

## 品詞体系の変換を考慮した中間品詞体系と品詞変換表の設計

望主雅子 亀田雅之

(株)リコー 情報通信研究所

## 1 品詞体系変換の試み

近年、多くの形態素解析システムや形態素解析を使ったアプリケーションが開発されてきている。また、構文解析システムもツールとして使えるものが徐々に増えてきている。多くのシステムが存在する現在、これらを相互に接続して実験するには、接続時の品詞体系の相違が大きな問題となる。システムの使用する品詞体系は、その目的や用途によって異なるため、システムの数だけ存在するといえる。また設計者の言語に対する思想が反映されるものでもある。

品詞体系の変換を単純に行なうことは極めて困難であり、問題があるが、ここでは、既存のツール類の接続によるメリットの方を考え、品詞体系の変換を試みた。接続の一例として、大規模な辞書を用いずに日本語解析を行なう QJP[1]<sup>1</sup> の構文解析部が他の形態素解析と比較的容易に接続できると考え、形態素解析と QJP 構文解析部 (以下、QJP と呼ぶ) の接続を検討した。

品詞体系の比較や変換について、以前に羽田ら [2]、出羽ら [3] が行なっている。主に複数辞書を利用して、言語的な知識を獲得するために品詞の比較、変換を行なったものである。

品詞体系の変換には、ユーザ (形態素解析システムの作成者、使用者) の現在使用する体系と接続するシステム (QJP) の体系とを直接変換することが考えられるが、我々は、両システムの間に中間的な品詞体系を設定し、この体系を介してユーザが接続したいシステムを使用する方法をとった (図 1)。ユーザが自身の使用する体系と中間的な品詞体系の変換表を設定する。

## 2 中間品詞体系の設計

中間的な品詞体系は、ユーザの形態素解析系と似た体系の方がユーザの記述量を削減できる。中間品詞に、現存するすべての品詞分類を入れることも考えられるが、煩雑でわかりにくくなることや実際、困難なので、後続の構文解析系で必要とする点や体系の共通点を中心に中間品詞体系に盛り込む。

また、接続する構文解析系 (QJP) に独自で、他の形

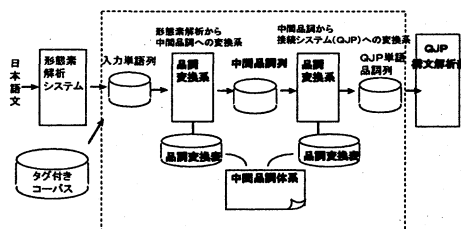


図 1: 品詞変換系の構成図

態素解析が持たない分類観点や品詞は、中間品詞体系から構文解析系の品詞への品詞変換表で対応することにした。

まず、現在広く知られている形態素解析系やコーパス、QJP の品詞体系の共通点、相違点を調査し、それに基づき品詞の設定を行なった。

## 2.1 品詞体系の比較対象

今回比較対象とした体系は以下である。

- JUMAN[4]: 田窪・益岡文法をベースにした形態素解析処理システム。
- EDR 辞書 [5]: 日本語品詞名を比較対象とした<sup>2</sup>。学校文法に近い。
- RWCP コーパス [6]: コーパスに付与したタグを対象とした。出現状況を考慮した分類が階層的に付与されている。
- 社内形態素解析システム hamo[7]: 素性による形態素解析システムで音声合成等に使用。
- QJP: 学校文法ベースで大規模な辞書を用いない。形態素品詞と単語品詞があるが構文解析系で用いる単語品詞を対象とした。

<sup>1</sup> 日本語解析系 QJP は形態素解析部と構文解析部を持つがこの二つは構文解析部を接続対象とした

<sup>2</sup> [3] で比較対象としていた分類である

## 2.2 品詞体系の設計方針

### 2.2.1 品詞名の設定

品詞名は品詞体系の性質を表わし、品詞変換にも大きく関わると考える。以下の二つのタイプがある。

(1) 修飾句+大分類名 (EDR,JUMAN の非活用語)

例：形容詞性述語接尾辞

(2) 品詞+分類観点 (RWCP,JUMAN の活用語 QJP,hamo) 例：形容詞 見出し形 非自立 (RWCP), ヤスイ=形:ク終 (QJP)

省略の少ない、分類観点が明確なものが品詞名としてわかりやすいと考える。

### 2.2.2 階層的な品詞体系の設計

中間品詞体系として大分類に分類観点を付与する観点付与型の階層的な品詞体系を設定し、以下の性質を持たせた。

#### ● 品詞の詳細さに応じた変換

できるだけ細かい分類での品詞の対応を目指しつつ、該当する品詞がない場合に少なくとも大分類品詞と対応がつくようにした。大分類レベルでも後続の構文解析が動作することを目指す。

#### ● 分類観点の優先順位付け

分類観点の重要な順に大分類、分類観点と後ろに付与し、本品詞体系の分類観点とその重要度を明確にした<sup>3</sup>。

例 副詞的名詞.時.数

動詞.終止連体.カ行5段.文語

優先する分類観点によって体系が変わるが、本体系では以下の順に優先的に分類観点を設定した。

#### － 構文解析での重要性和体系の共通点

連用修飾性、連体修飾性など係り受けの基本となる概念を優先し、各体系の共通点も考慮して大分類、観点(第1番目)を設定する。

#### － 品詞分類の境界に位置する語

先に規定した品詞の性質から逸脱する性質を持つ語、境界的な語(例：形動名詞)、連体修飾、連用修飾の性質を破る性質を持つ語について設定する(相対名詞、副詞.程度)。

#### － 意味、機能、品詞独自の性質

今後の詳細な解析のために必要な観点(例：固有、修飾、指示、転成、熟語)。品詞独自

の性質(名詞では複合語になった際の性質、活用語の活用タイプなど)。

- － 形態素解析の接続を考慮した部分の削除  
形態素解析の精度向上のための分類(後接語の違い)は今回は構文解析系接続を主目的とするので盛り込まない。

分類観点には複数の品詞に現われるものもあり、将来的には分類観点付与の順番の入れ替え可能な、柔軟な品詞体系を考えている。

## 2.3 品詞分類の概要

### 2.3.1 名詞類

各体系で、サ変動詞、副詞、接続助詞、修飾を受けて句をなしうる機能、数、時間など、他品詞の機能を持つ境界的な品詞が共通している。これらと、名詞複合語を解析する際に必要と思われるもの<sup>4</sup>を設定した。また、係り受け上特殊な副詞の修飾を受ける名詞(相対名詞)、格をとる下位分類(格、例：中心)を設定した。

品詞名	下位分類観点、説明
名詞	他に該当しない場合付与される
形式名詞	構文的な機能をもつ境界的な品詞類
副詞的名詞	下位の分類観点は[副詞、修飾、
形動名詞	[様態、和語、熟語、時、数、
サ変名詞	転成、形容詞、転成、動詞、指示、格]
相対名詞	相対名詞は副詞の修飾を受ける名詞
接辞名詞	名詞複合語の末に位置し、前部分を修飾。例：未満、以外 形式名詞的機能を持つものに[句]
固有名詞	下位分類は設定しない
数詞	下位分類は設定しない

### 2.3.2 形容動詞の扱い

形容動詞はものごとの状態を表し、他の語の様態を修飾する機能が形容詞と似ている一方、名詞複合語を構成する点、「だ」に接続する点で名詞にも似ている。各体系でもその設定が異なる。

	JUMAN	RWC	EDR	QJP	hamo
形容動詞	×	×	○	○	○
	(形容詞)	(名詞)			
活用	有	無	有	有	無

本体系では、形容動詞を設定し<sup>5</sup>、助動詞「だ」が接続するとした。

### 2.3.3 連体詞

JUMAN では指示詞を大分類にたて、下位分類に名詞、連体詞、副詞をたて、機能的に分類している。

<sup>4</sup> QJP 構文解析系では辞書を持たないため、名詞複合語は解析しないが、将来可能性があるため設定した

<sup>5</sup> 特殊な分類だが他品詞の分類観点(様態)にも現われている

<sup>3</sup> 分類観点として検討が必要なものは下位分類にすることで情報の精度を保證できる

本体系では、連体修飾、指示性という点で分け、代名詞を「名詞・指示」、連体修飾し指示的な語を「連体詞・指示」とした。それ以外の連体修飾語を「連体詞」とした。

## 2.3.4 助動詞

従来、助動詞とされた語には以下の性質が含まれる。

- 自立性の差  
動詞や形容詞に同表記語を持ち、補助的に使用されるようになった語(例：にくい、かかる)
- 機能、活用タイプ(類似する品詞)  
形容詞、動詞、形容動詞と類似する。
- 接続する語(名詞、動詞)
- 述部での位置、意味  
テンス、ムードなど(例：た、らしい)

JUMAN、RWCP では機能や接続する語の違いからいくつかの語を接尾辞、動詞などの別品詞にしている。

本体系では、語ごとに性質が大きく異なるので、細分類せずに、一語一品詞にした。

例) 助動詞、レル、終止連体

自立性の差のある、もともと動詞、形容詞だったものに対しては補助動詞、補助形容詞を設定する。また、活用、機能が形容動詞的な語(例：がち)を接尾辞に、連体修飾をうける語を形式名詞(例：よう)に分けた<sup>6</sup>。

## 2.3.5 動詞

形態素解析では活用語を展開するため、品詞名には活用タイプが先立って付与されていることが多いが、構文解析では活用タイプより活用形の情報が重要なので活用形を上位の分類観点とした。少なくとも活用形が変換できればよいと考える。

例) 動詞、終止連体、力行5段

活用形名も従来の命名をベースに省略せず(○終止連体、×終体)に、同機能の場合は意味的な観点(文語、口語、丁寧)と表記(命令、ロ、命令、ヨ)を設定した。

また、多くの体系が同表記語に一活用形名をあてている(例えば終止形と連体形)ので、それに合わせた。ただ、実際の構文解析ではどちらか判別する必要があるため、後述の品詞変換表で吸収する。

## 2.3.6 助詞

一般的な細分類(格助詞、準体助詞、接続助詞、副助詞、終助詞)以外では体系間に以下の差がある。(○

<sup>6</sup> JUMAN で判定詞とする「だ」は分けないが、一語一品詞なので必要によっては特別な処理ができる

はある、×はない、-はそもそも細分類がない)

助詞	QJP	JUMAN	EDR	RWCP	hamo
引用助詞	○	○	-	○	×
格助詞相当	×	×	○	○	○
係助詞	○	×	-	○	×
並列助詞	×	×	-	○	×

助動詞同様、一語ごとの機能が異なるので、細分類の下に語表記そのものを置いた。構文解析上重要な並列助詞を設定し、格助詞の機能(引用、連体、主格)を下位分類として設定した。また、係助詞は、副助詞との境界が難しいので、ここでは副助詞としてまとめた。

## 2.3.7 接頭辞、接尾辞

各体系では接続の性質や全体としての機能で分類している。本体系では機能的な観点を設定した。また、接尾辞の一部に従来助動詞とされた「だ」の系列の付く「みたい」等を含むようにした。

品詞	分類観点、説明
接頭辞	分類観点は[助数詞、固有、様態化修飾、指示、時]
接尾辞	分類観点は[助数詞、固有、名詞化、副詞サ変化、様態化(例：よう、みたい)]

## 2.3.8 その他(副詞、接続詞、感動詞、記号類)

体言を修飾する副詞の下位観点(程度)が重要である。

品詞	分類、説明
副詞	下位分類に(情態、程度、量、時制、注釈、陳述)を設定
形容詞	下位分類を設定しない。活用形は広く取り入れる(例：大きゅうございます)
接続詞	構文解析上重要な「等位」を下位の分類とした
感動詞	下位分類を設定しない
記号類	句点、読点、開き括弧、閉じ括弧を設定した

## 3 品詞変換表

基本的に品詞間の対応を記述するが、単純な変換ではうまく扱えない現象に対応できる機能を付加した。

- 出現環境によって品詞が異なる  
構文解析系では語が修飾か被修飾か、体言か用言かでは大きく異なる。両性質を持つ語について、出現環境によって判別する機能を持たせる。  
例： 食べる(連体形)人(名詞)  
食べる(終止形)。(句点)

- 語の単位が異なる

体系によって一語とする範囲が微妙に異なる。

JUMAN	書いた(動詞のタ形)
その他	書いた(動詞の音便形) た(助動詞)
RWCP,hamo	身軽(形容動詞) だ(助動詞)
その他	身軽だ(形容動詞の終止形)

これらを分割、併合できるような機能を持たせる。

- 品詞に属する語が異なる

同じ品詞分類であっても属する語が異なる場合が多くある。機能語は特に解析結果に影響する。語自身の表記、表記の一部を指定して、特定の語について別の品詞に変換できるようにする。

例) とても (副詞) → 相対的な副詞

- 変換表の記述を削減する

品詞名の一部や太分類名を部分一致させることで変換表の記述を削減する。また、活用形の展開による膨大な記述パターン数を、活用型、活用形の部分変換ルールによって記述の削減をする。

例えば、名詞全般を「\*名詞」で表現したり、品詞 (動詞、未然、力行5段) から品詞 (動:五カ未a) への変換を (動詞; 動)(未然; 未) といった部分変換ルールによって記述できるようにする。

- 品詞名からの品詞推定

変換表の記述が不備な場合に品詞名から品詞の太分類程度を推定し、変換する。

## 4 接続

JUMAN、社内の形態素解析システム hamo について品詞変換表を記述し、接続を行なった<sup>7</sup>。

変換表の規模を表1に示す。変換数の括弧内は品詞名の局所変換ルールによる記述削減を行なった場合の数である。品詞変換の例を以下に示す。

変換元	変換先	変換数
hamo	中間品詞体系	510(313)
JUMAN	中間品詞体系	754(127)
中間品詞体系	QJP	1014(673)

表1: 品詞変換数

例) とても身軽に歩く人だ。

### JUMAN 形態素解析結果 (表記、品詞のみ)

とても	副詞
身軽に	形容詞 ナ形容詞 ダ列基本連用
歩く	動詞 子音動詞力行 基本形
人	普通名詞
だ	判定詞 判定詞 基本形
。	句点

### 中間品詞体系変換結果

とても	副詞、程度
身軽	形容動詞
に	助動詞、ダ、連用二
歩く	動詞、終止連体、力行5段
人	名詞
だ	助動詞、ダ、終止
。	句点

<sup>7</sup> ユーザは構文解析の結果と中間品詞体系、ももとの品詞体系を参照できる

## QJP 単語品詞

とても	相対副詞
身軽に	形動:ダ用b
歩く	動:五カ体
人	名詞
だ	ダ=終
。	句点

## QJP 構文解析結果

### 【文節詳細情報リスト】

[1] [とても{相対副詞}] {副詞句 連用 副詞連体 程度} → 副詞句:副詞句 → [2]

[2] [身軽に{身軽だ|形動:ダ用b}] {副詞句 連用 A 従属} → 連用:動詞句 → [3]

[3] [歩く{歩く|動:五カ体}] {動詞句 連体形連体} → 連体形連体:体言句 → [4]

[4] [人{名詞}だ{だ|ダ=終}。{句点}] {コピュラ句 体言句 用言終止}

### 【係り受け木構造】

- 1[ 4]( 3) :人だ。
- 2[ 3]( 2) :「歩く
- 3[ 2]( 1)>: 「身軽に
- 4[ 1]( 0)>: 「とても

## 5 まとめ

階層的な中間品詞体系と機能拡張した品詞変換表によって、品詞体系の異なるシステムの接続を確認した。構文解析独自の観点を中間品詞体系からQJP品詞体系で吸収したため、他体系から中間品詞体系への変換数はおさえられた。

本変換系を音声合成システムの言語処理部に接続し、構文解析結果を反映したポーズ挿入を行なっている[8]。今後は他体系との変換、大量実文による実験を行ない、中間品詞体系、変換系を改良していく。

## 参考文献

- [1] 亀田 1995 「軽量・高速な日本語解析ツール『簡易日本語解析系 QJP』」 言語処理学会第1回年次大会
- [2] 羽田、松本 1996 「複数辞書の統合的利用のための汎用日本語辞書の構築」 自然言語処理研究会 116-1
- [3] 出羽、西野、辻井 1997 「知識獲得に向けての品詞体系の考察」 言語処理学会 第3回年次大会
- [4] 松本、黒橋、山地、妙木、長尾 1997 「日本語形態素解析システム JUMAN version 3.3」
- [5] 日本電子化辞書研究所 1996 「EDR電子化辞書1.5版仕様説明書」
- [6] 技術研究組合 新情報処理開発機構 テキストグループ データベースワークショップ 1995 「RWCテキストデータベース報告書」
- [7] 望主、伊藤 1991 「日本語形態素解析における素性の導入」 情報処理学会第43回大会
- [8] 佐藤 1997 「簡易日本語解析系QJPを用いたポーズ位置設定」 情報処理学会第55回大会