

語法規則方式による日本語文の構文・意味解析†

池田 尚志††

本論文は、隣接する2項間の係り受け関係の解析を基本とした実時間的な日本語文の構文・意味解析の手法について述べている。この手法では係り受けのための短期記憶としてプッシュダウンスタックを用いており、日本語の係り受けの構造は解析システムの中に自然に埋め込まれている。すべての規則は語および文節カテゴリの属性として記述されており外部規則としての文法規則は持たない。文の構造を記述する文法規則ではなく、語句の用法・機能を記述する規則という意味で、ここで用いている規則を語法規則とよんでいる。この手法に基づく解析システムを UTIlisp 上に作成し機械翻訳への応用も含め現在実験を行っている。この方式によって、多くの個別的な事項を含むいろいろの言語事実に対処するための拡張や修正にも辞書記述の工夫で柔軟に対処することが可能となった。

1. はじめに

本論文では日本語文の構文・意味解析の手法について述べる。日本語文の構文解析の方法としては、格構造をベースにした係り受け解析の枠組みによるものが多い²⁾⁻⁴⁾。日本語は膠着語であって、文節という構文的単位が存在し、文節内の語の並びには強い制約があるが文節相互間の並びには係りの文節が受けの文節に先行することおよび言語一般の原則としての非交叉条件のほかには文体/文脈/語用論上の要因を除けば構文上の制約はないなどの特徴がある。また文節間の係り受け関係は文節の順序関係ではなく文節の構成要素である機能語(助詞の類)および/または直接意味的な関係から決まり、意味的に省略可能な文節は構文上も省略可能であるなど、日本語文は意味主導型の構造をしているといえる。語順が基本的な文法関係を支える欧米語の場合は句構造文法をベースにした構文解析がほとんどであるが、このような特徴をもった日本語文に対しては意味関係の記述に適した依存構造(係り受け構造)によって統語関係と意味関係を同時に処理していく方式が自然である。

しかし従来提案されている係り受け解析の枠組みによるシステムでは格構造のみを係り受けに関する語の属性としており、語の記述という点で不徹底である。このために、“まず動詞を抽出してつぎに…”とするなど入力文全体にわたって何段かの処理を繰り返す方式のものも多く、これは人間が文を認識する振舞いと比較すると不自然である。(文献4)は本論文の手法と

類似の発想のシステムであるが、格構造以外の係り受け関係の扱いについては詳しい記述がない。)人間は、文の入力と同時進行的に文認識の処理を進め、通常は、文の入力が終わった時点で認識処理を完了している。

本論文で述べる手法は、人間は、文の構造を抽象的にとらえた外部規則としての文法規則を用いて文を認識しているのではなく、語の意味・用法・機能など個々の語そのものに属する知識(語法規則とよぶ)を順次発動することによって実時間的な認識を行っており、文の一般的な構造に関する知識は文認識の機構そのものとして存在しているという想定を近似するものである。この方法は、隣接する2項間の係り受け関係の解析を基本とするもので、すべての規則は語および文節カテゴリの属性として記述している。この方式によって、多くの個別的な事項を含むいろいろの言語事実に対処するための拡張や修正にも辞書記述の工夫で柔軟に対処することをねらっている。

この手法に基づく解析システムを M 380 上に UTIlisp を用いて作成した。また、この解析システムの応用として機械翻訳システムを作成し実験している¹⁾。

文の認識は本来、場面・文脈・世界に関するもろもろの知識を背景とした「理解」という行為として位置づけなければならないが、本論文で扱うのは構文・意味解析の範囲であり、このシステムをそのような広い枠組みの中に埋めこむことは今後の課題である。

2. 解析の手法

解析システムの基本的な構成を図1に示す。

文節部門は入力文字列の先頭側から順次文節を切り出し整形処理、格構造処理を経て文節単独の構文・意味構造である基本依存構造方式を得、係り受け解析部

† Syntactic and Semantic Analysis of Japanese Sentences by Lexico-Grammatical Method by TAKASHI IKEDA (Electrotechnical Laboratory).

†† 電子技術総合研究所

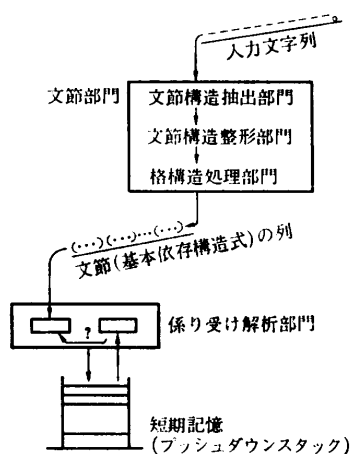


図 1 解析方式の概念図

Fig. 1 Overview of the method of analysis.

門に送り出す。

係り受け解析部門では、係り受け解析が局所的に終了し受けの依存構造式を待っている状態にある依存構造式を記憶しておく短期記憶としてプッシュダウンスタックを用いる。係り受け解析部門は、文節部門から送られてくる基本依存構造式に対し、順次次のようにして先行する語句との間の係り受け関係を調べる。

- D1. まず基本依存構造式を受けの項に置き、短期記憶上の依存構造式を係りの項に置いて、2項間の係り受け関係を調べる。
- D2. 係り受け関係が認められればその2項を新たな依存構造式として統合して受けの項に置き、短期記憶上の次の依存構造式を係りの項に置いて、同様に係り受け関係の処理を繰り返す。
- D3. 係り受け関係がなければ係りの項の依存構造式を短期記憶に戻し、受けの項の依存構造式を短期記憶に積む。

このように文頭側から順に認識を進め、最後の文節の処理が終わったときに短期記憶上に唯1個の依存構造式が残されていれば、それが入力文に対する構文・意味構造である。

以上が解析方式の基本であるが、このように短期記憶としてプッシュダウンスタックを用いることによって、日本語の係り受けの構造を自然に埋め込むことができる。人間は、文節の認識において若干の先読みをしていることや、「理解」の枠組みの中で認識を行っていることなどによって、通常局所的にも一意な解釈を得ているものと思われる、また係り受け関係もあいま

いさなく解釈され、文末の文節の処理の終了時に、一意な解析結果を短期記憶上に得ているものと思われる。(したがって短期記憶の深さも一つの文節に対する係りの文節の個数程度となる。)しかし本システムでは「理解」の枠組みを設定しておらず、また一般に機械処理の実際においては「完全な辞書記述」を前提できないとすれば、文節の切り出しにおいても係り受け解析においても複数の解釈の可能性を排除するわけにはいかない。

本システムでは複数の解釈に関して次のように対処している。

- ①仮定：自立語は最長一致で切り出されるものが正しい。
- ②仮定：最長の文字列を切り出す文節が正しい文節である。
- ③文節の付属語部分は通常一意的に解析できると前提する。一意的に解析できなかった場合には、整形処理、格構造処理を保留したまま係り受け解析部門に送り出し、係り受け解析部門で適切な文節構造を選択した後(4.2節③参照)、あらためて整形処理、格構造処理を経て、係り受け解析を行う。
- ④一つの文節に対して複数の解釈を許す。すなわち同形異語、一つの観念語に対する複数の格構造、一つの格構造に対する複数の変形操作を許す。
この結果、文節部門の出力は基本依存構造式の集合ということになる。
- ⑤係り受け解析部門では、二組の依存構造式の集合の間ですべての組み合わせについて係り受け関係を調べる。係り受け関係は、係り受け規則との適合度がある値以上であるとき成立するとみなす。
この適合度をもとに、統合生成された依存構造式の評価値を算出し、生成された依存構造式の集合の中で相対的にある値より低い評価値をもつものは棄却する。(適合度および評価値の算出については4.2節①参照)
- ⑥仮定：最も多くの係りをもつ依存構造式(の集合)が正しい依存構造式(の集合)であり、いったん成立した係り受け関係は破棄されることはない。

[すなわち後戻り処理は行わない。ただし、連体埋め込み構造に関しては別途の扱いとする。]

- ⑦文末の文節の処理が終了した段階で、短期記憶上に唯1個の要素が積まれていれば解析成功である

が、複数の要素が積まれているなら解析は完結しなかったことになる。本システムでは次のように、係り受け規則との適合度に関するしきい値を操作して「最も良い係り受け関係」を追求する処理を行う。(これを弛緩分析法とよぶ。(図2))

④しきい値を $T=T_0$ ($=0.9$) として前記D1～D3の係り受け処理を行う。

(短期記憶の要素が1個になっていれば終了。)

⑤短期記憶の内容を係り受け解析部門の入力側に逆順に積み戻し、しきい値を $T=T-d$ ($=T-0.1$) と下げてD1, D3 (D2は省く)の係り受け処理を行う。つまり、係り受けの成立した部分はそのままとし、成立しなかった隣接項間の係り受け関係をしきい値 $T-d$ で再度調べる。ただし、受けの項の再左端の係りとの間も調べる(図2)。

(短期記憶の要素が1個になっていれば終了。)

⑥⑤で新たな係り受け関係が成立した場合には、同様に短期記憶の内容を入力側に積み戻して⑤の手順に戻り係り受け関係を成長させる。

そうでなければ⑤の手順に戻しさらにしきい値を下げる。(ただし、しきい値が負になったら解析不能として終了する。)

解析が成功し短期記憶に残された1個の要素が複数の依存構造式の集合であるなら複数の解釈があったことになる。それらは評価値によって順序づけられている。

⑦補助手段として係り受けの範囲を指定する方法を提供する。すなわち、前編集として括弧を挿入しておけば、括弧で囲まれた部分を優先して一つの構造にまとめあげる。

弛緩分析法によれば、諸規則に完全には合致しない場合「たとえば、(昨日 彼 石 食べた)」のように助詞条件、意味条件が合わない文)でも「もっともらしい解釈」をひきだす。⑦の処理を繰り返す回数、機械にとっての理解のしづらさの度合となる。

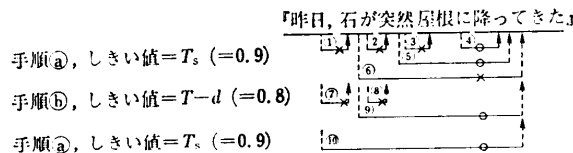


図2 弛緩分析法による係り受け関係の検査手順の例
Fig. 2 Determination of dependent-governor relations by the relaxation method.

仮定②は、たとえば、「僕の物」からは「僕の」を切り出すが、「僕のもの」からは「僕のもの」を切り出してしまうから一般には成立しない。しかし漢字仮名混じり文に関してはそのようなケースはまれであるから、個々に規則をたてて対処するか、そのような部分は分かち書きすることを想定してこの仮定を置いた。このような避難処置と比較して、この仮定を置かない場合の処理の負担は未登録語の処理法ともからんで大きすぎる。なお②は漢字仮名混じりの通常の表記での入力を前提としての仮定だが、カナまたはローマ字入力とする場合には、文節以下の単位に分かち書きしておけばよい。また将来の音声入力を想定したとしても、音声情報は文節分かち書きに相当する以上の情報を含んでいるはずだと考える。

仮定⑥は、弛緩分析の手順下での処置であるから、係り受け関係の評価値が適切であるなら、弱い結合をはやまって成立させてしまうことはない。ただ、連体埋め込み構造に関しては係り受け関係の評価値にかかわらず、埋め込みの範囲を誤る可能性が大きいから別途の扱いとした。

3. 文節部門

3.1 文節構造式の抽出

すべての語は、語幹を外見出しとして登録しており、語幹は、その語幹をもつ語の内部見出しの集合を値としてもつ。同形異語は内部見出しが異なることになる。内部見出しが語のいっさいの属性を保持している。

文節部門では、活用規則、接続規則、文節カテゴリ規則を用いて、自立語とそれに後接する付属語を切り出し文節構造式を抽出する。(句読点や疑問符等は付属語、接頭語の類は自立語として扱っている。なお、本システムでは、品詞という属性は存在しない。)

活用規則はすべての語が備えている属性であり、評価されて次のような活用リストを与える。

((活用語尾1 活用形1 付属語1 付属語1'...)
...)

活用リスト中の付属語は、後の処理のためにその活用形に伴って付属語を導入しておきたい場合に記述するもので、たとえば形容動詞語幹の活用リストを：

(... (nil -T *KD>T) ...)

として、形容動詞の語幹が体言化する場合にそれをマークするための擬機能語 *KD>Tを導入しておくことができる。

接続規則はすべての付属語が備えている属性であり、評価されて活用形のリストを与える。前接する語の活用形がこのリストに含まれている場合にのみ、その付属語は前の語に後接する付属語として認められる。

文節カテゴリ規則は活用形の属性であり、評価されて文節カテゴリ名または nil を与える。すなわち、活用条件と接続条件を満たす最大の文節が；

((活用形 付属語 1 付属語 2 …) 自立語 1 自立語 2 …)

{ここで、付属語 1 は最後尾の付属語であり、活用形は付属語 1 の活用形である。自立語 1、自立語 2 は同形異語である。}

のように切り出されたとき、この活用形の文節カテゴリの値が nil ならば、この文節は正当な文節として認定されない（たとえば未然形で終る文節は棄却される）。また、文節カテゴリ名を与えるならば、活用形をその文節カテゴリ名で置きかえたものを文節構造として出力する。

助詞の類の多くは、接続関係によって、つまり文節内での前後関係によって異なる機能をもつものが多いが、その場合にはすべて同形異語として扱っている。また、助詞の類の活用形は接続関係と文節カテゴリを考慮して決める。付録 1 に現在設定している文節カテゴリを示す。

未登録語の扱い

自立語が切り出せなかった場合には、次のようなヒューリスティックスで未登録自立語を含む文節構造式を抽出する。

- 付属語を 1 個以上含むような文節構造式で自立語部分が最短のものが妥当である。(ただし、文末にまで至る場合には付属語を含まなくてよい。)
- 未登録自立語は、体言/サ変/形容動詞の活用をする。

3.2 文節構造式の整形

付属語は意味・役割の上からは文構成上の機能を担う機能語であり、自立語は何らかの観念を指し示す観念語である。(接続詞、指示詞などは高次の機能語というべきであるが、ここでは便宜上一括して観念語とよぶ。)

ところで付属語の中には、観念語の役割を併せもつとみなすべきものがあり、構文・意味解析の立場からは、適当な観念語で置き換える、あるいは適当な観念語を前後に補って処理を進めると都合がよいものがあ

る。たとえば、「これは本だ」の「だ」は付属語であるが、構文・意味解析の上からは「これ」と「本」を係りとして文を構成する用言とみなしたほうが都合がよい。「これは本らしい」や「これは本か」の「らしい」や「か」の直前にはこの用言「だ」が省略されているとみなすべきである。また、「彼のはつまらない」の「の」の直後には、体言「モノ」が省略されているとみなすべきであり、「行くのはつまらない」の「の」は体言とみなすべきである。

付属語の属性である整形規則は、このように文節構造を変更したり追加したりするプログラムである。文節構造整形部門は、文節構造式中の付属語を出現順に走査し、整形規則を持っていればそれを評価する。

整形規則としては、上述のような観念語への置き換え、観念語の導入だけでなく、機能語の挿入/削除/置換の他任意の処理を組み込むことができ、入力文表層の字面処理と以後の構文・意味処理の間にたって規則記述の柔軟性に寄与することができる。

3.3 格構造式への展開および格構造式の変形

3.3.1 格構造規則

格構造規則は観念語がもつ属性であり、その観念語がどのような語句(依存構造式)を係りとしてとるかということ、すなわち受けの観点からの観念語の用法を記述するものであって係り受け規則の一部を成すものである。係る可能性のある語句なら何でも格構造規則として記述しておいてさしつかえないが、通常は、その観念語の基本的な用法にかかわるものを格構造規則として記述する。

格構造規則は次のように定義される格構造の集合であって、一つの格構造が観念語の一つの用法に対応する。

((ptype s1 s2…) SP1 SP2…)

ptype ……格構造のタイプを示す。

s ……その格構造にまつわる意味特徴。

SP ……係りの項にある依存構造式が係りとして認定されるための条件を記述した式(格条件式)。

SP :=(rtype EC SC)

rtype ……格役割名

EC ……特記条件

SC ……意味条件

ptype は用法の分類として使用することができ、また、ptype および s によって、観念語の属性として記述してある意味特徴のほかに、その用法に伴って現れ

る意味特徴を追加することができる。

格役割名は、受けに対する係りの構文・意味的な位置・役割を表示すると同時に、係りの語句がその係りの位置にあること（ないしその可能性）をマークする機能語（助詞の類）あるいは文節カテゴリ名の集合を値としてもつ。これには適合度を付随させることができる。

意味条件は、係りの依存構造式が備えているべき意味特徴の集合であり一般には和、積、否定を用いて記述できる。意味条件にも適合度を付随させることができる。

特記条件は通常は空であるが、係り受けに関するその他の特別な条件がある場合に、適当なりスプ関数としてそれを記述するものである。また省略のされにくさ、連体修飾のされにくさなども特記条件として記述できる。

格構造は元来、文の意味的構造を表現するものとして用言に関連して議論されてきたものであるが、本システムでいう格構造は観念語の基底的な係り受け関係を定式化したものであって、用言に限らずすべての観念語がもつ属性として一般化している。本システムでは格役割は係り受けの関係一般に関してそのクラスを表示すると同時に係りの依存構造式を識別する名前であって、格構造に現れる格役割はその一部を成すものである。付録2, 3に現在設定している ptype, rtype について示す。

3.3.2 変形規則

機能語の中には格構造を変形する機能をもつものがある。（受身「れる」、使役「られる」など。また「～的だ」、「～化する」や3.1節であげた擬機能語 *KD>T による体言化変形なども含む。）つまり、日本語では、機能語（助動詞の類）を観念語に後接することによって、その観念語が表現する基底的な内容と派生的に関連する内容を表現することが行われる。機能語の属性である変形規則はこのように格構造を変形する（格条件式を変更／削除／追加し、また格構造のタイプを変更する）規則であり、それが付属している動詞の意味特徴なども考慮してプログラムすることができる（付録4参照）。この変形は何段か連続することもある。格構造処理部門では、文節構造式中の機能語を出現順に走査して変形規則をもつものがあればそれを評価する。

3.3.3 格構造処理部門での処理

格構造処理部門は、文節構造式中の観念語の格構造

規則をとりだし、必要な変形処理を行い基本依存構造式を作って係り受け解析部門に送り出す。

依存構造式は次のように定義される。（図3中の例参照）

(H (type fwl fw2...)(cw s1 s2...) D1 D2...)

H——係り受け解析の履歴を保持しておく式

type——文節カテゴリ／格役割名

fw——機能語

cw——観念語

s——意味特徴

D——格条件式／依存構造式

意味特徴としては、cw の属性として記述されている意味特徴と、格構造のタイプおよび格構造にまつわる意味特徴、さらにそれらの上位の意味特徴すべてが並べられる。上位の意味特徴は意味特徴の属性として辞書に記述されている。（なお、意味特徴の照合時には、cw 自身も意味特徴と同列の扱いを受ける。）

基本依存構造式では、Hは nil, type は文節カテゴリ、Dはすべて格条件式である。係り受け解析部門で、依存構造式 Dm が依存構造式 Df の係りとして認定されると、Dm は Df の最左端の係りとして付加され、それが格条件式と照合の結果成立した係り受けである場合にはその格条件式は削除される。その際、Dm の文節カテゴリ名は格役割名で置き換えられ、係り受け成立に関する諸情報が Df の履歴保持式に書き加えられる。

4. 係り受け解析部門

4.1 係り受け規則

（係り受け解析処理の例を図3に示す。）

係り受け規則には係りの規則と受けの規則があり、文節カテゴリ、機能語、観念語の属性として記述されている。これらはリスプのプログラムであり、係り受け解析部門で次に示す①～⑦の順序で評価される。ただし、評価した結果、値が非 nil である場合には、係り受け関係が確立したか、または係り受け関係のありえないことが判明したものと解釈して以後の係り受け規則は評価しない。

（以下で Dm は係りの項にある依存構造式、Df は受けの項にある依存構造式である。）

- ①Df の文節カテゴリの受けの規則
- ②Dm の文節カテゴリの係りの規則
- ③Df の観念語の受けの規則
- ④Dm の観念語の係りの規則

『昨日は、太郎と花子だけが市場に出かけた』

- A B C D E
- “出かける”の格構造規則を ($\langle T \text{ I.E} \rangle (:T \text{ 0 !OA}) (.TT \text{ 0 !OP})$) とする。
 - I) DとEの係り受け処理が終わった時点では;
Df \Rightarrow ($\langle \langle \text{市場に...} \rangle \rangle$ (\$SYUSHI \text{ タ1} \rangle) (出かける $\langle T \text{ I.E} \rangle$ (NIL (.TT =0) (市場 !PC)) (: T0 !OA))
短期記憶 \Rightarrow ($\langle \langle \text{花子だけが} \rangle \langle \text{太郎と} \rangle \langle \text{昨日は} \rangle$)
 - II) Dm \Rightarrow (NIL (\$T > Y \text{ ガ0} \text{ ダケ0} \rangle) (花子 !OHH))
短期記憶 \Rightarrow ($\langle \langle \text{太郎と} \rangle \langle \text{昨日は} \rangle$)
 - ダケ0 とガ0 の結合性規則が評価され、ガ0 が係りの機能語とみなされる。
 - 4.1 節に示した順序で係り受け規則が評価され、この場合連用性体言文節カテゴリ $\$T < Y$ の係りの規則 (4.1 節に例示した) によって Dm が Df 中の格条件式 (: T 0 !OA) に適合することが確認され Df は次のように変わる。{ 係りの機能語ガ0 は、4.1 節 ⑤ でいう係りの規則 (ガ0 の場合は無し) をとりだす対象となるだけでなく格構造中の機能語条件 (今の場合: T で示されている) の対象ともなる。 }
Df \Rightarrow ($\langle \langle \text{花子だけが...} \rangle \langle \text{市場に...} \rangle \rangle$ (\$SYUSHI \text{ タ1} \rangle) (出かける ...) (NIL (: T \text{ ガ0} \text{ ダケ0} \rangle) (花子 ...)) (NIL (.TT =0) (市場 ...)))
 - III) Dm \Rightarrow (NIL (\$T > \text{ ト0} \rangle) (太郎 !OHH))
短期記憶 \Rightarrow ($\langle \langle \text{昨日は} \rangle$)
 - 係りの機能語は ト0 である。
 - 同様に係り受け規則が順次に評価されるが、ここでは ト0 の係りの規則を (HRULE & AND 1) としてあれば、Dm は $\langle \text{花子だけが} \rangle$ と並立することが確認される。
Df \Rightarrow ($\langle \langle \text{花子だけが...} \rangle \langle \text{市場に...} \rangle \rangle$ (\$SYUSHI \text{ タ1} \rangle) (出かける ...) ($\langle \langle \text{太郎と...} \rangle \rangle$ (: T \text{ ガ0} \text{ ダケ0} \rangle) (花子 ...) (NIL (&AND \text{ ト0} \rangle) (太郎 ...))) (NIL (.TT =0) (市場 ...)))
 - IV) Dm \Rightarrow (NIL (\$T > Y \text{ ハ0} \rangle) (昨日 !TD))
短期記憶 \Rightarrow NIL
 - “昨日”の係りの規則を、
(GRULE [係りの機能語=(ハ0 モ0 NIL) かつ、Df が機能後タ1 を含む \rightarrow 1.0] _T 1)
としてあれば、これによって Dm は Df に係ることが確認され解析は終了する。
{ この規則は “昨日” の係り構造規則として次のように記述することもできる。
(_T (RC [Df はタ1 を含む] <)
ここで、_T は機能語条件 (ハ0 モ0 NIL) を持つ。RC は、特記条件として係り受け条件を追加する関数。< は任意の用言が持つ意味特徴。 }

図 3 係り受け解析処理の例

Fig. 3 Example of the dependency analysis procedure.

⑤Dm の係りの機能語の係りの規則

⑥Df の文節カテゴリの受けの規則

⑦Dm の文節カテゴリの係りの規則

⑤でいう係りの機能語は、Dm に付属している機能語のうち結合性規則の値が最も高いもの (等しい場合はより外側のもの、またどの機能語も結合性規則の値が nil なら nil) である。結合性規則は機能語の属性であって結合に関与する機能を持つか否かを記す規則であり、たとえば格助詞や接続助詞の類の場合 1.0、副助詞の類の場合 0.5 を与えておけばよい。

文節カテゴリに属する係り受け規則には、最初に評価されるもの (①, ②) と最後にされるもの (⑥, ⑦) とがある。

以上の係り受け規則は係り受け関係の有無を判定し処理する包括的な手続きであるが、より具体的に係り受け関係を記述する規則として、格構造規則、係り構造規則、一般結合規則の三つがある。格構造規則と次に述べる係り構造規則は、通常、上記の①または②の

係り受け規則の中で評価される。

たとえば、連用性体言文節の文節カテゴリの係りの規則 (②で評価される) は次のように書くことができる。

“Df が用言性の依存構造式でないか、Dm の係りの機能語が「は」で Df が連体形であれば値 T を返す。そうでなければ格構造規則を評価する関数、つまり『Df 中の格条件式に Dm が適合するか否かを調べ、適合するなら Dm を Df に統合する』関数を評価する。また、Dm の観念語の係り構造規則も評価する。”

係り構造規則は観念語の属性であって、格構造規則が、受ける側から係りの語句を規定する規則であるのと対照的に、係る側から受けの語句を規定する規則である。この規則は、副詞や連体詞あるいは連体性体言文節などの係りの条件を記述するのに有用である。係り構造規則は格条件式と同じ形式: (rtype EC SC) で定義される係り構造の集合である。ただし、意味条件

SC は格構造の場合とは逆に、受けの依存構造式が備えているべき意味条件となる。

一般結合規則は、格構造規則や係り構造規則による係り受け関係としては把握しにくい構造を処理するための規則で、次のような関数を用いて記述する。

(GRULE condition rtype index)

ここで、condition は、係り受けの条件を記述する関数であって、値は nil、または係り受け関係の適合度を表す数値である。

rtype は係り受け関係が認められ Dm を Df に統合するときに付加される格役割名であるが、格構造規則や係り構造規則のときのような条件記述には関与しない。

index は、この一般結合規則を識別するための任意の記号であり、係り受けの履歴保持項中の情報の一つとして、どの規則で Dm が Df に統合されたかを記しておくためのものである。(なお、格構造規則と係り構造規則の場合には、何番目に記述されている格構造あるいは係り構造であるかということが、この識別記号に対応するものとなっている。)

たとえば、連用性体言文節の文節カテゴリの係りの規則(⑦で評価される)を GRULE を用いて、(GRULE 0.05 -T>Y 1) と書いておけば、弛緩分析法によって、“連用性体言文節は、最終的には無条件で、後統する用言文節に係る (低い評価値 0.05 で)” という規則を記述したことになる。(ただし、先に例示した連用性体言文節の係りの規則(②で評価される)を前提している。)

このほか、一般結合規則の特殊な場合として、並立構造を扱う関数:(HRULE rtype index) と、連体埋め込み構造を扱う関数:(RRULE) がある。

RRULE は次のような処理を行う関数であり、「と」や「や」等の機能語の係りの規則の中で評価されることを想定している。

“Df に係っている依存構造式の中で最左端のものを D としたとき、Dm が D と並立して Df に係るか否かを調べる。すなわち、D と Df の係り受けを成立させたのと同じ条件 (Df の係り受けの履歴保持項に記されている) で Dm と Df の間に係り受け関係が成立するか否かを調べ (ただし機能語条件は無視する)、さらに Dm と D の意味特徴の類似度をも加味してこの係り受けの適合度を定める。係り受けが成立するなら格役割名 rtype をもって Dm を D に統合する。ただしこの判定で係り受けが否定されても、Df が文末の

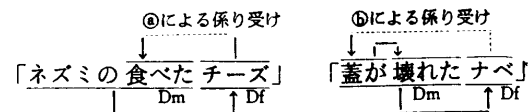
語句である場合には、Dm は Df に並立すると認定する。”

RRULE は、Dm が連体形用言の依存構造式、Df がそれに対する受けの体言の依存構造式である場合に、表層の係り受けの方向とは逆に、Df が Dm の係りとして成立するか否かを調べる次のような関数である。

“④Df が Dm の格構造規則に適合するなら (この場合、機能語条件は無視する) Df を Dm に統合し、さらにその Dm を Df に統合する。

⑤適合しないなら、Dm は既に係っている依存構造式の格構造規則に Df が適合するか否かを調べ同様の処理を行う”

連体埋め込み構造の解析については次節で述べるが、RRULE は連体形用言文節カテゴリの係りの規則の中で評価されるものである。この RRULE によって次のような同一名詞連体修飾の表現が分析される。



4.2 係り受け解析部門の処理

係り受け解析部門での基本的な処理手順については(2)で述べた。ここでは、いくつかの補足を行う。

①係り受け関係の適合度と依存構造式の評価値

一般結合規則の場合には係り受け関係の適合度は関数 GRULE の第 2 引数 condition を評価した値そのものである。格構造規則と係り構造規則の場合には、格役割名 rtype で示される機能語条件に付随して記されている適合度を A、意味条件に付随して記されている適合度を B としたとき (3.3.1 項参照。記されていないならば 1.0 とみなす。)、機能語条件を重視した加重平均 $0.6 \times A + 0.4 \times B$ を適合度としている。

また、依存構造式 G の評価値 h は、G への係りが D1, D2, ..., Dm であり、Di の評価値が hi、Di と G との適合度が gi であったとし、また、G の格構造中で格条件式 SP1, SP2, ..., SPn が省略項として残されているとすると、 $h = \frac{\sum_{i=1}^m (h_i + g_i)}{2 - \sum_{j=1}^n p_j}$ とする。ここで pj は省略に対するペナルティであり、既定値として 0.1 を与えているが、格構造中の特記条件の一つとして (px) と記しておけば、その項の省略に対するペナルティは x となる。また G への係りが無い場合 h の第 1 項は 1.0 とする。

②連体修飾構造の解析

Df が連体形用言文節の場合、次のような後戻り法

によって Df の係り受けの範囲を定める。

まず, Df を受けとして通常の係り受け解析を行い (結果を Df' とする), 次に後続する体言文節を解析する (結果を Dn とする). Df' を係り, Dn を受けとして係り受け解析を行う. 係り受けが成立しなければ, Df' の最左端の依存構造式 (つまり, Df に最も近い係り) との係り受け関係を解除してもとに戻したうえで, 再度 Dn との係り受け解析を行う. この手順を繰り返すことによって Df の受けの範囲を定める。

Dn が連体性体言文節で後続する体言文節に係る場合には, Df' とその体言との係り受け関係も調べ, 評価値がより良くなるならば, Df' が係る範囲をその体言文節にまで広げる。

連体修飾関係には次の 2 種がある*。

④ 同一名詞連体修飾 (Dn が Df' への係りとなる)

例: 「ネズミの食べた チーズ」
 Df' Dn
 「蓋が壊れた ナベ」
 Df' Dn

⑤ 付加連体修飾 (Dn は Df' への係りとならない)

例: 「魚の焼ける におい」
 Df' Dn

“におい” の格構造はたとえば,

(N>)(=E)

= E は連体形文節を要求する格役割名。

特記条件無し, 意味条件無し。

したがって, 通常の係り受け規則の適用を行う (⑥ の場合を解析する) と同時に, Dn を係り, Df' を受けとした係り受け規則の適用 (前節で述べた関数 RRULE を評価する) も行わなければならない。

③ 文節構造式が一意に定まっていない場合の処置

この場合には (そのようなケースはまれであるという前提だが), 文節構造式のリストがそのまま係り受け解析部門にもちこまれる。係り受け解析部門では文節カテゴリの属性である係り文節選択規則, 受け文節選択規則を評価して文節構造式の候補をしばらくこむことを試みる。

たとえば, 「泳ぎ,」 に対しては, 連用形用言文節とする解釈と動詞が体言化した文節とする解釈がありうるが, 連用形用言文節カテゴリと体言文節カテゴリの受け文節選択規則として, それぞれ, “連用性体言文節, 副用言文節”, “連体性体言文節, 連体形用言文節, 副体言文節” と記述しておけば, 係りの側にたとえば「彼は」があれば連用形用言文節と解釈され, 「彼

の」とあれば体言文節と解釈される。

受け文節選択規則で一意にしばらくこめない場合には, そのまま短期記憶に積んで解析を進め, 後段で係り文節選択規則を評価する。それでも一意にしばらくこめない場合は現在はユーザに問い合わせさせて確定するようにしている。

④ 係り受け解析終了時の処理

Df の係りとなって取り込まれた各依存構造式に対してノードをわりふるとともに, 各依存構造式の機能語列を走査し, 機能語の属性である法情報規則を評価する。法情報規則は任意のプログラムであるが, 機能語の様相的働きに関連した情報をノードに書き込むことを想定して設けた規則である。この規則によってたとえば, 「は」/「も」を伴っている場合に『主題』/『共鳴』の標識をつける, 「行かないでいる/行っていない」の「ないでいる/ていない」の働きをそれぞれマークしておくなどの処理をすることができる。法情報規則のより積極的な活用は文脈処理ともからんだ今後の課題である。

5. おわりに

個々の語および文節カテゴリの属性である語法規則を評価することによって, 実時間的に日本語文の構文・意味解析を行う手法について述べた。

この手法に基づく解析システムを M380 上に UTIlisp を用いて作成し, 現在, 機能語約 400 語, 基本的語彙を中心とした自立語約 2000 語を対象として辞書記述を試み, 機械翻訳への応用も含めて実験を行っている。(解析システムは約 1500 行程度である。) 広範な言語規則の整理は今後の積み重ねが必要であるが, ここで述べた手法によって多くの個別的事項を含む言語諸規則を比較的容易にかつ体系だって記述し解析できるようになったと考える。

問題点および今後の課題としては;

① 語の見出しの配列法を検討すること。(現在は他の語との関連無しに単に機械的に配列しているため, 最長一致法で自立語を切り出す過程は実時間的になっていない。)

② 意味特徴の有無ではなく, より有効な意味的距離/類似度の尺度を見出すこと。

③ 解析システムを「理解」の枠組みの中に埋め込んで, 文脈解析を行い, またそこからのフィードバックを得ること。

④ 日本語の有用な制限文法を定めること。

* 奥津: 生成日本文法論, 大修館, 東京 (1974)。

⑤他の言語への適用について検討すること。
などがある。

謝辞 最後に研究推進に激励をいただいている若松清司部長、柿倉正義研究室長に、また研究室の諸氏に感謝いたします。

参 考 文 献

- 1) 池田：語法規則変換方式による機械翻訳—翻訳の表現式と翻訳文生成関数—, 情報処理学会「自然言語処理技術シンポジウム」(1984).
- 2) 日高, 吉田：格文法による日本語の構文解析, 情報処理学会「自然言語処理技術シンポジウム」(1983).
- 3) 島津, 内藤, 野村：機械翻訳システム LUTE の日本語意味解析, 情報処理学会自然言語処理研究会, 33 (1982).
- 4) 市山, 村木：日本語処理の観点から見たプッシュダウンオートマタの再評価, 情報処理学会「自然言語処理技術シンポジウム」(1983).
- 5) 池田：一般化された格構造による意味表現を用いた日本語文の構文解析法について, 信学論, Vol. 60-D, No. 10, pp. 814-821 (1977).
- 6) 池田：日本語文における格の種類についての考察, 情報処理学会自然言語処理研究会, 41(1984).

文の種類

1. 事実の知覚に関して述べる文

①行為に関して述べる文

- | | |
|-------------------------------|--------------------------|
| イ. 意志・思考などの精神作用に関して述べる文…<W | 私は彼を憎む |
| ロ. 観念上の行為に関して述べる文 | …<D 点の集りを線と呼ぶ |
| ハ. 往来に関して述べる文 | …<T 首相が全国を遊説する |
| ニ. 相手を伴う共同的行為に関して述べる文 | …<C 不動産業者に土地を売る |
| ホ. 動作・労働などイ〜ニ以外の一般の行為に関して述べる文 | …<L 窓を壊す・歌を歌う
文書を翻訳する |

②現象に関して述べる文

…<P 台風で雨が降る

③状態に関して述べる文

- | | |
|----------------------|------------------|
| イ. 感情・感覚などに関して述べる文 | …<F 私は本を／本が読みたい |
| ロ. ものの属性に関して述べる文 | …<A 電子は質量が小さい |
| ハ. 可能性・必要性などに関して述べる文 | …<Z 植物には文が必要だ／要る |
| ニ. 存在に関して述べる文 | …<E ブラックホールがある |
| ホ. 存在の状態を述べる文 | …<Y 山がそびえている. |
| ヘ. 結果・状態に関して述べる文 | …<X 彼は棒が頭に当たった |

2. 対象間の関係などの認識に関して述べる文

…<R 分子は原子を含む. 猿は動物だ

3. 事実の評価・解釈などについて述べる文

- | | |
|---------------------|------------------|
| ①事実の影響を被るものについて述べる文 | …<S 彼は雨に降られた |
| ②事実の原因・発動者について述べる文 | …<M 台風が雨を降らせる |
| ③行為を受け取るものについて述べる文 | …<V 子が親に本を読んでもらう |
| ④行為を受け取るものについて述べる文 | …<B 親が子に本を読んであげる |

付録 1 文節カテゴリーの種類

語を体言、用言、副言（副詞／連体詞／接続詞）に大別し、文節カテゴリーを次のように設定している。

文節カテゴリー	例
1 \$T>Y	連用性体言文節 太郎が
2 \$T>T	連体性体言文節 太郎の
3 \$T>	体言文節 太郎,
4 \$RENYOU	連用形文節 歩き
5 \$RENTAI	連体形文節 歩く
6 \$KATEI	仮定形文節 歩けば
7 \$SYUSHI	終止形文節 歩く
8 \$MEIREI	命令形文節 歩け
9 \$TOIKAKE	疑問形文節 歩くか
10 \$Y>Y	連用性用言文節 歩くので
11 \$F>T	連体詞文節 その
12 \$F>Y	副詞文節 ゆっくり, しかし
13 \$F>	副言文節 とても, もっと

付録 2 格構造のタイプ (ptype) について

- 副言には格構造は持たせていない。
- 体言には現在, ptype を区別しておらずすべて N> としている。
- 用言には、以下のように文の種類（すなわち用言の用法）と対応して ptype を設定している⁶⁾。

ptype 文の例

付録 3 格役割 (rtype) の種類について

● 格役割は次のようにその種類に応じてその先頭の記号で大別している。

種類	先頭の記号	説明
①体言 → 体言	@	「の」に係るもの
	&	「での」、「からの」など
	*	②の用法が名詞化された場合
②体言 → 用言	:	主格
	.	目的格, 対象格など
	-	時, 所, その他の一

		一般格
③用言 → 体言	=	連体形で係るもの
④用言 → 用言	\$	論理的な関係
	¥	事象の連接
	%	格関係として係るもの
⑤副言 → 体言	-	「その本」
⑥副言 → 用言		「ゆっくり歩く」
⑦副言 → 副言	/	「もっとゆっくり」

● 用言の格構造中に現れる格役割は、各 ptype ごとにそれぞれ数個の格役割を設定している⁶⁾。以下にいくつかの ptype についての格役割を示す。

ptype	格役割名	意味	基本の機能語条件	例
<T	: T	往来する主体	ガ	首相がアメリカに行く
	. TP	往来する場所	ヲ	野原を歩く。大学を卒業する
	. TF	出発点	カラ, ヲ	日本を出発する
	. TT	到着点	ニ, ヘ	日本に到着する
	. TM	手段	デ	飛行機で行く
<C	: C	やりとりする主体	ガ	議員が演説する
	. CO	やりとりする対象	ヲ	資料を渡す。話を聞く
	. CA	やりとりする対象	ニツイテ, ニカンシテ	天候について話す
	. CT	やりとりする相手	ニ	彼に話す
	. CW	やりとりする相手	ト	彼と話す。彼女と結婚する
	. CF	やりとりする相手	カラ, ニ	彼に貰う
	% CQ	やりとりする内容	ト, ヨウニ	行くように言う
<A	: A	属性を述べる対象	ハ	水は密度が大きい
	. AO	属性を述べる対象	ガ	役場は屋根が新しい。彼は妻が病気だ。春はあけぼのが良い。蠣は広島が本場だ。
	. AN	属性	ニ	この水は飲料に適する。日本は資源に富む
	. AC	比較の対象	ト	彼は体格が父親と似ている

付録 4 変形規則の例

変形規則は格構造の種類 (ptype) を参照して記述している。以下に「~てもらう」に対応する変形規則の例を示す。XA は格役割を追加する関数、XX は格役割を変更する関数である。この規則によって、機

能語“テモラウ”が後接している場合、ptype: <W, <T, <C, <L, <E, <M は <B に変換されて格役割: B (主格) が追加され (<M の場合は目的語 [使役されるもの] が: B に変更される。 <C の場合も同様の変更が為される場合もある。), 各文の主格が: B に変更される。

変形規則の記述形式

(新しい ptype

(現 ptype の並び リスブ関数の並び)

(現 ptype の並び リスブ関数の並び)

…)

“テモラウ”に対応する変形規則

```
( <B(( <W <T <C <L <E)
(XA(:B !OHH))
(XX(:W ..B) (:T ..B)(:C ..B)(:L ..B)(:E ..B)))
(<M(XX (.MW. :B)(.MC. :B)(:M ..B)))
(<B(<C(XX(.CT. :B)(.CW. :B)(:C ..B))))
```

(昭和 59 年 12 月 21 日受付)

(昭和 60 年 5 月 9 日採録)