

「五国史」宣命のコーパス化

池田 幸恵 (長崎大学)

須永 哲矢 (昭和女子大学)

Construction of the Corpus of *Gokokushi-Senmyo*

Yukie Ikeda (Nagasaki University)

Tetsuya Sunaga (Showa Women's University)

1. はじめに

天皇の和文詔勅である宣命は、奈良時代から平安時代にかけての日本語を知る貴重な資料であるものの、校本・総索引のある続日本紀宣命を除き、日本語史研究において利用される機会は多くない。その理由としては、宣命が国史や古記録の記事中に分散して存在することや、続日本紀宣命以外の宣命には定本がなく、またそれぞれの宣命の読みが確定していないことなどが挙げられる。

発表者は、宣命の中でも、『続日本紀』から『日本三代実録』までの「五国史」宣命を日本語史の資料として広く利用できるようにすることを目指し、「五国史」に収められた宣命に読みを与え、全体をコーパス化することを計画している。

本発表では、宣命のコーパス化の指針と基本仕様を紹介し、コーパス化により宣命の表記や語彙の研究にどのような利点が生じるのかを示すことにより、研究利用という観点からの現時点における処理方針案を示すこととする。

2. 「五国史」宣命のコーパス化の意義

2. 1 既存の主な電子資料とその限界

宣命は、一見すると漢文文献のようであるものの、原則として日本語の語順に従い、付属語や用言の活用語尾を万葉仮名で小書しているため、その背後にある当時の日本語をある程度再現できる点で貴重な資料である。宣命の起源は漢文詔勅にあり、語彙・語法に漢文訓読の影響が見られる一方で、儀式の際に口頭で宣布されるという性格上、その語彙には口語性があることも指摘されており、日本語史の資料としての価値は高い(小谷 1986、山口 1993)。

宣命を含む「五国史」(「六国史」)の電子テキストとしては、星野聰・水野柳太郎氏により作成された電子データを XML 化し検索可能にした渡瀬茂氏の「XML による六国史検索の試み(試行版)」や日本文学電子図書館の「六国史[全文]」などがある。

しかしながら、これらのデータにおいては、本文テキストの仕様および、データ形式や検索条件の面で限界があり、幅広い利用には至っていない。まず、本文テキストの仕様では、原文の漢字文字列がそのままデータ化されているのみであり、読み下し文は作られていない。そもそも続日本紀宣命以外の宣命には定本が存在せず、読みが確定できていないものも多いため、各所に不確かさを抱えたまま読み下すよりは、もとの文字列をそのまま電子化するという処理は妥当であるといえる。しかし、このような形式のテキストは宣命

にふれる機会の多い専門研究者以外には扱いにくく、同時代の別資料を扱う中で、あるいはより広く通時的研究を行う中で宣命を参照したいという需要には応えられていない。

また、データ形式の面でも、構造化がなされていない単純なプレーンテキスト形式であるために、検索条件も文字列検索のみが可能となっている。漢字と万葉仮名の区別のために、万葉仮名部分を小さく表示したり、〈〉に入れて表示するデータも存在するが、そのような処理はあくまで書籍同様、読む際の見た目を考えた処理に過ぎず、データ処理で語を取り出すという検索においては無力である。結局、既存の電子テキストでは、例えば漢字の「天」(「天皇・天津神」など)と万葉仮名の「天」(助詞の「て」)などは検索段階では区別することができず、検索結果を個々に確認していくしかない。

2. 2 「五国史」宣命コーパス化計画のめざす形式とその意義

本研究では、データ化の中心となる宣命本文の表記を漢字仮名交じり文に改め、それぞれの語に読みを与えたうえで、現在構築が検討されている『日本語歴史コーパス』と共通性をもった形態論情報を付与したコーパスを構築することを計画している。

従来の漢字列のテキストではなく、漢字仮名交じりの「専門外でも読みやすい」テキストにすること、また、形態論情報等の、本文以外の様々な情報をも持たせた構造化テキストにすること、の2点が本計画の特徴である。

読みの定まっていない箇所もある本文に対し、全ての漢字に読みを与えたうえで漢字仮名交じり文にするという方式は、その処理の妥当性・確実性において不安を抱えることにもなる。しかし、そのような不安点を差し引いても、読み下し文を本文とした、専門外の研究者にも利用しやすい形式のデータを提供する意義は大きいと判断した。

また、読み下し文にすることの意義は、単に読みやすだけにとどまるものではない。国立国語研究所『現代日本語書き言葉均衡コーパス』の大きな利用価値の一つである、形態論情報の付与も、前提として読みが与えられていればこそ可能になるものである。読み下し文を中心に据えたテキストを構築することにより、専門外の研究者にも利用しやすくなり、かつ、形態論情報など種々の情報を付与することが可能となる。それにより、研究資源としての宣命利用が宣命研究者以外にも広がり、日本語史研究に大きな成果をもたらすことが期待できると考えている。

読み下し文にすることで宣命の原表記を変える点や読みの確実性への不安は生じようが、これについても、構造化テキストとして原表記の情報を付与する予定であり(3. 2 参照)、さらに将来的には原文との対応を図ることにより、解決可能だと考えている。

3. 「五国史」宣命コーパスの基本仕様

3. 1 本文テキストの基本仕様

前節でも述べたとおり、「五国史」宣命コーパスでの本文は、読み下し文にすることを考えている。

以下に示す<原文>は、宣命の原文の表記そのままに近い状態である。一見して明らかのように、宣命の表記形式は、「原則として日本語の語順に従い、自立語を大書し、付属語や用言の活用語尾を万葉仮名で小書する」と言っても、小書部分を平仮名に代えると、そのまま現代の漢字仮名交じり文になるというものではない。

宣命は、天皇の和文詔書である内容面の特殊性だけでなく、このような表記面での特殊性により、従来、専門の研究者以外には扱いにくい資料であった。宣命は、たとえ電子化

されたとしても、宣命小書体表記に慣れていない利用者が読み下すことは困難であり、また、語彙研究の一環としてその語彙を利用しようとしても、語を取り出すこと自体に習熟を要する資料であるといえる。

コーパス化の意義の一つは、その資料を利用しやすくすることにより、利用層を拡大することである。宣命の専門研究者のみならず、語彙や文法を研究する中で、さまざまな資料を横断的に検討するその一つとして宣命を利用できるようにするためには、原表記のままではなく、漢字仮名交じり文に読み下した方が有用であると判断した。

「五国史」宣命コーパスでの本文は、その次に掲げる<漢字仮名交じり文>のような形を基本とする。<原文>と比べれば、一般ユーザーにとっては、かなり読みやすいものである。このような形式で本文を整備することで、より利用しやすいコーパスになると考える。

例)『日本後紀』巻22、延暦23年10月12日条

<原文>

天皇詔旨良万止勅命乎紀伊国司郡司公民陪従司々人等諸聞食止宣此月波閑時尔之豆国風御覧須時止奈毛常母聞所行須今御坐所乎御覧尔磯嶋毛奇麗久海漱毛清晏尔之豆御意母於多比爾御坐坐故是以御坐坐世留名草海部二郡乃百姓尔今年田租免賜比又国司国造二郡司良尔冠位上賜比治賜布目已下及郡司乃正六位上乃人尔波男一人尔位一階賜布又御座所尔近岐高年八十已上人等尔大物賜波久止詔布勅命乎衆聞食止宣

<漢字仮名交じり文>

天皇が詔旨らまと勅りたまふ命を紀伊国司郡司公民陪従へる司々の人等諸聞き食へと宣りたまふ。此の月は閑ある時にして国風を御覧す時となも常も聞所し行す。今御坐す所を御覧すに磯嶋も奇しく麗しく海漱も清く晏かにして御意もおだひに御坐坐す。故是以て御坐坐せる名草海部二郡の百姓に今年の田租免し賜ひ又国司国造二郡の司らに冠位上げ賜ひ治め賜ふ。目已下及び郡司の正六位上の人には男一人に位一階賜ふ。又御座す所に近き高年八十已上の人等に大物賜はくと詔りたまふ勅命を衆聞き食へと宣りたまふ。

また、原文を漢字仮名交じり文に読み下す際、また、その過程で漢字に読みを与えていく際にも、一定の指針を設け、テキスト全体としては均質な処理を施したものとなるよう努める。以下に、読み下し、および読みを与える際の基本指針を示す。

<漢字仮名交じり文に読み下す際の指針>

- (1) 一文ごとに句点を付ける。
- (2) 倒置表記は日本語の語順に改める。
- (3) 万葉仮名は平仮名にする。

例) 詔旨良万止→詔旨らまと 於多比爾→おだひに

- (4) 読み添える助詞や活用語尾も平仮名で表記する。

例) 天皇大命→天皇が大命 勅→勅りたまふ 聞食→聞き食へ

<読みを与える際の指針>

個々の語に読みを与える際には、以下の読みを採用する。

- (1) 北川(1982)にある読み。

- (2) 「六国史」の古写本にある読み。
 (3) 『類聚名義抄』『色葉字類抄』などの古字書にある読み。

・北川 (1982) にある読み

天皇 (スメラミコト)、詔旨・命 (オホミコト)、勅・宣 (ノリタマフ)、国司 (クニノミコトモチ)、郡司 (コホリノミヤツコ) 公民・百姓 (オホミタカラ)、司々人等 (ツカサヅカサノヒトドモ)、諸聞食・衆聞食 (モロモロキキタマヘヨ)、此 (コノ)、月 (ツキ)、時 (トキ)、御覧 (ミソナハス)、常 (ツネ)、聞所行 (キコシメス)、今 (イマ)、御坐・御坐坐 (オホマシマス)、所 (トコロ)、奇 (アヤシ)、麗 (ウルハシ)、清 (キヨシ)、御意 (ミココロ)、故 (カレ)、是以 (ココヲモチテ)、二郡 (フタコホリ)、今年 (コトシ)、田租 (タヂカラ)、免賜 (ユルシタマフ)、又 (マタ)、冠 (カガフリ)、位 (クラキ)、上賜 (アゲタマフ)、治賜 (ヲサメタマフ)、目 (サクワン)、已 (ヨリ) 下 (シモツカタ)、及 (オヨビ)、正六位上 (オホキムツノクラキノカミツカタ)、男 (ヲノコ)、一人 (ヒトリ)、一階 (ヒトシナ)、近 (チカシ)、高年 (トシタカキヒト)、八十 (ヤソヂ)、上 (カミツカタ)、大物 (オホミモノ)

・「六国史」の古写本、『類聚名義抄』『色葉字類抄』など古字書にある読み

紀伊 紀伊 岐 (元和三年古活字本倭名類聚抄、巻5、10ウ)
 陪従 陪 ソフ マサシ サカヌ アリ ハムベリ マコト ツカフ マシフ クスク イヨタツ
 (観智院本類聚名義抄、法中23オ5)
 従 シタカフ ヨリ ユルス ヲフ ホシイマ、 ソヒク ウツス ヨル キタル ツカフ
 コノム コトモナシ トモ アレイツ (観名、仏上23ウ1)
 陪従 ベイジウ (三卷本色葉字類抄、上53ウ4)
 陪従 ソヘ (日本書紀、巻2、鴨脚本・弘安本)
 シタカヘ (日本書紀、巻2、乾元本・丹鶴本)
 閑 閑 シヅカナリ ミヤビカナリ ホノカナリ ウヤヒカナリ ヒラク トラフ ウルハシ ヲシ
 フ ナラフ フセリ ノリ ホノメク イタヅラ ヒソカニ イトマ ナヲシ ナホナリ
 禾ケン (観名、法下39オ6)
 国風 跡さきとおなしやとりに行あひてかたるにこそは国ふりもきけ (言継卿集・435)
 磯 磯 イソ カト (観名、法中7オ6)
 嶋 嶋 シマ (観名、法上55オ6)
 海 海 ウミ 禾カイ (観名、法上2オ5)
 激 激 ミゾ (観名、法上23ウ)
 渚 ナキサ 激 同 (色葉字類抄 中32ウ1)
 晏 晏 オソシ ハル ヤハラカナリ タケヌ ウレシ ヒタク クラシ シツカ ヤスラカ ヨ
 ロコフ (観名、仏中51ウ6)
 名草 名草 奈久佐 (倭名類聚抄、巻5、24ウ)
 海部 海部 阿末 (倭名類聚抄、巻5、24ウ)
 国造 ^{クニ} 国 ^{ツコ} 造 (日本書紀、巻22、岩崎本)

3. 2 構造化テキストとしてのデータ形式

3. 2. 1 基本方式

「五国史」宣命コーパスはXML形式の構造化テキストとする。3. 1に示した本文テキ

スト（読み下し文）の文字列のみのデータではなく、本文テキストに対しマークアップ言語 XML でタグ付けをし情報をもたせることで、コーパスとしての利便性を高める。

例えば読み下して平仮名となった助詞や助動詞などの原表記（万葉仮名・漢文助字）の情報は、タグとして埋め込むことで、表面的には手の加わった漢字仮名交じり文であっても、データとしては手の加わる前の原表記を辿れるようにする。宣命本文を利用しやすい読み下し文に統一することも、宣命資料を構造化テキストタグ付きコーパスとすることも初の試みであるが、原表記の情報も保持するという構造化テキストであることをもって初めて、原文を読み下し文に大幅改編するという処理も許されよう。

3. 2. 2 宣命コーパスに付与される情報

正式なタグ名や XML タグセットの詳細は、今後の検討を経て確定させなければならないが、現時点では、各宣命を「詔」としてまとめ、「詔」の内部を「文」に区切り、さらに「文」を BCCWJ、『日本語歴史コーパス』等と同様の言語単位である「短単位」（単語）に区切る、という階層を主な構造として考えている。

タグによって付与される情報としては、以下に挙げるものを計画している。なお、以下のタグ例の書式は、理念の説明のために簡略化したものであり、実際の書式とは異なる。

(1) 詔

宣命を詔単位でまとめる。詔番号、宣布日、国史名、宣命の型（宣下・奏上・僧綱宣）を付与。

- ・ 詔番号…続日本紀宣命は本居宣長『続紀歴朝詔詞解』の詔番号。『日本後紀』から『日本三代実録』までの「四国史」宣命は、馬場（1993）で付された詔番号。
- ・ 宣布日…当該宣命が収められた国史の記事の年月日。
- ・ 国史名…国史名と当該宣命の収められた巻数。

例) 続日本紀 1...『続日本紀』巻 1

- ・ 宣命の型

宣下…天皇から臣下へ宣布するもの。宣命を「ノリタマフ（詔・勅・宣）」で結ぶ。

奏上…天皇から神社・山陵に奏上するもの。宣命を「マヲス（奏・申）」で結ぶ。

僧綱宣…僧官の補任に関するもの。宣命を「マヲス（白）」で結ぶ。

〈タグ例〉

<詔 詔番号="1" 宣布日="文武 1 年 8 月 17 日" 国史名="続日本紀 1" 型="宣下">

(2) 文

詔の内部を文ごとに区切った単位。

〈タグ例〉

<文>天皇が詔旨らまと勅りたまふ命を紀伊国司郡司公民陪従へる司々の人等諸聞き食へと宣りたまふ。</文><文>此の月は閑ある時にして国風を御覧す時となも常も聞所し行す。</文>

(3) 形態論情報

BCCWJ、『日本語歴史コーパス』と同様に、全ての文を短単位に区切って形態論情報を付与する。付与する形態論情報は品詞・活用型・活用形などである。コーパス内の全ての語

に対し情報を付与することになるが、この作業に関しては形態素解析辞書 Unidic による自動解析結果をもとに、人手で修正する予定である。

〈タグ例〉

<短単位 語彙素="治める" 品詞="動詞" 活用型="文語下二段" 活用形="連用形">治め</短単位>

<短単位 語彙素="給う" 品詞="動詞" 活用型="文語四段" 活用形="終止形">賜ふ</短単位>

(4) 読み

漢字に与えた読み。

〈タグ例〉

<読み 読み文字列="アキツミカミ">現御神</読み>

<読み 読み文字列="オホ">大</読み><読み 読み文字列="ヤシマ">八嶋</読み><読み 読み文字列="グニ">国</読み>

(5) 原表記

電子テキスト化に際して万葉仮名・漢文助字を平仮名表記に改めた場合の原表記。

〈タグ例〉

詔旨良万止→詔旨<原表記 原仮名="良万止">らまと</原表記>

天皇之→天皇<原表記 原助字="之">の</原表記>

令有尔→有ら<原表記 原助字="令">しむる</原表記><原表記 原仮名="尔">に</原表記>

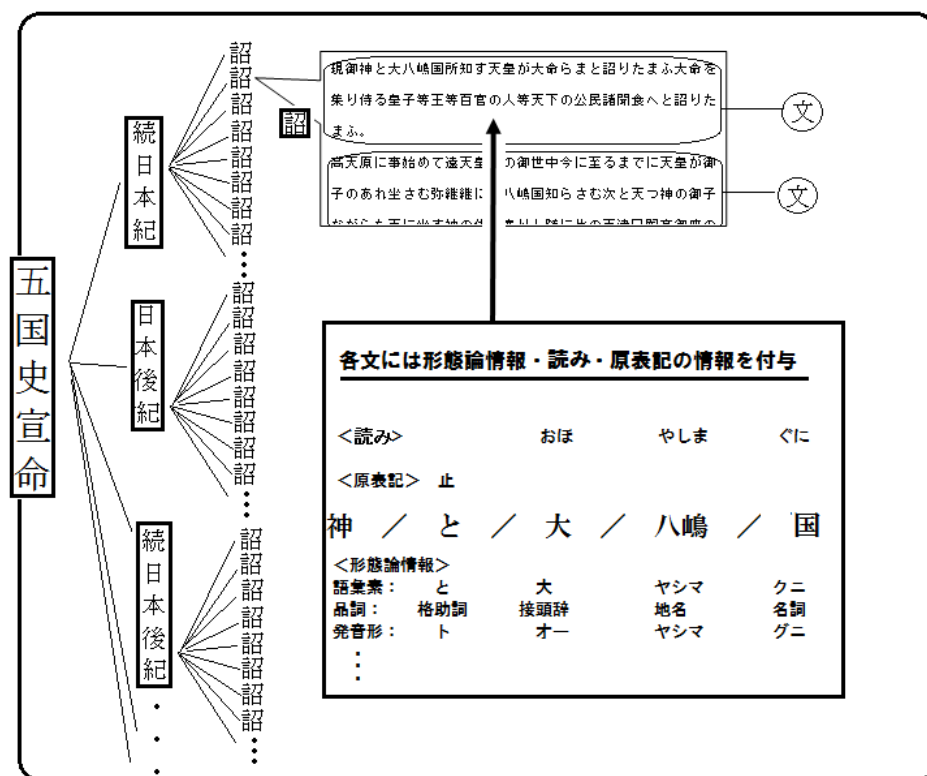


図1 宣命コーパスの階層構造のイメージ

```

<詔詔番号="66" 宣布日="延暦 23 年 10 月 12 日" 国史名="日本後紀 12" 型="宣下">
<文><読み 読み文字列="スメラ">天皇</読み><読み 読み文字列="オホミコト">詔旨</読み>
<原表記 原仮名="良万止">らまと</原表記><読み 読み文字列="ノ">勅</読み>りたまふ
<読み 読み文字列="オホミコト">命</読み><原表記 原仮名="乎">を</原表記><読み 読み文字列="キノ">紀伊</読み>
<読み 読み文字列="クニノミコトモチ">国司</読み><読み 読み文字列="コホリノミヤツコ">郡司</読み>
<読み 読み文字列="オホミタカラ">公民</読み><読み 読み文字列="シタガ">陪従</読み>へる<読み 読み文字列="ツカサヅカサ">司々</読み>
<読み 読み文字列="ヒト">人</読み><読み 読み文字列="ドモ">等</読み><読み 読み文字列="モロモロ">諸</読み>
<読み 読み文字列="キ">聞</読み>き<読み 読み文字列="タマ">食</読み>へ<原表記 原仮名="止">と</原表記>
<読み 読み文字列="ノ">宣</読み>りたまふ。</文>

```

図2 XML データの例 (試作段階、簡略化のため短単位タグは除いたもの)

4. 宣命コーパスの利用例

これまで述べてきたように、宣命のコーパス化の意義の一つは、専門の研究者以外にも宣命資料を利用しやすくすることであるが、宣命研究者にとっても、用例採集や集計の効率化をはじめとして、多大な効果が期待される。以下では、宣命研究の立場から考えうる、コーパスの利用例を挙げる。

4. 1 表記研究

- (1) 宣命ごと国史ごとの万葉仮名の種類と使用頻度 (簡単な字母表作成)
- (2) 宣命による漢文助字と万葉仮名の使用頻度の違い
- (3) 宣命における読み添え
- (4) 仮名書き自立語

4. 2 語彙研究

- (1) 宣命における高頻度語と低頻度語
- (2) 使用される形容詞・形容動詞の語形の変遷
- (3) 形容詞・形容動詞をどのくらい重ねるか 例) 明_支浄_支直_支誠_支之心
- (4) 動詞をどのくらい重ねるか 例) 聞_食驚_伎悦_備尊_備念_久波

5. 公開形式

本コーパスの公開形式としては、以下の2種を考えている。

5. 1 XML ファイルの利用形態

本文テキストに XML タグによって文章構造・形態論・表記に関する情報を付与した形式が、「宣命コーパス」本体としての基本データとなるが、XML ファイルそのものを扱えるユーザーはまだ限られているのが現状である。そこでまずは利用時の利便性、他資料利用との連動性を鑑み、『日本語歴史コーパス』の一部として、BCCWJ、『日本語歴史コーパス』同様、検索システム「中納言」で検索可能な形での公開を検討している。

5. 2 総索引

特に古典語研究においては、このような電子データの利用そのものに不慣れなユーザーもいまだ多いのが現状であり、使い慣れた紙媒体での索引形式の需要も大きい。それにもかかわらず、続日本紀宣命以外には語彙総索引が存在していない。そこで形態素解析結果を用いた総索引作成システム（小木曾・須永 2010）を利用し、宣命コーパスのデータをもとに前後文脈つき総索引を作成し、電子データを使い慣れていないユーザーにも利用しやすい形で公開することも計画している。

付記

本研究は、日本学術振興会科学研究費基盤研究（B）「和漢の両系統を統合する平安・鎌倉時代語コーパス構築のための語彙論的研究（24320086、研究代表者：田中牧郎）及び、基盤研究（C）「宣命に使用される字音語についての再検討」（25370521、研究代表者：池田幸恵）による成果の一部です。

文献

- 小木曾智信・須永哲矢（2010）「近世文語 UniDic」「中古和文 UniDic」を利用した総索引作成システムの開発」（『じんもんこん 2010 論文集』15、pp.119-124）
- 北川和秀（1982）『続日本紀宣命 校本・総索引』吉川弘文館
- 小谷博泰（1986）『木簡と宣命の国語学的研究』和泉書院
- 田中牧郎（2005）「言語資料としての雑誌『太陽』の考察と『太陽コーパス』の設計」（『国立国語研究所報 122 雑誌『太陽』による確立期現代語の研究 『太陽コーパス』研究論文集 pp.1-48 博文館新社）
- 馬場治（1993）『五国史所載宣命の国語史的研究』（『telos』11、金沢経済大学人間科学研究所）
- 山口佳紀（1993）『古代日本文体史論考』有精堂出版

参考 URL

XML による六国史検索の試み（試行版）

<http://www013.upp.so-net.ne.jp/wata/rikkokusi/index.html>

日本文学電子図書館 六国史〔全文〕

<http://www.j-texts.com/sheet/rikkoku.html>

日本語歴史コーパス

http://www.ninjal.ac.jp/corpus_center/chj/