

日本語における動詞の造語モデルの作成

1 G-9

永井 秀利[†] 中村 貞吾[†] 日高 達[‡]九州工業大学[†] 九州大学[‡]

1 はじめに

べた書きされた日本語文の形態素解析においては、多くの曖昧さや単語辞書未登録の語の存在が大きな問題となる。これらの問題に対処するために、我々は日本語における単語の造語モデルの研究を進めてきた。造語モデルは、日本語における漢字の造語能力に着目し、その結合による単語の造語のモデル化を試みたものである。

従来の造語モデルはその実現性を重視し、機械的に構築可能なモデル化を目指した。そのため従来のモデルでは造語を正規文法の範疇でとらえていたが、モデルの精密化を進め、さらに複合語の造語モデル構築を目指すには、造語を文脈自由文法の範疇でとらえる方が望ましい。本稿では、日本語における動詞について分析を行い、より精密な造語のモデル化を試みた結果について述べる。

2 従来の造語モデル

日本語は表意文字である漢字を用いる言語である。造語モデルはこの漢字の結合による造語能力に着目し、単語の造語を確率モデル化したものである [1]。従来の造語モデルでは、基本的には漢字1文字とそれに続く一連の平仮名列（送り仮名を考慮したもの）を造語の最小単位である造語単位とし、その意味的・音韻的結合による造語の過程を1重マルコフモデルでモデル化した。この際、単純なマルコフモデルでは日本語の単語とは認めがたいような短い未知語を多数造語する恐れがあるため、造語単位を位置的性質で細分し、これを一つの内部状態としたモデルを採用していた。

ところがこの従来形式の造語モデルを動詞に対して適用しようとした場合、短い登録語が多いためうまく機能しない。サ変動詞（サ変名詞）には長い語も多く存在するが、他の動詞に関しては漢字2文字以下の語がほとんどである。また、動詞「春めく/harumeku」のように、従来の造語単位定義では1造語単位の語と解釈されてしまうが、実際には名詞「春/haru」に接

尾辞「-mek-u」が接続することにより、品詞を変化させつつ造語が行われているものもある。動詞として登録されている語には複合語的な語も多い。特に和語動詞に関しては根幹から新たに造語される可能性に乏しく、和語動詞を含めた造語は、通常、複合語的な様相を呈するであろう。したがって、日本語で語として用いられるものに関してより精度の高いモデル化を行うためには、従来のモデルでは扱いが不十分であった複合語的な構造をも見据えた造語を考える必要がある。

3 語構造の分析とモデル化

単語辞書を見ると、語源を同じくして、そこから派生的に発生した動詞も数多く登録されていることがわかる。例えば、「動く」、「動かす」、「動ける」のような組がそれである。従来の造語モデルの定義では、これらは異なる造語単位であり、相互関係は全く考えられていなかった。しかし、これらの語は同じ語根「動k/ugok」に対して、派生接尾辞「-(空列)-」、「-as-」、「-e-」が接続（「空列」は扱いの便宜のため）して造語され、さらに活用語尾「-(r)u」が接続して語としての形態を整えたものであると言える。つまり、これらの語の使用に際しては、同じ基礎概念に基づいて用いられていると考えられるため、語の出現頻度を考える場合には、基礎概念となる「動k/ugok」を基準として扱うべきである。

ここでは、「-as-」や「-e-」のように、それ自身が何らかの概念を担うというよりは接続する対象の品詞的属性や語性（自他、相など）を変化させる性質を持つ構成要素を「機能語基」と呼び、それ以外の「動k/ugok」のような構成要素を「概念語基」と呼ぶ。従来の造語単位は細分され、これらの語基が新たな造語単位として定義される。

概念語基は品詞的属性により次のように分類される。

- ・体言類：事物の意味を表すもの
- ・用言類：動作・作用を表すもの
- ・相言類：体言（類）を修飾し、その性質・状態を表すもの
- ・副言類：用言（類）を修飾し、その状態・程度・限定を表すもの

造語は、中心となる概念語基に別の語基が結合する

An Approach to Construct a Japanese Verb Model
Hidetoshi Nagai[†], Teigo Nakamura[†] and Toru Hitaka[‡]

[†]Kyushu Institute of Technology

680-4 Kawazu, Iizuka, Fukuoka 820, Japan

[‡]Kyushu University

6-10-1 Hakozaki, Higashi-ku, Fukuoka 812, Japan

ことにより行われる。この結合により構築されたもの(複合語基)も造語過程において概念語基として扱われ、その品詞的性質により上記の分類が適用される。

複合語基の造語過程は、中心となる概念語基に対して結合する語基が(1)概念語基である場合と、(2)機能語基である場合とに大別される。

(1)の造語過程には、

体言類 + 体言類 ⇒ 体言類
 相言類 + 体言類 ⇒ 体言類
 用言類 + 用言類 ⇒ 用言類
 用言類 + 体言類 ⇒ 体言類
 体言類 + 用言類 ⇒ 用言類
 副言類 + 用言類 ⇒ 用言類

などがある(活用形に関する条件は略)。これらの造語においては、構成要素である語基の間に修飾関係や格関係といった関係の存在を認めることができる。

(2)の造語過程は、機能語基の種類により、(a)品詞的屬性を変化させる機能語基である場合と、(b)語性を変化させる機能語基である場合とに分けられる。

(a)の品詞的屬性を変化させる機能語基は、否定漢語系接頭辞(無, 不, 未など)が体言類を相言類化する例以外はすべて接尾辞である。相言類を体言類化する「-sa」(明るさ, 楽しさ)や、体言類を用言類化する「-mek-」(春めく)などがこの例と言える。

(b)の語性を変化させる機能語基は、「動かす」における「-as-」などである。この「-as-」は、自動詞としての性質を持つ語根「動k」に接続して他動詞化する働きを示している。他の例としては、自動詞「付く」に対する他動詞「付ける」の「-e-」(可能の態を加える性質も持つ)や自動詞「残る」、他動詞「残す」の「-r-」, 「-s-」などがある。「-e-」は他動詞「焼く」、自動詞「焼ける」のように自他逆の変化にも用いられる。自他の変化方向の差異は、語根「焼k」の自他性に依存すると考えられる。

前述した造語過程(1)に示した結合規則は一般化しすぎており、このままで造語規則として用いるには造語能力が強すぎる恐れがある。例えば「用言類 + 用言類 ⇒ 用言類」の形式では多くの組合せが可能であり、その多くに何らかの意味解釈が可能であるが、すべての組合せが受理可能なわけではない。用言類の自他や態、相といった語性が、この造語規則に対する制約として機能する。

ここで自動詞「 α_i : 降る(降 r-u)」, 「 β_i : 散る(散 r-u)」, 他動詞「 α_i : 降らす(降 r-as-u)」, 「 β_i : 散らす(散 r-as-u)」, および自動詞に使役の助動詞「せる(-se-ru)」が結合したものである「 α_c : 降らせる」, 「 β_c : 散らせる」について、 α , β の組での複合動詞

の造語を見る。

造語されるもののうち、 $\alpha_i \cdot \beta_i$ である「降らし散る」は複合語としては認めがたい。これは他動詞と自動詞との結合であり、語性の統一が取れていないためである。 $\alpha_c \cdot \beta_i$ である「降らせ散る」も同様である。 α_c の概念語基は自動詞としての語性を持つが、これに使役の助動詞が付加することにより他動詞化を生じている。このことは複合語造語モデルにおいては少なくとも一部の助動詞は機能語基として組み入れねばならないことを示している。

それに対し、 $\alpha_i \cdot \beta_i$ である「降り散らす」や $\alpha_i \cdot \beta_c$ である「降り散らせる」は、語性の統一が取れていないにもかかわらず、複合語として受理可能であるように見える。これは「降り散らす」や「降り散らせる」の構造が実際には「用言類 + 用言類 ⇒ 用言類」ではないためである。これらの語は $\alpha_i \cdot \beta_i$ である「降り散る」(自動詞複合語)に対し、機能語基「-as-」や使役の助動詞が接続することで他動詞化したものである。よって語性の統一は保たれており、語としての不自然さを有しない。

この造語機構における性質を利用することにより、例えば「~降らせ散る」のような文が存在した場合、その文は「~(降らせ散る)」のような複合語を含むのではなく「(~降らせ)散る」のような構造を持つことが決定可能となる。これは係り受け解析における曖昧さ解消に有効であると言える。

4 おわりに

造語モデル強化へ向けての動詞造語のモデル化について述べた。本稿では非サ変動詞について扱ったが、サ変動詞(名詞)の場合は造語の機構が異なるため、さらなる分析とモデルの拡張が必要となる。非サ変動詞とサ変動詞との間での造語機構の差異の分析は、類似した差異を有すると思われる形容詞と形容動詞との分析にも役立つであろう。

本稿では造語過程における音韻面での変化(音便, 転音など)については触れなかったが、これについてもモデル化を行わねばならない。ただし、音韻面での変化は規則性が弱く、同一条件下においても変化を生じる場合と生じない場合とがあるため、確率的なモデル化が不可欠と言える。

参考文献

- [1] 永井, 日高: 日本語における単語の造語モデルとその評価, 情処論文誌 Vol.34 No.9, pp.1944-1955(1993)