

日本語文法 - 水谷文法と結合価文法を基盤にして -

3K-10

丸山 直子

日本アイ・ビー・エム株式会社

サイエンス・インスティテュート

0.はじめに

筆者は現在、日英機械翻訳のための日本語文法を書いている。本稿では、その特徴と概要を述べる。この文法は、Lispで書かれたパーサ PLNLP (bottom up parallel)^[1]の上にインプリメントされている。特徴は以下の三点である。

- 1) 水谷文法を基盤にしていること
- 2) 結合価文法の考え方を採用していること
- 3) いわゆる助動詞について、それを一種の関数と考えることによって、様々な情報を抽出しようとしていること

1.で水谷文法の概要とその導入の仕方、2.で結合価文法の概要とその導入の仕方を述べ、3.で助動詞を関数として扱うとはどういうことか、4.で出力の形について、述べる。

1.水谷文法の概要と導入

1.1.水谷文法の概要

ここで、水谷文法と言っているのは、概ね文献[2]に記述されているものを指している。山田文法と時枝文法を統合発展させたものである。基本的にCFG(文脈自由文法)で書かれている。大枠を図示すると、以下のようになる。

文	-->	喚態句 述態句
述態句	-->	述素 述態辞
述素	-->	体連語 相連語 用連語 体連語【格表示】述素 情況語【情況化】述素
用連語	-->	(用連語O 用連語S 用連語R) [動詞特類1 動詞特類2]
用連語O	-->	動詞
用連語S	-->	述素 セル
用連語R	-->	述素 レル

これは、全くおどろきなものにすぎない。ここで、述態辞とは助動詞(「ダ」「アル」「ナイ」「デス」「ゴザル」「マス」「ベシ」「タ」「ウ」「ヌ」「マイ」「ラシイ」)の並び(「セル」「レル」「ソウダ」「ヨウダ」「タイ」などは別扱い)、動詞特類とは補助用言(1はテを介さないもの:「ハジメル」「ツツケル」の類、2はテを介するもの:「テイル」「テアル」の類)をそれぞれさす。体連語は体言相当の表現、相連語は形容詞相当の表現、用連語は動詞相当の表現、情況語は副詞相当の表現である。

水谷文法の特徴として私が考える点を列挙すると、次のようになる。

- ・ 枠組がきちんと形式化されており、除外仕様も銘記されている。
- ・ 文組成論と文成立論を区別して考え、係助詞・陳述副詞を文の成立にかかわるとする。
- ・ 表現の構文要素としてのレベルを、語の水準、句の水準、文の水準というように設定している。
- ・ 文の構造を入れ子構造として考えている。
e.g. ((それが犬だ) ((雨が降ら)ない)

1.2.水谷文法の導入

PLNLPでは、ルールをボトムアップに記述する(つまり、右辺・左辺を通常の書き替え規則とは逆に書く)ようになっている。また、構文的・意味的情報を、素性としてもつことが可能であり、ある条件が成立したときにそのルールが起動するように書ける。基本的に左辺がconditionで右辺がactionである。左辺の条件が整ったときに右辺のノードが生成されるのである。

機械翻訳の対象とする文献に喚態句に相当するものが出てこないことから、喚態句を導入しない、とか、無形の辞としての零記号を導入しない、また、水谷文法では三分木で記されているものを二分木で記述するなどの変更はあるものの、概ね水谷文法を忠実に模している。[]で示されているものについては、前接する要素の素性として記述したり、それぞれに対応する非末端記号を設けたりして、具現化している。

除外仕様については、殆どがまだ手付かずだが、広義の係り結びについては、呼応関係がはっきりしているものだけを具現化している。

e.g. ~しか~ない, 決して~ない

これは否定のスコープとも絡む問題で、格表示の種類や、情況語の種類によって、述態辞「ナイ」がついたものに対して、さらに「体連語【格表示】」や「情況語【情況化】」がつくようにしている。つまり、ある条件を満たす「体連語【格表示】」・「情況語【情況化】」に対して、次のようなルールを設けているのである。

(〈体連語〉【格表示】 | 〈情況語〉【情況化】) <述態句O>
--> <述態句>

2.結合価文法の概要と導入

2.1.結合価文法の概要

結合価文法は、テニエル他にその端を発するものだが^[2]、この文法は、動詞(実際には動詞に限らない。述語の核になっているもの)を中心にして、それがどのような要素を支配するかを記述するものである。これには、表層のレベルと深層のレベルが考えられるが、特に表層に着目したものを結合価文法、深層レベルで捉えるものを格文法とよぶのが一般的である。

2.2.結合価文法の導入

水谷文法では、〈体連語〉【格表示】が無制限に述素に付きうる。これは、「太郎が次郎が犬を学校に帽子を家から食べた」などの表現を許すことを意味し、述語になる要素を多く含む文が入力された場合には、その係り受けに関して多くの曖昧さを認めることにもなる。例えば、「彼女に自然食品を食べることを勧めた。」という文があった場合、「彼女に」「自然食品を」はともに、「食べる」にも「勧めらる」にも係りうる。従って、次の三つの解釈が成り立つ。

- 1) (((彼女に(自然食品を食べる))ことを勧め)た。
 ...「彼女に」->「食べる」,
 「自然食品を」->「食べる」
- 2) (彼女に((自然食品を食べる)ことを勧め)た。
 ...「彼女に」->「勧める」,
 「自然食品を」->「食べる」
- 3) (彼女に(自然食品を(食べる)ことを勧め)))た。
 ...「彼女に」->「勧める」,
 「自然食品を」->「勧める」

ここに、結合価文法の考え方を導入するとどうなるか。それぞれの動詞(勿論、名詞にも意味素性をふっておく必要がある)に次のような情報を与えておけば、曖昧さは解消され、2)の解釈のみがとられる(3)の解釈がとれないのは、「勧める」が「食べることを」と「自然食品を」の両者をとることができないからである)。

動詞 食べる ... N[hum]ガ N[con]ヲ V
 勧める ... N[hum]ガ N[abs]ヲ N[hum]ニ V

但し、ここでは、結合価文法を、文解析のもっともらしさの目安として用いている。つまり、結合価をもって、すべてを判断することには限界があるとみて、もっともらしさの重み付けに使うという方法を取っている。ある程度ゆるい文法にしておくことで、解析できる文の範囲を大きくしておこうという思想である。

3. 助動詞を関数として扱う

助動詞を関数として扱うとはどういうことなのかについて、ここでは述べる。

大筋としては、ある性質をもった要素を引項としてとると、ある性質をもった要素を値として返す、そういうものとして考えるのである[3]。例えば、「テイル」の場合(これは動詞特類2であるが)、原則として、動作性(+act)・持続性(+cont)・過程性(+prog)をもった用連語を引項としてとると、「進行」というアスペクトをもった用連語を値として返す、そういう関数として記述するのである。一方、変化(+ch)の属性をもった用連語をとると、「結果の残存」というアスペクトをもった用連語を値として返す。「テイル」にはいろいろな意味があって、前にくる動詞によって、その意味が異なる」というところを関数として捉えなおすのである。「テイル」の意味が複数あるのではなくて、「テイル」の意味は、関数としてのマッピングそのもの、つまり、

用連語 [+ch] --> 用連語 [結果の残存] 経験, 反復

用連語 [+act, +cont, +prog]

--> 用連語 [進行] 経験, 反復

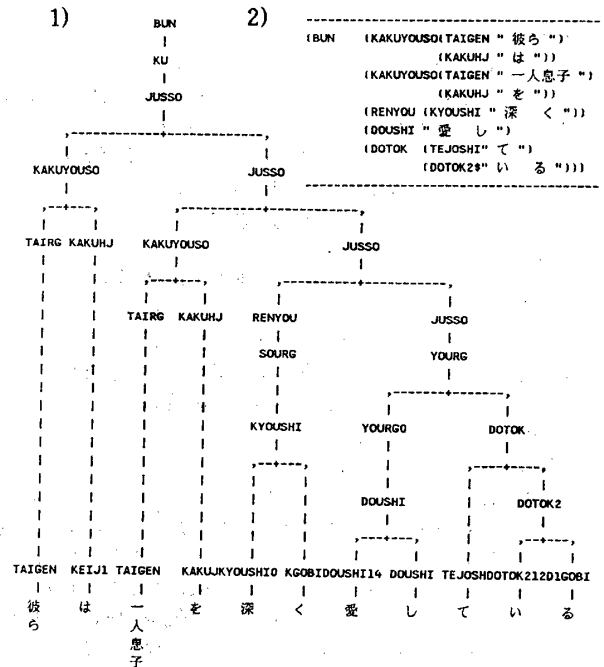
用連語 [+act, -cont] --> 用連語 [経験] 反復

となる、これ全体であるとするのである。ここで、左辺の [] の中は、用連語の意味属性、右辺の [] の中は、新たに生まれた、アスペクト情報である('様々な情報' と言っているなかには、アスペクト情報の他に、テンス情報や、ムードの情報がある)。アスペクト情報中の '>' は、強さを表しており、他に何も規定するものがなければ、最も左に位置するものが読み手(聞き手)に意識される。例えば、「歩く」という用連語に「テイル」がついた場合を考えてみよう。「歩く」の意味属性は [+act, +cont, +prog] である。「歩いている」は、[進行] 経験, 反復 をアスペクト情報としてもつ用連語ということになる。「歩いている」をそれだけで聞いた場合、それは、現在「歩く」という動作を行っているという「進行」の意味を担うものとして解釈される。ところが、これに、「すでに」という副詞がついたらどうだろうか。「すでに歩いている」というのは、「経験」を表す表現である。

また、「毎日」という副詞がついたらどうだろうか。「毎日歩いている」というのは、「反復」を表している。ある種の副詞は、このようにアスペクト情報の候補を絞る働きをする。筆者の文法では、このような副詞の働きも反映させている。また、述態辞に関しても、例えば、「ウ」の場合、
 述素 [+意志性] --> 述態句 [意志] 勧誘
 述素 [-意志性] --> 述態句 [推量]
 というような記述を施している(右辺の [] の中は、ムードの情報である)。

4. 出力の形

出力としては、木を出力するが、これには二種類ある。一つは、rule structure tree と呼ばれるもので、ルールの動きどおりに生成される木であり、一つは、computed structure tree と呼ばれるもので、よりフラットな構造として、捉えるものである。例えば、「彼らは一人息子を深く愛している。」という文の場合、rule structure tree としては、1)が、computed structure tree としては、2)が出力される。2)の形のものが、さらに依存構造に変換され、transfer 部門に渡される。



5. おわりに

本稿では、筆者が現在開発中の、水谷文法と結合価文法を基盤にした日本語文法について、その概要を述べた。水谷文法を詳細化し、実用(この場合は日英機械翻訳)に耐え得るものにするのが、目標である。水谷文法で除外仕様になっているところをどう扱うか、意味的な要素をどこまでいれるか、チューニングアップの中でアドホックでないルールを作り上げていくには、どうすればよいか、などさらに検討を重ねていくつもりである。

参考文献

- [1] Heidorn, George E. (1986) forthcoming. "PLNLP: the Programming Language for Natural Language Processing." IBM Journal of Research and Development.
- [2] 水谷静夫編(1983)『朝倉日本語新講座3 文法と意味I』朝倉書店。
- [3] 賀来直子(1985)『日本語のアスペクト』情報処理学会第31回全国大会論文集 pp.1361-1362.