N-037

文型パターン解析を用いた日本語 e-learning コンテンツ管理システムの構築

Construction of an E-learning Contents Management System using Grammar Pattern Analysis

山本　樹† **芝野　耕司**†

Tatsuki Yamamoto　Kohji Shibano

一般の e-learning システムとは異なり，語学用では，素材としての単語，文及び会話は，特定の教材に依存せず，教材を超えて共有可能である。また，一文ごとに，その文を構成する単語及び文型の情報を与えることによって，教材をより効果的なものにすることができるとともに，再利用可能な素材コンテンツデータベースを構築することができる。

日本語教育で用いられている形態素概念は，一般の日本語処理での概念と異なる。また，文型パターンには，形態素情報と表層情報の両方が含まれており，これらを含めた解析が必要となる。この論文では，こうした日本語教育固有の日本語処理ニーズを取り入れた日本語解析について報告する。

1.　はじめに

東京外国語大学では留学生日本語教育センターと情報処理センターとの共同で，留学生日本語教育センターで執筆し，世界中で広く利用されている『初級日本語』[1]をもとに，初級学習者向けの日本語 e-learning システムを開発している[2]。

『初級日本語』は本冊をはじめ，『文法解説』，『単語帳』など 7 分冊，合計 1441 頁及び音声テープ 2 巻からなる日本語入門教材であり，300 時間の授業時間を想定し，開発されている。また，タマサート大学(タイ)，マラヤ大学(マレーシア)，マラ工科大学(マレーシア)，バルセロナ自治大学(スペイン)，チュービンゲン大学(ドイツ)などの海外の大学でも用いられている。

今回のシステム開発に当たっては，これらの既存教材を利用するとともに，ひらがな・カタカナ・ローマ字素材を追加するとともに，新たにイラスト，練習問題解答の追加，追加音声録音を行い，同時に，海外からの利用を想定し，英語，中国語（繁体字版，簡体字版），朝鮮語，タイ語，マレー語，モンゴル語版の作成も行った。

また，2004 年 4 月から留学生日本語教育センターで実際の授業での利用を開始している。

語学教材は，単語，文及び会話からなる。そして，これらの基本素材に対して，音声，翻訳文，画像，ビデオなどの素材が関連づけられている。一つの教科書などの教材作成には，カリキュラム，言語教育観などに応じて，様々な教科書が執筆されているが，基本素材そのものは，同じ日本語文であることから再利用可能である。

†東京外国語大学　地域文化研究科

Graduate School of Area and Culture Studies,
Tokyo University of Foreign Studies

すなわち，300 時間（一般の大学でのコマ数換算では 6 コマ程度）の素材から 50 時間（週 1 コマ）の教材や教育目的に合わせた特定目的用の教材を作成することができる。

こうした教材の作成には，語彙及び文型を基本素材に関連づけることが必要となる。また，従来の印刷物の教科書を WBT（Web Based Training）でのハイパーテキストリンクを用い，学習者が疑問に思ったその場で，単語の意味や文型に関する文法情報を参照できるようにすることは，WBT の特徴を活かしたハイパーテキスト教材には，必須である。

しかし，既存の印刷教材には，これら語彙及び文型情報は付加されていない。また，文単位では 1 万以上に及ぶことから，日本語解析を利用した自動処理が必要とされる。

使用している教材『初級日本語』，『単語帳』及び『文法解説』では，約 2000 単語と約 650 の文型パターンを基に、文を構成している。この「単語」及ぶ「文型」は，日本語教育を目的とし，茶筅[3]などの一般の日本語解析で用いられている形態素とは異なる。このため，一般の日本語解析をそのまま用いることはできない。

また，『単語帳』は，新出語彙一覧と各国語訳との対応を目的とし，『文法解説』での文法記述に正確には対応していない。このため，まず文法記述での品詞情報を付加した単語帳の作成が必要である。

表1　文型パターン中の形態素と茶筅の形態素との対応

品詞名	文型パターン	茶筌
名　詞	N　N (Plain Form)	名詞　一般
	Person	名詞　固有名詞　人名
	Place	名詞　固有名詞　地名
	Time	名詞　数＋名詞　一般
	Thing	名詞　一般から変換
	Occupation	名詞　一般から変換
	Reason	名詞　一般
	Number	名詞　数
イ形容詞	Aい (Plain Form)	形容詞　基本形
	Aい	形容詞　基本形から変換
	Reason	形容詞　基本形＋接続助詞「ので」
ナ形容詞	Aな (Plain Form)	名詞　形容動詞語幹＋助動詞「ダ」
	Aな	名詞　形容動詞語幹から変換
	Reason	名詞　形容動詞語幹＋助動詞「ダ」＋接続助詞「ので」
動　詞	V (Plain Form)　Vdic	動詞　基本形
	Vt	動詞から変換
	Vi	動詞から変換
	VN	名詞　サ接続
	V (Potential Form)	動詞から変換
	Vます	動詞＋助動詞「ます」
	Vない	動詞＋助動詞「ない」
	V	動詞　基本から変換

また，日本語教育での形態素認識と一般の形態素認識との対応をとる必要がある。

2. 手法

2.1. 単語帳と形態素対応表の作成

『単語帳』の 2000 語のうち固有名詞 50 語を除く単語に『文法解説』に添う形の品詞を与える。

一方既存の形態素解析ソフト[3]を利用し、単語へ品詞情報を与える。この処理によって，一般の日本語処理のための概念での形態素分類と語学教育で必要とする形態素分類との対応を得ることができる。対応を表 1 に示す。

表 1 の対応表をもとに，文型パターン解析で必要とされる形態素分類への変換を行う。この変換によって，日本語教育で必要とされる単語を各文に付加することができる。

表2 文型パターンと形態素･表層情報

文型パターン	形態素情報	表層形
NがViています	<N> <Vi>	「が」「ています」
Nが見えます	<N>	「が」「見えます」
<Place>に行っています	<Place>	「に」「行って」「います」
<Person>に<Thing>をあげます	<Person> <Thing>	「に」「を」「あげます」
N1はN2が<Number> あります	<N1><N2><Numbe	「が」「あります」
<Time>までにV	<Time><V>	「まで」「に」
Aい-くV	<Aい-> <V>	「く」
Aい(Plain Form)とおもいます	<Aい(Plain Form)>	「と」「思います」
Aな-ですから、Vてください	<Aな-> <V>	「ですから」「てください」
Aな(Plain Form)かもしれません	<Aな (Plain Form)	「かもしれません」
V(Plain Form)だろうと思います	<V(Plain Form)>	「だろう」「と」「思います」
Vdicのに必要です	<Vdic>	「のに」「必要です」
Vtて行きます	<Vt>	「て」「行きます」
Viて行きます	<Vi>	「て」「行きます」」
VNのあとで、Vました	<VN> <V>	「のあとで」「ました」
わたしはNがVますたいです	<N> <Vます>	「わたしは」「が」「たいです」
Vないでおきます	<V ない>	「で」「おきます」
Vているところです	<V>	「て」「いる」「ところです」
XはYをViないせます	<X><Y><Viない>	「は」「を」「せます」
YはXにVないれます	<Y><X><Vない>	「は」「を」「せます」
XはYにNをVtないせます	<X><Y><N><Vtない>	「は」「に」「せます」
YはXにViないられます	<Y><X><Viない>	「は」「を」「られます」
<Reason>。それで、<Result>	<Reason> <Result>	「それで」
なにかありますか		「なにか」「あります」「か」

2.2. 文型パターン解析

最も基本的な文型パターン「<N１>は<N２>です」では、<N>は名詞という形態素情報をあらわし，「は」及び「です」は表層形を示す。すなわち，文型パターンは，形態素情報と表層情報の両方を用いて表現されている。

このため，文型パターンの解析では，単純な一つの文字列に対するパターンマッチではなく，複数の情報を取り扱うことが必要となる。

各文に対して，文型パターン解析を行うと，一つの文に対して，複数の文型が対応することもある。

このようにして，『文法解説』で提供される約 650 の文型の表層情報に形態素情報を加えた文型パターンデータベースを構築する。

文型パターンの解析では、その他に以下のような考慮をする必要がある。

A, 一つの文型パターンで、同じ品詞を二つ以上使用するとき、同じ単語をそれぞれに組むとエラーになる。

B, 形容詞＜A い・A な＞では、文型パターン中の「ー」マークで、語幹以外の部分を除く。

C, 動詞＜V＞では、動詞のあとに続く助詞「て」「たら」「たり」、助動詞「た」を持つもの、受身（れる）・使役（せる）可能（る）の文型パターン時には動詞の活用を行う必要がある。

「て」「た」「たら」「たり」が接続するものは、連用形である。また音便によっては語尾が「っ」「ん」で接続し、促音便には「っ」が接続し、撥音便では「ん」に接続する。

動詞 Vi・Vt は未然形、可能は連体形で接続する。

文型パターンによっては動詞の活用を行い、さらに接続する形態素をみて、音便による接続語の違いがある場合にはそれも変換する。

また「ー」マークでは語幹以外を除く。

これにより、適切な形で、適切な品詞が解析できる。

2.3. e-learning コンテンツ管理システムの構築

各文に対する単語情報及び文型情報の e-learning システムへの組み込みに当たっては，既に作成している教材の XML ファイルに対して，各文毎の単語情報及び文型パターン情報を，XML タグを拡張し，追加した。

本来は「かいわ」や「れんしゅう」の各文に単語情報及び文型情報を追加し，参照を可能とするだけではなく，逆に単語や文型から関連する文の参照も可能にすることが望ましいが，これを実現する方法に関しては，まだ検討中であり，今後の課題といえよう。

3. 今後の課題

この研究では，形態素解析結果を変換し，日本語教育で必要とされる単語及び文型パターンを抽出した。

研究の目的としては，ここで生成した単語及び文型パターンを利用し，様々な目的に応じた日本語教材を作成することを目指している。

基本的には，教材作成には，文型をどの順で導入していくのか，語彙をどの順で導入していくのか，また，コミュニケーション機能を会話でどのように導入していくのかなど指定することが必要である。また，日本語教育の専門家にとっての使いやすいユーザインタフェースの検討も必要である。

こうした要求を盛り込んだ日本語教材開発システムに発展させることが今後の課題である。

参考文献

[1]東京外国語大学留学生日本語教育センター，『初級日本語』，凡人社，2002

[2]佐野洋，林俊成，藤村知子，芝野耕司，「初級学習者向け日本語 e-Learning システム」，言語処理学会第 10 回年次大会併設ワークショップ「e-Learning における自然言語処理」論文集，pp.33-36，言語 処理学会，2004

[3]松本裕治，北内啓，山下達雄，平野善隆，松田寛，高岡一馬，浅原正幸，形態素解析システム『茶筌』version 2.3.3 使用説明書，2003