

範疇単一化文法に規則を併用した日本語の構文解析

関洋平 飯島正 原田賢一

慶應義塾大学大学院理工学研究科

e-mail: yohei@hara.cs.keio.ac.jp

1 はじめに

本研究の目的は構文・意味レベルにおいて日本語を理解する計算機上のシステムを構築することにある。このような方向で近年取り組まれている文法の枠組の一つとして、**単一化文法**と呼ばれる一連の文法形式がある。単一化文法は、属性と属性値の組合せをリスト化したデータ構造である素性構造を用いることにより、単語の係り受け関係ならびに意味情報を単語に対応させて表現を行っている。すなわち、文脈自由文法における書換え規則のようなルールを、できるだけ辞書に埋め込む形で取扱うことを目標としている文法枠組である。

本研究では、係り受け関係の取扱いに**範疇文法**の考え方を取り入れた単一化文法の枠組の一つであり、日本語に対する適用例のあまりない**範疇単一化文法** [6](CUG)を日本語に対して適用することで、その有効性を検証する。範疇単一化文法は、範疇文法を基本概念としていることから、より辞書情報に重点をおいた構文解析を行うことが可能になっている。同じ方向性をもった提案としては、主辞駆動句構造文法 [4](HPSG) が有名である。HPSG と CUG の主題化などの文法現象に対する取扱いの違いについては、[3] を参照されたい。

2 範疇文法

範疇文法では Joachim Lambek の提案による Lambek Calculus の

$$/ \text{ 除去規則; } \frac{\Delta \Rightarrow B/A \quad \Gamma \Rightarrow A}{\Delta, \Gamma \Rightarrow B}$$

$$\backslash \text{ 除去規則; } \frac{\Gamma \Rightarrow A \quad \Delta \Rightarrow A \backslash B}{\Gamma, \Delta \Rightarrow B}$$

を基礎としている。具体的には、2~3の基本範疇を /, \ を用いて組み合わせることにより、品詞の表現を行い、上記の除去規則を適用することにより、構文解析が

行われる。

3 範疇単一化文法

範疇文法と単一化文法を融合させる最初の試みとして**範疇単一化文法** (CUG) は提案された。具体的な融合の方法は、トップレベルの素性として value(値), direction(方向), argument(引数) を与えることで, /, \ を素性構造に埋め込むことにより実現している。これは、範疇文法を単一化文法として定式化したものであるといえる ([6] では PATR 形式の dag 表現を用いて表されている)。以下に、具体例として、名詞「太郎」と後置詞「が」の素性構造を示す。

太郎 (n)

値 :	範疇 :	n(名詞)
	分類 :	人間
	意味 :	太郎

が^(n\ppnom)

値 :	方向 :	'\'					
	引数 :	値 : <table border="1"> <tr> <td>範疇 :</td> <td>n(名詞)</td> </tr> <tr> <td>分類 :</td> <td>人間</td> </tr> <tr> <td>意味 :</td> <td>X</td> </tr> </table>	範疇 :	n(名詞)	分類 :	人間	意味 :
範疇 :	n(名詞)						
分類 :	人間						
意味 :	X						
値 :	範疇 :	pp(後置詞句)					
	格 :	が					
	役割 :	主格					
	意味 :	X					

このような考え方に基づく融合の例としては、他に、単一化範疇文法 (UCG), 単一化に基づくコンビネータ範疇文法 (Unification-based CCG) がある。これとは逆に、範疇文法の考えを基礎として、素性構造を基本範疇に対応させた例としては、Type Logical Grammar や Typed Applicative Grammar[2] がある。

4 日本語の範疇

日本語の基本範疇としては、名詞 (N), 文 (S), 後置詞句 (PP) の三つを設けることにする。その三つを /, \ を用いて組み合わせることで、以下のような品詞を表現することができる。

A Japanese Syntax Analysis Method based on Categorical Unification Grammar with additional rules.

Yohei SEKI, Tadashi IJIMA and Ken'ichi HARADA
Faculty of Science and Technology, Keio University
3-14-1 Hiyoshi, Kouhoku-ku, Yokohama 223, JAPAN

	品詞分類	範疇
文	文	S
動詞	動詞 (格三つ, あげる)	PP\ (PP\ (PP\ S))
	動詞 (格二つ, 会う)	PP\ (PP\ S)
助詞	動詞 (格一つ, 歩く)	PP\ S
助詞	必須格助詞 (が, を, に)	N\ PP
	係助詞 (は)	N\ (S/ (PP\ S)), N\ (S/ S)
	任意格助詞 (時間格に, 場所格で)	N\ ((PP\ (PP\ (PP\ S))) / (PP\ (PP\ (PP\ S))))
	終助詞 (ね, か)	S\ S
助詞	接続助詞 (が)	S\ (S/ S)
助詞	時制助動詞 (た, ている)	(PP\ (PP\ (PP\ S))) \ (PP\ (PP\ (PP\ S)))
	断定助動詞 (です, だ, である)	N\ (PP\ S)
名詞	名詞	N
	連体詞 (この, その, あの)	N/ N
	接尾辞 (時, 君)	N\ N
名詞	時間を表す名詞 (昨日, 3時, 夕方)	(PP\ (PP\ (PP\ S))) / (PP\ (PP\ (PP\ S)))
	接続詞	S/ S
その他	副詞	(PP\ (PP\ (PP\ S))) / (PP\ (PP\ (PP\ S)))
	感動詞	S

また、この他に動詞・形容詞・形容動詞については語幹と活用語尾を分離し、語幹には基本範疇としてVを与え、活用の種類を接続情報の素性として持たせるようにしている。

具体的には、形容動詞語幹「嫌い」に範疇 $PP\ V_{\text{形容動詞}}$ を、活用語尾「な」に範疇 $(PP\ V_{\text{形容動詞}})\ (N/ N)$ を、断定助動詞「だ」に範疇 $V\ (PP\ S)$ を与えることにより、「太郎はあの子が嫌いだ」「太郎が嫌いな子がいる」というように、形容動詞の語幹「嫌い」に「な」や「だ」が接続することで、どの位置にどの格役割を持つ品詞が来るかという区別ができるようになる。また、同様にサ変動詞語幹「愛」に活用語尾「している」が接続すると、「愛する主体」と「愛される対象」が格後置詞句として堀起こされる。「している」が「勉強」に接続した場合は、格後置詞句は「勉強する主体」の一つの役割しか堀起こさない。この区別は、サ変動詞語幹に

格数素性を持たせて、活用語尾「している」との間で共起関係を調べることにより行っている。

5 日本語の諸現象に対する取扱い

日本語特有の文法現象に対する取扱いは、範疇の設定とルールを導入により行う。ルールを導入することで、辞書の繁雑さを避けることができる。

● 範疇による取扱いの具体例

1. 助動詞「せ」「させ」による使役化

助動詞「せ」「させ」には、動詞に接続して文の述部の意味を「その行為の強制」に変化させ、構文的には動詞がとる後置詞句に、「強制者」の役割をもつ格後置詞句の一つを追加する。例えば、「花子が料理を食べる」という文の動詞語幹「食べ」に接続して「太郎が花子に「花子が料理を食べる」という行為を強制する」というように意味を変化させる。これを範疇によって表すと、以下のようになる。

● 「させ」の範疇

$$(PP_{sem_x} \setminus (PP_{nom, が^s, sem_y} \setminus S_{sem_z})) \\ \setminus (PP_{sem_x} \setminus (PP_{acc, に, sem_y} \setminus (PP_{nom, 花子^s, sem_w} \\ \setminus S_{force_{agt:w, pa:a:y, obj:z} }))))$$

2. 助動詞「れ」「られ」による受動化

助動詞「れ」「られ」は、「太郎が花子に料理を食べさせ」というような文章に接続して、行為主体者と行為受動者の意味役割を入れ替えるという機能をもつ。範疇によって表すと、以下のようになる。

● 「られ」の範疇

$$(PP_{acc, sem_x} \setminus (PP_{nom, が^s, sem_y} \setminus X)) \\ \setminus (PP_{acc, sem_y} \setminus (PP_{nom, 花子^s, sem_x} \setminus X))$$

3. 助詞「は」による主題化

以下の二つの例文による「は」の役割の違いを扱うために、以下の範疇を「は」に対して与える。

● 例文

太郎はあの子を愛している。

太郎はあの子が愛している。

● 「は」の範疇

$$N \setminus (S/ (PP \setminus S))$$

※ N と PP は 情報を共有

これにより上の例文の「は」が動詞句「あの子を愛している」の欠落格である主格であり、下の例文

の「は」は動詞句「あの子が愛している」の欠落格である目的格であることが扱えるようになる。

4. 係り受けの非交差制約

本研究で採用している範疇文法は、引数を前後どちらに取るかをはっきり区別しているため、係り受けの非交差制約を自然な形で表現できる。以下、「美しい日本の自然」と「日本の美しい自然」という二つの文に対して、本研究で設定した範疇を用いた解析結果を示す。

(a) 美しい日本の自然

「美しい日本の自然」という文の意味は2通り考えられる。一つは、「美しい」が「日本」を修飾する場合、もう一つは「日本の自然」を修飾する場合である。本研究で設計した範疇を用いて構文解析を行った例を以下に示す。

解析結果 (a.1)

「美しい」が「日本」を修飾する場合

美しい	日本の	自然	
$V_{\text{形容詞}}$	$N \setminus (N/N)$	N	
引数	関数		\ 除去
$PP \setminus S$			
(ルール適用)			
N/N	N		
関数	引数		/ 除去
N			
引数	関数		\ 除去
N/N			
関数	引数		/ 除去
N			

解析結果 (a.2)

「美しい」が「日本の自然」を修飾する場合

美しい	日本の	自然	
$V_{\text{形容詞}}$	$N \setminus (N/N)$	N	
引数	関数		\ 除去
$PP \setminus S$			
	引数	関数	\ 除去
	N/N		
	関数	引数	/ 除去
	N		
(ルール適用)			
N/N	N		
関数	引数		/ 除去
N			

以上に示したように、「美しい」が「日本」

あるいは「日本の自然」を修飾することを自然な形で解析できる。本研究で試作しているシステムは、複数の解析結果がある場合、バックトラックして全ての解析結果を検出する。

(b) 日本の美しい自然

「日本の美しい自然」における「美しい」は、「日本」を修飾することができない。これは、日本語が前から後ろへ向けて修飾する言語であることによる。本研究で提案した範疇で「日本の美しい自然」を解析すると、以下のようになり、この説明にあった解析を行っていることが理解できる。

日本の	美しい	自然	
N	$N \setminus (N/N)$	$V_{\text{形容詞}}$	$V_{\text{形容詞}} \setminus (PP \setminus S)$
引数	関数		
N/N			
	引数	関数	\ 除去
	$PP \setminus S$		
	(ルール適用)		
	N/N	N	
	関数	引数	/ 除去
	N		
関数	引数		/ 除去
N			

● ルールによる取扱いの具体例

1. 動詞と形容詞による連体修飾

日本語の動詞と形容詞は連体形と終止形が同じ活用をもつという特徴がある。これにより格後置詞句が欠落した文が、後方に名詞を補って連体修飾を行う、ということが可能になっている。この現象を取扱うために、以下のルールを導入する。

(a) $PP \setminus S \ N \rightarrow N/N \ N$

※ $PP \setminus S$ の PP と $N/ \ N$ の N は 情報を共有

これにより「私は望遠鏡を持つその少年をみる。」のような文章が与えられた際、望遠鏡を持っているのは「私」ではなくて「その少年」であることが取扱えるようになる。

2. 語順の自由度

語順の自由度を取扱う方法としては、引数の集合を取ることが出来るように枠組を拡張するという

方法がある ([1][5] など)。この考え方は、HPSG においても応用されている。しかし、これは範疇文法の根本的な利点を損なう感触も受ける。ここではルールにより取扱う方法について検証を行う。以下の7つのルールを導入することにより、格後置詞句の入れ替えが可能になる。

● 必須格後置詞句 (が, に, を 格) の入替えルール

- (a) $PP_{が} PP_{を} \setminus (PP_{が} \setminus S) - PP_{が} PP_{が} \setminus (PP_{を} \setminus S)$
- (b) $PP_{が} PP_{に} \setminus (PP_{が} \setminus S) - PP_{が} PP_{が} \setminus (PP_{に} \setminus S)$
- (c) $PP_{に} PP_{を} \setminus (PP_{に} \setminus S) - PP_{に} PP_{に} \setminus (PP_{を} \setminus S)$
- (d) $PP_{が} PP_{を} \setminus (PP_{に} \setminus (PP_{が} \setminus S))$
 $- PP_{が} PP_{が} \setminus (PP_{を} \setminus (PP_{に} \setminus S))$

● 任意格後置詞句 (時間格に, 場所格で) の入替えルール

- (a) $(PP \setminus (PP \setminus (PP \setminus S))) / (PP \setminus (PP \setminus (PP \setminus S))) PP \setminus (PP \setminus S)$
 $- (PP \setminus (PP \setminus S)) / (PP \setminus (PP \setminus S)) PP \setminus (PP \setminus S)$
- (b) $(PP \setminus (PP \setminus (PP \setminus S))) / (PP \setminus (PP \setminus (PP \setminus S))) PP \setminus S$
 $- (PP \setminus S) / (PP \setminus S) PP \setminus S$
- (c) $(PP \setminus (PP \setminus (PP \setminus S))) / (PP \setminus (PP \setminus (PP \setminus S))) S$
 $- S / S S$

6 おわりに

範疇単一化文法にルールを併用して用いることで、以下のような日本語特有の現象に対応できる。

	日本語の文法現象	アルゴリズム	実現
構文的特徴	係り受けの曖昧性	◎	◎
	語順の自由度	◎	△
	入れ子構造	◎	△
	並置構造	◎	△
	同じ品詞の連続	△	×
意味的特徴	遊離数量詞	△	×
	副詞の呼応	◎	△
	「自分」の照応先特定	△	×

対応済み…◎, 一部対応…△, 未対応…×

前章までで言及していない項目がいくつかある。並置構造は「～が～が」「～に～に」のように共通の助詞を持った後置詞句が連続する文の取扱いについて検討している。同じ品詞の連続は、複合助詞や助動詞の連続などについてであり、これは後ろの範疇を高階化するルールを導入したり、接続制約素性を導入することで取扱いの柔軟性を増している。遊離数量詞は「3匹の子豚」「子豚3匹」「子豚が3匹」のような句における3匹が子豚を計量している現象についてであるが、これについては範疇を動的に変化させるルールを導入して取り扱っている。副詞の呼応現象は、「少しも～ない」長距離の依存関係を持つものであるが、この現象については、「～」の動詞の部分に「ない」が否定のアスペクト素性を追加するような範疇および素性構造を与え、「少しも」が後方に否定のアスペクト素性をもつ範疇を必要とするかたちに対応している。再帰代名詞の「自分」の照応については、範疇よりも素性の制約をベースにして補完を行っている。

これより、範疇単一化文法の日本語への有効性を示すことができた。

参考文献

- [1] Beryl Hoffman: "A CCG APPROACH TO FREE WORD ORDER LANGUAGES", Proc. of 31th Annual Meeting of ACL(1993), pp.300-302.
- [2] 飯島正, 関洋平, 柳原正秀他: "範疇文法のための関数的な計算機構", 情報処理学会研究報告 96-NL-111-6(1996), pp.33-40.
- [3] Carl J. Pollard: "Categorial Grammar and Phrase Structure Grammar: An Excursion on the Syntax-Semantics Frontier", in Richard T. Oehrle, Emmon Bach and Deirdre Wheeler: "Categorial Grammars and Natural Language Structures", Dordrecht Reidel Pub(1988), pp.391-415.
- [4] Carl Pollard and Ivan A. Sag: "Head-Driven Phrase Structure Grammar", The University of Chicago Press, CSLI(1994).
- [5] Satoshi Tojo: "Free-ordered CUG on Chemical Abstract Machine", Proc. of COLING 1994(1994), pp.870-874.
- [6] Hans Uszkoreit: "Categorial Unification Grammars", Proc. of COLING 1986(1986), pp.187-194.