

事例を用いた日本語解析
- 「知恵の泉」のインタフェース部の構想 -

井上 秀行, 折原 良平, 長尾 加寿恵, 大須賀 昭彦, 楠井 洋一

株式会社 東芝 システム・ソフトウェア技術研究所

ある制限分野において、体系的な意味辞書や構文規則を用いることなく、事例を用いて日本語文を解析する方法について述べる。この事例を用いた方法では、過去の解析結果である事例に分野固有の知識やその使い方が含まれているため、その分野固有の知識に基づく解析が可能になる。また、事例を用いたことにより、格の省略されている文や格をあいまいにする語を含む文においても解析が可能になる。さらに、解析の度に事例が増大していくといった学習効果も加わることになる。

A case-based parsing as a user interface of abduction support system "CHIE-NO-IZUMI"

Hideyuki INOUE, Ryohei ORIHARA, Kazue NAGAO, Akihiko OHSUGA, Yoichi KUSUI

Systems and Software Engineering Laboratory, TOSHIBA corporation

This paper describes a model of case-based parsing. Traditional parsers use syntactic rules and meaning lexicon in their analyzing processes. The case-based parser differs from them in point of making use of past cases. The case-based parser can recognize input sentences based on the knowledge of target domain by employing past cases. Furthermore, a learning mechanism can be realized by adding results of analysis to the case-base where past cases are collected.

1. はじめに

ユーザが計算機システムを利用するため、何かの指示をシステムに与えようとした場合、ユーザがシステムにどのように要求を伝えるかという問題がある。これは、ユーザが物事を理解し、伝達する場合には、日本語（自然言語）を用いて行っているのがほとんどであり、システムが何かの処理をする場合には、システム自身が処理を行いやすいシステム固有の内部表現を用いて行っているためである。従来は、この要求の伝達をユーザがシステムの内部表現を用いて行っていたため、ユーザにとってシステムを使いこなすのは困難な作業であった。そのため、ユーザが理解や伝達に用いている日本語を直接の入出力とし、システムが処理を行う際に用いるシステム固有の内部表現に変換することができれば、そのシステムはユーザにとって利用しやすいシステムと言えるであろう。

このような目的のための日本語解析は、一般の日本語解析とは異なり、システムが対象とする分野内においてのみ解析が行えればよい。システムにとって必要なのは、その分野固有の知識に基づく解釈によって解析された結果であり、一般的な解析結果ではないからである。しかし、従来の解析方法では、一般的な解釈を基に作成されている体系的な辞書や構文規則を用いて解析を行うため、一般的な解析結果しか得ることができなかった。そのため、対象とする分野を制限して日本語解析を行う場合、体系的に作成されている辞書や構文規則を用いて解析を行うよりも、その分野の類似した過去の解析結果（事例）を用いて解析を行った方が、ユーザの求めている解析結果が得られやすいであろう。これは、過去の解析結果である事例には、その分野固有の知識が含まれているためである。

そこで、筆者らは、事例を用いた日本語解析の実現を試みている。本稿では、この事例を用いた日本語解析方法について考察する。

2. 日本語解析のアプローチ

従来より、日本語解析については多くの研究がなされている。それら従来の日本語解析の多くでは、解析は次のような処理手順によって行われている。

- 1) [形態素解析] 入力された文字列の単語の位置を同定する。
- 2) [構文解析] 形態素解析で得られた単語や文節を基本として、文節間の係り受け関係や係り受けの非交差などの規則を基に、それら文節間の関係を解析し文の構造を得る。
- 3) [意味解析] 構文解析によって得られた結果を基に、意味上の単語や文節の関係を各々の単語の意味的属性を用いて明らかにし、文の内容を表す構造を作り出す。

この3つの処理手順を経ることによって、文内容の理解を行い、次処理である応用処理（ここでは、この解析結果を利用して行うシステムの処理を言う）へ進む[浮田90]。

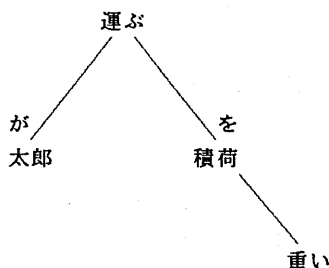
例えば、『太郎が重い積荷を運ぶ。』という文をこの方法で解析すると、まず、形態素解析によって、次のような日本語リストを作成し、この文に用いられている単語を決定する。

《日本語リスト》

太郎ーが 重い 積荷ーを 運ぶ

次に、この結果を基に次処理である構文解析を行い、次のような構文木を作成し、この文の構造を得る。

《構文木》



最後に、この構文木を基に意味解析を行うことにより、『運ぶ』という動作の動作主が『太郎』であり対象が『積荷』である、といった文の内容を表す次のような構造を得て、応用処理へと進む。

動 詞：運ぶ
動作主：太郎
対 象：積荷

通常この3つの処理は、辞書および構文規則を用いることによって行われている。しかし、人間が思考過程において日本語文を解析しているとき、このように辞書や構文規則を使いながら解析しているとは考えられない。むしろ、子供が母国語を学習するときと同じように、最初から体系的な辞書や構文規則を持っているのではなく、解析しようとしている日本語文と類似している過去の事例（解析結果）を基にしながら解析を行い、徐々に体系的なものを作っていく、と考えられる。また、ユーザは、現在解析しようとしている日本語文の対象分野を知った上で、文章上に表れない様々な対象分野固有の知識を用いて解析を行っている。それ故、ある対象分野内で使われるシステムにおいて日本語文の解析を行う際には、体系的な辞書や構文規則を用いるだけでなく、その対象分野固有の知識も用いるべきである。このような方法として、その分野の類似した過去の事例を用いて解析する方法が考えられる。これは、事例には、体系的な辞書や構文規則と同等なものが含まれており、さらに、その分野固有の知識やそれがどのように使われたかも示されているからである。それ故、事例を用いて日本語解析を行えば、よりユーザの意図に近い解析結果が得られると考えられる。

3. 日本語解析の方法

筆者らが言う「事例を用いた日本語解析」の方法とは、入力される日本語文の形態素解析を従来通りの方法で行い、その解析結果の日本語リストと過去の事例の類似点を基に、以降の解析を行っていくとするものである。これは、日本語では、英語のように単語間をあけるという「分かち書き」がされないため、[Martin89]のように最初から事例に基づく手法を用いることができないためである。

また、この方法で解析を行う場合、日本語文を入力する前に、どのような事例を参照しながら解析を

行うかについて、ユーザに対象とする分野を選択してもらうことにしている。これは、ユーザは、解析を行おうとしている日本語文の対象分野に関する知識（状況）を認識していると考えられるため、参照する事例の対象分野もユーザに選択させる方が、ユーザの求めている解析結果が得られやすいと考えられるからである。将来的には、この分野の判別については[Martin89]のように自動的にできるようにしたい。

3-1. 全体構成

事例を用いた日本語解析方法は、従来の体系的な辞書の代わりとなる事例を格納しておく「事例ベース」、ユーザの入力を基に実際に解析を行う「解析系」の大きく2つの構成要素によって実現できると考えている（図1）。その個々の構成要素の処理について、以下に記述する。

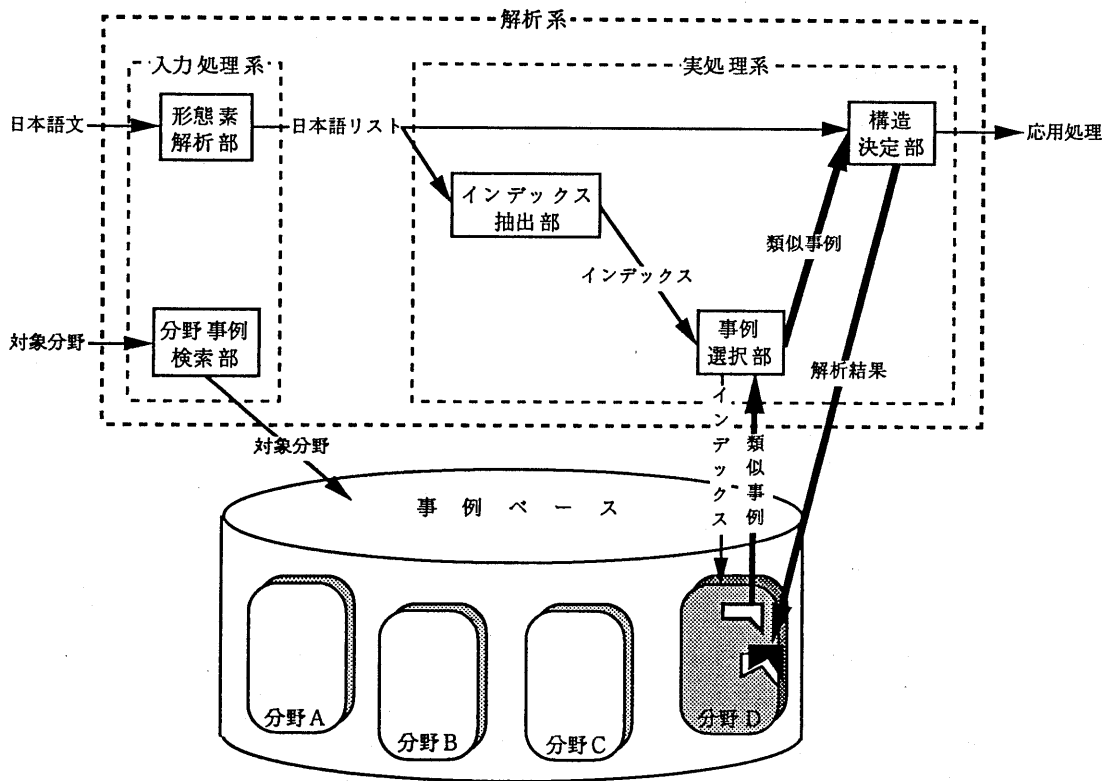


図1：全体構成

3-1-1. 事例ベース

事例ベースは、既に解析された日本語文の解析結果である過去の事例を格納しておく場所である。この事例ベースにある事例は、解析の際に参照されるため、どの分野に関する事例であるか、また、どのようなインデックスが使われている事例であるか、などが分かるように図2のような意味ネットワーク構造で構成されている。ここで、「源泉格」や「目的格」は、移動する対象である「対象格」の始点と終点を表し、「道具格」は、その移動の際に使用される道具を表す。

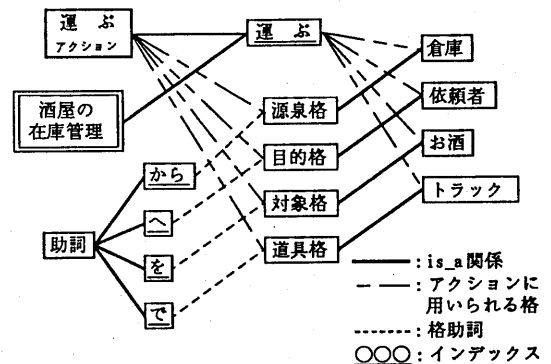


図2：過去の事例（1）

3-1-2. 解析系

解析系の内部は、図1にあるように入力処理系と実処理系の2つの処理系から構成され、それぞれの処理系はさらにいくつかの処理部より構成されている。

入力処理系は、解析の際に対象とする分野や日本語文など、ユーザの入力に対する処理を行う。この入力処理系は、図1にあるように2つの構成要素から構成されている。この2つの構成要素の処理はそれぞれ以下の通りである。

- ・形態素解析部：従来通りの方法で形態素解析を行い、日本語リストを作成する。
- ・分野事例検索部：ユーザによって選択された対象分野の全ての事例を事例ベースより検索する。

実処理系では、入力処理系によって得られた対象分野の全事例と日本語リストを基に、日本語文を解析していく。この実処理系は、図1にあるように3つの構成要素から構成されている。この3つの構成要素の処理はそれぞれ以下の通りである。

- ・インデックス抽出部：日本語リストを基に、用言（動詞など）や付属語などの構造を決定する上でキーとなるインデックスを抽出する。
- ・事例選択部：対象分野の全事例の中から、最も類似した事例を選択する。
- ・構造決定部：日本語リストとインデックス、最も類似した事例を基に構造を決定する。

3-2. 解析の手順

3-2-1. 処理の流れ

ここでは、事例を用いた日本語解析の手順を解析系の処理の流れに従って記述していく。

最初に、入力処理系においてユーザの入力に対する処理を行うのだが、まず、ユーザに対象とする分野を選択してもらい、その対象となった分野の全ての事例を事例ベースの中から検索する。次に、ユー

ザに日本語文を入力してもらい、従来と同様な処理によって形態素解析を行い日本語リストを作成する。

次の実処理系では、ここまでの入力処理系によって検索または作成された「対象分野の全事例」と「日本語リスト」を基に処理を行っていく。まず、日本語リストを基に用言や付属語など、過去の事例との照合の際にキーとなる語をインデックスとして抽出する。次に、対象分野の全事例のうち、マッチング関数Mの値が最大となる事例を最も類似した事例として選択する。このマッチング関数Mは、解析する日本語文のインデックスである用言や付属語を基に定義する。最後に、類似した事例の構造を参照しながら「対象格」や「道具格」などに該当する語を、類似事例のインデックスを基に当てはめていき構造を決定する。この構造を決定する際、それぞれの格に当てはまる語が決定できない場合には、ユーザにその格に当てはまる語を補充・決定してもらうことにより、構造を決定する。また、この処理によって新しく得られた構造を、逐次事例ベースに保存していく。

以上の処理によって全ての解析が終了する。この解析の処理手順をまとめると図3のようになる。

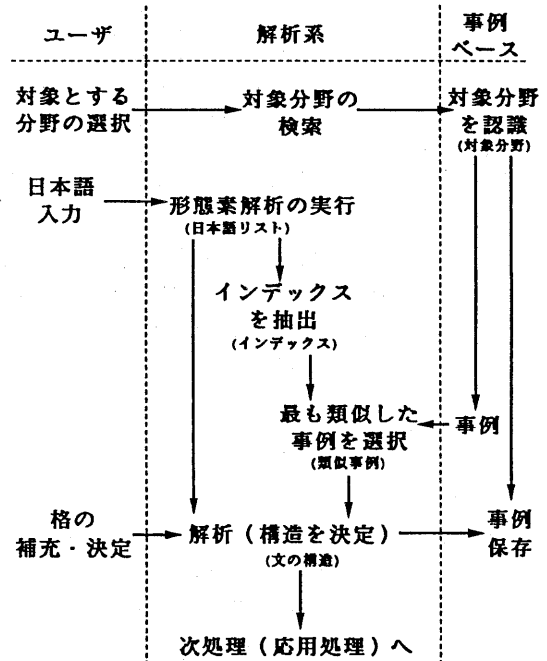


図3：処理の流れ

3-2-2. 処理の実例

実際に、処理の流れに従って『酒造所から倉庫へトラックで積荷を運ぶ。』という日本語文の解析を行うと、次のようになる。

まず、ユーザに対象とする分野を選択してもらい、その対象分野の全事例を検索するのだが、ここでは、「酒屋の在庫管理」が選択されたとし、この対象分野の全事例としての次の2つの事例が保存されているとする。

《対象分野の全事例》

『倉庫から依頼者へお酒をトラックで運ぶ。』（図2）

『お酒が酒造所から倉庫へ搬入される。』（図4）

次に、日本語文『酒造所から倉庫へトラックで積荷を運ぶ。』を入力してもらい、形態素解析を行うことにより次のような日本語リストを作成する。

《日本語リスト》

酒造所ーから 倉庫ーへ トラックーで 積荷ーを 運ぶ

ここまでが入力処理系の処理であり、この全事例と日本語リストを基に以降の実処理系の処理を行う。

まず、日本語リストを基にインデックスとして、用言や付属語を抽出する。

《インデックス》

「から」、「へ」、「で」、「を」、「運ぶ」

次に、インデックスを基に対象分野の全事例の中からマッチング関数Mを用いて最も類似した事例を選択するのだが、この例文の場合、次の事例が最も類似した事例として選択される。

《類似事例》

『倉庫から依頼者へお酒をトラックで運ぶ。』

最後に、最も類似した事例の構造(図2)を参照しながら、インデックスを基に「源泉格」「目的格」「対象格」「道具格」に当てはまる語を決定していき、図5のようなこの文の構造を得る。さらに、この文の構造を新たに事例ベースに加える。

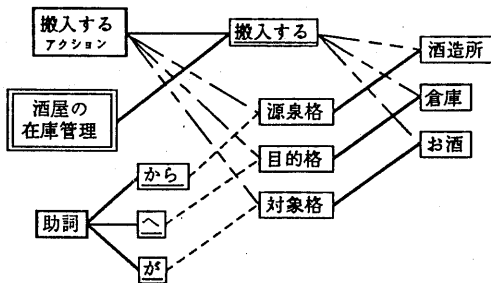


図4：過去の事例(2)

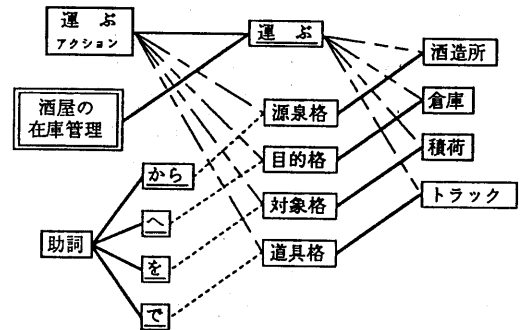


図5：解析結果(1)

4. 事例を用いた場合の利点

このように対象分野を制限して解析を行う場合、この事例を用いた解析方法は以下のような点で効果的である。

まずはじめに、対象とする分野固有の知識を用いた解析結果が得られることである。従来の解析方法で用いていた体系的な辞書や構文規則は、どのような分野においても解析が可能ないように一般的な知識を用いて作られている。そのため、従来の解析方法では、その分野固有の知識を用いた解析結果を得ることができなかった。しかし、事例を用いた解析方法の場合、過去の解析結果である事例に対象分野固有の知識やその使い方までが含まれているため、その分野固有の知識を用いた解析結果を得ることができる。

次に、格の省略されている文において、省略されている格が分かり、完全な解析結果を得ることが可能である。従来の方法で格の省略されている文を解析した場合、その格が省略されているのか、この文で用いられている動詞がその格を用いないのか分からないまま解析を行っていた。そのため、このような文では、その格に当てはまる語が補充されないまま解析を行ってしまい、完全な解析結果を得ることができなかった。しかし、事例を用いた解析方法の場合、事例の参照によって格が省略されているのではないかと推測できるようになる。これは、事例にはその動詞がどのような格を用いるかが含まれているからである。そのため、事例を基にその格に当てはまる語の候補をユーザに示し、決定または補充し

てもらうことにより完全な解析結果を得ることができる。

最後に、格をあいまいにする語を含む文において対象分野にあった解析結果が得られることである。従来の方法では、格をあいまいにする語を含むためその文に対する解析結果がいくつも考えられた場合、常に対象分野にあった解析結果が得られるとは限らなかった。しかし、事例を用いた方法の場合、事例にはその語の使い方が含まれているため、事例を参照することによって品詞や格に当てはまる語が決定できるようになる。そのため、対象分野にあった解析結果を得ることができるようになる。

ここでは、2、3番目の効果点について、実際の処理例を以下に述べていく。

4-1. 格の省略

格の省略によって格が決定できない場合として、例えば、『酒造所から積荷が搬入される。』という文が考えられる（この例では「道具格」については省略して考えることにし、対象分野は「酒屋の在庫管理」とする）。この文が入力された場合、従来の方法で解析を行うと、

動 詞：搬入する

源 泉：酒造所

対象物：積荷

という解析結果が作成される。このように、この解析方法では「目的格」に対応する語が分からないまま解析が終わってしまい、この文は「目的格」が省略されている文なのか、この文に用いられた『搬入する』という動詞が「目的格」をとらないのか、のどちらなのかが不明確なままになってしまう。

一方、事例を用いる方法でこの文の解析を行った場合、この文において、格が省略されているのか、この動詞には用いられない格なのか判断できるようになる。例えば、事例ベースに『お酒が酒造所から倉庫へ搬入される。』という文の解析結果が保存されていると仮定する。この文では、『が、から』という助詞が使われ、さらに、『搬入する』という動詞が使われており、図4のような形で『が、から、搬入する』という語をインデックスとした事例が保存されていることになる。ここで『酒造所から積荷が搬入される。』という文が入力された際に、最も類似した事例として先の事例を参照し解析することによって、『搬入する』という動詞が「目的格」をとるであろうことが推測できる。このように、類似した事例を参照することによって、この文では「目的格」が省略されていると推定することができ、その分野の事例を基に「目的格」となる語の候補をユーザに示し、選択または補充してもらうことにより図6のような解析結果を得ることができる。

このように、この事例を用いた方法で解析を行えば、省略されている「目的格」をユーザに補充して貰うことによって、格の省略が行われていた文の解析が可能になる。

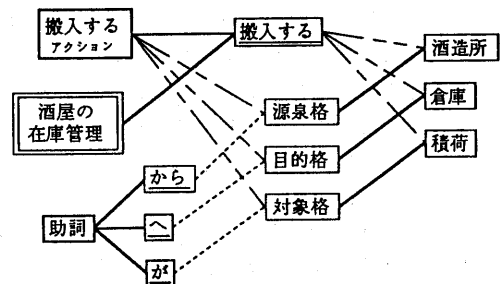


図6：解析結果（2）

4-2. 格をあいまいにする語

格をあいまいにする語を含んでいるために構造を決定できない場合として、例えば、『積荷はトラックで倉庫から販売店へ運ばれる。』という文があげられる。この文が入力された場合、『で』で文が一度切れると解釈したトラックが積荷であるという意味の「対象格」なのか、トラックを使ってという意味の「道具格」なのか、どちらの意味も含まれてしまうため、『トラック』という語がどちらの格となるのかあいまいである。従って、この文が入力された場合、従来の方法で解析を行うと、

- | | |
|--------------|--------------|
| (1) 動 詞 : 運ぶ | (2) 動 詞 : 運ぶ |
| 源 泉 : 倉庫 | 源 泉 : 倉庫 |
| 目 的 : 販売店 | 目 的 : 販売店 |
| 対象物 : トラック | 対象物 : 積荷 |
| | 道 具 : トラック |

の2つの解析結果が候補となってくる。しかし、この解析方法では、トラックを「対象格」とする(1)の解析結果と、「道具格」とする(2)の解析結果のどちらの結果がより好ましいかを決めるのは困難であるため、必ずしもユーザが求めている解析結果が得られるとは限らない。

このような格をあいまいにする語を含む文に対しても、蓄積されている事例を用いることによって、ユーザが求めている解析結果を得ることが可能となる。これは、現在解析されている文において、選択された分野に適した解析結果を候補としてあげることができるようになるからである。例えば、酒屋の在庫管理に関する分野において解析を行っているのであれば、インデックスを基に最も類似した事例として選択される事例(図2)から分かるように、『で』という助詞によって決定されるトラックという語が「対象格」でなく「道具格」とされるため、解析結果は(2)に対応する(図8)となることが推定できる。また、車の販売店に関する分野で解析を行っているのであれば、最も類似した事例である事例(図7)から分かるように、『で』という形容動詞(用言)の活用語尾によって決定されるトラックという語は「道具格」ではなく「対象格」となるため解析結果は(1)に対応する(図9)となることが推定できる。

このように、事例を用いた方法で解析を行えば、格をあいまいにする語を含んだ文も、ユーザの支

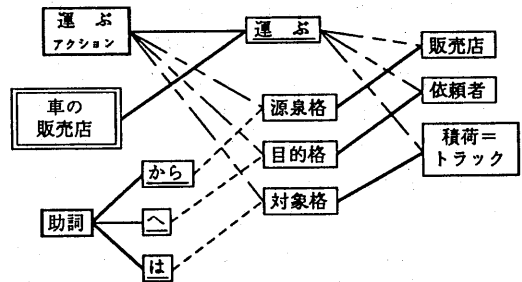


図7 : 過去の事例(3)

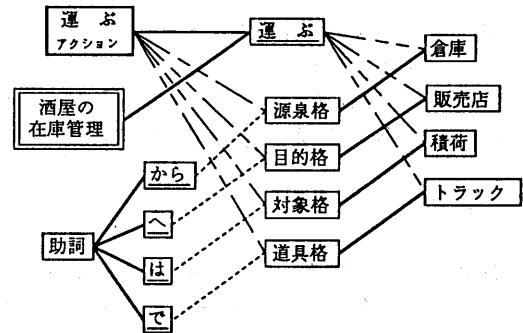


図8 : 解析結果(3)

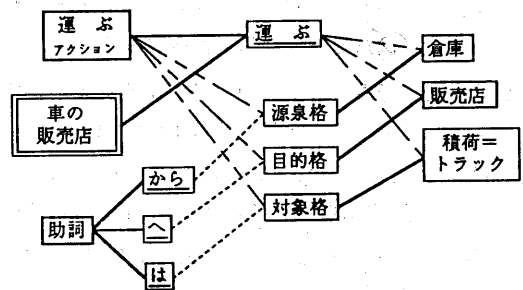


図9 : 解析結果(4)

援を得ることによって、解析が可能になる。

5. 終わりに

事例を用いて日本語解析を行った場合、1) 対象とする分野固有の知識を用いた解析が行える、2) 格の省略されている文において完全な解析が行える、3) 格をあいまいにする語を含む文において対象分野にあった解析が行える、といった利点がある。

また、このように事例を用いた解析方法では、「用言が含まれていない」といったような不十分な日本語文が入力されたときでさえ、対象分野やインデックス（この場合は用言以外のインデックス）を基に類似した事例を選択し、同じインデックスを用いている事例ではどのような用言が使われているかをユーザに提示することが可能である。それ故、従来の方では解析できなかった「用言がない」といった不十分な日本語文が入力された場合においても、ユーザに最も適している用言を選択（または補充）してもらうことが可能となる。

以上のことより、従来の解析方法は全ての分野を対象にしているため、対象分野固有の解釈が行えなかったのに対し、事例を用いた解析方法では、対象分野の事例を基に解析を行うので、対象分野固有の知識を用いた解釈が可能になる。

さらに、この解析方法では、1つの文を解析する度に解析結果を事例として保存していくため、従来方法で用いられていた体系的な辞書や構文規則の代わりとなる事例が、解析を行う度に増大する、といった学習効果も加わることになる。

今後は、現在検討中であるユーザインタフェース部を用いて、我々が現在開発中である発想支援システム「知恵の泉」[折原89a][折原89b]の知識獲得との相乗効果を確認・評価していく。

【参考文献】

- [浮田90] 浮田 輝彦, 自然言語インタフェース
人工知能学会全国大会 チュートリアル講演テキスト, pp. T5-1~T5-17, 1990.
- [Martin89] C. E. Martin. Case-based Parsing. In Riesbeck, C. K., Schank, R. C., editor,
Inside case-based reasoning, pp. 319~392. Lawrence Erlbaum Associates, Inc., 1989
- [折原89a] 折原 良平 他, 発想支援システムの構想
情報処理学会第38回全国大会, pp466~467, 1989.
- [折原89b] 折原 良平 他, 発想支援システム「知恵の泉」の類推機構,
ソフトウェア科学会第6回大会, pp. 85~88, 1989.
- [二村85] 二村 良彦 他, 新しいプログラミング・パラダイムによる共通問題の設計,
情報処理 Vol. 26, No. 5, pp. 458~459, 1985.
- [山崎84] 山崎 利治, 共通問題によるプログラム設計技報解説,
情報処理 Vol. 25, No. 9, pp. 934, 1984.
- [水谷83] 水谷 静夫 他, 文法と意味, 朝倉書店, 1983.