  
 お父さんとお母さんは、びっくりして、木の陰に隠れました。  
 しかし、すぐに、子供達の事が心配になりました。  
 もう、じっとしていられません。  
 一緒に、夢中で子供達の所へ走って行きました。  
 「あ、お帰りなさい」と、おうむが木の陰から首を出して言いました。

図3 例題文章

- 1) の「孔雀のお父さんはどこに住んでいますか」に対する応答「森の中」は、文章中の1),2),6)の文「森の中に孔雀のお父さんとお母さんと子供達が住んでいました。」に語彙レベルでマッチングしたので、確信度3の濃い青で表示される。
- 2) の「母さんはどこに住んでいますか」に対する応答「森の中」は、「母さん」と「お母さん」が別々の語であり、意味分類レベルでマッチングしたので確信度1の薄い青で表示されている。
- 3) の「孔雀のお父さんは、どんな目的で、遠くまで行きましたか。」に対する応答「孔雀のお父さんが食べ物を捜しに。」は、確信度3の濃い青で表示される。なお、字面レベルでのマッチングによる確信度2による中間の青の表示は、図4の例にはないが、辞書中のエントリの重複や、語義選択の誤りを吸収するために使用している。
- 4),5) の「お母さんが留守番を頼んだのは誰?」「お父さんは何の事が心配になりましたか。」の質問は、省略補完がうまく利用できた例である。本文中では、4),8)の「留守番をお隣のおうむに頼んで行きました。」や5)の「しかし、すぐに、子供達の事が心配になりました。」と、主語が省略されているが、文脈表現では省略補完処理により、前文の主題である「お母さん」や「お父さんとお母さん」がそれぞれ「頼んだ」「心配した」の主語であることが分かっており、質問文に対して正しい答えを返すことができる。
- 6) の「子供達は何をしましたか。」という質問に対しては、本文中にある子供達の行動をリストアップし、1),2),6)から「孔雀のお父さんとお母さんと子供達が森の中に住んでいた。」6)から「子供達は、小さくちばしをばくばくして動物の鳴き真似を喜んだ。」という応答を返す。ここでは、文生成でボイス表現を扱っていないので「ばくばくさせて」という使役表現は生成できていない。
- 7) の「おうむは鳴き真似が上手ですか?」というYN質問に対しては、「おうむは鳴き真似が上手です。」という応答を返す。

質問文入力部

Question & Answer

お父さんが帰らなかった時、お母さんは何をしましたか。

1) □ 孔雀のお父さんはどこに住んでいますか。  
 森の中。  
 2) □ 母さんはどこに住んでいますか。  
 森の中。  
 3) □ お父さんは、どんな目的で、遠くまで行きましたか。  
 孔雀のお父さんが食べ物を捜しに。  
 4) □ お母さんが留守番を頼んだのは誰?  
 お隣のおうむ。  
 5) □ お父さんは何の事が心配になりましたか。  
 子供達。  
 6) □ 子供達は何をしましたか。  
 孔雀のお父さんとお母さんと子供達が森の中に住んでいた。  
 子供達が小さくちばしをばくばくして動物の鳴き真似を喜んだ。  
 7) □ おうむは鳴き真似が上手ですか?  
 おうむは鳴き真似が上手だ。  
 8) □ お父さんが帰らなかった時、お母さんは何をしましたか。  
 お母さんがお父さんを捜した。  
 お母さんが留守番をお隣のおうむに頼んでいた。

質問応答ログ

図4 アプリケーションの実行例

<sup>5</sup> 小学校2年生用の教材文([15]から引用)に、かなを漢字に直すなどの修正を加えたもの。

- 8)の「お父さんが帰らなかった時、お母さんは何をしましたか。」という質問は、時間関係の推定が利用できた例である。「ある日、暗くなっても、お父さんは帰ってきませんでした。」という文章と時間帯を同じくする文の中から、「お母さん」がしたことをリストアップし、8)から「お母さんがお父さんを捜した。」4),8)から「お母さんが留守番をお隣のおうむに頼んでいた。」という応答を返す。

### 4.3 解析精度と実行時間

解析精度については、定量的評価を行っていない。実行例程度の文章に対して、応答が的確である可能性は、多く見積もって5割程度である。これには、語彙辞書の精度、形態素解析・句関係解析・意味解析の精度が大きく関係している。文脈解析においては、概念辞書や推論機構の欠如がその精度に影響していると思われる。

実行速度は、NeXT Startion で実行した場合、実行例の「おうむの留守番」で文脈解析終了までに48分かかる。質問に対する応答は、質問の種類によって異なるが、平均して約40秒程度である。この遅さは、意味解析と文脈解析をプロダクションシステムでインプリメントしていることに起因する。ルール数は意味解析1087ルール、文脈解析123ルール、質問応答93ルール、計1303ルールである。

## 5 まとめ

対象分野についての知識や対象世界モデルを持たなくとも、10万語の語彙辞書と、言語学的にオーソドックスな解析技術の組み合わせで、どの程度の文章理解システムが実現できるかの試みを行なった。

今回作成した質問応答システムでは、的確な応答が得られる可能性は多く見積もっても5割程度という結果を得た。この結果は満足できるものではない。しかし、これは、対象分野についての知識や対象世界モデルの欠如によるものではなく、語彙辞書・形態素解析・句関係解析・意味解析すべてにおける全体的な精度の低さに起因するものと考えており、解析精度が向上すれば、ある程度汎用的な文章理解システムが実現可能であると考えている。

また、今回保留した名詞句や名詞述語の解析を含め、意味・文脈解析の向上には、概念辞書や推論機構が必要であると考え、次の段階での試みとして、概念辞書の作成を行ないその利用を検討する予定である。

## 参考文献

- [1] 田中裕一, 向井国昭, 三吉秀夫, 安川秀樹, 平川秀樹, 横井俊夫: 状況意味論に基づく談話理解システム DUALS—その基本原理—, 情報処理学会研究報告 自然言語処理 50-6, 1985.
- [2] 三吉秀夫, 安川秀樹, 平川秀樹, 向井国昭, 田中裕一, 横井俊夫: 状況意味論に基づく談話理解システム DUALS—そのインプリメンテーション—, 情報処理学会研究報告 自然言語処理 50-7, 1985.
- [3] 江良浩一, 甲斐郷子, 中村順一, 吉田将: テクス・アスペクトを考慮した物語文章における事象間の時間関係の抽出, 電子情報通信学会技術研究報告 NLC93-6, 1993.
- [4] 佐野洋: 統語文型と文のレベル, 情報処理学会研究報告 自然言語処理 82-1, 1991.
- [5] 奥津敬一郎, 沼田善子, 杉本武: いわゆる日本語助詞の研究, 凡人社, 1986.
- [6] 森山良行, 松木正恵: 日本語表現文型, アルク, 1989.
- [7] 仁田義雄: 動詞に係る副詞的修飾成分の語相, 日本語学 vol.2, no.10, 1983.
- [8] 情報処理振興事業協会技術センター: 計算機用日本語基本動詞辞書 IPAL(Basic Verbs)—解説編—, 情報処理振興事業協会, 1987.
- [9] 情報処理振興事業協会技術センター: 計算機用日本語基本動詞辞書 IPAL(Basic Verbs)—辞書編—, 情報処理振興事業協会, 1987.
- [10] 井上和子: 変形文法と日本語(下), 大修館書店, 1976.
- [11] 森山卓郎: 日本語動詞述語文の研究, 明治書院, 1988.
- [12] 酒井桂一, 池田裕治, 藤田稔: ガイダンス機能を備えた自然言語処理インタフェースシステム(2)—発話意図抽出処理—, 情報処理学会第42回全国大会, 4C-6, 1991.
- [13] 金水敏, 木村秀樹, 田窪行則: 指示詞, 日本語文法 セルフ・マスターズシリーズ4, くろしお出版, 1989.
- [14] 桃内佳雄: 物語文章における事態間の時間的関係の表現について, 情報処理学会研究報告 自然言語処理, 62-4, 1987.
- [15] 永野賢: 文章論総説, 朝倉書店, 1986.
- [16] 池田裕治, 藤田稔, 柵木孝一: 意味解析における言語知識の体系的活用, 情報処理学会第39回全国大会, 5F-8, 1989.