

形態素解析辞書「中古和文 UniDic」を利用した古典学習教材の作成

須永 哲矢 (昭和女子大学) †

Preparation of Teaching Materials in Classical Literature using an Electrical Dictionary for Morphological analysis "Chuko-Wabun UniDic"

Tetsuya Sunaga (Showa Women's University)

1. はじめに

国立国語研究所では形態素解析辞書「中古和文 UniDic」が公開されている。「中古和文 UniDic」は平安仮名文学作品を対象とした画期的な形態素解析辞書であり、あくまで現代語を対象とした従来の解析辞書では無力であった古典資料を高精度で解析することが可能となった。これを利用することで、日本語の歴史的研究の幅が大いに広がることが期待され、実際、国立国語研究所「日本語歴史コーパス」構築に際してもこの「中古和文 UniDic」が利用されている。今後さまざまな方面での利用が期待できる「中古和文 UniDic」であるが、その利用の在り方はなにも研究利用にとどまるものではない。教育面においても、主に高等学校での古典学習等、活躍の場面は多用に考えうる。本稿では、形態素解析辞書「中古和文 UniDic」の教育転用の一つとして、古典学習教材の作製法を紹介する。

2. 高校の古典学習の在り方と、従来の一般的な学習資料

高等学校の古典学習の要点としては、助動詞の意味や識別、敬語、そして現代語訳などが定番である。文法教育に関しては、学習者にとって「文法」というもの自体が苦手意識を引き起こすせいか、このところ扱いが軽くなる傾向にあるようであり、そのため、「この語の並びでなぜこういう訳になるのかわからない」というまま学習を重ねるといった学習者も多く見受けられる。これに対し現代語訳に関しては、高校で扱うような主要作品の訳なら容易に手に入る時代になっている。その結果、文法への理解がなく、自力では意味を取ることができないまま、訳文を暗記して試験に臨むことを繰り返し、学習が進んでも知識が積み重ならない、というような状態の学習者も多い。

高校生向けの訳としては、助動詞や敬語を意識した、単語ごとの逐語訳に近い形の学習用の訳文も以前から出回っているが、知識のある訳者自身が単語ごとの1対1対応を意識して訳出したにせよ、知識をこれから学んでいく学習者側は、原文と訳文を見比べて、単語間の対応がすぐにとれるわけではない。学習用逐語訳の訳文を単語ごとに切って原文と対応をとっていけば学習効果は高まるであろうし、学習用の訳の理念も本来はそこにあると考えられるが、そのような面倒な作業をする学習者は現実には少ない。学習のための逐語訳は、当初の理念・目的を理解されないまま、実際にはただ話の内容を理解するためだけに読まれ、「わかりにくい文章」との印象をもたれて終わってしまう。

単語に区切る、文法を意識する、という観点からは、いわゆる教科書ガイドと呼ばれるような、原文を全て品詞分解し、その隣に活用の種類や活用形を付した学習資料も以前からあるが、こちらは本文に上記のような品詞情報を付与することが主眼であり、結局訳は別掲であることが多く、文法知識が訳に結び付くには至らない。紙面構成からも、従来の教科書ガイドは本文の隣に品詞情報を小書で書き込んだ時点で非常に読みにくく、さらに逐語訳を付与したうえで、見やすい紙面を構成するというのは極めて困難であろう。また、原文を品詞分解してしまった以上、単語レベルではない、より大きな表現単位に対する注釈などは付与しにくくなってしまふ。

古典学習においては、品詞分解、活用形や助動詞といった文法知識を前提に、単語ごと

† tetsuyasunaga@gmail.com

の対応関係を意識した逐語訳を与え、最終的な解釈としての訳文にたどりつく、という理解のしかたが理想的であると考えますが、従来の一般的な学習資料では、品詞分解は品詞分解だけ、訳文は訳文だけにならざるを得ず、各要素が結び付いた総合的な学習資料の作成は考えにくい。これに対し、「中古和文 UniDic」を用い、本文テキストを形態素解析した Excel ファイルを利用すれば、各分野の情報をを総合的に表示した学習教材が作成可能になると考える。

3. 形態素解析辞書「中古和文 UniDic」

高校での学習という視点から言えば、形態素解析辞書は「機械が自動で品詞分解して、活用の種類や活用形を書き出す」という作業をしてくれる、ということになる。機械に品詞分解をさせ、その結果をもとに、必要・目的に応じて逐語訳や注釈を書き加えていくことで、必要な情報をひとつにまとめた学習教材を自由に作成することができる。

現在公開されている「中古和文 UniDic」は中古和文 UniDic ホームページより無償でダウンロードできる。ホームページでの指示に従ってダウンロード・インストールを行えば、特に機械処理に関する詳しい知識がなくとも、誰でも手軽に形態素解析を行う事ができる。

操作画面は図 1 のとおり。操作用ツール「和文茶まめ」が用意されており、ユーザはマウス操作で簡単に解析が行えるようになっている。古典本文を txt 形式で用意しておけば、あとはこの操作画面でファイルを指定してやれば、自動で品詞分解が完了する。



図 1 操作画面（「和文茶まめ」）

今回は txt ファイルを Excel に出力して加工する、という手順を想定している。実際の解析手順は次の通り。

- ・「ファイル (XML/TXT) を解析」にチェック
- ・「参照」をクリックし、解析対象とするファイル名を指定
- ・「Excel に出力」にチェック
- ・「実行」をクリック

以上の操作のみで図 2 に示すような、品詞分解済みの Excel ファイルが出力される。

もとのTXT ファイル (ここでは例として「花山天皇の退位」)

「ファイル (XML/TXT) を解析」

「参照」で解析対象ファイルを指定 (ここでは「花山天皇の退位」)

「Excel に出力」

「実行」

品詞分解が自動で行われた Excel ファイル

	A	B	C	D	E	F	G	H	I	J	K
1	出典	文境界	書字形	発音形	語彙素読み	語彙素	品詞	活用型	活用形	語形	語種
2	花山天皇の	あはれ	アワレ	アワレ	哀れ	名詞-普通名詞-形状詞可能	アワレ			和	
3	花山天皇の	なる	ナル	ナリ	なり-断定	助動詞	文語助動詞	連体形-一	ナリ	和	
4	花山天皇の	こと	コト	コト	事	名詞-普通名詞-一般			コト	和	
5	花山天皇の	は	ワ	ハ	は	助詞-係助詞			ハ	和	
6	花山天皇の	、				補助記号-読点				記号	
7	花山天皇の	おり	オリ	オリル	下りる	動詞-一般	文語上二段	連用形-一	オル	和	
8	花山天皇の	おはしまし	オワシマシ	オワシマス	おわします	動詞-非自	文語四段-	連用形-一	オワシマス	和	
9	花山天皇の	ける	ケル	ケリ	けり	助動詞	文語助動詞	連体形-一	ケリ	和	
10	花山天皇の	夜	ヨ	ヨ	夜	名詞-普通名詞-一般			ヨ	和	
11	花山天皇の	は	ワ	ハ	は	助詞-係助詞			ハ	和	

図2 操作画面「和文茶まめ」での操作と、出力される Excel ファイル

4. 形態素解析結果をベースにした学習教材の作成

4. 1 教材作成の全体概要

以上の手順で出力した Excel ファイルをもとに、教材として不要な列を削除、必要な情報を追加することにより、目的に応じた教材を自由に作成することができる。

例えば一般的な教材作成としては「書字形」列(もとの本文)、「品詞」列、「活用型」列(学校教育での「活用の種類」)、「活用形」列を残し、新たに逐語訳の列や漢字の読みの列、適宜助動詞の意味や注釈を記入する列を追加する。これにより、いわゆる教科書ガイドの特色であった品詞情報と、語と語を1対1対応させた逐語訳を合わせることが可能となる(図3)。さらに作業自体は一般的な Excel ファイルの編集作業であるため、セルの結合等を利用すれば、品詞分解を行った後でも、複数単語に渡るまとまりに注釈を与えることもでき、学習上、要注意箇所の色付けを行って目立たせることなども容易である。そして最大の特徴は、市販のもとと違い、資料の内容を自由に設定できることである。

教科書

教材例

黒地: 自動解析結果を利用

白地: 目的に応じ手作業で追加

読み	本文	逐語訳	品詞	活用の種類	活用形	注釈等
けんしやう	眼証	丸見え	名詞			
に	で		助動詞	助動詞-ナリ-否定	連用形	否定
こそ			係助詞			
あり	ある		動詞	3行変格	連用形	
けれ	なあ		助動詞	助動詞-ケリ	已然形	詠嘆
。	。					
いかが	どう		副詞			出家することへのためらいがある。ここでの「丸見えだし…」のあとも、いろいろ理由をつけてまためらって見せる。
す	す		動詞	3行変格	終止形	
べから	べき		助動詞	助動詞-ベシ	未然形	
む	だろう		助動詞	助動詞-ム	連体形	
」	」					
と	と		格助詞			
仰せ	おっしゃ		動詞	下二段-サ行	未然形	尊敬
られ	られ		助動詞	下二段-ラ行	連用形	尊敬
ける	たの		助動詞	助動詞-ケリ	連体形	過去
本	本		格助詞			

図3 形態素解析結果を利用した教材のイメージ

4. 2 作業手順例

形態素解析結果そのままの Excel ファイルから、学習用教材を作成する作業手順の一例を紹介する。活用の仕方と目的次第で、作成方法はいくらでもありうるはずであり、ここで示すのはあくまで一例である。

(1) 不要な列の削除、必要な列の追加、列名の書き換え

図3に示したように、「読み/本文/訳/品詞/活用の種類/活用形/注釈等」という情報をもった教材を作成するとする。解析結果そのままのファイルでは図4のように「出典/文境界/書字形/発音形/語彙素読み/語彙素/品詞/活用品/活用形/語形/語種」となっているので、このうち「出典/文境界/発音形/語彙素読み/語形/語種」列(図4中、黒地にしてある列)を削除する(図5)。

出典	文境界	書字形	発音形	語彙素読み	語彙素	品詞	活用品	活用形	語形	語種
花山天皇の	日	あはれ	アワレ	アワレ	哀れ	名詞-普通名詞-形状詞可能			アワレ	和
花山天皇の		なる	ナル	ナリ	なり-断定	助動詞	文語助動詞	連体形-一般	ナリ	和
花山天皇の		こと	コト	コト	事	名詞-普通名詞-一般			コト	和
花山天皇の		は	ワ	ハ	は	助詞-係助詞			ハ	和
花山天皇の		、			、	補助記号-読点				記号
花山天皇の		おり	オリ	オリル	下りる	動詞-一般	文語上二段	連用形-一般	オル	和
花山天皇の		おはしまし	オワシマシ	オワシマス	おわします	動詞-非自立	文語四段	連用形-一般	オワシマス	和

図4 解析結果そのままの状態

書字形	語彙素	品詞	活用品	活用形					
あはれ	哀れ	名詞-普通名詞-形状詞可能							
なる	なり-断定	助動詞	文語助動詞	連体形-一般					
こと	事	名詞-普通名詞-一般							
は	は	助詞-係助詞							
、	、	補助記号-読点							
おり	下りる	動詞-一般	文語上二段	連用形-一般					
おはしまし	おわします	動詞-非自立	文語四段	連用形-一般					

図5 不要な列を削除

「書字形」列を「本文」、「活用品」を学校教育に合わせ「活用の種類」等と書き替える。また、「語彙素」列を「訳」と書き替え、逐語訳を記入する列として利用する。逐語訳は手作業で入力すべき列であるが、新規に空白の列を作って一から記入していくよりは、「語彙素」列の表示を利用して適宜書き替えたほうが作業が早い。「語彙素」とは本来辞書見出しの代表表記が表示される部分であるが、現代語形(の、活用がある場合には終止形)が表示される列であるため、名詞・助詞等はそのまま利用できる場合が多く、既存の列の中では最も逐語訳に流用しやすい。また、本文の横に漢字の読みの列を追加する(図6)。

読み	本文	訳	品詞	活用の種類	活用形	注釈
	あはれ	哀れ	名詞-普通名詞-形状詞可能			
	なる	なり-断定	助動詞	文語助動詞	ナ	連体形-一般
	こと	事	名詞-普通名詞-一般			
	は	は	助詞-係助詞			
	、	、	補助記号-読点			
	おり	下りる	動詞-一般	文語上二段	ラ	連用形-一般
	おはしまし	おわします	動詞-非自立	文語四段	サ	連用形-一般

図6 列の調整完了

(2) 品詞情報の書き換え

「中古和文 UniDic」での品詞情報は、例えば「名詞-普通名詞-形状詞可能」など、「(大分類) - (中分類) - (小分類)」という形で細かく出力される。これらの細かい分類は高校の古典では必要ないことが多いため、見やすさも考慮して適宜削除する。Excelの置換を

用いれば一括で変換可能である。例えば「名詞-普通名詞-形状詞可能」を「名詞」だけ書き換えたい(「-普通名詞-形状詞可能」を削りたい)なら、検索文字列を「-普通名詞-形状詞可能」、置換文字列を空欄にすればよい(図7)。

	C	D	E	F	G
1	訳	品詞	活用の種類	活用形	注釈
2	哀れ	名詞-普通名詞-形状詞可能			
3	なり-断定	助動詞	文語助動詞-なり-断定	連体形-一般	
4	事	名詞-普通名詞-一般			
5	は	助詞-係助詞			
6	、	補助記号-読点			
7	下りる	動詞-一般	文語上二段-ラ行	連用形-一般	
8	おわします	動詞-非自立可能	文語四段-サ行	連用形-一般	
9	けり	助動詞	文語助動詞-ケリ	連体形-一般	
10	夜	名詞-普通名詞-一般			
11	は	助詞-係助詞			
12	、	補助記号-読点			
13	藤壺	名詞-固有名詞-一般			
14	の	助詞-格助詞			
15	上	名詞-普通名詞-副詞可能			
16	の	助詞-格助詞			
17	御	接頭辞			
18	局	名詞-普通名詞-一般			
19	の	助詞-格助詞			
20	小戸	名詞-普通名詞-一般			
21	より	助詞-格助詞			
22	出でる	動詞-一般	文語下二段-ダ行	未然形-一般	
23	させる	助動詞	文語下二段-サ行	連用形-一般	
24	給う-尊敬	動詞-非自立可能	文語四段-ハ行	連用形-一般	
25	けり	助動詞	文語助動詞-ケリ	連体形-一般	
26	に	助詞-接続助詞			

図7 不要な品詞情報の削除

名詞・動詞等の中分類以下や、「補助記号-読点」「補助記号-読点」「補助記号-括弧開」「補助記号-括弧閉」などを削除する。また、助詞に関しては「助詞-格助詞」などと出力されるが、助詞の場合は「格助詞」等のレベルでの判別が望ましいため、「助詞-」の方を削除する。活用の種類につく「文語」、活用形につく「-一般」を削除する。以上のような工程を経て、図7の状態から、不要な品詞情報を削除した後の状態が図8である。

読み	本文	訳	品詞	活用の種類	活用形	注釈
	あはれ	哀れ	名詞			
	なる	なり-断定	助動詞	助動詞-なり-断定	連体形	
	こと	事	名詞			
	は	は	係助詞			
	、	、				
	おり	下りる	動詞	上二段-ラ行	連用形	
	おはしまし	おわします	動詞	四段-サ行	連用形	
	ける	けり	助動詞	助動詞-ケリ	連体形	
	夜	夜	名詞			
	は	は	係助詞			
	、	、				
	藤壺	藤壺	名詞			
	の	の	格助詞			
	上	上	名詞			
	の	の	格助詞			
	御	御	接頭辞			
	局	局	名詞			

図8 不要な情報を削除した後の状態

以上は手作業と言っても、Excel の置換機能を使えば大した手間はかからない(4.3 参照)。あらかじめ削除すべき情報を知っている必要もなく、ファイルを上から眺めて、これは消そう、と思ったものを一括で消していく、という作業で十分対応できる。

また、「中古和文 UniDic」での品詞・活用形判定は、学校教育とやや異なるところがあるので、その点に関しては注意が必要である。学校教育に関わりそうな大きな違いは以下の2つである。1つは、学校教育での「形容動詞」は「形状詞」(語幹部分) + 断定の助動詞「なり」として出力されること、もう1つは、完了の助動詞「り」が下接する際の活用形が命令形として出力されること(学校教育では已然形扱いすることが多い)である。これらはいずれも手動で書き替えなければならないが、「形状詞」「命令形」を検索し、チェックして書き替えればよいだろう。ともに一般的なテキストであればたいした量は出ず、検索して逐一目視確認という作業は、作業量としては十分現実的である。

(3) エラー修正と、読み・訳の付与、注釈の記入

図 8 の状態にまでできれば、あとは「訳」の列を逐語訳に書き換えていき、適宜読みや注釈を自由に記入していけば、完成である。

本教材は「中古和文 UniDic」が自動で品詞分解し、品詞情報を与えた結果をもとに作成しているが、自動解析の結果はもちろん全て正しいとは限らない。機械が自動判別には一定数の誤りもあるので、解析エラーがどこにあるかの確認と修正は必要である。ただし、「中古和文 UniDic」は平安仮名文学作品に対しては高い解析精度を実現しており、平安仮名文学作品であれば、教育上どうしても直さなければならないエラーはさほど多くは生じない。また、高校の古典では平安仮名文学作品以外にも扱われるが、学校教科書での表記・仮名遣いであれば、平安期以外の資料でもある程度の精度が得られそうである(4.4 参照)。教材という性質上、エラーが無いかの確認は必須であるが、訳の付与をしながら同時に確認していけばたいがいエラーは見つけ出せる。(1) (2) に示した下準備を終えたのち、訳を付与しながら読み・注釈の適宜付与、エラーの確認と修正を同時進行で行う、というのが効率的であろう。

この他、「中古和文 UniDic」で品詞分解された言語単位(「短単位」と呼ばれる)は、学校教育の品詞分類と比べてやや短い、という点もないではない。例えば「花山寺」は「花山」/「寺」に分割して出力されるが、高校教育の感覚で品詞分解するなら「花山寺」で一語、というほうが一般的であろう。そこでこれらも修正し、もう少し長めの単位につないでいく、という工程もありうるが、このあたりの単位の切り方は高校教育の要点ではなく、解析結果のままにしておいても「間違い」というわけでもない。「これも直さなければ使えない」というのではなく、「余力があれば、直すとよりよくなる」という程度の問題だと位置づけられよう。この他、要注意箇所を色づけして目立たせたり、新たな列を追加して逐語訳ではない、最終的な「きれいな訳」を加えていく等、手の加え方は無限に拡張できるが、基本工程はここに記した(1)～(3)の作業だけで完了する(本稿末尾に作成した学習教材のサンプルを掲げる)。

以上、「中古和文 UniDic」を用いた教材作成の作業手順を紹介したが、このようにして作成が可能といっても、ここに示した作業を実際に行う場合には多大な時間がかかるというのであれば、この手法での教材作成は現実的とは言えまい。また、「中古和文 UniDic」の解析結果には、実際エラーがどのくらい出るのかも検証すべきである。エラーが余りに多く、逐一修正せねばならないようであれば、自動解析の意味がない。そこで最後に、作業時間および解析精度の検証を行った結果を簡単に紹介する。

4. 3 作業コストの検証

上記作業に実際どのくらいの時間がかかるのかを検証した。サンプルは『大鏡』より教科書によく採用される「花山天皇の出家」。タイトル、リード文、系図等も含んだ教科書の版組で4ページ分、教科書に掲載される1話分の分量としてはよく見られる量である。解析結果による語数は687語であった。これを、本文を冒頭から打ち込んでテキストデー

タを作成、「中古和文 UniDic」で形態素解析後、4.1 に示した作業工程に従ってサンプルを作成した。工程（3）に関しては、注釈の付与は用途・目的・作業者の意図次第でいくらかでも時間がかけられてしまうため、標準的な必須作業としてのエラー修正、読み・訳の付与までを行った。結果を表 1 に示す。

表 1 作業時間（『大鏡』より「花山天皇の退位」、687 語、教科書 4 ページ分）

テキストデータ作成	17分
形態素解析	1分
列の調整、品詞情報書き換え(工程(1)~(2))	5分
エラー確認・修正(工程(3))	16分
読み付与(工程(3))	8分
逐語訳付与(工程(3))	33分
総計	70分

テキストデータの作成（単純な打ち込み作業）に 17 分ほどかかっており、これを差し引いて、テキストデータさえあれば、逐語訳付与までの作業は 53 分で完了することが明らかになった。この後、必要に応じて注釈の記入等の作業を追加しうるし、もちろん作業に対する慣れ・不慣れによる差も、結果的な時間に反映されようが、教科書 4 ページ程度であれば、基本作業自体は 1 時間前後で完了するとみてよかろう。また、今回は検証のためエラー確認・修正と読み付与、逐語訳付与の 3 工程を別個に行ったが、実際には同時進行が可能な作業であり、作業に慣れたうえでこの工程を同時に行えば、さらなる時間短縮が見込める。

4. 4 教材作成目的から見た「中古和文 UniDic」の解析精度

中古和文 UniDic は、平安仮名文学作品に対しては高い解析精度を実現しており、中古和文 UniDic Ver0.5 の段階で、単位境界（品詞の切れ目が正しいか）で 99.3%、品詞認定で 97.8% という解析精度が報告されている（中古和文 UniDic ホームページほか）。ただしこれはあくまで平安仮名文学作品に対してのものであり、高校の古典学習においては、中世以降のテキストもさまざまに扱われるため、「中古和文」からやや離れたテキストに対しても実用に耐えるのかは、確認したいところである。そこで、サンプルデータとして作成した『大鏡』の外に、高校の教科書の定番であり、かつ、「中古和文」から外れるものとして『方丈記』『平家物語』を中古和文 UniDic を用いて解析し、解析精度を調査した。今回解析に使用したのは中古和文 UniDic Ver1.3(2012 年 12 月公開)である。

中古和文 UniDic での解析結果には、語種や発音形など、教材作成には使用しない情報も出力されるが、今回の調査では教材作成に使用する情報のみに関し、精度調査を行った。調査項目は単位境界、語彙素、品詞、活用型（活用の種類）、活用形の 5 つである。これらの項目に対し、教材作成にあたって修正すべき箇所がいくつあるかを調査した。要修正箇所として数えられた「誤り」の中では、UniDic の品詞体系・規程上は誤りでないもの（「形状詞」という品詞認定や、完了の助動詞「り」の上の活用形「命令形」など）も含むため、今回の精度調査はあくまで高校の教材作成という用途から見た精度調査であり、中古和文 UniDic の、形態素解析辞書そのものの精度調査とは異なるものである点、注意されたい。

結果は表 2 のとおり。『大鏡』で精度 98~99% 台であるのに対し、『方丈記』では 97~98% 台、『平家物語』では 94~96% 台と、「中古和文」から遠ざかるにつれ、精度がやや落ちるといった予想通りの結果となった。『平家物語』あたりになると、擬態語や音便形に対応できない場合が増えてくるようである。ただそれでもなお、教材作成の素材として十分利用できる精度であることは間違いなく、『平家物語』の教材作成も実際に行ったが、中古和文と比

べるとやや手間がかかったという程度差で、作成可能であった。学習者が読みやすいよう、程よく漢字表記が用いられ、仮名遣い等も正規化されている学校教科書のテキストは、形態素解析というシステムにとって相性がよく、学校教科書というのは形態素解析辞書の活躍の場のひとつと考えてよさそうである。

表2 各テキストにおける要修正箇所数と解析精度

		単位境界	語彙素	品詞	活用型	活用形	
大鏡 「花山天皇の出家」	総語数	要修正	10	11	10	10	3
	687	精度	98.5%	98.4%	98.5%	98.5%	99.6%
方丈記 「大火とつじ風」	総語数	要修正	14	14	13	12	17
	1394	精度	97.7%	97.7%	97.9%	98.1%	97.2%
平家物語 「木曾の最期」	総語数	要修正	50	72	58	48	46
	617	精度	96.4%	94.8%	95.8%	96.6%	96.7%

なお、表2の単位境界・語彙素・品詞等の要修正箇所の半数ほどは「形状詞」という品詞認定によるものであり、活用形の要修正箇所の3分の1ほどは完了の助動詞「り」の上の「命令形」である。これら UniDic と学校文法の差異による修正箇所は、規則的であり、検索・置換で簡単に修正できる。また、活用形の要修正箇所の残りの大部分は、文末の終止形・連体形の誤りである。要修正箇所の過半数は上記のタイプに尽くされるので、現実的な作業上は、①「形状詞」を検索してチェック、書き換え②「命令形」を検索し、下に助動詞「り」が続いていたなら已然形に書き換え、③文末の終止形／連体形は要確認、という3点さえ注意すれば、他の誤りはそれこそ「見れば気付く」というような個別の誤解析ばかりであり、気付かないようなところで細かい品詞情報が誤っている、というようなケースはほとんどない。上記①②③だけ気を付け、あとは本文を流し読みしてだけで、修正作業は完了させてよく、逐一個別に全ての項目に目を通さなくても問題ない精度である、というのが、実際に教材作成を試して得た感想である。

5. おわりに

以上、本稿では形態素解析辞書「中古和文 UniDic」の教育転用の在り方として、古典学習用の教材作成を紹介した。「中古和文 UniDic」は研究面でもさまざまな利用の可能性はあるはずであり、今後ともその可能性を探っていきたい。

末尾に「中古和文 UniDic」を利用して作成した教材サンプルを掲げる。

文 献

小木曾智信・小椋秀樹・田中牧郎・近藤明日子・伝康晴(2010)「中古和文を対象とした形態素解析辞書の開発」『情報処理学会研究報告 人文科学とコンピュータ』

Vol.2010-CH-85(No.4) pp.1-8

小木曾智信・小椋秀樹・近藤明日子・須永哲矢(2010)「形態素解析辞書「中古和文 UniDic」とその活用例」『日本語学会 2010 年度秋季大会予稿集』 pp.243-248

小椋秀樹・須永哲矢(2012)『中古和文 UniDic 短単位規程集』平成 21 (2009) - 平成 23 (2011) 年度科学研究費補助金基礎研究(C)「和文系資料を対象とした形態素解析辞書の開発」研究成果報告書 2 (課題番号 21520492、代表者 小木曾智信)

関連 URL

日本語歴史コーパス「中納言」 <http://maro.ninjal.ac.jp/>

中古和文 UniDic <http://www2.ninjal.ac.jp/lrc/index.php?UniDic>

資料：教材サンプル 『大鏡』より「花山天皇の出家」(一部)

読み	本文	訳	品詞	活用の種類	活用形	注釈
	あさましく	驚きあきれ	形容詞	形容詞-シク	連用形	
	候ひ	まし	動詞	四段-ハ行	連用形	丁寧→読者
	し	た	助動詞	助動詞-キ	連体形	
	こと	事	名詞			
	は	は	係助詞			
	、	、				
	人	人	名詞			
	に	に	格助詞			
	も	も	係助詞			
	知ら	知ら	動詞	四段-ラ行	未然形	
	せ	せ	助動詞	下二段-サ行	連用形	使役
	給は	なさら	動詞	四段-ハ行	未然形	尊敬→花山
	で	ないで	接続助詞			
	、	、				
	みそかに	密かに	形容動詞	ナリ活用	連用形	
	花山寺	花山寺	固有名詞			
	に	に	格助詞			
	御座まし	いらっしやっ	動詞	四段-サ行	連用形	尊敬→花山
	て	て	接続助詞			
	、	、				
	御	御	接頭辞			
すけ	出家	出家	名詞			
	入道	入道	名詞			仏門に入ること
	せ		動詞	サ行変格	未然形	
	させ	なさっ	助動詞	下二段-サ行	連用形	尊敬→花山
	給へ		動詞	四段-ハ行	已然形	尊敬→花山
	り	ってしまっ	助動詞	助動詞-リ	連用形	
	し	たの	助動詞	助動詞-キ	連体形	
	こそ	こそ(驚きだ)	係助詞			結びは省略。「あさましく候ひし」など。
	。	。(当時)				
	御	御	接頭辞			
	年	年(は)	名詞			
	十九	十九(歳)	数詞			