

# 単一化文法による日本語文の解析 - 形容詞の取り扱いについて -

川尻渉太\* 乙武北斗\*\* 吉村賢治\*\*  
(福岡大学 大学院工学研究科\* 工学部\*\*)

## 1 はじめに

従来、日本語の解析は形態素解析、構文解析の順に行っている。しかし、形態素解析を行う際に使われる品詞の接続条件を表す文法規則は、構文解析でも使用する規則を簡略化したものであり、逐次的に解析を行った場合、同様の規則を重複して適用することになる。また、形態素解析の対象に接辞等による語形成が含まれているが、語形成の中には統語部門で行われるものもあり、構文解析の中で扱う必要があることがわかっている。そこで本研究では、単一化文法を用いて形態素解析と構文解析を同一の枠組みの中で行う解析手法について提案する。本稿では特に、形容詞の扱いについて述べる。

## 2 2 提案手法

### 2.1 文法モデル

本研究では従来の日本語解析で一般に用いられている学校文法は用いず、動詞などの活用を音韻変化で説明する文法モデルを用いる[1]。例えば、学校文法では「書く」という動詞は「書いた」や「書かれる」というように活用するが、本文法モデルにおいては、動詞の語幹の「書 k」に接辞の「ita」や「rare」「ru」が付き、音韻変化により「書 ita」「書 kareru」となる。学校文法における動詞の5段活用とそれに対応した本文法モデルでの取り扱いを表1に示す。

表 1 学校文法における5段活用の取り扱い

学校文法		本文法モデルにおける取り扱い
未然形	書かない 書こう	書 k + ana + i 書 k + you
連用形	書きます	書 k + imas + ru
終止形	書く	書 k + ru
連体形	書くとき	書 k + ru + とき
仮定形	書けば	書 k + reba
命令形	書け	書 k + e

### 2.2 解析システムの構成

本文法モデルに適した形態素解析システムに KIMMO[2]がある。KIMMO は音韻変化の規則を表現した FST(Finite State Transducer)と文法規則を表現した FSA(Finite State Automaton)で構成されている。本研究では、形態素解析と構文解析を同一の枠組みで行うことを目的とし、FSA の代わりに構文解析で用いる単一化文法[3]を用いる。本システムの構成を図1に示す。

また、KIMMO における音韻変化の規則は、文字列レベルのみで記述された規則であるが、日本語の音韻変化はほとんどのものが特定の品詞の組み合わせで生じることから、品詞等の情報を規則に取り入れる必要があることがわかっている[4]。

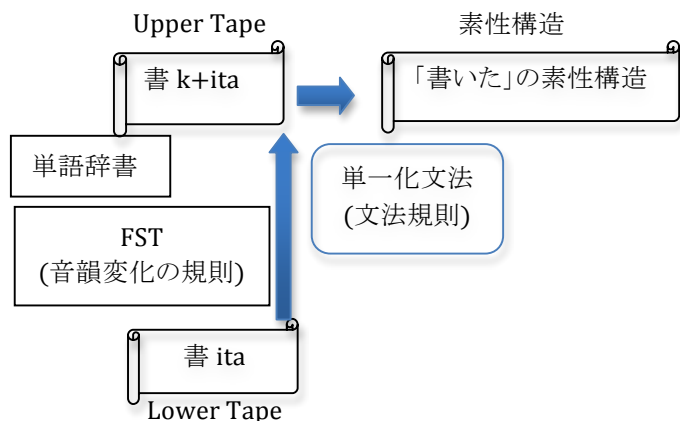


図 1 本解析システムの構成

### 2.3 形容詞の取り扱い

形容詞は最近の言語学の研究[5]に従って、形容詞の語根「青」に述語化繫辞「k」とダミー繫辞「ar」が接続した構造を持っているとする。形容詞「青い」における学校文法での活用とそれに対応した本文法モデルでの取り扱いを表2に示す。

表 2 形容詞の取り扱い

学校文法		本文法モデルにおける取り扱い
未然形	青かろう	青 + k + ar + you
連用形	青かった 青く	青 + k + ar + ita 青 + k + u
終止形	青い	青 + i
連体形	青いとき	青 + i + とき
仮定形	青ければ	青 + k + ar + reba

単一化文法における形容詞文「空が青かった」の解析例を次に示す。例文には 6 つの形態素が含まれており、それぞれの素性構造は次のようになる。

$$(1) \text{空: } \left[ \begin{array}{l} \text{head: [pos: n]} \\ \text{sem: } \left[ \begin{array}{l} \text{denote: X} \\ \text{constraint: sky(X)} \end{array} \right] \end{array} \right]$$

$$(2) \text{が: } \left[ \begin{array}{l} \text{head: } \left[ \begin{array}{l} \text{pos: p} \\ \text{gr: ガ格} \end{array} \right] \\ \text{sem: X} \\ \text{subcat: } \left\{ \left[ \begin{array}{l} \text{head: [pos: n]} \\ \text{sem: X} \end{array} \right] \right\} \end{array} \right]$$

$$(3) \text{青} : \left[ \begin{array}{l} \text{head} : [\text{pos} : \text{adj\_root}] \\ \text{sem} : \left[ \begin{array}{l} \text{denote} : \left[ \begin{array}{l} \text{property} : \text{青い} \\ \text{arg} : X \\ \text{darg} : \text{color} \end{array} \right] \\ \text{constraint} : \text{XC} \end{array} \right] \\ \text{subcat} : \left\{ \left[ \begin{array}{l} \text{head} : \left[ \begin{array}{l} \text{pos} : \text{p} \\ \text{gr} : \text{ガ格} \end{array} \right] \\ \text{sem} : \left[ \begin{array}{l} \text{denote} : X \\ \text{constraint} : \text{XC} \end{array} \right] \end{array} \right\} \end{array} \right]$$

$$(4) \text{k} : \left[ \begin{array}{l} \text{head} : [\text{pos} : \text{adj}] \\ \text{sem} : X \\ \text{subcat} : Y \\ \text{adjacent} : \left\{ \left[ \begin{array}{l} \text{head} : [\text{pos} : \text{adj\_root}] \\ \text{sem} : X \\ \text{subcat} : Y \end{array} \right] \right\} \end{array} \right]$$

$$(5) \text{ar} : \left[ \begin{array}{l} \text{head} : [\text{pos} : \text{v}] \\ \text{sem} : X \\ \text{subcat} : Y \\ \text{adjacent} : \left\{ \left[ \begin{array}{l} \text{head} : [\text{pos} : \text{adj}] \\ \text{sem} : X \\ \text{subcat} : Y \end{array} \right] \right\} \end{array} \right]$$

$$(6) \text{ita} : \left[ \begin{array}{l} \text{head} : [\text{pos} : \text{s}] \\ \text{sem} : \left[ \begin{array}{l} \text{tens} : \text{past} \\ \text{process} : X \end{array} \right] \\ \text{subcat} : \left\{ \left[ \begin{array}{l} \text{head} : [\text{pos} : \text{v}] \\ \text{sem} : X \\ \text{subcat} : \{\} \end{array} \right] \right\} \end{array} \right]$$

次に各形態素の単一化を行う。(1) (2)は下位範疇化素性の原理に従って単一化可能であり、同様に(3)(4)も下位範疇化素性の原理に従って単一化可能である。また、(3)(4)の単一化によってできる「青 k」は同様の原理に従って(5)と単一化可能である。単一化後、「空が」と「青 kar」の素性構造は次のようになる。

$$(7) \text{空が} : \left[ \begin{array}{l} \text{head} : \left[ \begin{array}{l} \text{pos} : \text{p} \\ \text{gr} : \text{ガ格} \end{array} \right] \\ \text{sem} : \left[ \begin{array}{l} \text{denote} : X \\ \text{constraint} : \text{sky}(X) \end{array} \right] \\ \text{subcat} : \{\} \end{array} \right]$$

$$(8) \text{青 kar} : \left[ \begin{array}{l} \text{head} : [\text{pos} : \text{v}] \\ \text{sem} : \left[ \begin{array}{l} \text{denote} : \left[ \begin{array}{l} \text{property} : \text{青い} \\ \text{arg} : X \\ \text{darg} : \text{color} \end{array} \right] \\ \text{constraint} : \text{XC} \end{array} \right] \\ \text{subcat} : \left\{ \left[ \begin{array}{l} \text{head} : \left[ \begin{array}{l} \text{pos} : \text{p} \\ \text{gr} : \text{ガ格} \end{array} \right] \\ \text{sem} : \left[ \begin{array}{l} \text{denote} : X \\ \text{constraint} : \text{XC} \end{array} \right] \end{array} \right\} \end{array} \right]$$

上記(7)(8)は単一化可能であり、単一化後の素性構造は

次のようになる。

$$(9) \text{空が青 kar} : \left[ \begin{array}{l} \text{head} : [\text{pos} : \text{v}] \\ \text{sem} : \left[ \begin{array}{l} \text{denote} : \left[ \begin{array}{l} \text{property} : \text{青い} \\ \text{arg} : X \\ \text{darg} : \text{color} \end{array} \right] \\ \text{constraint} : \text{sky}(X) \end{array} \right] \\ \text{subcat} : \{\} \end{array} \right]$$

最後に(6)(9)を単一化することで例文の素性構造が次の様に得られる。

$$\left[ \begin{array}{l} \text{head} : [\text{pos} : \text{s}] \\ \text{tens} : \text{past} \\ \text{sem} : \left[ \begin{array}{l} \text{process} : \left[ \begin{array}{l} \text{denote} : \left[ \begin{array}{l} \text{property} : \text{青い} \\ \text{arg} : X \\ \text{darg} : \text{color} \end{array} \right] \\ \text{constraint} : \text{sky}(X) \end{array} \right] \end{array} \right] \\ \text{subcat} : \{\} \end{array} \right]$$

### 3 おわりに

本稿では FST と単一化文法を用いて形態素解析と構文解析を同一の枠組みの中で行う日本語文の解析手法を提案した。今後は、品詞情報を含めた音韻規則の作成や各形態素の素性構造の決定を行い、システムの完成を目指すとともにその有効性を検証していきたい。

### 参考文献

- [1] 三浦, 吉村, 首藤. 日本語形態素解析のための派生文法. 福岡大学工学集報. (1995).
- [2] K.Koskenniemi. A general computational model for word-form recognition and production. Proc. of 10th Coling. pages 178-181.(1984).
- [3] 吉村賢治. 自然言語処理の基礎. サイエンス社.(2012).
- [4] 吉村, 三浦, 首藤. 2レベルモデルに基づく日本語の形態素処理. 言語処理学会 第3回年次大会 発表論文集. (1997).
- [5] Nishiyama, Kunio. Adjectives and Copulas in Japanese. Journal of East Asian Linguistics 8. pp.183-222.(1999).