古文の用語索引作成に関する基礎的研究

○上田英代　上田裕一
今西祐一郎　樺島忠夫
仲川隆弘　村上征勝

Creating a Vocabulary Index for *Genji Monogatari Taisei*

○H. Ueda, Y. Ueda,
Y. Imanishi, T. Kabashima
T. Nakagawa, M. Murakami

Indexes of the words in Japanese literary classics are generally arranged in a-i-u-e-o order, under which headings the words are given with various prefixes, suffixes, inflections, and so forth. Thereafter the volume, page, and line numbers of each entry are given. Very occasionally, particles, auxiliary verbs, and other associated words are also listed.

Such indexes are useful in defining where and in what context each word of a classic is used. However, if you want to know whether some particular word is always used in the same context, you must check every page on which it appeares and note with what other words it appears in conjunction. This can become a very cumbersome process for words frequently used.

The current study aimed at creating a software program that could list up words (in boldface type) in a-i-u-e-o order, together with five to six preceding and the same number of following words that revealed the context. This was achieved using our previously completed full-text database of *Genji Monogatari Taisei* in which all words had been identified by parts of speech.

1）　用語索引作成の目的

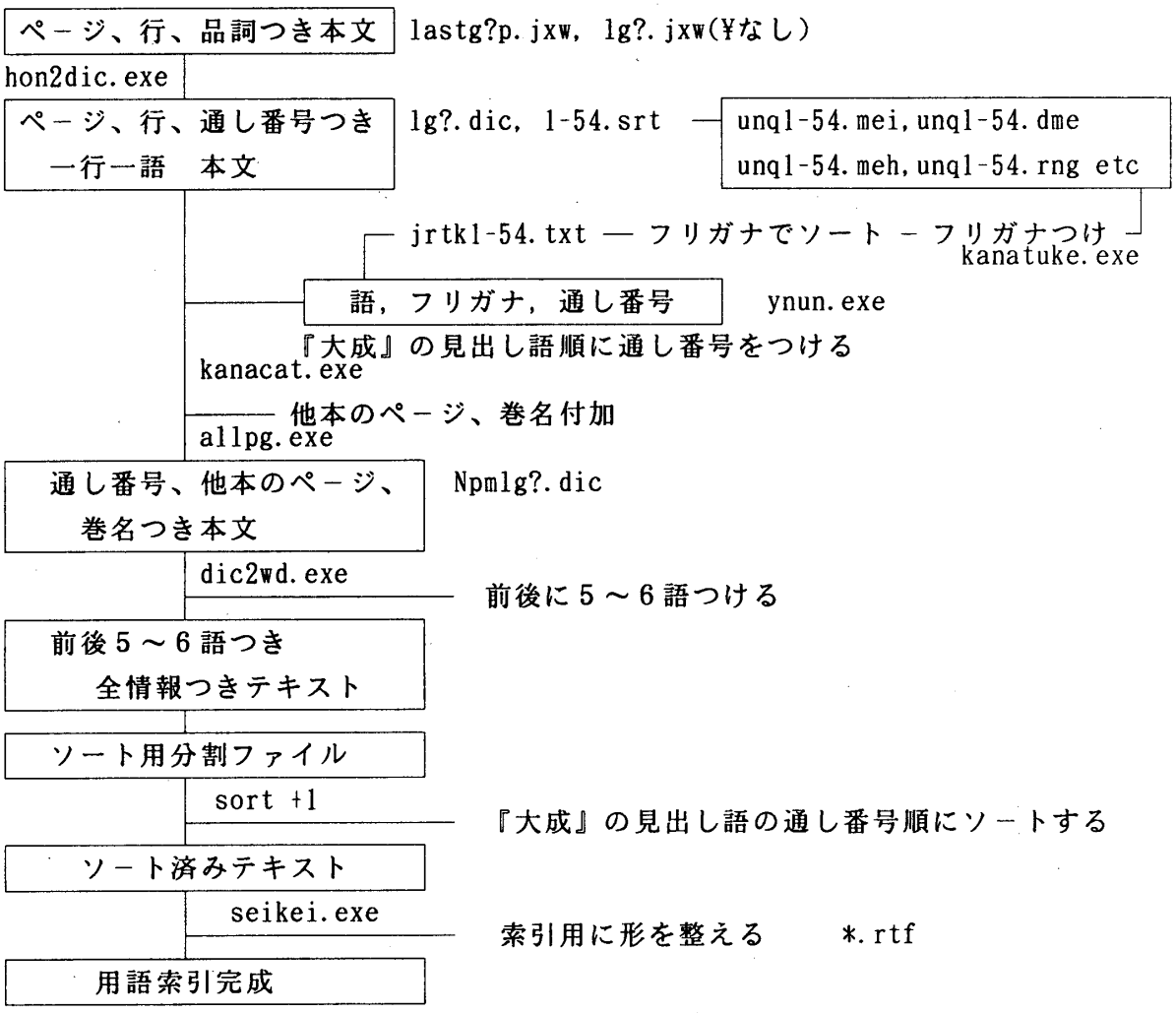
日本の古典索引の多くは、見出し語がアイウエオ順に並び、その下に接頭語や接尾語などが付いた形、活用語ならば活用語尾を付けた形を載せ、そこに本文の巻番号やページ番号、行番号などが出現順に配列されている。見出し語に助詞、助動詞などの付属語も付けた形で載せている索引も僅かながらある。

ある作品の本文中で、それぞれの語がどのような役割を担っているかを検討しようとするとき、索引は欠かせないものである。索引の利用方法は様々あるが、例えば使用頻度の高い語がどのような文脈の中で使われているのか、その使われ方は一定なのか、どんな特徴があるのかを調べようとするとき、従来の形式のものだとその本文のページや行にあたって、前後関係を含めて抜き書きしなければならない。使用頻度が高ければ高いほどその作業を何度もしなければならず、意味の類似した語や、正反対の意味の語なども含めて比較検討しようとすると、この作業はかなり繁雑なものとなる。

こうした作業を省力化し、見出し語を引くだけで前後関係も含めた本文が参照できれば非常に便利であるし、これがコンピュータ上で検索できれば更に高速化できる。筆者等は、『源氏物語大成』の品詞情報付きフルテキストデータベースを利用して、すべての語に前後それぞれ５～６語の文脈を付加したデータを作り、その語をアイウエオ順に並べて、用語総索引（ＫＷＩＣ）を作成することにした。

２）作成の手順と作業

作成の手順を流れ図で表すと以下のようになる（図１）。



（図１）

１．『源氏物語大成』に出現する語の一つ一つにページ、行、行中順番号を付ける。本文をこの形式にしたのは、語単位に様々の情報を付加したのち、前後５～６語の文脈をつけて本文形式に戻す時に必要だからである（図２）。

２．必要情報の付加

この語一つ一つに『大成』の見出し語（項目）の下にあげられた用例に通し番号を付け、その通し番号、および他系列本文の参考ページ、巻名等をつける（図３）。

3．すべての語の前後に5～6語ずつ付加する（図4）。

```
0387-01 00010 ひとしれぬ 連語 0
0387-01 00020 御心つから 副詞 0
0387-01 00030 の 助詞 0
0387-01 00040 物おもはしさ 名詞 0
0387-01 00050 は 助詞 0
0387-01 00060 いつとなき 形容 0
0387-01 00070 こと 名詞 0
0387-01 00080 な 助動 0
0387-01 00090 めれ 助動 0
0387-01 00100 と 助詞 0
0387-01 00110 かく 副詞 0
0387-01 00120 おほかた 名副 0
0387-02 00010 の 助詞 0
0387-02 00020 よ 名詞 0
0387-02 00030 に 助詞 0
0387-02 00040 つけ 連語 0
0387-02 00050 て 助詞 0
0387-02 00060 さへ 助詞 0
0387-02 00070 わつらはしう 形容 0
0387-02 00080 おほしみたるる 動詞 0
0387-02 00090 こと 名詞 0
0387-02 00100 のみ 助詞 0
0387-02 00110 まされ 動詞 0
0387-02 00120 は 助詞 0
0387-03 00010 もの心ほそく 形容 0
0387-03 00020 世中 名詞 0
0387-03 00030 なへて 副詞 0
0387-03 00040 いとはしう 形容 0
0387-03 00050 おほしなら 動詞 0
0387-03 00060 るる 助動 0
```

（図2）

```
0387-01 00010 ひとしれぬ 連語 0 1-417 2-415 散
0387-01 00020 御心つから 副詞 0 1-417 2-415 散
0387-01 00030 の 助詞 0 1-417 2-415 散
0387-01 00040 物おもはしさ 名詞 0 1-417 2-415 散
0387-01 00050 は 助詞 0 1-417 2-415 散
0387-01 00060 いつとなき 形容 0 1-417 2-415 散
0387-01 00070 こと 名詞 0 1-417 2-415 散
0387-01 00080 な 助動 0 1-417 2-415 散
0387-01 00090 めれ 助動 0 1-417 2-415 散
0387-01 00100 と 助詞 0 1-417 2-415 散
0387-01 00110 かく 副詞 0 1-417 2-415 散
0387-01 00120 おほかた 名副 0 1-417 2-415 散
0387-02 00010 の 助詞 0 1-417 2-415 散
0387-02 00020 よ 名詞 0 1-417 2-415 散
0387-02 00030 に 助詞 0 1-417 2-415 散
0387-02 00040 つけ 連語 0 1-417 2-415 散
0387-02 00050 て 助詞 0 1-417 2-415 散
0387-02 00060 さへ 助詞 0 1-417 2-415 散
0387-02 00070 わつらはしう 形容 0 1-417 2-415 散
0387-02 00080 おほしみたるる 動詞 0 1-417 2-415 散
0387-02 00090 こと 名詞 0 1-417 2-415 散
0387-02 00100 のみ 助詞 0 1-417 2-415 散
0387-02 00110 まされ 動詞 0 1-417 2-415 散
0387-02 00120 は 助詞 0 1-417 2-415 散
0387-03 00010 もの心ほそく 形容 0 1-417 2-415 散
0387-03 00020 世中 名詞 0 1-417 2-415 散
0387-03 00030 なへて 副詞 0 1-417 2-415 散
0387-03 00040 いとはしう 形容 0 1-417 2-415 散
0387-03 00050 おほしなら 動詞 0 1-417 2-415 散
0387-03 00060 るる 助動 0 1-417 2-415 散
```

（図3）

```
ひとしれぬ/御心つから/の/ ○○○○○  0387-01 00040 物おもはしさ 0 /は/いつとなき/こと/ 1-417 2-415 散
物おもはしさ/は/いつとなき/ ○○○○○ 0387-01 00070 こと 0 /な/めれ/と/ 1-417 2-145 散
いつとなき/こと/な/めれ/と/かく/おほかた/の/ ○○○○○ 0387-02 00020 よ 0 /に/つけ/て/ 1-417 2-145 散
よ/に/つけ/て/さへ/わつらはしう/おほしみたるる/ ○○○○○ 0387-02 00090 こと 0 /のみ/まされ/は/ 1-417 2-145 散
こと/のみ/まされ/は/もの心ほそく/ ○○○○○ 0387-03 00020 世中 0 /なへて/いとはしう/おほしなら/ 1-417 2-145 散
いとはしう/おほしなら/るる/に/さすかなる/ ○○○○○ 0387-03 00090 こと 0 /おほかり/ 1-417 2-145 散
おほしなら/るる/に/さすかなる/こと/おほかり/ ○○○○○ 0387-04 00010 れいけいてん 0 /と/きこえ/し/ 1-417 2-145 散
こと/おほかり/れいけいてん/と/きこえ/し/は/ ○○○○○ 0387-04 00060 宮たち 0 /も/おはせ/す/ 1-417 2-145 散
れいけいてん/と/きこえ/し/は/宮たち/も/おはせ/す/ ○○○○○ 0387-04 00100 院 0 /かくれ/させ/たまひ/ 1-417 2-145 散
宮たち/も/おはせ/す/院/かくれ/させ/たまひ/て/ ○○○○○ 0387-04 00150 のち 0 /いよいよ/あはれなる/御ありさま/ 1-417 2-145 散
かくれ/させ/たまひ/て/のち/いよいよ/あはれなる/ ○○○○○ 0387-05 00020 御ありさま 0 /を/たた/この/ 1-417 2-145 散
あはれなる/御ありさま/を/たた/この/ ○○○○○ 0387-05 00060 大将殿 0 /の/御心/に/ 1-417 2-145 散
あはれなる/御ありさま/を/たた/この/大将殿/の/ ○○○○○ 0387-05 00080 御心 0 /に/もてかくさ/れ/ 1-417 2-145 散
もてかくさ/れ/て/すくし/たまふ/なる/へし/ ○○○○○ 0387-06 00010 御をとうと 0 /の/三のきみ/うちわたり/ 1-417 2-145 散
すくし/たまふ/なる/へし/御をとうと/の/ ○○○○○ 0387-06 00030 三のきみ 0 /うちわたり/に/て/ 1-417 2-145 散
たまふ/なる/へし/御をとうと/の/三のきみ/ ○○○○○ 0387-06 00040 うちわたり 0 /に/て/はかなう/ 1-417 2-145 散
```

（図4）

4．索引形式に整える

1．2．3．までの作業は、巻毎に行なう。前後の本文をつけた巻毎のテキストから、『大成』の見出し語の通し番号順にソートすると、同じ語が出現順に集まって来る。

3）作成過程での問題点

1．『大成』の索引は、見出し語がひらがなでアイウエオ順に並べられ、本文中で漢字混じりで表記されていても、その表記は示されていない。即ち同じ語に異表記があるのかないのか、あるとすればどんな形のものが何個かなどということはわからない。また、コンピュータで機械的にソートすると同じ語であっても、漢字混じりの語や漢字から始まっている語はひらがな表記の語の後のほうに並んでくるし、同表記異義語がある場合も、表記が同じならば異なる語でも出現順に並んでくる。この問題は見出し語をどういう順序で並べてゆくかの問題とも重なってくる。即ち、語の意味を重視して同じ語を一箇所にまとめて並べるか、表記のみを重視して類似した形の語のまとまりで並べるかの問題である。

同じ語で異表記の例　　（表１）

見出し語	フリガナ	本文中に出現する形
おほとの	オホトノ	おほとの，おほ殿，大との，大殿
かくもん	ガクモン	かくもむ，御かくもん，御かくもむ
こころ	ココロ	こころ，心，御心，御心とも
みきのおとと	ミギノオトト	右のおとと，右大臣
みきのおほいとの	ミギノオホイトノ	みきの大殿，みきのおほゐとの， 右のおほいとの，右のおほい殿， 右の大との，右の大殿，右大殿， 右大との
みきのおほとの	ミギノオホトノ	みきの大殿，右のおほとの，右の大殿 右の大との，右のおほ殿，右大との

同じ表記で意味の異なる語　　（表２）

見出し語	フリガナ	意味の違い
あし	アシ	悪し，足，葦
あかし	アカシ	明石，明かし，赤し
こと	コト	事，言，琴

2．　今回は、同じ語は一箇所に集め、本文に出現する順に並べることにした。そのため、『大成』の見出し語の下で同じ語を示すための通し番号をつけ、番号順にソートして同じ語で異表記のもの、複数形、接頭語接尾語がついたものを一箇所に集めた。この番号を正確につけるために、番号付け用辞書をつくった。各品詞ごとに５４帖の異なり語集を作り、同じ語であれば複数形や接頭語接尾語のついたものにも同一フリガナをつけ、フリガナ部分でソートし一箇所に集めた。同じフリガナでも意味が異なる語があるときは、同じ語であることを示すための通し番号を、異なる意味の数だけつける。この番号付け用辞書を使って各巻のページつき一行一語形式本文の語に通し番号をつけた。すべての情報が付いた語に前後５～６語ずつの文脈を付加する。この文脈付き語を番号順にソートすると『大成』の見出し語順に語が並んでくる。このようにして『大成』の見出し語順の用語索引（ＫＷＩＣ）ができる。

この用語索引では、同じ接頭語が付いた語を調べたり、複合した語の後部の語が同じものを比較することはできないので、巻末に機械的に表記の形でソートした表と、語の後ろからソートした逆引き用の表とを添付するつもりである。