

形態素分析の基礎から現代日本語の形成まで

佐野 洋^{*} 赤坂 宏二^{**} 久保 幸弘^{**} 杉村 領一^{**}^{*}(株)東芝 総合研究所^{**}ICOT

1. はじめに

一般に形態素解析を行なう場合には、文の分析によって得られるなんらかの単位を仮定する。多くの場合、文を解析する際に、分析あるいは分割の単位として単語を立てるのが通例である。そういった場合、その独立性は暗黙の内に仮定されている。本稿では分析の単位を語基と接辞に求めた。そして、語基と接辞から語が構成される過程を考慮して、①語形成の結果得られる語に独立性を仮定し、②この語を文分割の単位とすることによる形態素解析を試みた。

2. 語の構成要素

語の構成要素としては「語根」、「語基」、「語幹」、「接辞」があるとされる〔1〕。本稿では分析の単位を語基と接辞に求めるが、語基とは一般に語の意味的な基幹部分とされる。接辞とはいわゆる助詞、助動詞をはじめ、「春めく」の「めく」、「お父さん」の「さん」などの語の基幹部分以外の総称である。そうして、語構成の手段を森岡〔4〕の考えを前提として、語基の派生と屈折においた。ここでの屈折とは、狭い意味での体言語基が格助辞を作って格表示を示す手段のみならず、用言が活用する場合の語形変化も含めた広義の屈折とする。

屈折現象は自立語基に限られる。例えば、指示語基「こ」は自らは屈折出来ない指示語基である。派生とは「この」のように自立語基を形成する過程である。

語構成の結果得られる語（文分析の単位）は、文中の成分として図1に示す機能を持つものとして分類できる。

語 :-	述語		(文の叙述成分)
	連体語		(連体修飾成分)
	補語		(連用修飾成分)
	副用語		(連用修飾成分)
	独立語		(独立成分)

図 1 文の成分としての語

形態素解析の結果、文は図1の五つの機能の語に分解されることになるが、構文を分析する際の一次的な文の分析単位としては十分と考えられる。

語基による分析の例を表2に示す。形容語基「高」を例として挙げる。同表の単語分析が一般に行われている考え方である。

表 2 語基分析の例

語基分析		単語分析(従来の考え方)	
高-い	(述語)	高(語幹)-い	(形容詞)
高-さ	(補語)	高さ	(名詞)
高-ま-る	(述語)	高ま(語幹)-る	(動詞)
高-め-る	(述語)	高め(語幹)-る	(動詞)
高-々	(副用語)	高々	(副詞)

直感的には「高い」、「高さ」、「高まる」等は同じ「語幹要素」を持つ語であると考えるのが自然であろう。従来の単語分析によると形容詞語幹「高」、自動詞語幹「高ま」、他動詞語幹「高め」、そして名詞「高さ」を区分して処理をしていた。

それに対し、ここでは表2の語基分析のようにこれらの語はすべて一つの語基から派生されるものとする。

3. 活用

語基分析によると従来の活用語処理が変わる。従来の方法に比べて、いわゆる語幹部分が異なることと基本的に活用語/無活用語の区分がなくなるためである。

さて、文が表わす内容に2分する要素を仮定する考え方が多くの学者により主張されている〔2〕。伝統的には「詞」と「辞」をはじめ、例えば「言表事態」と「言表態度」、「命題」と「モダリティ」などの言葉が使われている。

日本語では文の基幹となる要素は叙述能力を持つ述語にある。そこで用言の活用表の修正に際し、述語に上述の考え方を導入し、ここでは寺村〔3〕に従い、述語を「コト」を表わす部分と「ムード」を示す部分から構成されるという前提をおいた。そして語基分析のために「コト」を構成する部分を派生により作られものとし、「ムード」を示す部分を屈折により表わされものとし、すなわち格支配を中心としたヴォイス、アスペクト、否定表現等を含む部分を「コト」とし、話者の意志判断を「ムード」とした。

表3 に活用形態を示し、表4 には「咲(ク)」を例とした語基分析を示す。

表 3 活用形態

語基分析	活用形態	例
屈折に關与する 形態変化	現在形	咲ク
	完了形	咲イタ
	現在推量形	咲コウ
	完了推量形	咲イタダロウ
	否定意志形	咲クマイ
	命令形	咲ケ
	現在仮定形	咲ケバ
	完了仮定形	咲イクラ
派生に關与する 形態変化	連用形	咲イテ
	中立形	咲キ
	連接形	咲カ

従来の活用表は、たしかに形態的に全ての接続の形態を網羅している。しかし、語基分析によれば派生と屈折は語構成の上から機能的には基本的に相違がある。そこで本稿の活用表ではこの区分を明示した。

4. 実現

語構成に基づく形態素解析はLAX [5] と呼ばれる形態素解析プログラムでおこなわれる。LAX はICOTで開発された言語解析ツールである。このプログラムは逐次推論マシンPSI上で動作する。LAX の文法表記は、この語構成による分析の手段を容易に受け入れる能力を持っている。図5 に一部の表記例を示す。

```
def-macro(カ,[ウ系強変化カ,ウ系強変化イク]),
く::[ 派生(macro(カ),[]),
屈折([],end)]],
いた::[ 派生(macro(カ),[]),
屈折([],end)]],
...
```

図 5 LAX での記述例

5. まとめ

文の分析の単位を語基と接辞に求め、派生と屈折という手段で語構成の立場から形態素解析を試みた。これまで暗黙の内に仮定されて分析の単位を単語から語基と接辞へと変更し、それに随する活用形態の考察を行った。

活用の基本形態については佐野、天野 [6] に拠る。

[6] では話者の意志表示として活用を考察したが、派生現象についての扱いが明確ではなかった。本稿では基本的機能の違いからこれを区分した。

解析のための規則は、現在、基本的な語解析に止まっている。重要な問題としては、合成語の構成と転成する語基の扱いである。転成とは「願いがかなう」のように用言語基「願(ウ)」が派生途中で体言語基へ機能変化したものである。合成語については派生以前の現象として、また転成については語基の分類を詳細化することで対処する予定である。今後はこの語基分析をもとにした形態素解析規則の充実に計ってゆく予定である。

[参考文献]

- [1] 日本文法事典 有精堂
- [2] 芳賀綏 『日本文法教室』 東京堂出版
- [3] 寺村秀夫 『日本語のシンタクスと意味』 vol 1,2 くろしお出版
- [4] 森岡健二 『形態素論』
- [5] 杉村、赤坂、松本『並列形態素解析システムLAXの実現』 情報処理学会35回全国大会論文集
- [6] 佐野、天野 『日本語構文分析のための品詞分類について』 情報処理学会34回全国大会論文集

表4 「咲(ク)」の派生と屈折

語基	現在形	推量形	否定 推定形	命令形	仮定形	連用形	接続表現	中立表現
(ask) 咲き、か	く いた	こう いたろう	くまい	け	けば いたら	いて	ながら	(き)
咲かす、そ	る	そう	るまい	よ	れば	て	ながら	(せ)
咲か仕	る	よう	るまい	よ	れば	て	ながら	(れ)
咲きた	い	かろう	×	×	×	くて	×	×
咲きたか	る	ろう	るまい	れ	れば	って	×	×
咲かな	い	×	×	×	ければ	くて	×	×
咲きま	す	しょう	すまい	せ	すれば	して	×	×
咲きそう	体言語基を構成し、ダネ活用助辞により自立する(「だ、です」)。							
咲くよう	だ だった	だろう だったろう	×	×	ならば だったら	で に		
咲くそう	だ だった	だろう だったろう	×	×	ならば だったら	で に		

括弧が屈折を示す。縦並びは派生を示す。「×」は存在しない形態を示す。