

兼永本古事記・出雲国風土記抄データベースの構築

Construction of “Kanenagabon Kojiki · Izumonokuni Fudokisyo-Database”

松本 智子

Tomoko Matsumoto

国文学研究資料館 研究情報部

東京都品川区豊町1-16-10

National Institute of Japanese Literature

Yutakamachi 1-16-10, Shinagawa-ku, Tokyo 142-8585, Japan

あらまし：

『兼永本古事記・出雲国風土記抄』データベースの構築過程を報告する。これは、国文学研究資料館原本テキストデータベース事業の成果であり、これまでも諸作品のデータベース化を手掛けているが、今回のデータベースでは、上代の文献に見られる表記の多様性を考慮し、特に固有名詞についてはシソーラスを設け検索の便に備えた。データベース・シソーラスともに利用者参加型の形態を目指して作成されたものである。

Summary：

I will report the method for proceeding construction of “Kanenagabon Kojiki · Izumonokuni Fudokisyo”-database. This is a result of the project at the National Institute of Japanese literature. In this database, we made thesaurus, allowing for diverse expression with same meaning in ancient literature. This database and thesaurus was made with the aim of the participation model.

キーワード：兼永本古事記、出雲国風土記抄、データベース、シソーラス

Keywords：kanenagabonkojiki, izumonokunifudokisyo, database, thesaurus

1. はじめに

国文学研究資料館では、原本テキストデータベース事業の成果を「国文学研究資料館データベース古典コレクション」として、これまで『源氏物語（絵入）』『二十一代集』『吾妻鏡』『歴史物語（栄花物語・大鏡・今鏡・水鏡・増鏡）』を、そして今回紹介する『兼永本古事記・出雲国風土記抄』をいずれもCD-ROMの形で岩波書店から刊行してきた。

この事業の目的は、信頼できる古典作品のデータベースを、研究者をはじめ学生や一般読者に広く提供することを第一とする。同時に、本データベースは、全データの閲覧・再利用が可能であり、データ記述仕様を公開する開放型のデータベースとして、また、利用者によるデータの改変・追加等の編集、任意の作品データを追加登録しての利用を許容する参加型のデータベースとして企画されており、研究補助ツールとし

て利用することが十分可能である。

本データベース事業は、原則として3年からなる基本作業の後、4年目に公開という形をとっている。『兼永本古事記』『出雲国風土記抄』も、①原本からのテキスト入力・②専門の研究者による監修作業・③総監修（各監修員間の調整及びデータベースとしての仕上げ）といった作業を2000年から1年ずつ行い、4年目にあたる今年3月の刊行に至った。

今回は『兼永本古事記』『出雲国風土記抄』データベースの構築に関わった一人として、その構築過程を報告したい。また、本データベースの特徴といえる固有名詞一覧（シソーラス）の作成とその展開についても述べたいと思う。

シソーラスの作成は、国文学研究資料館データベース室と、監修員として今回の事業に携わってくださっ

た上代文学の専門家とのコラボレーションによって完成した。このことは、研究が個人による研究に特化されがちな国文学界において注目すべき研究成果のひとつと言える。

2. 底本について

・『兼永本古事記』

『古事記』は、和銅5（712）年に成立した日本最古の典籍である。3巻からなり、上巻は神話、中巻は神武天皇から応神天皇までの時代、下巻は仁徳天皇から推古天皇までの時代で構成される。

兼永本は『古事記』諸本のうちト部系とされる1本であり、本文決定において真福寺本と同じく最も重要な伝本である。成立は不明であるが、諸資料により大永二年が基点と考えられている。現在翻刻された諸テキストの多くは真福寺本を底本とするが、その真福寺本と同様に『古事記』の本文として重要な兼永本のデータベース化は、『古事記』原本に迫るための資料提供として大きな意味を持っている。

本データベースでは鈴木鹿仁氏所蔵の貴重な写本を底本として用い、全冊のカラー画像を提供している。



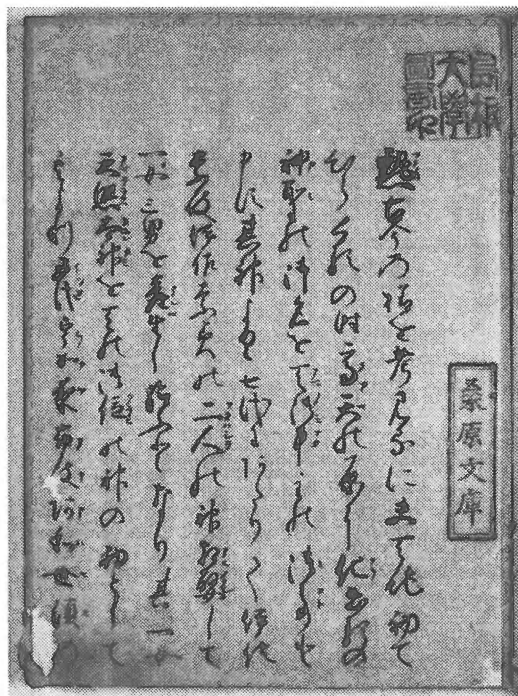
〈『兼永本古事記』第一冊 一丁表の画像〉

・『出雲国風土記抄』

『出雲国風土記抄』は、天平5（733）年成立した地誌『出雲風土記』の最初の注釈書であり、天和3（1683）年に岸崎時照が著したものである。地名の詳しい考証に特徴があり、今日なお省みられるべき点が多い。本データベースで使用した島根大学図書館

桑原文庫蔵本は伝本のうち三番目に古い本文を持つ。

このような資料的価値の高い島根大学図書館桑原文庫蔵の2冊本・4冊本を底本として用い、全冊のカラー画像を提供する。4冊本の注釈部分および2冊本文はこれまで未紹介のものであり、本データベースでの本書全体の紹介は大きな意味を持っていると言えるであろう。



〈『出雲国風土記抄』第一冊 一丁表の画像〉

3. 検索システムについて

古典コレクションデータベースは、古典作品のデータ（マスターデータ）とそのデータを検索利用するためのシステム（散文検索システム・和歌検索システム）で構成される。本データベースでは散文検索システムを利用。

マスターデータは次の4つの基本領域から成っており、散文検索システム上での各領域とも自在な閲覧・検索が可能である。各領域の詳しい内容は次頁の〈データベースの基本的な構成〉参照。

1. 本文領域
2. 標準化領域
3. 注記領域
4. メモ領域

〈データベースの基本的な構成〉

¥A …	作品名	古事記・出雲風土記抄（四冊本）・出雲風土記鈔（二冊本）
¥T …	大見出し等	『兼永本古事記』の場合は「序」「神代」「歴代」、『出雲国風土記抄』の場合は郡名が巻名となる。
¥M …	小見出し	『兼永本古事記』の場合は「別天神五柱」など、『出雲国風土記抄』の場合は郡を構成する地名などが小見出しとなる。巻名の下位項目に見出しにあたるものがない場合は「小見出しなし」とする。
¥B…	本文領域	底本本文をできるだけ忠実に翻字する領域。訓点・連辞符・傍記を除いた白文の形で翻字する。誤字もそのままとする。
V…	冊数名	その巻が底本の何冊目にあたるかをV 0 1のかたちで示す。
P…	頁数	1丁表をP 0 0 1、同裏をP 0 0 2というように3桁で表現する。
L…	行数	L 0 1というように2桁で表現する。
¥C…	標準化領域	本文領域の漢字を書き下し文に整えた領域。『兼永本古事記』では、真福寺本との主要な異同を示している。また、両作品とも必要に応じて意味内容を整えるための校訂を行った。
V…	冊数名	¥BのVに同じ。
P…	頁数	¥BのPに同じ。
L…	行数	¥BのLに同じ。
¥D…	注記領域	底本にある傍記を書き込む領域。漢字の振り仮名にあたる部分の仮名（付訓）や注記などを底本のまま翻字する。注記的な傍記には#印を付け、振り仮名と区別した。
V…	冊数名	¥BのVに同じ。
P…	頁数	¥BのPに同じ。
L…	行数	¥BのLに同じ。
¥E…	メモ領域	校異や注釈等を書き込む領域。本データベースでは、主要注釈書等の頁情報を付加している。
V …	冊数名	¥BのVに同じ。
P …	頁数	¥BのPに同じ。
L …	行数	¥BのLに同じ。

マスタデータおよび検索結果はファイル出力が可能であり、利用者の自在な利用に対応している。そして、最たる特徴は参加型のデータベースであるということである。これまで市販されているデータベースは、既存データの改変など利用者の介入できる仕様になっていないのが一般的である。しかし、本データベースは、利用者がデータを成長させていく参加型のデータベースを提供するという立場に立ち、システム自体がそれを可能にする仕様となっている。すなわち、本システムには、標準・注記・メモの各領域において「編集」機能が付いており、利用者自らがデータを編集することが可能である。参加型のデータベースという所以のひとつがここにあると言える。また、

『兼永本古事記』『出雲国風土記抄』に加え、たとえば利用者が『日本書紀』を同時に検索できるようにしたいと思えば、『日本書紀』データを本データベースのフォーマットに即して作成し、マスタ登録を行えば、本システム上で利用可能となり、横断検索もできるようになる。実際に多種多様な本文をマスタにして同時に検索することが簡単にできるような支援システムもある。

〈合字について〉

シフト J I S 第一水準・第二水準の範囲内に収まらない字体については、「■」とし、その後 | | の中に偏傍等を記す形でもとの字体を示した。篇傍を＋、

冠脚を／等の記号で示した。

上代の文献には特殊な文字が多く使われており、Uniコードでの入力も不可能である。そういった文字を篇や旁などの部品でひけるようにするための工夫である。

4. 完成マスターデータ

〈完成マスターデータ（古事記）の一部〉

¥A古事記

¥T序

¥M小見出しなし

¥B本文領域

V01

P001

L01古事記上巻〈并序〉

L02臣安萬侶、言、夫混元、既凝、氣象、未効、無

L03名、無為、誰知其形、然、乾坤、初分、參神作

L04造化之首、陰陽斯開、二靈為群品之祖、

L05所以、出入幽顯、日月彰於洗目、浮沈海

L06水、神祇呈於滌身、故、太素杳冥、固本教、

L07而識孕土産嶋之時、元始綿■〔之＋（壘一土）〕

、頼先聖、

L08而察生神立人之世、寔懸鏡吐珠、而

〈中略〉

¥C標準領域

V01

P001

L01古事記ノ上巻〈并ハセテ序〉

L02臣安万侶、言ス、夫レ混元、既ニ凝リテ、氣象、未ダ*効レズ（*敦）、

L03名モ無ク、為モ無シ、誰カ其ノ形ヲ知ラム、然レドモ、乾坤、初メテ分レテ、參神

L04造化ノ首メヲ作シ、陰陽斯ニ開ケテ、二靈群品ノ祖ト為ル、

L05所以ニ、幽顯ニ出入シテ、日月ヲ洗フニ彰ハル、海

L06水ニ浮ヒ沈ミテ、神祇身ヲ滌クニ呈ル、故、太素ハ杳冥レドモ、本教ニ因リテ、

L07土ヲ孕ミ島ヲ産ム時ヲ識リ、元始ハ綿■〔之＋貌〕ナレドモ、先聖ニ頼リテ、

L08神ヲ生ミ人ヲ立テル世ヲ察ル、寔ニ知ル鏡ヲ懸ケ珠ヲ吐キテ、

〈中略〉

¥D注記領域

V01

P001

L01（3）△ノ／（7）アハセテ

L02（2）ヤスマロ／（6）マウス／（8）ソレ

L03（8）ノ／（16）△テ

L04（4）ハシメヲ

L05

L06

L07

〈中略〉

¥Eメモ領域

V01

P001

L01新全集【16】集成【17】思想大系【10】
校異【并序：序并】

L02新全集【16】集成【17】思想大系【10】
校異【效：敦】

L03新全集【16】集成【17】思想大系【10】

L04新全集【16】集成【17】思想大系【10】

L05新全集【16】集成【17】思想大系【10】

L06新全集【16】集成【17】思想大系【10】

校異【呈：里】【杳：杳】【固：目】注記【「杳」を真福寺本により「杳」と改める。】【「固」を道祥本により「因」と改める。】

L07新全集【16】集成【18】思想大系【10】

注記【「■〔之＋（壘一土）〕」を寛永版本により「■〔之＋貌〕」と改める。】

—以下略—

5. 検索システムでの利用

〈特徴〉

- ・本文の閲覧検索（不一致文言検索可能）
- ・表示されたテキストに対応する部分の画像がボタン1つで参照できる。
- ・マスターデータ・検索結果を出力して利用可能
- ・検索システム上での標準・注記・メモの各領域のデータの変更・追加が可能であり、利用者自身によりデータ編集ができる。
- ・マスター仕様に従ってデータを作成することにより、任意の作品を検索システム上で利用できる。たくさんの作品を登録して利用すると、さながら電子化された図書館ようになる。

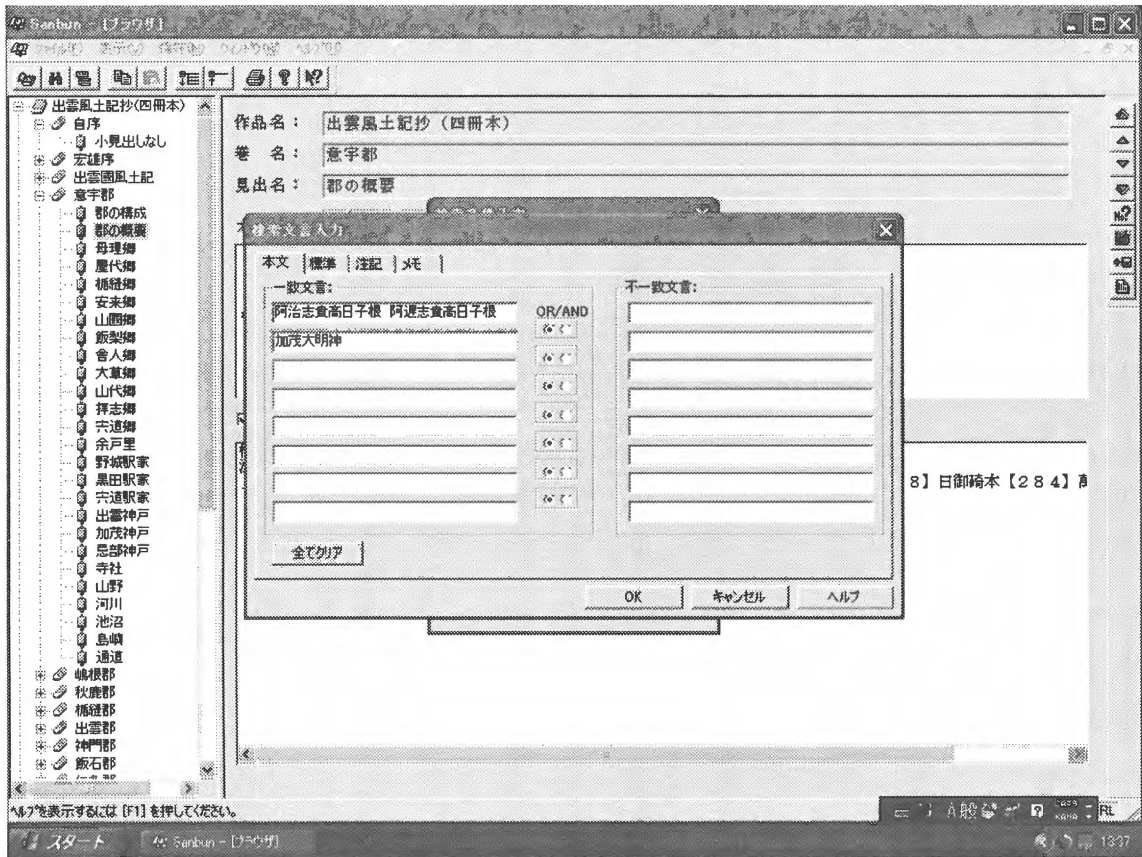
〈ブラウザを開いた画面〉



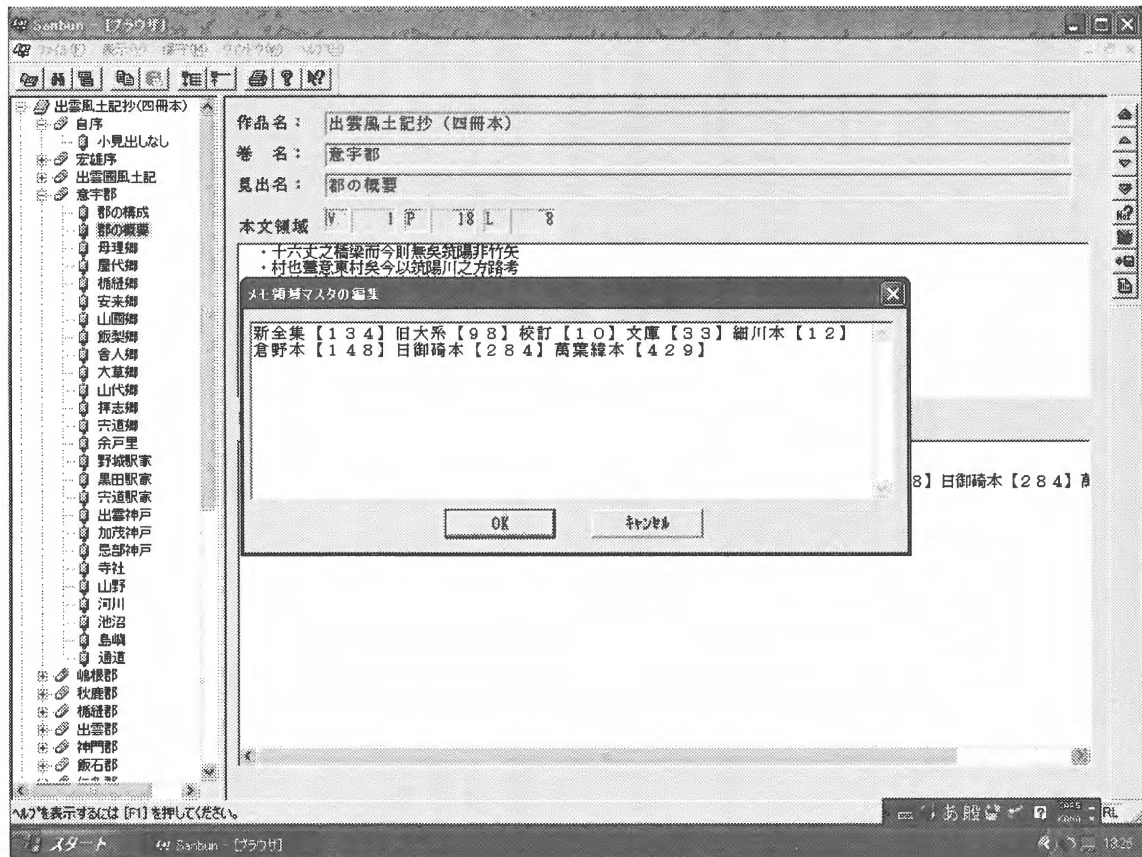
〈本文に対応した画像を開いた画面〉



〈検索文言入力画面〉



〈メモ領域の編集画面〉



6. シソーラスについて

これまでの古典コレクションデータベースでは、標準領域において表記や仮名遣いを統一することで検索の便宜を図ってきた。しかし、文学作品に登場する人名は特に、場面や時間の流れに沿って、人名の呼称は異なってくるのである。従って、そうした様々な異称に対応可能で、より効率的な検索方法が模索された。

【源氏物語（絵入）】で例えて言うと、作品中、光源氏に対し、源氏の君・大将殿・内大臣・大臣殿・源氏の大内大臣・六条の院といった呼称が用いられており、作品中に登場する光源氏を全て検索したいという場合には、上に記した複数の呼称を全て検索文言として入力しなければならない。また、その前提として、上に記した複数の呼称全てを利用者側が認知していることが必要となってくる。呼称が多くなればなるほど、利用者にもみえづいた検索では正確な検索結果が得られないということになる。ただし、【源氏物語】のような作品の場合、固有名詞に関しては、上のように複数の呼称がある場合でも、ひとまずはひとつの呼称に限定することができるため、その呼称で統一するか、併記することで検索漏れを最小限に抑えることが出来る。

今回の古事記・風土記など上代文献に登場する人名や地名などの固有名詞に関していうと、本文が万葉仮名と不完全な漢文とを混用した本文であることから、その表記法が多岐にわたっている。例えば、スサノオにしても、本データベース中では須佐ノ鳥・須佐鳥・須佐乎・須佐男・須佐乃鳥・須佐乃乎・須佐乃衰・須佐之男・須佐能鳥・須佐能乎・須佐能男・須佐能衰・須佐衰・素尊・素戔鳥・素戔鳥・素戔鳴・素盞鳥のように17種類の表記が見られる。それにも拘わらず、作品によってあるいは研究者によってどの表記を主とするかについては意見が分かれるのである。このように上代の文献では、複数の表記に関して各専門分野においても定説を持たないものは少なくないため、表記の統一を図ることには無理が生じてしまう。こうした状況下で、全ての表記を漏れなく検索するためにはどうしたらよいかという検討がなされた。その結果、無理に表記の統一を図るのではなく、同音異表記および異音異表記で同人・同神・同地名を一覧にしたシソーラスを作成しようということになった。

(1) 作成方法

監修員によるデータ校正の際、人名・神名・地名に関して、監修員の判断で語頭・語末に次のようなマーキングを行ってもらった。

全角・大文字で本文領域に

J 人名・神名 N

S 地名 X

の形式でマークする。

〈マーキングの実際〉

注記L03

本文L03 J 番仁岐命N、初降于S高千嶺X、J 神倭天皇N、経

標準L03 番仁岐命、初メテ高千嶺ニ降り、神倭天皇、メモL03

注記L04

本文L04 歴于S秋津嶋X、化熊出爪、天劔獲於J高倉N、

標準L04 秋津嶋ニ経歴リ、化熊川ヲ出デテ、天劔ヲ（*劔）高倉ニ獲、メモL04 校異【劔：劔】注記【爪：川（岩波大系）】

まず、上のようなデータをもとに、プログラム処理によって、J-Nで囲まれた部分を人名・神名、S-Xで囲まれた部分を地名として抽出した。次に、それぞれをエクセルの表にし、表中のデータで並べ替えを行った後、同じ表記のものは一件のみ残し、あとのデータは削除した。

更に、同音異表記のものを寄せるために、この表において固有名詞の読みを入力するという作業を行った。これは、たとえば「淡島」と「阿波志摩」のような同音異表記の地名はコード順に並び替えを行っても離れたところに位置するため、同定作業を行う際見落としてしまう可能性があるからである。こうした下準備を経て同定作業に入った。同定作業を終えたデータをHTMLファイルにし、古典コレクションに添付のRead meの中に挿入した。

〈読みを入力し、同定作業を終えたデータ〉

アイカヒメ 秋鹿日女 秋鹿比売
アオヌウマヌオシヒメ 青沼馬沼押比売

アオハタノサクサヒコ	青幡佐久佐日古 青幡佐草 日古 青幡佐草日子
アオミノイラツメ	青海郎女 飯豊 忍海郎女
アオミノイラツメ	飯豊
アオミノイラツメ	忍海郎女
アカイコ	赤猪子
アカイメノイラツメ	赤比売郎女
アガタヌシハエ	県主波延

アクシチベエカゲキヨ	悪七兵景清
アクトヒメ	阿久斗比売
アケタツ	曙立
アザミツヒメ	阿耶美都比売
アザミノイリヒメ	阿耶美能伊理毘売
アサヤマシ	朝山氏
アシイノイナキ	葦井之稻置

以下は、read me に示した凡例と固有名詞の一覧表である。

《人名・神名一覧凡例》

1. 『兼永本古事記』『出雲国風土記抄』両マスタデータの標準化領域から神名・人名を抜き出したものである。
2. 〈見出し〉・〈テーブル（＝同神・同人の異表記を集めた領域）〉・〈備考〉の三項目を並べて、一覧できるようにした。
3. 〈見出し〉はコード順に並べた。
4. 〈テーブル〉内は、空白一字を入れて横に並べた。テーブル内の項目全てをそのままコピーし、検索文言入力画面に貼り付けて利用できる。
5. 「伊奘冊」と「伊佐奈美」（共にイザナミ）のような同音異表記のものについては、それぞれを見出しに立てることはせず、同テーブル内にまとめた。
6. 異音異表記であるが、同神名・同人名をさすと考えられる場合は、その中でより一般的と見なされる呼称を上位項目として見出しに立て、そのテーブルに、下位項目の表記をも含む全ての表記パターンを掲げた。下位項目についても、それぞれ見出しを立て、あればその項目の同音異表記を掲げた。さらに「(上位項目)を参照」という意味で、備考に「→(上位項目)」と記した。
7. 天皇については、見出しに通行の呼称を用いた。よって天皇については、本文中にない表記が見出しに立っていることもある。
8. ブラウザ左上の「編集」→「このページの検索」によって、一覧内の検索も可能である。

《地名一覧凡例》

1. 『兼永本古事記』『出雲国風土記抄』両マスタデータの標準化領域から地名を抜き出したものである。
2. 〈見出し〉・〈テーブル（＝同地の異表記を集めた領域）〉・〈備考〉の三項目を並べて、一覧できるようにした。
3. 見出しはコード順に並べた。
4. 〈テーブル〉内は、空白一字を入れて横に並べた。テーブル内の項目全てをそのままコピーし、検索文言入力画面に貼り付けて利用できる。
5. 「棕見」と「久良弥」（共にクラミ）のような同音異表記のものについては、それぞれを見出しに立てることはせず、同テーブル内にまとめた。
6. 備考に「→○○」とある場合、○○は該当見出し語と異音異表記であるが、同地と考えられる地名である。この時、○○を見出し語とするテーブルには、当該見出し語を含む全ての表記パターンを掲げている。
7. 「郷」「村」「川」「社」等の前に同地名がついている場合は、地名部分のみ全ての表記を掲げ、テーブル化した。

〈例〉

クタミ	久多美 玖潭 忽美
クタミノサト	久多見郷 久多美郷 玖潭郷
クタミノムラ	久多見村 久多美村
クタミノヤシロ	久多美社 久多美山大明神
クタミヤマ	久多美山 久多美乃山

とある場合、テーブルには、
久多美 玖潭 忽美 久多見 久多弥
のように掲げる。

(3) 問題点

今回作成したシソーラスは『兼永本古事記』『出雲国風土記抄』のデータから作成したシソーラスであるため、本データベースの中でのみ有効であるが、逆に言うと汎用性に欠けるシソーラスということになる。すなわち、他作品のマスタデータを登録して横断検索を行う際には、検索漏れを生じてしまうことが考えられ、シソーラスを用いた検索の有効性を失ってしまうのである。

この問題点を解決するためにはどのような方法が考えられるであろうか。

(4) 本シソーラスの利点

一般的なシソーラス辞典は、そのみで完結しており、利用者がそこに介入する、すなわち改変・追加するという発想を持たないため、そのような仕様にもなっていない。しかし、本シソーラスは本文データと同様に参加型シソーラスになっている。先の(1)に解説したように、マスタデータの固有名詞にマーキングをし、それを抜き出すという比較的容易な方法を取っており、この作業自体は利用者にとっても問題はないであろう。またシソーラスはHTMLファイルのため、ソースを開き、抜き出した固有名詞を用いて、データの改変・追加を行うことも随時可能である。

従って、マスタデータを更新・作成するのと同じように、利用者自らがシソーラスを更新・作成することを提案したい。

7. 今後の課題

ただし、そうした改変・追加に際して、シソーラス中のどの項目に対して行えばよいかに関しては、ある程度の専門知識が要求されるであろう。

また、利用者がマーキングを施したデータから固有名詞を抜き出し、既存のシソーラスに改変・追加を現在のシソーラスに反映するためには、件数が多くなればなるほど手作業では困難をきたし、プログラミング等何らかの技術が必要となる。

今後こうした問題解決のための支援ツールの開発を検討したい。

謝辞

『兼永本古事記・出雲国風土記抄』データベースの構築は国文学研究資料館原本テキストデータベース

事業（研究情報部室長 中村康夫教授代表）として、平成11年度から14年度にかけて行われたものである。

『兼永本古事記』の監修員をして下さった青木周平氏・神田典城氏、『出雲国風土記抄』の監修員をして下さった岩下武彦氏・荻原千鶴氏・木村龍司氏にこの場をかりて御礼申し上げる。