

黑板モデルを導入した階層的な知識分割に基づく日本語意味解析システム

非会員 一本木真史<sup>†</sup>      非会員 浅野 晃<sup>†</sup>  
 正員 上原 邦昭<sup>†</sup>      正員 豊田 順一<sup>†</sup>

Japanese Language Understanding System Using Blackboard Model

Masashi IPPONGI<sup>†</sup>, Akira ASANO<sup>†</sup>, Nonmembers, Kuniaki UEHARA<sup>†</sup> and Jun'ichi TOYODA<sup>†</sup>, Members

あらまし 本報告では、格文法ベースにした係受け解析による意味解析システムについて述べる。一般に、意味解析は階層的な処理と仮説の生成および検証を伴う。これは問題解決の典型的なパターンである。意味解析を問題解決の一種とみなすと、パーザは問題解決器に相当するものと考えられる。我々が実現したシステムは、黑板モデルをパーザの基本機能として採用し、解析過程は文の先頭から解析を進め、後戻り処理を伴わないように構成している。また、意味解析の階層的な処理内容に従って、必要な知識を細分化し、それぞれ独立した知識源としてまとめている。この結果、細分化された知識の管理が容易になっており、自然言語特有の例外的な表現にも対処しやすくなっている。また、動詞の文型・述語素などを抽出するための辞書には、情報処理振興事業協会の開発した「日本語基本動詞辞書 IPAL」を採用している。

1. ま え が き

一般に、日本語文には文節内部の語順に関して強い制約がある反面、文節同士の順序に関しては非交差の原理などの緩い制約しか存在しない。従って、日本語文を意味解析する場合、英文のように統語情報を利用することができず、名詞の意味と助詞から文節間の係受け関係を抽出して、文全体としての意味構造を得るという方法を採用する必要がある。このような方針で設計されたシステムでは、既に解析の終了した文節をプッシュダウンスタックなどの短期記憶に蓄えておき、動詞から係受け関係を解析するための手掛りを発見した時点で、初めて文節と動詞の係受け関係を抽出するようにしている<sup>(1)</sup>。このため、係受け関係を決定する際に、必ず後戻り処理を伴うという問題点がある<sup>(2)</sup>。特に、複文などを解析する場合、後戻り処理による解析効率の低下が著しいものとなる。

一方、日本語文の解析に通常の句構造文法を用いた

場合、構文上の制約が緩いため、文法規則の数が膨大なものとなり、解析効率が非常に低下するという問題点がある。このような問題は、文法規則を、文節を解析するための規則、文の構造を解析するための規則などのように、適用される対象ごとに分類し、各々が必要となった段階で適用すれば、解決できると考えられる。また、処理過程の各段階で用いられる文法規則をすべて同等に扱うのではなく、それぞれの処理の段階に合わせて規則をモジュール化すれば、規則の追加・修正が容易になる。

本報告で提案する解析システムは、文頭からあらゆる係受け関係の候補を考慮しながら並列的に処理を進め、後戻り処理を伴わずに意味解析を実行している。更に、黑板モデル<sup>(3)(4)</sup>を導入して意味解析の過程を明確に階層化し、個々の処理内容に合わせて知識を細分化している。また、処理対象として、一般的な日本語の記述方式である漢字仮名混じりべた書き文を扱っている。

本システムは KCL (Kyoto Common Lisp) で記述されており、日本 DG 社の MV/8000-II 上で稼働している。また、解析用の動詞辞書には「日本語基本動詞辞

<sup>†</sup> 大阪大学産業科学研究所, 茨木市  
 The Institute of Scientific and Industrial Research, Osaka University, Ibaraki-shi, 567 Japan

書 IPAL<sup>(7)</sup>を採用し、関係データベース DG/SQL で管理している。DG/SQL は KCL の C 言語インタフェースによって結合されており、Lisp 関数から直接 IPAL を検索することが可能になっている。

## 2. 動作原理

### 2.1 意味解析の手法

通常、日本語文では文の意味構造の中心を成す動詞が最終文節に出現する。従って、後戻りをすることなく意味解析を行うためには、各文節が取り得るあらゆる格構造を考慮して、既に解析されている文節の係受け関係と組み合わせながら(仮説の生成)、新たな係受け関係を設定したり(仮説の修正)、既に設定されている係受け関係を棄却する(仮説の破棄)といった処理を行わなければならない。このような仮説の生成と検証の繰返しは、典型的な問題解決パターンと一致するため、日本語文の意味解析を一種の問題解決とみなすことができる。このような考え方に基づき、本システムでは、日本語文の意味解析過程を、「単語の意味を得る」、「文節の意味構造を得る」、「文節間の関係を得る」、「文全体の意味を得る」などをサブゴールとして、階層的な問題解決過程とみなし、個々のサブゴールを達成するために仮説の生成と検証を繰り返しながら解析を進めるという手法を採用している。

以上のような処理を実現するためには、処理内容に合わせた階層的な知識をもつシステムを構築しなければならない。本システムでは、パーザの基本機能として黒板モデルを利用し、各処理過程に必要な知識をそれぞれ知識源としてまとめて利用している。また、システムの制御部が黒板上のデータを常に監視し、機会主義的に知識源中の知識を発火させながら解析を進めるようにしている。

### 2.2 格構造候補の生成および絞り込み

本システムでは、文頭から文のタイプを予測するために、日本語文を意味内容により 12 種類に分類し、それぞれの文のタイプごとに、共起しうる格構造を設定している(表 1)<sup>(8)</sup>。例えば、「～は～と～について～」という文を解析する場合について考える。まず、システムが文節「～は」を読み込むと、表 1 を参照して入力文が①精神的行為に関する文、②観念的行為に関する文、③往来に関する文、④やりとりに関する文、⑤一般的行為に関する文、⑥現象に関する文などのタイプであると予測できる。続いて「～と」という文節を読み込むと、既に設定された予測は①精神的行為に関

表 1 文のタイプと設定格

文のタイプ	設定格構造
精神的行為に関して述べる文	主体, 内容, 対象
観念的行為に関して述べる文	主体, 対象, 目標, 相手, 手段/道具
往来に関して述べる文	主体, 場所, 出発点, 到着点
ものやりとりなどの共同的行為に関して述べる文	主体, 対象, 相手
動作・労働などの一般的行為に関して述べる文	主体, 対象, 目標, 相手, 対象の始状態, 対象の終状態, 目標物の材料, 手段/道具, 方法
現象に関して述べる文	対象, 原因, 変化の始点, 変化の終点, 所属の始点, 所属の終点
感情・感覚などに関して述べる文	主体, 源
ものの属性に関して述べる文	対象, 属性, 比較対象
可能性・必要性などに関して述べる文	主対象, 副対象
存在に関して述べる文	対象, 場所
存在の状態を述べる文	対象
対象間の関係などの認識に関して述べる文	対象, 相手

する文、②観念的行為に関する文、③やりとりに関する文、④一般的行為に関する文に絞り込むことができる。更に「～について」という文節を読み込むと、最終的に①やりとりに関する文であると文のタイプを決定できる。このように、既に解析された文節から格構造の候補を予測すれば、文の先頭から解析を行っても、動詞を認識せずにある程度解析文のタイプを決定することができる。また、あらゆる可能性を常に考慮しながら解析しているために、後戻り処理が必要なくなり、解析効率が向上するという利点がある。

本手法のもう一つの特徴として、複文などの解析を効率よく行うことができるという点が挙げられる。例えば、文

「私は彼が窓を壊すのを見た。」

が与えられた場合、「私は彼が」の部分まで読み込んだ時点で、システムは文節「私は」と「彼が」がどちらも文の主体となる可能性があることを発見し、解析文が複文であると予測できる。このように、解析の初期段階で文体を予測できるために、以後の解析で無駄な処理を省くことができる。

しかしながら、以上のような手法では、助詞のみから文体を予測しているために、必ずしも一意に文のタイプを決定できるとは限らない。例えば、「～が～を～に～」という文を解析する場合、①往来に関する文、②

やりとりに関する文、③一般的行為に関する文という文のタイプを予測できるが、これらを一つに絞り込むことはできない。このような場合は、以下の章で述べる単語の意味的な情報を利用する必要がある。

### 3. 意味解析に必要な知識

#### 3.1 チェック

自立語と付属語、付属語と付属語の結合の合文性を調べるために、結合の正当性を判断するための有限状態オートマトンをチェックと呼ぶ。チェックは、自立語の後に続く付属語の合文性を調べると同時に、各自立語に言語情報(品詞・活用型・活用形など)を与えるものである。本システムで使用したチェックは、文献(9)で提案された辞結合チェックを拡張し、助詞・助動詞などの品詞も同定できるようにしたものである。

#### 3.2 文節知識

文節知識とは、チェックで処理された結果から、自立語と付属語の順序関係を調べ、一文節を抽出するための知識である。文節知識は、基本的に「一文節内には自立語が1個だけ存在し、自立語は文節の先頭に現れる」という日本語の特徴を利用して構成している。現在、文節知識として約30個の規則を用意している。

#### [文節知識の例]

前提条件：名詞+助動詞+助詞+名詞

結果：一文節

#### 3.3 意味素性辞書

係受け関係の抽出においては、文中に最も多く存在する自立語(名詞)の意味を正確に把握する必要がある。本システムでは、動詞辞書IPALの意味記述方式に基づき、「人間」、「組織」、「動作」など、19個からなる名詞の意味素性を用意している。このような知識を意味素性辞書と呼ぶ。

#### 3.4 格構造知識

2.2で述べたように、本システムでは日本語文を意味内容により12種類に分類し、それぞれの文のタイプごとに共起しうる格構造を設定している。例えば、表1の文のタイプ「往來に関して述べる文」の格構造「出発点」は、「場所を表す名詞+から」、「場所を表す名詞+より」などの文節から予測することができる。このように、名詞の意味素性と助詞の結合状態から各文節が構成要素となりうる文のタイプと格構造候補を予測する規則を格構造知識と呼ぶ。

#### [格構造知識の例]

前提条件：場所を表す名詞+「から」

結果：文のタイプ→往來に関して述べる文  
格構造 → 出発点

#### 3.5 係受け範囲決定知識

解析文が複文の場合は、名詞から得られる意味的な情報や助詞を用いて、各文節がいずれの動詞に係るかを決定しなければならない。このような係受け関係を決定するために、各単語の意味・機能・用法などに基づいて、格構造候補から係受け関係を決定するための知識を係受け範囲決定知識と呼ぶ。係受け範囲決定知識は以下の基本方針に基づいて構成されている。

(a) 非交差の原理

(b) 1文1格の原理

(c) 一つの係受け範囲内に、複数のタイプの文の格構造は混在しない

#### 3.6 変形規則

文を先頭から解析する場合、受身や使役なども考慮して、すべての格構造候補を予測すると、格構造候補の数が膨大なものとなり、解析効率が極度に低下する可能性がある。このため、本システムでは受身や使役については前もって考慮せず、助動詞からそれらであることがわかった段階で、初めて変形規則を適用し、既に得られた格構造候補を書き換えるようにしている。例えば、受身文の場合は主格と目的格を入れ換える規則、使役文の場合は使役格を追加し、目的格を主格に戻すための規則などを用意している。これらの変形規則は2.4で示した文のタイプごとに設定されており、現在約20個用意している。

#### 3.7 動詞辞書

本システムでは、動詞辞書として情報処理振興事業協会より提供された「計算機用日本語基本動詞辞書IPAL」を利用し、動詞の文型・テンス・ヴォイス・アスペクト・述語素などを抽出している。述語素とは、名詞句と動詞、二つの名詞句と動詞、名詞句同士の結びつきなどに基づいて、名詞の意味と助詞から文の格関係を表したものである。IPALで用いられている述語素は約60種に分類されている。例えば、名詞句と動詞の結びつきを表す述語素には動作の主体や対象といったものがあり、二つの名詞句と動詞の結びつきを表す述語素には起点や着点などがある。また、名詞句同士の結びつきには部分-全体などの関係がある。本システムでは、これらの述語素を文の意味構造として利用している。

#### 4. 解析過程の詳細

一般に、黒板モデルは、①問題解決に必要な知識を処理対象に応じて蓄えた知識源、②問題解決過程の中間結果を階層的に保持しておく黒板、③黒板上のデータなどの知識源を適用するかを決定する制御部の三要素から構成されている。本システムでは、このような構成をもつ黒板モデルをパーザの基本機能として利用している。以下に解析手順の詳細について説明する。

##### 4.1 文字解析部

システムが入力文を読み込むと、まず形態素解析が行われる。形態素解析には、通常、自立語辞書が用いられるが、項目数が数万と非常に大きいため、2次記憶装置の使用が不可避となり、主記憶上での項目同士のつき合せに要する時間よりも、主記憶装置と2次記憶装置の間でデータ転送に要する時間の方が大きくなるという問題がある。また、従来の最長一致法や文節数最少法に基づく形態素解析では、一つの自立語を抽出するために上記のような辞書検索が数回行われるため、処理効率が非常に悪くなっている。このような無駄な処理過程をなくするために、本システムでは、「通常の日本語文では、漢字列・片仮名列は一つの自立語を構成し、平仮名列は助詞・助動詞列を構成する」といった字種情報を利用して、辞書を検索することなく自立語を抽出している。自立語が抽出されると、更にチェッカを利用して付属語との結合の正当性が調べられる。

##### 4.2 単語解析部

システムの制御部は、黒板上のデータを参照しながら形態素解析の過程を常に監視し、黒板上のデータが変化することに文節知識を適用して一文節を抽出している。更に、抽出された文節に名詞が含まれている場合は、意味素性辞書を利用して個々の名詞に意味素性を与えている。名詞が複数の意味素性をもつ場合は、係受け範囲決定知識を適用して解消している。文節知識および意味素性辞書が適用されると、意味素性が付与された1個の文節が黒板上に生成される。

##### 4.3 文節解析部

黒板上に文節が生成されると、その文節に格構造知識が適用され、格構造候補が与えられる。また、助動詞から解析文が使役文であると決定できる場合は、変形規則が適用され、文節に与えられたすべての格構造候補が書き換えられる。しかしながら、「られる」などの助動詞を伴う文の場合は、受身のはかに、自発・可

能・尊敬などの場合が考えられるので、黒板上のデータに新たな格構造候補を追加するという方法を採用している。以下の例は、やりとりに関する文の場合の変形について示したものである。

##### [変形例]

本が(主体)彼から(与える者)彼女に(受ける者)渡された。

↓(変形規則)

本が(対象)彼から(主体)彼女に(受ける者)渡された。

この例では、「渡された」という文節を読み込んだ時点で、解析文が受身であると予測され、「主体」と「与える者」という格構造が、それぞれ「対象」と「主体」に書き換えられている。

##### 4.4 文節間解析部

黒板上に格構造候補が生成されると、係受け範囲決定知識を適用して、文のタイプを絞り込みながら、係受け関係が決定される。また、1文1格の原理を満たさない係受け関係が発生する場合は、解析文が埋込み文であると予測して、その可能性が黒板上に書き込まれる。

この係受け範囲決定知識の適用によって、名詞の意味素性の多義性が解消される場合もある。例えば、「会社」という名詞は、組織を表す場合と場所を表す場合がある。もし、「会社に」という文節を含んだ解析文が「もののやりとりに関して述べている文」とであると予測されると、「会社」は組織を表していると決定できる。また、解析文が「往来に関して述べる文」ならば、場所を表していると決定できる。

##### 4.5 文解析部

解析文中に動詞が現れると、動詞辞書 IPAL から動詞の文型が取り出され、可能性のあるいくつかの係受け範囲のうち、動詞の文型と正しく対応づけられているものに絞り込まれる。但し、解析文が埋込み文の場合には、その種類に応じて上記の処理を以下のように分類している<sup>10)</sup>。

##### (a) 単純埋込み文の場合

埋込み文の動詞(副動詞)を含む文節が、主動詞の必須格の一つとなっている埋込み文を、単純埋込み文と呼ぶ。単純埋込み文を解析する場合は、全体構造が二つの単文から成り立っているとみなして、個々の動詞の文型と係受け範囲を対応づけている。

##### [単純埋込み文の例]

解析文:「私は 彼が 窓を 壊したと 考えた。」

文型：「壊す」：(人間 が)(生産物 を)  
 「考える」：(人間 が)(\_\_\_ と)

↓

文構造：「壊す」：(彼 が)(窓 を)  
 「考える」：(私 が)  
 (彼が窓を壊した と)

(b) 埋込み文の動詞直後の名詞(被修飾名詞)が埋込み文中の格要素に対応する、一般的連体埋込み文の場合

被修飾名詞を文型中の格要素の一つにあてはめて、動詞の文型と係受け範囲を対応づけている。

[一般的連体埋込み文の例]

解析文：「私は 彼が 弟に 渡した 資料を 読む。」

文型：「渡す」：(人間 が)(人間 に)(生産物 を)  
 「読む」：(人間 が)(生産物 を)

↓

文構造：「渡す」：(彼 が)(弟 に)(資料 を)  
 「読む」：(私 が)(資料 を)

(c) 埋込み文の被修飾名詞が埋込み文中の格要素に対応しない、特殊連体埋込み文の場合

この種の埋込み文の被修飾名詞には、「こと」、「とき」など、その語自身が明確な意味をもたない特殊な語が多い。本システムでは、これらの被修飾名詞が現れると、特殊連体埋込み文であるとみなして解析している。

[特殊連体埋込み文の例]

解析文：「私は 彼が 弟に 資料を 渡したことを知らない。」

文型：「渡す」：(人間 が)(人間 に)(生産物 を)  
 「知る」：(人間 が)(\_\_\_ を)

↓

文構造：「渡す」：(彼 が)(弟 に)(資料 を)  
 「知る」：(私 が)(こと を)

最後に、IPALからテンス・ヴォイス・アスペクトなどを取り出し、文の深層構造、すなわち述語素の各要素が決定される。IPALを用いても格構造候補が一つに絞り込めない場合は、その文自体にあいまい性があるとして、複数の解析結果を抽出している。

## 5. む す び

本稿では、黒板モデルを利用して知識を細分化し、各語の意味・用法・機能などを利用して解析を行う日本語意味解析システムについて述べた。本システムと同様に、後戻り処理を伴わずに文の先頭から解析を進

めるシステムに Connectionist Model を用いた連想記憶に基づくものが提案されている<sup>(1)</sup>。このシステムは、従来の逐次型計算機ではなく、並列型計算機をベースとして開発されている。このシステムは並列的な処理を行い、本システムの処理過程と類似したものであるが、並列型計算機は我々が即座に手にすることができるものではなく、システムの実用性という点において劣っていると思われる。また、基本メカニズムの設計が行われているのみで、比較的容易な文しか解析できず、システムの拡張にはかなりの時間を要すると思われる。

本システムは、各動詞に付随するほとんどの必須格を解析することができるが、IPALには任意格や修飾語句などに関する言語情報が記載されていないために、現在これらの処理については無視している。今後は、任意格用の格構造知識を追加して、システムの解析能力を拡張する予定である。

一方、以上のような制限を課しても、並列句の問題が残されている。例えば、格助詞「と」を伴う文節を含む文を解析する場合、本システムの特徴である「後戻りをしない」という原則を適用すると、並列句の可能性も同時に考慮する必要がある。更に、並列句の処理は、省略語の存在を仮定しなければならないという問題とも密接に絡んでおり、本システムの枠組みとは別のメカニズムで対処する必要がある。本システムでは、並列句が予想される「と」のような助詞を発見すると、通常解析と同時に、データ駆動型(デーモン)で意味素性辞書を用いた並列句処理用アルゴリズム<sup>(2)</sup>を起動することを考えている。しかしながら、黒板モデルとの制御の受渡しをどのように行うか、あるいは両者間の整合性をどのようにとるかなど、いくつかの課題が残されている。

また、現在は1文単位で解析を進めているため、指示代名詞を含む文はその名詞の意味内容を得られず、文のあいまい性が増加するという問題がある。更に、必須格が省略されている文から係受け関係を得ることはできず、このような文を解析することは不可能になっている。以上のような問題を解決するためには、文脈情報を利用する必要があり、その手法についても現在検討中である。

## 文 献

- (1) 日高, 吉田: "格文法による日本語の構文解析", 情処学自然言語処理技術シンポジウム, pp. 41-46 (1983).
- (2) H. P. Nii: "Blackboard systems: The blackboard

model of problem solving and the evolution of black-board architectures", The AI Magazine, 7, 2, Summer, pp. 38-53 (1986).

- (3) 池田尚志：“日本語文における格の種類についての考察”，情知学自然言語処理研究，41-2 (1984)。
- (4) 池田尚志：“語法規則方式による日本語文の構文・意味解析”，情知学論，26, 6, pp. 1079-1088 (昭60)。
- (5) 一本木，浅野，上原，豊田：“黑板モデルの導入により後戻り処理を伴わないべた書き日本語文の解析”，情知学自然言語処理研究，60-2 (1987)。
- (6) 井佐原，田中：“日本語埋め込み文の構文解析における諸問題”，情知学自然言語処理研究，26-4 (1981)。
- (7) 情報処理振興事業協会：“計算機用日本語基本動詞辞書説明書”(1986)。
- (8) L. D. Erman, F. Hayes-Roth, V. R. Lesser and D. R. Reddy：“The Hearsay-II speech-understanding system: Integrating knowledge to resolve uncertainty”, ACM Comput. Surv. 12, 2, pp. 213-254 (1980)。
- (9) 西村，水谷，尾上，大野：“日本語基本文法 複文篇”，電総研究報告，784 (1978)。
- (10) 田中穂積：“計算機による自然言語の意味処理に関する研究”，電総研究報告，797 (1979)。
- (11) 田村，安西：“Connectionist Model を用いた自然言語処理システム”，情知学論，28, 2, pp. 202-210 (昭62)。  
(昭和62年3月30日受付，5月27日再受付)



一本木真史

昭60 阪大・工・応物卒。昭62 同大大学院前期課程了。現在，住友海上火災勤務，自然言語理解に興味をもつ。情報処理学会会員。



浅野 晃

昭62 阪大・工・応物卒。現在，同大大学院前期課程在学中。自然言語理解に興味をもつ。

上原 邦昭



昭53 阪大・基礎工・情報卒。昭58 同大大学院博士課程退学。同年，大阪大学産業科学研究所勤務。現在，同研究所助手，工博。現在，人工知能，特に自然言語理解，および自動プログラム合成の研究に従事。ACM，計量国語学会，情報処理学会，人工知能学会，日本ソフトウェア科学会各会員。



豊田 順一

昭36 阪大・工・通信卒。昭41 同大大学院博士課程単位取得退学。同年阪大・基礎工・助手，昭44 助教授，昭57 大阪大学産業科学研究所教授。工博。現在，自然言語理解，画像理解，文書画像処理，およびICAIシステム等の研究に従事。日本認知科学会，人工知能学会，情報処理学会各会員。