

## 中古和文を対象とした形態素解析辞書の開発

小木曾智信<sup>†</sup> 小椋秀樹<sup>†</sup> 田中牧郎<sup>†</sup> 近藤明日子<sup>†</sup>  
伝 康晴<sup>††</sup>

現在開発を行っている「中古和文 UniDic」を紹介する。これは平安時代の仮名文学作品を典型とする和文系の資料を対象とする形態素解析辞書であり、すでに公開中の「近代文語 UniDic」同様、日本語の歴史的資料の形態素解析を可能にするものである。

### Development of an Electrical Dictionary for Morphological Analysis of Classical Japanese

Toshinobu Ogiso<sup>†</sup> Hideki Ogura<sup>†</sup> Makiro Tanaka<sup>†</sup>  
Asuko Kondo<sup>†</sup> and Yasuharu Den<sup>††</sup>

In this paper, we present "Chuko-Wabun UniDic", which is an electrical dictionary for morphological analysis of classical Japanese. The dictionary is especially designed for the analysis of literary texts in the Heian period, and is an effective means for examining historical texts, like "Kindai-Bungo UniDic" for modern Japanese.

## 1. はじめに

現代日本語を対象として始まったコーパス日本語学は、その対象を歴史的研究にまで広げつつある。人間文化研究機構国立国語研究所では新たに通時コーパスを構築することとなっており、日本語の歴史的資料を対象とした自然言語処理技術が求められつつある。また、通時コーパスの完成を待たずとも、すでに古典文学作品をはじめとする多くの日本語の歴史的資料のテキストデータが集積されている。これらの資料を高度に活用するためにも、自然言語処理の基礎となる技術である形態素解析の必要性は高まっている。

なかでも、『源氏物語』をはじめとする中古和文は日本語の古典テキストの代表的なものであり、これを対象とした形態素解析は多くの研究者にとって利用価値の高いものとなるはずである。平安時代語の解析の試みとしては、和歌を対象とした山元(2007)などがあるものの、散文を含むテキストを対象とした形態素解析辞書は公開されていない。本発表では新たに開発した中古和文を対象とした形態素解析辞書「中古和文 UniDic」の概要を紹介する。

## 2. 形態素解析辞書 UniDic

### 2.1 UniDic の特長

現代語の形態素解析辞書の一つとして、発表者らを中心に、主として国立国語研究所で開発を行っている UniDic がある。UniDic は次のような日本語研究に適した特長をもった形態素解析辞書である。

1. 「短単位」という揺れが少ない齊一な単位を見出し語に採用している。
2. 語彙素・語形・書字形・発音形の階層構造を持ち、表記の揺れや語形の変異にかかわらず同一の見出しを与えることができる。
3. 話し言葉のテキストの解析に対応しているほか、アクセントや音変化の情報を付与することができ、音声処理の研究に利用することができる。

1., 2.により、単位の揺れや見出し付与の方法など、これまでの形態素解析辞書が抱えていた問題を解決している(伝ほか2007)。「短単位」については単位の認定方法を規程集に詳細に定め、揺れを防いでいる(小椋ほか2009)。

齊一な単位や階層化された見出し構造は、歴史的な資料を対象とした形態素解析辞書の作成時にもたいへん有用である。

<sup>†</sup> 国立国語研究所  
National Institute for Japanese Language and Linguistics

<sup>††</sup> 千葉大学文学部  
Faculty of Letters, Chiba University

揺れの少ない齊一な単位は、テキストの解析結果を用いた語彙比較を可能にする。古典文学作品についてはすでに多くの索引が作られているが、単位や見出し付与の方針の違いにより相互の比較が難しかった。UniDicをベースとすることで、作品間の比較を可能にするだけでなく、時代の違いをも超えた各種のテキスト間で相互に語彙を比較可能なデータベースを構築することが可能になる。

また、階層化された見出しを用いることで、文語形や旧字・旧仮名遣いの表記語を同一見出しの元にまとめることができるため、さまざまな時代のテキストに出現する語形・表記を統一的に扱うことができる。

## 2.2 近代文語 UniDic と中古和文

発表者らは、現代語用の UniDic をもとに、歴史的な資料を対象とした最初の形態素解析辞書として「近代文語 UniDic」を開発、公開した(小木曾・小椋・近藤 2008, 小木曾 2009)。これは主として近代の文語論説文(明治普通文)を対象とした解析辞書であり、文語の活用・歴史的仮名遣い・旧漢字・踊り字などに対応し、文語文を正しく解析することが可能になっている。これにより、「太陽コーパス」(国立国語研究所 2005)における文語記事など、近代文語文で書かれたテキストを解析して研究に利用することができるようになった。

近代文語 UniDic が対象とする近代文語文は、中古以来の漢文訓読文の流れにある。したがって、漢文訓読文に対しては、近代よりも古い時代の文章であっても、近代文語 UniDic を適用することで比較的良い解析結果を望むことができる。一方、中古以来の和文(仮名文)は近代文語文とは性質の異なるものである。例えば、漢文訓読文では漢語が多く用いられるのに対して、和文では和語がほとんどであり、和語であっても和文と漢文訓読文では使用する語彙が助動詞等にいたるまで大きく異なっている(築島 1963)。したがって、和文系の資料のためには、別の解析辞書を用意することが必要である。

和文系の資料は、中古の仮名文学作品以降、中世の擬古物語や近世・近代の擬古文に至るまで長い期間にわたって書かれており、これらの解析のために新たな解析辞書を作ることは十分に意味のある試みといえる。そこで、和文系資料を対象とした専用の形態素解析辞書「中古和文 UniDic」を新たに開発することとした。

## 3. 中古和文 UniDic の開発

### 3.1 形態素解析辞書作成の流れ

形態素解析を行うには、解析に用いる見出し語のリストに、当該語の出現しやすさ(生起コスト)、当該語と他の語・品詞との間のつながりやすさ(接続コスト)の情報

を付与した形態素解析用の辞書が必要となる。現在使われている主な形態素解析システムでは、生起コスト・接続コストを機械学習の方法によって統計的に取得しており、形態素解析辞書は図 1 のような流れで作成されることになる。

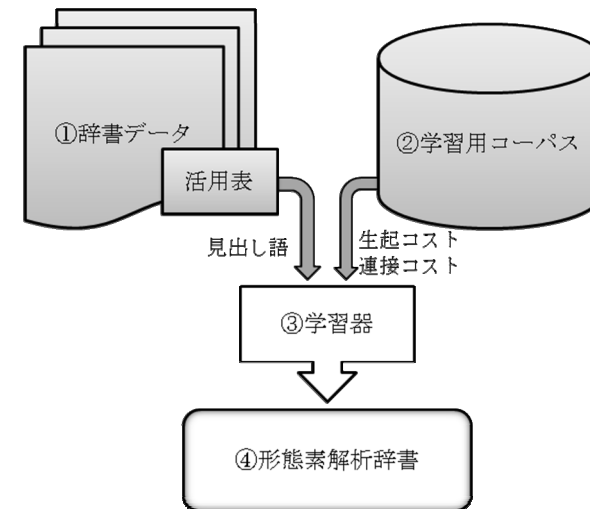


図 1 形態素解析辞書作成の流れ

形態素解析辞書を新たに作成するには、解析に用いる見出し語の一覧(①辞書データ)と、その見出し語を用いて文章に正しく情報を付与した文章のデータ(②学習用コーパス)が必要となる。辞書データと学習用コーパスから、プログラム(③学習器)によって形態素解析辞書が作られる。辞書データは活用表によって各活用形に展開できるようにしておく必要がある。できあがった形態素解析辞書は、辞書に応じた形態素解析器と組み合わせることでテキストの形態素解析が可能になる。

現代語用の UniDic は形態素解析器として ChaSen と MeCab に対応しており、ChaSen 版と MeCab 版の解析辞書を作成している。それぞれの解析器で用いられている統計モデルが異なっており(ChaSen は HMM, MeCab は CRF)、新しく開発された MeCab の方が一般に高い精度で解析を行うことができる。そのため、今回作成する中古和文 UniDic では MeCab 版のみを作成した。

### 3.2 データ管理システム

辞書データとコーパスの整合性を保ちつつ、辞書データへの見出し語追加やコーパスの修正作業を行っていくためには、データを管理するためのデータベースとデータ修正用のツールが必要である。このシステムとして、『現代日本語書き言葉均衡コーパス』の構築のために開発した形態論情報データベースを利用している。Microsoft SQL Server 2005 で構築したデータベースサーバと Microsoft Access で開発した修正用クライアントにより、ネットワークを通じた多人数による修正作業が可能となっており、このデータベースから解析辞書の作成に必要なデータを出力することができる(小木曾・中村 2009)。

### 3.3 辞書データの整備

中古和文の解析のためには、これまでに登録されていない見出し語を辞書データに追加する必要がある。しかし、一般的な語の多くはすでに UniDic に登録されているため、既存の国語辞書の見出し語などから追加していく方法はあまり効率的ではない。むしろ、送り仮名の違いによる異表記語など、実際に出現した形を登録することが効果的である。そこで、今回は見出し語をあらかじめ追加することをせず、学習用のコーパスに実際に出現し、未登録であるために正しく解析されなかった語を追加していく方法をとった。

新たに追加した語として、たとえば「あなり(あんなり<ありなり)」「べかめり(べかんめり<べかるめり)」のような撥音無表記形がある。これらは現代における古文の読みの慣習にあわせ、UniDic の見出し語構造を用いて、語彙素「有り」、語形「アン」(ラ行変格活用、連体形-撥音便)、書字形(表記)「あ」、発音形「アン」のように登録している。

見出し語の単位認定については、通時的な比較ができるようにするため、可能な限り現代語と共通の枠組みで処理を行ったが、語の用法変化などに伴い、やむを得ず扱いを変えたものがある。たとえば、指示詞について、現代語においては「この」「その」などは連体詞として扱っている。しかし「こ」「そ」が単独で指示代名詞として使われる中古語では、これらは代名詞+格助詞として扱った方が適切である。このほか、いくつかの副詞や接続詞についても同様の扱いとした。これらの点は、今後、中古和文向けの規定として文書化して整備していく予定である。

### 3.4 学習用コーパスの整備

学習用コーパスの整備は、典型的な中古和文として、作り物語・歌物語・日記文学から電子化テキストが入手可能なものを選び、これを既存の形態素解析辞書で解析したのち修正を行うという方法で行った。いずれのテキストも、校訂済みで漢字仮名交じりの読みやすい本文として整備されたものを利用している。

### (1) 学習用コーパスとした文学作品

中古和文 UniDic 0.3 の作成に利用した学習用コーパスを表 1 に示す。

このうち『源氏物語』はインターネット上のサイト「源氏物語の世界 再編集版(XML形式)」で公開されているデータによる。これは、渋谷栄一氏が公開している「源氏物語の世界」のデータを、宮脇文経氏が XML 形式に再編集したものである(2004, 2005 年度 IPA 未踏ソフトウェア創造事業)。表 1 の章割りもこのデータにしたがっている(「夕顔」の巻は全 4 章のうちの一部)。

また、『大和物語』は西端・木村(1994)、『土佐日記』『紫式部日記』は西端・木村・志甫(1996)に付属の本文テキストデータを利用している。

表 1 学習用コーパス

作品名		短単位語数
源氏物語	桐壺_1 章	1724
	桐壺_2 章	2189
	桐壺_3 章	2653
	帚木_1 章	3834
	帚木_2 章	4204
	帚木_3 章	4763
	空蝉_1 章	3046
	夕顔_1 章	1923
	夕顔_2 章	520
大和物語		26483
土佐日記		7936
紫式部日記(一部)		4771
合計		64046

### (2) 近代文語 UniDic の利用

学習用コーパスの初期の解析は、近代文語 UniDic をもとに、生起コスト等を調整した形態素解析辞書を作成して行った。近代文語文では漢語割合が高いため、たとえば「県(あがた)」は「県(けん)」と解析されてしまう。そこで、漢語のコストを上げ、和語のコストを相対的に下げることにより、「あがた」を優先させるようにしたものである。UniDic の見出し語には「語種」が付与されているためこのような調節をおこな

うことができる。こうした工夫により、当初より近代文語 UniDic より高い精度で中古和文を解析することが可能であった。

## 4. 中古和文 UniDic の解析精度

### 4.1 評価データ

表 1 の学習用コーパスのうち、約 1 割 (5653 語) を文単位でランダムサンプリングして精度評価用のデータとし、残りの 9 割を機械学習用のデータとした。したがって、評価データは学習に用いていないデータであるが、解析辞書にとって未知の語は含まれていないことになる。

### 4.2 解析精度

中古和文 UniDic 0.3 の解析精度を、現代語用の UniDic 1.3.12、近代文語 UniDic 1.1 で同一のデータで評価した結果とともに図 2・表 2 に示す。解析器は MeCab 0.98 を用いた。

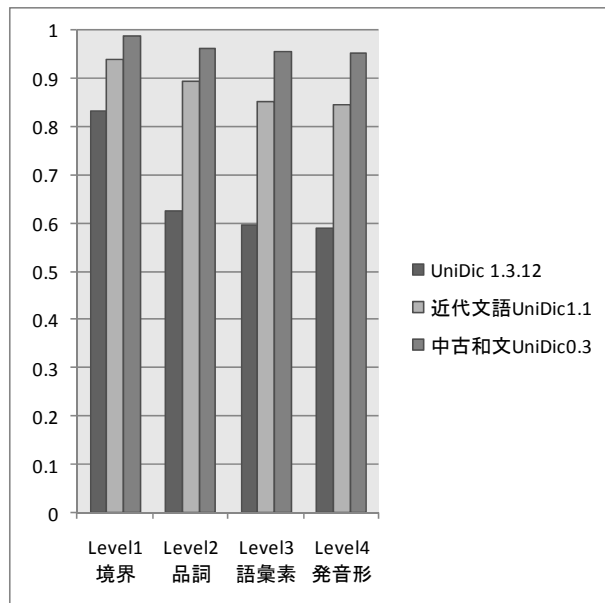


図 2 各種 UniDic による中古和文の解析精度の比較 (F 値)

表 2 各種 UniDic による中古和文の解析精度

		(現代語用) UniDic1.3.12	近代文語 UniDic1.1	中古和文 UniDic0.3
評価コーパス語数		5653	5653	5653
Level 1 境界認定	出力語数	6011	5670	5640
	正解語数	4856	5322	5580
	Recall	0.859012	0.941447	0.987086
	Precision	0.807852	0.938624	0.989361
	F 値	0.832646	0.940033	0.988221
Level 2 品詞認定	出力語数	6011	5670	5640
	正解語数	3650	5065	5432
	Recall	0.645674	0.895984	0.960905
	Precision	0.60722	0.893298	0.96312
	F 値	0.625856	0.894638	0.96201
Level 3 語彙素認定	出力語数	6011	5670	5640
	正解語数	3470	4814	5392
	Recall	0.613833	0.851583	0.953829
	Precision	0.577274	0.849029	0.956028
	F 値	0.594991	0.850303	0.954926
Level 4 発音形認定	出力語数	6011	5670	5640
	正解語数	3441	4787	5372
	Recall	0.608703	0.846807	0.950291
	Precision	0.57245	0.844268	0.952482
	F 値	0.59002	0.845535	0.951385

中古和文 UniDic の精度は発音形認定でも 95%を超えており、現代語の UniDic や近代文語 UniDic とは異なり、中古和文を高い精度で解析することが可能になっている。ただし、この数字は他の時代のテキストを対象とした辞書が、それぞれの目標とする

テキストを解析したときの精度（約 98%）と比較すると必ずしも十分とは言えない。

現在、中古和文 UniDic の学習用コーパスは約 6 万語に過ぎないが、近代文語 UniDic で約 40 万語、現代語 UniDic では 200 万語以上の学習用コーパスを用いている。また、中古和文 UniDic の学習用コーパスは、整備をはじめて日が浅いため誤りが少なくない。そのため、評価データに誤りが入り込み、これが見かけ上の精度を下げている。今後、見出し語の追加を行う一方で、学習用コーパスの量を増やすとともに質を高めて行く必要がある。

なお、現代語用の UniDic で中古和文を解析した場合の精度が極端に低いが、実際の解析結果を見るとこの数字よりはずっとよく解析できている印象を受ける。これは、文語の活用語を口語とは別の活用型としているために、動詞の四段活用→五段活用、形容詞の文語形容詞活用→形容詞型活用といった、事実上同形の語であっても全て誤りであると判定されているためである。先に述べた指示詞のように単位認定の方法を変えた語もあり、これらを正解と見なせば、もう少し精度は上がることになる。

#### 4.3 解析例・解析誤り例

中古和文 UniDic 0.3 の解析精度の参考資料として、精度評価時に確認した解析誤りの例を挙げる（付録 A）。この中には、実際には解析結果が正しいのに評価データが誤っていたため誤りと判定された事例も含んでいる（※印付きで掲載）。また、未知語を含むテキストの解析精度をはかる目安として、『竹取物語』の冒頭の実際の形態素解析結果を挙げる（付録 B）。

## 5. おわりに

中古和文 UniDic の開発により、平安時代の仮名文学などの形態素解析が可能になった。現代語の各ジャンルの書き言葉・話し言葉（現代語版 UniDic）、近代の文語文（近代文語 UniDic）に加えて、和文系資料の解析が可能になったことで日本語の多くの歴史的資料に形態素解析を施し、高度な利用へとつなげていくことが可能になりつつある。いずれの解析辞書も UniDic をベースとし、短単位という同じ単語認定基準にしたがっているため、通時的な比較も可能となっており、今後進めていく通時コーパスの構築に威力を発揮するはずである。

今後の課題として、中古和文の学習用コーパスを質・量ともに充実させて精度を向上させる必要がある。また、通時コーパスの構築に向け、他の時代・ジャンルの解析辞書を整備し、解析対象のテキストに最適な辞書を選択して形態素解析を行うことができるようにしていくことも課題の一つである。発表者らは、すでに歴史的仮名遣いの口語文を対象とした形態素解析辞書の開発を進めているが、今後、多ジャンルのテ

キストにきめ細かく対応するために、既存の解析辞書と少量のデータをもとに新しいジャンルのテキストを解析できるようにする領域適応の方法を探っていく必要がある。また、形態素解析結果の利用の面では、構文解析や意味解析などのより高度な自然言語処理技術の開発や、コーパス言語学的応用が期待される場所である。

**付記** 本発表は科研費・特定領域研究「日本語コーパス」および科研費・基盤研究 C（課題番号 21520492）による成果の一部を含むものである。

## 参考文献

- 1) 山元啓史（2007）「和歌のための品詞タグづけシステム」『日本語の研究』3 巻 3 号
- 2) 伝康晴・小木曾智信・小椋秀樹・山田篤・峯松信明・内元清貴・小磯花絵（2007）「コーパス日本語学のための言語資源：形態素解析用電子化辞書の開発とその応用」『日本語科学』22
- 3) 小椋秀樹・小磯花絵・富士池優美・原裕（2009）『現代日本語書き言葉均衡コーパス』形態論情報規程集改定版（LR-CCG-08-03）
- 4) 国立国語研究所（2005）『太陽コーパス—雑誌「太陽」日本語データベース』博文館新社
- 5) 小木曾智信・小椋秀樹・近藤明日子（2008）「近代文語文を対象とした形態素解析辞書・近代文語 UniDic」『言語処理学会第 14 回年次大会予稿集』pp.225-228
- 6) 小木曾智信（2009）『近代文語文を対象とした形態素解析のための電子化辞書の作成とその活用』国立国語研究所・科研費報告書 19720110
- 7) 築島裕（1963）『平安時代の漢文訓讀語につきての研究』東京大学出版会
- 8) 小木曾智信・中村壯範（2009）『現代日本語書き言葉均衡コーパス』形態論情報データベースの設計と実装』国立国語研究所内部報告書（LR-CCG-08-04）
- 9) 渋谷栄一「源氏物語の世界」 <http://www.sainet.or.jp/~eshibuya/>
- 10) 宮脇文経「源氏物語の世界 再編集版（XML 形式）」 <http://www.genji-monogatari.net/xml>
- 11) 西端幸雄・木村雅則・志甫由紀恵（1996）『平安日記文学総合語彙索引 土佐日記・蜻蛉日記・和泉式部日記・紫式部日記・更級日記』勉誠出版
- 12) 西端幸雄・木村雅則（1994）『歌物語総合語彙索引 伊勢物語・平中物語・大和物語』勉誠出版
- 13) 「形態素解析辞書 UniDic」 <http://download.unidic.org>
- 14) 「近代文語 UniDic」 <http://www.kokken.go.jp/lrc/index.php?Unidic>
- 15) 「MeCab: Yet Another Part-of-Speech and Morphological Analyzer」 <http://mecab.sourceforge.net/>  
工藤拓，京都大学情報学研究所-NTT コミュニケーション科学基礎研究所 共同研究ユニット
- 16) 「ChaSen - 形態素解析器」 <http://chasen-legacy.sourceforge.jp/>  
奈良先端科学技術大学院大学松本研究室

付録

付録 A 中古和文 UniDic 0.3 の解析誤り例 (評価データから抜粋)

※印は正解 (評価データ) に誤りがあったもの.

	出現形	語彙素読み	語彙素	品詞	活用型	活用形
正解※	言ふ	イウ	言う	動詞-一般	文語四段-ハ行	連体形-一般
	なる	ナリ	なり-断定	助動詞	文語助動詞-ナリ-断定	連体形-一般
出力	言ふ	イウ	言う	動詞-一般	文語四段-ハ行	終止形-一般
	なる	ナリ	なり-伝聞	助動詞	文語助動詞-ナリ-伝聞	連体形-一般
正解	浦戸	ウラド	ウラド	名詞-固有名詞-地名-一般	*	*
出力	浦戸	ウラト	ウラト	名詞-固有名詞-人名-姓	*	*
正解	頭	アタマ	頭	名詞-普通名詞-一般	*	*
出力	頭	トウ	頭	名詞-普通名詞-一般	*	*
正解	に	ナリ	なり-断定	助動詞	文語助動詞-ナリ-断定	連用形-ニ
出力	に	ニ	に	助詞-格助詞	*	*
正解	止む	ヤメル	止める	動詞-一般	文語下二段-マ行	終止形-一般
出力	止む	ヤム	止む	動詞-一般	文語四段-マ行	終止形-一般
正解※	雨降り	アメフリ	雨降り	名詞-普通名詞-一般	*	*
出力	雨	アメ	雨	名詞-普通名詞-一般	*	*
	降り	フル	降る	動詞-一般	文語四段-ラ行	連用形-一般
正解	学べ	マナブ	学ぶ	動詞-一般	文語四段-バ行	已然形-一般
出力	学べ	マナブ	学ぶ	動詞-一般	文語四段-バ行	命令形
正解	入れ	イル	入る	動詞-一般	文語四段-ラ行	已然形-一般
	二十	ニジュウ	二十	名詞-数詞	*	*
	一	イチ	一	名詞-数詞	*	*
	日	ニチ	日	接尾辞-名詞的-助数詞	*	*
出力	入れ	イレル	入れる	動詞-一般	文語下二段-ラ行	連用形-一般
	二十	ハツ	二十	名詞-数詞	*	*
	一日	ツイタチ	一日	名詞-普通名詞-副詞可能	*	*
正解※	集まり	アツマリ	集まり	名詞-普通名詞-一般	*	*
出力	集まり	アツマル	集まる	動詞-一般	文語四段-ラ行	連用形-一般
正解	渡り	ワタル	渡る	動詞-非自立可能	文語四段-ラ行	連用形-一般
出力	渡り	ワタリ	渡り	名詞-普通名詞-一般	*	*
正解	二	ニ	二	名詞-数詞	*	*
	月	ガツ	月	接尾辞-名詞的-助数詞	*	*

出力	二月	キサラギ	如月	名詞-普通名詞-一般	*	*
正解	朝	アサ	朝	名詞-普通名詞-副詞可能	*	*
出力	朝	アシタ	朝	名詞-普通名詞-一般	*	*
正解	日	カ	日	接尾辞-名詞的-助数詞	*	*
出力	日	ヒ	日	名詞-普通名詞-副詞可能	*	*
正解	御	ミ	御	接頭辞	*	*
出力	御	オン	御	接頭辞	*	*
正解	おほせ	オオス	仰す	動詞-一般	文語下二段-サ行	連用形-一般
	たぶ	タブ	給ぶ	動詞-非自立可能	文語四段-バ行	連体形-一般
出力	おほせ	オオセル	果せる	動詞-非自立可能	文語下二段-サ行	連用形-一般
	たぶ	タブ	給ぶ	動詞-非自立可能	文語四段-バ行	終止形-一般
正解	翁	オウ	翁	名詞-普通名詞-一般	*	*
出力	翁	オキナ	翁	名詞-普通名詞-一般	*	*
正解	明け	アケル	明ける	動詞-一般	文語下二段-カ行	未然形-一般
	ぬ	ズ	ず	助動詞	文語助動詞-ズ	連体形-一般
出力	明け	アケル	明ける	動詞-一般	文語下二段-カ行	連用形-一般
	ぬ	ヌ	ぬ	助動詞	文語助動詞-ヌ	終止形-一般
正解	おひ	オウ	生う	動詞-一般	文語上二段-ハ行	連用形-一般
出力	おひ	オウ	負う	動詞-一般	文語四段-ハ行	連用形-一般
正解	て	ツ	つ	助動詞	文語助動詞-ツ	未然形-一般
出力	て	テ	て	助詞-接続助詞	*	*
正解	染め	ソメル	染める	動詞-一般	文語下二段-マ行	連用形-一般
出力	染め	ソム	染む	動詞-一般	文語四段-マ行	命令形
正解※	思ひ	オモウ	思う	動詞-一般	文語四段-ハ行	連用形-一般
	かはし	カワス	交わす	動詞-非自立可能	文語四段-サ行	連用形-一般
出力	思ひかはし	オモイカワス	思い交わす	動詞-一般	文語四段-サ行	連用形-一般
正解	音	オト	音	名詞-普通名詞-一般	*	*
出力	音	ネ	音	名詞-普通名詞-一般	*	*
正解	かかる	カカリ	斯かり	動詞-一般	文語ラ行変格	連体形-一般
出力	かかる	カカル	掛かる	動詞-非自立可能	文語四段-ラ行	連体形-一般
正解	思ひいで	オモイイデル	思い出でる	動詞-一般	文語下二段-ダ行	未然形-一般
	なむ	ナム	なむ	助詞-終助詞	*	*
出力	思ひいで	オモイイデル	思い出でる	動詞-一般	文語下二段-ダ行	連用形-一般
	なむ	ナム	なむ	助詞-係助詞	*	*
正解	かたみ	カタミ	互	名詞-普通名詞-一般	*	*
出力	かたみ	カタミ	形見	名詞-普通名詞-一般	*	*

付録B 中古和文 UniDic 0.3 による形態素解析結果 (『竹取物語』冒頭)

※網掛け部分は解析誤り.

書字形	発音形	語彙素読み	語彙素	品詞	活用型	活用形	語種
				空白			記号
竹取	タケトリ	タケトリ	竹取	名詞-普通名詞-一般			和
物語	モノガタリ	モノガタリ	物語	名詞-普通名詞-一般			和
				空白			記号
今	イマ	イマ	今	名詞-普通名詞-副詞可能			和
は	ワ	ハ	は	助詞-係助詞			和
昔	ムカシ	ムカシ	昔	名詞-普通名詞-副詞可能			和
				補助記号-読点			記号
竹取	タケトリ	タケトリ	竹取	名詞-普通名詞-一般			和
の	ノ	ノ	の	助詞-格助詞			和
翁	オキナ	オキナ	翁	名詞-普通名詞-一般			和
と	ト	ト	と	助詞-格助詞			和
いふ	ユウ	イウ	言う	動詞-一般	文語四段-ハ行	連体形-一般	和
者	モノ	モノ	者	名詞-普通名詞-一般			和
あり	アリ	アル	有る	動詞-非自立可能	文語ラ行変格	連用形-一般	和
けり	ケリ	ケリ	けり	助動詞	文語助動詞-ケリ	終止形-一般	和
				補助記号-句点			記号
野山	ノヤマ	ノヤマ	野山	名詞-普通名詞-一般			和
に	ニ	ニ	に	助詞-格助詞			和
まじり	マジリ	マジル	混じる	動詞-一般	文語四段-ラ行	連用形-一般	和
て	テ	テ	て	助詞-接続助詞			和
竹	タケ	タケ	竹	名詞-普通名詞-一般			和
を	オ	ヲ	を	助詞-格助詞			和
取り	トリ	トル	取る	動詞-一般	文語四段-ラ行	連用形-一般	和
つつ	ツツ	ツツ	つつ	助詞-接続助詞			和
				補助記号-読点			記号
よろづ	ヨロズ	ヨロズ	万	名詞-普通名詞-副詞可能			和
の	ノ	ノ	の	助詞-格助詞			和
こと	コト	コト	事	名詞-普通名詞-一般			和
に	ニ	ニ	に	助詞-格助詞			和
使ひ	ツカイ	ツカウ	使う	動詞-一般	文語四段-ハ行	連用形-一般	和
けり	ケリ	ケリ	けり	助動詞	文語助動詞-ケリ	終止形-一般	和
				補助記号-句点			記号

名	ナ	ナ	名	名詞-普通名詞-一般			和
をば	オバ	ヲバ	をば	助詞-格助詞			和
				補助記号-読点			記号
さぬき	サヌキ	サヌキ	サヌキ	名詞-固有名詞-地名-一般			固
の	ノ	ノ	の	助詞-格助詞			和
造	ツクリ	ツクリ	作り	名詞-普通名詞-一般			和
と	ト	ト	と	助詞-格助詞			和
なむ	ナン	ナム	なむ	助詞-係助詞			和
いひ	イー	イウ	言う	動詞-一般	文語四段-ハ行	連用形-一般	和
ける	ケル	ケリ	けり	助動詞	文語助動詞-ケリ	連体形-一般	和
				補助記号-句点			記号
そ	ソ	ソ	其	代名詞			和
の	ノ	ノ	の	助詞-格助詞			和
竹	タケ	タケ	竹	名詞-普通名詞-一般			和
の	ノ	ノ	の	助詞-格助詞			和
中	ナカ	ナカ	中	名詞-普通名詞-副詞可能			和
に	ニ	ニ	に	助詞-格助詞			和
				補助記号-読点			記号
もと	モト	モト	元	名詞-普通名詞-一般			和
光る	ヒカル	ヒカル	光る	動詞-一般	文語四段-ラ行	連体形-一般	和
竹	タケ	タケ	竹	名詞-普通名詞-一般			和
なむ	ナン	ナム	なむ	助詞-係助詞			和
一筋	ヒトスジ	ヒトスジ	一筋	名詞-普通名詞-一般			和
あり	アリ	アル	有る	動詞-非自立可能	文語ラ行変格	連用形-一般	和
ける	ケル	ケリ	けり	助動詞	文語助動詞-ケリ	連体形-一般	和
				補助記号-句点			記号
怪し	アヤシ	アヤシイ	怪しい	形容詞-一般	文語形容詞-シク	語幹-一般	和
がり	ガリ	ガル	がる	接尾辞-動詞的	文語四段-ラ行	連用形-一般	和
て	テ	テ	て	助詞-接続助詞			和
				補助記号-読点			記号
寄り	ヨリ	ヨル	寄る	動詞-一般	文語四段-ラ行	連用形-一般	和
て	テ	テ	て	助詞-接続助詞			和
見る	ミル	ミル	見る	動詞-非自立可能	文語上一段-マ行	連体形-一般	和
に	ニ	ニ	に	助詞-接続助詞			和
				補助記号-読点			記号
筒	ツツ	ツツ	筒	名詞-普通名詞-一般			和
の	ノ	ノ	の	助詞-格助詞			和
中	ナカ	ナカ	中	名詞-普通名詞-副詞可能			和

光り	ヒカリ	ヒカル	光る	動詞-一般	文語四段-ラ行	連用形-一般	和
たり	タリ	タリ	たり-完了	助動詞	文語助動詞-タリ-完了	終止形-一般	和
。			。	補助記号-句点			記号
それ	ソレ	ソレ	其れ	代名詞			和
を	オ	ヲ	を	助詞-格助詞			和
見れ	ミレ	ミル	見る	動詞-非自立可能	文語上一段-マ行	已然形-一般	和
ば	バ	バ	ば	助詞-接続助詞			和
、			、	補助記号-読点			記号
三	サン	サン	三	名詞-数詞			漢
寸	スン	スン	寸	接尾辞-名詞的-助数詞			漢
ばかり	バカリ	バカリ	ばかり	助詞-副助詞			和
なる	ナル	ナリ	なり-断定	助動詞	文語助動詞-ナリ-断定	連体形-一般	和
人	ヒト	ヒト	人	名詞-普通名詞-一般			和
、			、	補助記号-読点			記号
いと	イト	イト	いと	副詞			和
うつくしう	ウツクシュー	ウツクシイ	美しい	形容詞-一般	文語形容詞-シク	連用形-ウ音便	和
て	テ	テ	て	助詞-接続助詞			和
ゐ	イ	イル	居る	動詞-非自立可能	文語上一段-ワ行	連用形-一般	和
たり	タリ	タリ	たり-完了	助動詞	文語助動詞-タリ-完了	終止形-一般	和
。			。	補助記号-句点			記号
翁	オキナ	オキナ	翁	名詞-普通名詞-一般			和
、			、	補助記号-読点			記号
言ふ	ユウ	イウ	言う	動詞-一般	文語四段-ハ行	連体形-一般	和
やう	ヨー	ヨウ	様	形状詞-助動詞語幹			漢
、			、	補助記号-読点			記号
「			「	補助記号-括弧開			記号
我	ワレ	ワレ	我	代名詞			和
、			、	補助記号-読点			記号
朝	アシタ	アシタ	朝	名詞-普通名詞-一般			和
ごと	ゴト	ゴト	毎	接尾辞-名詞的-一般			和
夕	ユウ	ユウ	夕	名詞-普通名詞-一般			和
ごと	ゴト	ゴト	毎	接尾辞-名詞的-一般			和
に	ニ	ニ	に	助詞-格助詞			和
見る	ミル	ミル	見る	動詞-非自立可能	文語上一段-マ行	連体形-一般	和
竹	タケ	タケ	竹	名詞-普通名詞-一般			和
の	ノ	ノ	の	助詞-格助詞			和
中	ナカ	ナカ	中	名詞-普通名詞-副詞可能			和
に	ニ	ニ	に	助詞-格助詞			和

おはする	オハスル	オハス	おわす	動詞-一般			和
に	ニ	ナリ	なり-断定	助動詞	文語サ行変格	連用形-ニ	和
て	テ	テ	て	助詞-接続助詞			和
、			、	補助記号-読点			記号
知り	シリ	シル	知る	動詞-一般	文語四段-ラ行	連用形-一般	和
ぬ	ヌ	ヌ	ぬ	助動詞	文語助動詞-ヌ	終止形-一般	和
。			。	補助記号-句点			記号
子	コ	コ	子	名詞-普通名詞-一般			和
と	ト	ト	と	助詞-格助詞			和
なり	ナリ	ナル	成る	動詞-非自立可能	文語四段-ラ行	連用形-一般	和
給ふ	タマウ	タマウ	給う-尊敬	動詞-非自立可能	文語四段-ハ行	終止形-一般	和
べき	ベキ	ベシ	べし	助動詞	文語助動詞-ベシ	連体形-一般	和
人	ヒト	ヒト	人	名詞-普通名詞-一般			和
な	ナン	ナリ	なり-断定	助動詞	文語助動詞-ナリ-断定	終止形-一般	和
めり	メリ	メリ	めり	助動詞	文語助動詞-メリ	終止形-一般	和
。			。	補助記号-句点			記号
」			」	補助記号-括弧閉			記号
と	ト	ト	と	助詞-格助詞			和
て	テ	テ	て	助詞-接続助詞			和
、			、	補助記号-読点			記号
手	テ	テ	手	名詞-普通名詞-一般			和
に	ニ	ニ	に	助詞-格助詞			和
うち	ウチ	ウチ	打ち	接頭辞			和
入れ	イレ	イレル	入れる	動詞-一般	文語下二段-ラ行	連用形-一般	和
て	テ	テ	て	助詞-接続助詞			和
、			、	補助記号-読点			記号
家	イエ	イエ	家	名詞-普通名詞-一般			和
へ	エ	エ	へ	助詞-格助詞			和
持ち	モチ	モツ	持つ	動詞-一般	文語四段-タ行	連用形-一般	和
て	テ	テ	て	助詞-接続助詞			和
来	キ	クル	来る	動詞-非自立可能	文語カ行変格	連用形-一般	和
ぬ	ヌ	ヌ	ぬ	助動詞	文語助動詞-ヌ	終止形-一般	和
。			。	補助記号-句点			記号