

結合価構造に基づく日本文解析

林 良彦

N T T 電気通信研究所

本稿では、日本語用言の持つ結合価構造を文解析の色々な局面に適用する手法について述べる。結合価構造は、結合価パターンと呼ぶ構造パターンとして辞書中に記述される。この構造パターンは、単位文の構造決定を行なうためのパターンマッチングに適用される他、埋め込み名詞句の構造決定、複合用言におけるゼロ代名詞化された格要素の補完、複合名詞・複合用言の解析などにも適用される。すなわち、結合価構造に基づくことにより、これらの言語現象に対しても統一的な解析を行なうことが可能となる。本稿では、さらに日英機械翻訳への適用における問題点についても論じる。

On the Japanese sentences analysis
based on the valency structure

Yoshihiko HAYASHI

NTT Electrical Communications Laboratories
1-2356, Take, Yokosuka-City, Kanagawa, 238 Japan

In this paper, a method to apply the valency structure of Japanese predicates to Japanese sentences analysis is presented. In the dictionary, each of the valency structures is formulated as a structure pattern named "Valency-pattern". This structural pattern is applied to the analysis of unit-sentences, suppletion of zero-pronoun case-elements, analysis of embedded noun-phrases, analysis of complex words. These analysis processes are discussed. Taking this structure to be basis of the Japanese sentences analyzing process allows the uniform treatment of such language phenomena. The problems in applying this method to Japanese-to-English machine-translation are also discussed.

1. はじめに

日本語文の解析においては、依存文法あるいは格文法的な考え方が広く用いられている。このような考え方は、文構造を動詞などの用言を述語とした関数構造としてとらえようとする所にその特徴がある。この述語の持つ関数構造という考え方は、現在においては、生成文法やLFGなどにおいてもみられる考え方である。

格文法では、文の意味を述語とその支配要素である名詞句との間の深層格関係としてとらえようとする。ところが、実際の自然言語処理の場面においては、深層格の認定は必ずしも必要でなく、例えば機械翻訳では原言語と目的言語との間の対応関係を認定することのほうが重要である。機械翻訳においては、むしろ、深層格の認定を誤ることによって適切な訳出がなされなくなるということの方が問題である。また、設定する深層格のセット、1つ1つの深層格の定義についても明確な基準が存在しないことも、大語彙を対象とする実際の自然言語処理システムの構築において大きな問題となる。この点で、同じ述語の関数構造という考えを基にした文法理論でも、格文法よりも表層の情報重視する結合価文法のほうが現実的である([1])。

筆者らは、以上のような考えに基づき、用言とその支配要素である格要素との結び付きを記述した結合価パターンと呼ぶ構造パターンを日本語文解析の色々な局面に適用する方法について検討を行ってきた。すなわち、結合価パターンは、日本語単位文レベルの解析に用いられる他、これを用いることにより、名詞句における埋め込み構造の解析、複合文におけるゼロ代名詞化された格要素の補完、複合名詞・複合用言の解析等が実現される。

本稿では、まず2章で結合価構造に基づく日本語解析の概要について述べた後、3章で単位文解析におけるパターンマッチング、格要素の補完、複合用言解析への適用について述べる。また、4章では、日英翻訳における日本語解析に本方法を適用した時の問題点等について論じる。

2. 結合価構造と日本語解析

2.1 格文法と結合価文法

格文法と結合価文法は共に、依存文法をその基盤とした文法理論であると考えられる。その共通点の最大なのは、文構造をとらえるさいに動詞などの述語を中心にすえるという点である([1])。また、その相違点は、深層における意味関係を考えるか、表層における要素の結び付きを重視するかにあると考えられる。

格文法においては、いくつかの深層格を設定するが、その格のセットをどのようなものにするか、それぞれの格をどう定義するかについては、定まった基準がない。これは、実際の自然言語処理を格文法に従って実現しようとするさいに、大きな障害となる。すなわち、動詞などの用言に対して格フレームを与える必要があるが、それが統一性のないものになってしまう恐れがある。これに対し、結合価文

法においては、深層格というものは設定しない。問題となるのは、どのような意味素性を持つ名詞がどのような表層格マーカ（日本語であれば助詞など）を伴って注目している用言と共起するかという点である。

以上より、機械翻訳のような応用に対しては、深い意味理解が必ずしも常に要求されないこと、また、大語彙を対象とする場合の実現性から、格文法よりも結合価文法の考え方にに基づく方が現実的であると考えられる。

2.2 結合価構造パターンの記述

用言の持つ結合価構造をベースとする解析方式においては、それぞれの用言の持つ結合価構造を適当な枠組みに形式化して、それを辞書に記述しておくが必要になる。以下に、構造パターン記述における基本的な方針を示す。**複数の構造パターンの記述**：同一用言の持つ複数の用法に対するそれぞれの構造パターンは、別パターンとして記述する。格要素に対する制約条件の記述によって、それぞれのパターンが極力排他的になるように記述する。

格要素に対する制約条件：格要素を作る名詞句の中心名詞に対する意味的制約条件と、助詞表現部分に対する制約条件がある。前者については、階層的意味素性体系を用いて、とるべき意味素性、とってはならない意味素性を記述することができる。これによって、パターン間の関係のある程度まで排他的に構成することができる。後者については、助詞表現部分のバリエーションをどの程度まで構造パターンにおいて吸収し、どの程度を処理において救済するかという問題がある。パターンの記述においては、全ての助詞表現のバリエーションを吸収することはいたずらにパターンの数を増やすことになりかねない。そこで、極力格助詞を用いて記述を行ない、格助詞の還元が困難な特殊な格助詞的表現についてのみバリエーションを記述する。

記述する格の範囲：日本語では、必須格/任意格の境目は必ずしも明確でなく、記述する格の範囲は、応用によって定めるべきであると考えられる。現在、日英機械翻訳への適用を行なっているので、日英翻訳において必須的な要素、及び、デフォルトのレベルでは、日本語格助詞と英語の前置詞の対応が異なるような任意的な格については、パターンごとにそれを記述する。

態による結合価パターンの変化：受動、使役、可能などの態によって、結合価パターンは変化を受ける。これについては、変化の仕方をいくつかのパターンに分け、処理において、標準的な結合価パターンに還元する([4])ことによって対処する。

以上の基本方針に従って記述する結合価構造パターンの概念図を図1に示す。また、使用している階層的意味素性体系のトップレベルを図2に示す。

2.3 日本語解析の概要

日本語解析の前段において、分ち書き、文節関係受け解析が行なわれる。分ち書き処理において発生する多義は、

局所的なヒューリスティックによって極力解消が試みられる。また、係り受け解析によって生じる多義に対しては、評価値による順位付けが行なわれる〔3〕。文節間の係り受けとその係り受けの属性から、単位文、述部、格要素、名詞句、副詞要素などの範囲を決定する。次に、結合価パターンとのマッチングによる単位文解析により係り受け解析による曖昧さを解消し、単位文の構造を決定する。述部が複合用言から構成される場合は、その中の最後部に位置する単一用言（これを主用言という）の結合価パターンを用いて、派生要素（格要素及び副詞要素）を分離し、主用言の結合価パターンを用いて単位文解析を行なった後に派生要素をオリジナルな要素と同一のレベルに繰り上げる。複合文を構成する単位文において格要素の省略が行われている場合は、単位文間接続のある相手先の単位文の用言の結合価パターンを参照してその補完を試みる。また、名詞句における埋め込み構造についても埋め込み文の述部を構成する用言の結合価パターンを参照することによって、被修飾名詞の決定とその時の修飾関係の決定を行う〔5〕。このように、結合価パターンを日本語解析の色々な局面に適用することによって、以上のような言語現象が同時に出現するような複雑な文についてもその解析を統一的行なうことができる。

紹介する	
N 1 (主格)	SR : 「主体」 JR : 「が」, 「から」
N 2 (目的格 1)	SR : 「主体」 JR : 「を」
N 3 (目的格 2)	SR : 「主体」 JR : 「に」

* remark *
N i : 結合価要素のラベル
(深層格ではない)
SR : 名詞の意味的制約
JR : 助詞表現の制約

図1 結合価パターンの概念図(例:「紹介する」)

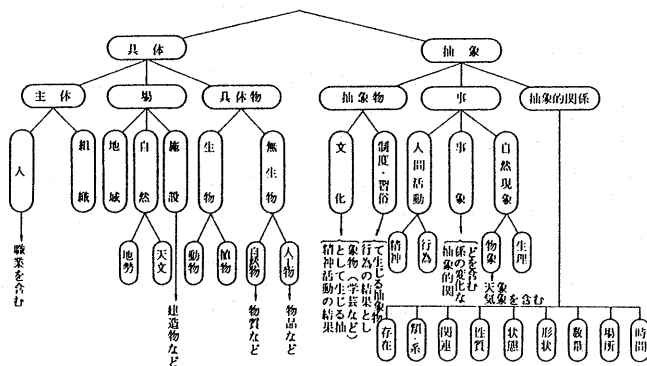


図2 階層的名詞意味素性体系〔12〕より転載

3. 結合価パターン日本語解析への適用

3.1 単位文解析におけるパターンマッチング

本節では、係り受け解析の結果である木構造と対応する結合価パターンとのパターンマッチング(例を図3に示す。)について説明する。このマッチングの基本は、木構造中の格要素をパターン中の格要素に対応付けることである。これは、格要素となる名詞句の中心名詞の意味素性と格要素の助詞表現がパターンに記述されたそれぞれの制約条件を満たすかどうかのチェックにより実現される。前者については、階層的な意味素性を使用しているため、入力中の名詞の意味素性は、パターン中の要素に記述されている意味素性より下位であれば良く、特に大きな問題はない。ここで問題なのは、後者の助詞表現のチェックである。なぜならば、結合価パターン中には、格助詞を中心とした標準的な助詞表現が原則として記述されている(どうしても格助詞に還元できないような場合を除く)ためである。

すなわち、実際の入力文においては、

(1) 格助詞以外の表現(副助詞、格助詞相当語などの連鎖)が頻繁に用いられる。

(2) 態により助詞表現が変化する。

という問題がある。(2)については、〔4〕に譲り、ここでは、(1)について述べる。(1)の問題は、格助詞、副助詞、格助詞相当語などの連鎖によって構成される助詞表現のなかから、表層格マーカとしてはたらいっているものを見いだす問題と、そのマーカが格助詞でない場合、それをどうやって格助詞へマッピングするかという問題に分けることができる。

表層格マーカの抽出: 助詞表現を構成する語連鎖を走査し、表層格マーカと副助詞的表現を分離する。この時、入力の助詞表現の文字列も別に保存しておく。これは、一部の結合価パターンにおいては、表層格マーカとして副助詞または、格助詞相当語が記述されていることがあるためである。語連鎖中に格助詞、格助詞相当語を含む場合は、連鎖中の最後に存在するそれを表層格マーカとして認定する。もし、

副助詞のみの連鎖で構成される場合は、とりあえず連鎖中で最も前にある副助詞を表層格マーカーとする。また、語連鎖に含まれる「は」以外の副助詞は、付加的な意味を持つものとして別にリスト化する。表層格の抽出と副助詞表現の分離の幾つかの例を図4に示す。

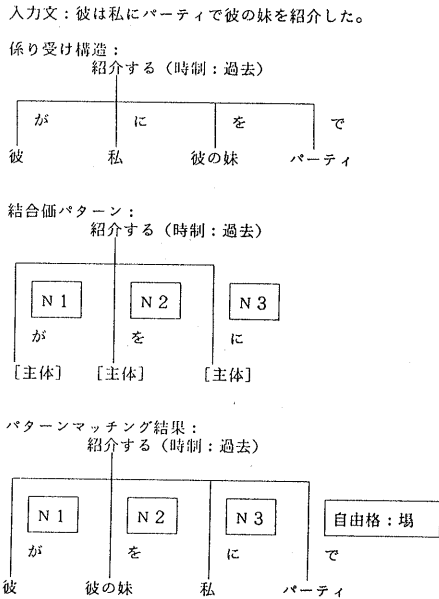


図3 係り受けの木構造とパターンマッチング

- 例1：彼は →
（表層格マーカー：「は」 副助詞リスト：nil）
- 例2：彼も →
（表層格マーカー：「も」 副助詞リスト：（「も」））
- 例3：彼にさえも →
（表層格マーカー：「に」 副助詞リスト：（「さえ」「も」））
- 例4：彼についても →
（表層格マーカー：「について」 副助詞リスト：（「も」））
- 例5：彼こそは →
（表層格マーカー：「は」 副助詞リスト：（「こそ」））

図4 表層格マーカーの抽出と副助詞表現の分離

格助詞への変換：上記に示した手法によって抽出された表層格マーカーが格助詞以外の場合、パターンマッチングを行うために格助詞への変換を行う必要がある。[6]においては、潜在格と格の強度という考え方でこれに対処している。本方式においては、1つ1つの副助詞、格助詞相当語に対して変換可能な格助詞を順位付で記述している。この順位は、言語データの分析により付与された静的な情報であり、テーブル化されている。この変換情報の例を表1に示す。格助詞への変換関係は、1：1の関係になるものと1：nの関係になるものがある。1：1の関係のものについては、パターンマッチングを行う前に対応する格助詞への変換をあらかじめ行なう。

表1 格助詞への変換例

表層格マーカー	変換格助詞
は	が を に
さえ	が を
が理由で	で
について	が を
によって	から により

パターンマッチング：パターンマッチング処理においては、入力中の格要素のうち、助詞表現がもともと格助詞であったもの、1：1の関係で格助詞に変換されたものを先に処理する。1：nの関係のものについては、まず別に待避してあったオリジナルの助詞表現を用いて結合価パターン中の要素へのマッチングを試みる。この時、対象となるパターン中の要素は、まだ入力文中の要素と結び付けられていない未バインドのものである。これによってマッチングが成功しなかった場合は、前節で説明した変換テーブルを基に上位の候補から順に格助詞へ置き換えてマッチングを試みる。すなわち、静的な情報である変換テーブルの内容を基に、現在のマッチングの状況を考慮しながら動的に処理を行う。

3.2 複合文における格要素の補完

次の例に示すように、日本語文の複合文においては、従文、主文に共通する格要素が省略されるというゼロ代名詞化の現象が多く現れる。

（例）私は彼と議論して、 設計方針を決定した。

↑ ↑
設計方針を 私は

次のような理由により、このような格要素はなるべく補完しておく必要がある。まず、複数の結合価構造が存在する場合、それらの中から最適なものを選ぶことが必要となるが、その際、なるべく多くの支配格要素が存在する方が最適選択における確度は高まる。また、機械翻訳への適用を考える場合、この補完を行っておかないと目的言語文生成において適当な代名詞化を行うことができない。

ゼロ代名詞化に対する詳しい考察が、[7]において行われている。ここでは、複合文を構成する文間の接続情報と主題の有無によって、先行詞の同定を行なっている。この補完処理は本質的には意味・文脈理解を必要とする問題であり、安易な対処では大きな副作用を招く。現在、[7]に述べられているような方法を検討中であるが、今のところ、格要素の補完を大略次のようなかなり限定された扱いによって処理している。

（1）複合文＝単位文（＝従文）＋単位文（＝主文）の形となっている部分構造を対象とする。

(2) 従文、主文の結合価構造を参照して、文中の格要素と結び付けられていない要素で、「ガ」格、「ヲ」格、「ニ」格にあたるもの(必須格的要素)を見つける。

(3) 上記の格要素を対象にまず、従文の格要素を主文に対して補完することを試みる。次に、主文の格要素を従文に対して補完することを試みる。

以上のような手順によって処理を行なっているため、補完を行なう範囲は限定されているし、必須格的な格要素以外の補完を行なうこともできない。また、結合価パターンを参照して、補完が可能かどうかを検査する際には、極力副作用を避けるため、同一の結合価ラベルを持つという制限を設けている。すなわち、従文と主文の間で主語や目的語などが一致しているとの前提にたっている。そのため、この前提が満たされないような文については、正しい補完を行なうことは難しい。

3.3 複合用言の解析

まず、本稿でいう複合用言の定義を与えておく。本稿では、「複合名詞+する」で全体としてサ変動詞としてはたらくもの(例:「研究開発する」、「新規開発する」、「原因分析する」)、及び、「複合名詞+だ」で全体として形容動詞としてはたらくもの(例:「栄養豊富だ」、「超高速だ」)を複合用言として扱う。ここでは、動詞と動詞によるもの(例:「持ち運ぶ」など)や、動詞と補助動詞によるもの(例:「走りだす」など)などは、複合用言としては扱わない。前者は、一語の動詞または、単位文接続として、後者は様相表現として扱う。

日本語においては、容易に複合語化が起こるため、たとえ分野をかなり限定してもこれらの複合用言の可能な出現形をあらかじめ全て辞書に登録しておくことは難しいと考えられる。また、文中の格要素との関係で意味が異なる場合も存在する(例:「電圧制御する」→「電圧を制御する」/「電圧で制御する」)。以上の2つの理由から、複合用言を辞書に登録しておいて静的に扱うには限界があり、日本語解析の処理の中でこれを解析する枠組みを持つておく必要がある。

複合用言の解析によって、複合用言は、主用言とその支配要素として派生される修飾要素(副詞要素および格要素)とに分けられる。その結果、オリジナルの単位文は、主用言を述語とし、オリジナルな修飾要素に加えて派生された修飾要素をその支配要素として持つ単位文と内部的に等価な形になる。

複合用言からの格要素の派生: 複合用言の解析は、複合名詞の解析と同じ考えで行なわれる。すなわち、複合名詞部分のなかから用言としてはたらいっている用言性名詞を抽出し、その他の語基をそれとの関係を用いて構造化する([8])。この際、その用言性名詞の用言形を持つ結合価パターンが参照され、格要素となるべきものが派生される。

ここでは、助詞表現の制約条件は関係ないので、格要素として派生しようとする語基名詞の意味素性が結合価パターン中のある要素の名詞の意味制約条件を満たすかがチェックされ、満たす場合、対応する結合価ラベルがその語基名詞に与えられる。一般の単位文レベルでのパターンマッチングに比べると助詞の条件が使えないため、一意に結合価ラベルを決定することはさらに難しいこと、また、上位の単位文レベルへの繰り上げに当たっては、単位文のオリジナルな格要素との整合性をとる必要があることから、意味素性のマッチングにおける度合いによって幾つかの候補を順位付けして出力する。複合用言解析による格要素の派生の概念図を図5に示す。

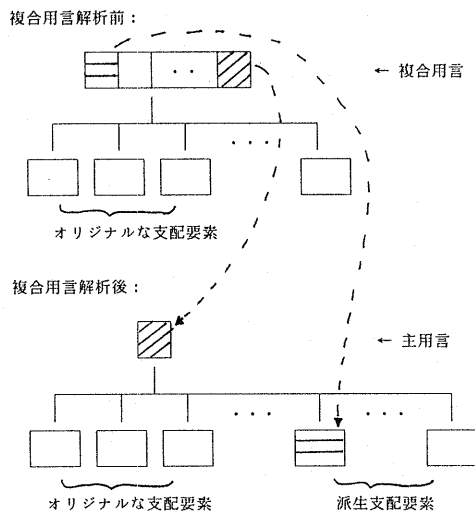


図5 複合用言解析による格要素派生の概念図

派生格要素の単位文への繰り上げ: 派生された格要素をもととの複合用言の主用言を述語とする単位文の格要素として繰り上げる。この際、オリジナルな格要素との整合性をとる必要がある。すなわち、1文1格の原則を満たすために、格が衝突しないように(厳密には、結合価ラベルが衝突しないように)繰り上げる必要がある。複合名詞解析によって出力される派生格の優先順位に従ってこの繰り上げを行う。

4. 日英翻訳への適用

本稿に述べた日本語文解析を現在日英翻訳へ適用している([9])。そこでは、日本語結合価パターンとそれに対応する英語の統語構造を対として与え構造変換を行なう。また、「格要素+用言」の形によって慣用的表現をなすものや機能動詞結合による表現についてもほぼ同様の枠組みで処理を行なっている([10])。すなわち、[9]で扱われる構造変換は、本稿に示した枠組みのより特殊化・個別現象化

されたものと考えることができる。本章の以下では、日英翻訳へ結合価構造の適用における問題点を考察する。

日英翻訳においては、結合価構造パターンとそれに対応する英語表現を対にして与えて英語表現の変換・生成を行っている。この方式によれば、構造変換規則体系は、個別的な変換規則の集まりとなるため木目の細かい翻訳が可能となるという長所を持つ。半面、変換規則辞書が大規模になる他、処理の上では用言に対して複数の結合価パターンが存在する場合、どのようにして最適なパターンを選ぶかが重大な問題となる。これは、どのパターンを選択するかによって、英語の基本的な構文構造と動詞表現が決定されるためである。

以上のような訳し分けという観点からすると、結合価パターンの記述と処理における最適パターンの選択がこの問題の鍵を握ることになる。この問題については、意味素性による選択制限が大きな役割を果たす。現在、図2にそのトップレベルを示した意味素性体系を基に、細かいレベルでの選択制限の付与や、場合によっては非選択制限をも記述するようにして、極力同一パターンに対する複数パターンが排他的になるようにしている。しかしながら、どこまで細かいレベルでこの制約を記述すれば訳し分けが可能か、あるいは、意味素性ベースでの訳し分けの限界がどの辺にあるのかについては、必ずしも明確でなく、今後、翻訳実験等を通して評価を重ねていく必要がある。

5. おわりに

本稿では、用言の結合価構造に基づく日本語解析、及び、日英翻訳への適用について報告した。機械翻訳などのように深い意味理解よりも大規模な語彙を持つことなどが要求される自然言語処理システムにおける日本語解析を実現するには、用言の持つ結合価構造に基本を置くことが有用であると考えられる。また、文が名詞句の形に変形されたと考えられる埋め込み構造の解析や、文が1つの名詞の形に集約されたと考えられる用言性名詞を含む複合名詞の解析においても用言の結合価構造はその基礎を与えるものと考えられる。

今後の課題としては、単位文レベルの解析においては、現在、テーブルに変換情報を記述しておくといった比較的静的な扱いをしている格助詞相当語の扱いをより精密にしていく必要があると考える。すなわち、どのような表現をどのような環境条件において格助詞相当語として扱って良いか、すなわち格助詞相当語の動的な側面をさらに検討しなければならない。結合価文法などの依存文法的な考え方においては、語順の自由さを扱いやすいところが日本語向きであると考えられているが、複合用言などにおいては、語基の順序も重要な情報を担っていると考えられる。今後は、複合語の語構成の研究成果を複合語解析に積極的に取り入れていく必要がある。

< 参考文献 >

- [1] 石綿, 荻野: 結合価から見た日本文法, in 朝倉日本語新講座3「文法と意味I」第2章, 朝倉書店, (1983).
- [2] 宮崎正弘: 係り受け解析を用いた複合語の自動分割法, 情報処理学会論文誌, Vol.25, No.6, (1984).
- [3] 宮崎, 白井, 林: 日英翻訳システムAL T-J/Eにおける日本語解析技術, 第33回情処全大講論集, 6J-2, (1986).
- [4] 河合敦夫: 日英翻訳システムAL T-J/Eにおける様相・時制処理, 第34回情処全大講論集, 5W-2, (1987).
- [5] 横尾, 林: 日本語埋め込み構造の解析, 第1回人工知能学会全国大会, 7-2, (1987).
- [6] 平井, 北橋: 格の強度と述語の構文および意味属性を用いた格構造の変換生成について, 情報処理学会論文誌, Vol.28, No.3, (1987).
- [7] 吉本啓: 談話処理における日本語ゼロ代名詞の扱いについて, 自然言語処理研究会, 56-4, (1986).
- [8] 石崎雅人: 日本語複合語の解析, 第35回情処全大講論集, 1T-1 (to appear), (1987).
- [9] 林, 奥, 石崎: 日英翻訳システムAL T-J/Eにおける日英変換技術, 第33回情処全大講論集, 6J-3, (1986).
- [10] 奥雅博: 日本語慣用表現の分析と日英翻訳への適用, 情報処理学会自然言語処理研究会資料, 62-2, (1987).