

事例を用いた日本語解析

4E-1

—「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

株式会社 東芝 システム・ソフトウェア技術研究所

1. はじめに

近年、システムが大規模化すると共に、システムに対する要求も多様化・複雑化する傾向にある。そのため、システム開発の初期工程である要求獲得工程では、ユーザの漠然とした問題意識やニーズを明確にし、ユーザが真に欲しているシステムイメージを要求として獲得しなければならない。我々は、この工程において、要求分析者がユーザ自身の認識していない潜在的な要求を含めて獲得する支援方法を研究しており、発想支援システム「知恵の泉」として開発中である [Ori89]。

このような工程を支援するシステムにおいて、ユーザインタフェースをどのように行うかが非常に重要なポイントとなってくる。一般に、ユーザから獲得される要求は、日本語(自然言語)のまま獲得されるのがほとんどである。従って、システムに日本語の要求をどのように理解させ、また処理させるかが重要な課題となってくる。そのため、システムにおけるユーザインタフェース機能として、日本語解析技術を用いることが大変効果的であると考えられる。

そこで本稿では、このような日本語解析機能の一実現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。従来の日本語解析では、一般に、解析は次のような処理手順によって行われている。1) 入力された文字列の単語の位置を同定し、それぞれの単語の品詞を決定する(形態素解析)。2) 形態素解析で得られた単語や文節を基本としてそれらの関係を解析し、文の構造を得る(構文解析)。3) 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明らかにし、文の内容を表す構造を作り出す(意味解析)。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へ進む [Uki90]。

通常これらの処理は、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、要求分析者やユーザ(以後、これらをまとめてユーザと呼ぶ)が思考過程において日本語文を解析しているとき、このように行っているとは考えにくい。むしろ、解析しようとしている日本語文と類似した過去の事例を基にしながら行っていると考えられる。それ故、要求分析工程において日本語解析を行う際には、辞書や構文規則を用いるよりも過去の事例を用いて行った方が、よりユーザの思考に近い解析結果が得られると考えられる。また「知恵の泉」は、過去の事例を用いて類推を行い対象領域の新しい知識を獲得していくため、日本語解析においても事例を用いた方が後の処理に有効である。

そこで、我々は、日本語解析の一部の処理において、過去の事例を用いて解析する方法について検討している。

3. 日本語解析の方法

我々が言う「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、日本語では通常「分かち書き」がされないため、[Mar89]のように初め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。

この方法で日本語解析をする場合、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させることにしている。これは、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)は、ユーザが認識していると考えられるため、参照とする事例の対象分野も、ユーザに選択させる方がより好ましい解析結果が得られると考えられるからである。また、このように参照事例とする対象分野の範囲を狭めることにより、解析しようとする分野に全く関係のない事例の選択が防止できるようになる。

3-1. 解析の手順

事例を用いた日本語解析の手順であるが、既に形態素解析が行われ、その解析結果である日本語リストがこの解析工程に入力されてくるものとする。

この解析工程では、まず入力されてくる日本語リストを基に、過去の事例を参照しながら解析を行う。その解析の際に参照する事例は、助詞、動詞、および対象分野を基に選択する。解析しようとする日本語文と、選択された事例とを照合しながら、事例で使われている格に該当する語を文中から決定することによって日本語解析を行う。また、省略されていた格や格のあいまいな語などを含んだ文においても、解析されなかった格をユーザに示し、その格の補充・決定などの助けを得ることで解析を行う。その際、事例を基に、各不透明点について最も頻度の高い候補をユーザに提示する。さらに、解析された事例を逐次保存していくことによって、事例ベースを増大させる。

3-2. 過去の事例

図1は、酒屋の在庫管理における事例の一部である。過去の事例は、このような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。

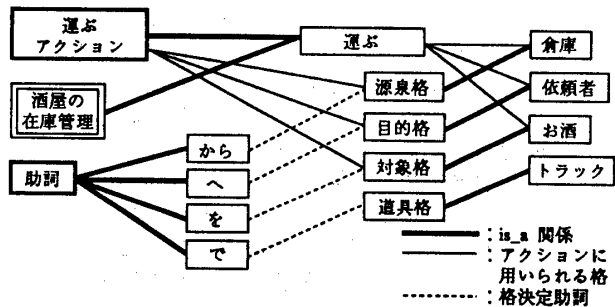


図1: 事例ベース

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

Hideyuki INOUE, Ryohei ORIHARA, Kazue NAGAO, Akihiko OHSUGA

TOSHIBA corporation

4. 事例を用いた場合の利点

辞書や構文規則を用いた解析方法では、1) 格の省略によって構造が決定できない、2) 格のあいまいな語を含んでいるために構造を決定できない、などの文（以後、不明確な格を含む文と呼ぶ）において、完全な構文解析が行われるとは限らない。しかも、日本語文においては、このような不明確な格を含んだ文が頻繁に使われている。

しかし、事例を用いた解析方法では、不明確な格を含んでいるために完全な構文解析が行えない文に対しても、それらの格をユーザに示すことにより補充して貰うことができるようになる。また、事例によればどんな構造が適当かを提示することができる。

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、

アクション：搬入する 源泉：酒造所 対象：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終わってしまい、この文では「目的格」が省略されているのか、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、どちらなのか不明確なままになってしまう。

一方、蓄積されている事例を用いる方法では、このような文において不明確な格が、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を判断できるようになる。例えば、『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事例に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2のような形で事例が保存されることになる。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、この事例を参照することによって、『搬入する』という動詞が「目的格」をとるであろうことが分かる。そこで、類似した事例の参照によって、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を、完全な文に修正することができるようになる。

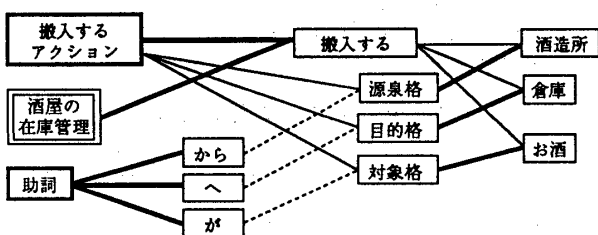


図2：参照事例（1）

4-2. 格のあいまいな語

格のあいまいな語を含んでいるために格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依頼者へ運ばれる。』という文がある。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、

アクション：運ぶ 対象：トラック ... (1)

アクション：運ぶ 対象：積荷 道具：トラック ... (2)
この2つの解析結果が候補となってくる。しかし、この方法で

は、トラックを「対象格」とするのか、「道具格」とするのか、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、ユーザが求めている解析結果が得られるとは限らない。

このような格のあいまいな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、事例（図1）から分かるように、トラックが「対象格」となることがなく「道具格」となることが可能なため、解析結果は（2）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、事例（図3）から分かるように、トラックは「道具格」ではなく「対象格」となるので解析結果は（1）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、格のあいまいな語を含んだ文も、完全な文に修正することができるようになる。

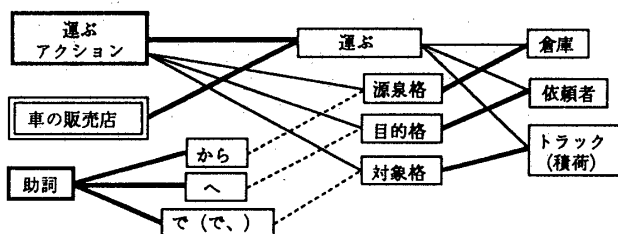


図3：参照事例（2）

5. 効果

事例を用いて日本語解析を行った場合、従来手法による解析方法では、不明確な格を含むために十分な解析が不可能であった日本語文に対しても、不明確な格の候補を解析の際に推定することができるようになる。

また、このように事例を用いた方法で解析を行った場合、「動詞がない」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、助詞と対象分野を基に事例を選択することによって、その文ではどのような動詞が使われる可能性があるかを推定し、候補となる動詞をユーザに提示することも可能である。

6. 終わりに

今後は、現在試作中のインタフェース部を用いて、「知恵の泉」の知識獲得と事例を用いた解析の相乗効果を確認・評価する。

参考文献

- [Mar89] C.E. Martin. Case-based parsing. In Riesbeck, C.K., Schank, R.C., editor, *Inside case-based reasoning*, pp. 319-352. Lawrence Erlbaum Associates, Inc., 1989.
- [Ori89] 折原 良平, 高柳 孝司, 菅野 成一郎, 長尾 加寿恵, 津田 淳一郎. 発想支援システムの構想. 情報処理学会第38回全国大会, pp. 466-467, 1989.
- [Uki90] 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.