

日本語・満族語の辞書作成のための 補助システム (Ⅷ)

本 田 道 夫

- I. はじめに
- II. 辞書データファイルとその形式
- III. 辞書作成までの処理手順
- IV. さいごに

I. はじめに

満族語文字を含む満族語・日本語の辞書作成の相談を受けて、その作成支援システムの開発を開始したのが1995年のことであり、それから約18年で辞書が完成することになった。現在、最終的な辞書の形で印刷して校正中であり、校正が完了次第、印刷・出版ということになる。なお、校正中ではあるが現時点で辞書の見出し語数は34,837、ページ数は用紙サイズA4、通常文字サイズ、2段組で1,218ページとなっている。

これまでかなり時間がかかったが、この時間はほとんど辞書作成者による辞書データの入力に要したものである。もちろん、その途中では、システムの動くOSがMS-DOSからWindowsになったのを始め、扱える満族語文字の追加・字形修正、シフトJISには入っていない文字のために文字鏡文字を利用できるような機能など、いくつかの機能の追加・プログラムの変更なども作成補助システムに対して行ってきた（[本田・今井1995]、[本田1998]、[本田2003]、[本田2006]、[本田2006]、[本田2008]、[本田2011]）。なお、本システムの編集サブシステム、印刷サブシステムなどは、[本田・山田1990]、[本田・吉

岡・山田 1991] のスラブ文字を主とした言語研究のために、スラブ文字、英文字、日本語文字をパソコンの画面上に表示して編集、また印刷もできるソフトウェアとして開発したものを元に、満族語文字を画面に表示して編集、印刷できるようにし、さらにシフト JIS には含まれていない多くの漢字も文字鏡で提供されている漢字を用いることにより扱えるようにしたものである。システムとしては、コマンド処理などのプログラム部分は共通の部分が多いが、満族語文字と追加の漢字の数が多いためスラブ文字用のものとは別のものとして開発し、さらに満族語文字用のシステムとして固有の部分もかなり存在するものである。

辞書作成のお話をいただいた最初の段階で、辞書データの入力から最終的に辞書として印刷するまでを、すべてパソコン上で行うこと考えて、辞書データの形式（以下の「形式 I」）を決めていた。その後、元となる資料ごと、あるいは大きな資料の場合は分割入力としたため複数のファイルが作成されてきた。ただし、中には、形式に合っていないファイルも作成されていたが、筆者はそのような別形式のものは目的としている辞書に取り入れられない別の目的のためのものだと思っていた。そのため、辞書データの入力はまもなく終了すると理解し、[本田 2012] ではまもなく辞書完成となると記していた。しかし、最初に辞書入力用に決めた形式のデータの入力終了後に、いよいよ辞書作成に取りかかれるかと思っていたが、辞書データの入力をされている方が、別形式のデータファイルの校正や追加、あるいは別形式の新規ファイルの作成にとりかかれたようなので、お尋ねしたところ、これら別形式で入力されているファイルも辞書に取り込むべきデータであることが分かった。そのため、さらに辞書データ入力に時間がかかることとなった。

しかし、これら別形式のファイルの入力もようやく完了したので、形式の異なるファイルを統合して、最終的な辞書として印刷できるものを作成する段階となった。

以下では、形式の説明と、このような形式の異なるファイルをどのようにして最終的に文書処理システム LaTeX に持ち込めるように処理したかについて説明する。

Ⅱ. 辞書データファイルとその形式

実際に入力・作成された辞書データファイルとその形式は、次の〔形式Ⅰ〕から〔形式Ⅳ〕の４種類であった。なお、MDAllSort.jmm は、もともとは入力の都合で 19 個のファイルとして作成されたものであったが、それを一つにまとめて見出し語（ローマ字表記部）をキーにアルファベット順にならべたものである。辞書作成当初は、パソコンは日本電気の PC 9801 で、OS は MS-DOS でありその OS 上では、メモリ上で扱えるデータ領域が 64 K バイトに制限されていたことから、64 K バイトまでの複数のファイルに分けて入力していた。

以下の記述は、最初の行に型式番号とその型式のファイル（拡張子 jmm）、その下がファイル内の各データの内容である。なお、たとえば〔形式Ⅰ〕の場合、最初の「番号」から最後の「日本語意味読み」までの項目を「1 グループ」として、ファイルは、複数のグループからなる。グループ間の区切りは空行としている。なお、以降では、他の形式についても、空行で区切られた複数の項目からなる一塊を「グループ」ということにする。

〔形式Ⅰ〕 MDAllSort.jmm

番号

満州文字表記

ローマ字表記

品詞

中国語意味〔出典情報〕

中国語意味読み

AA 番号


日本語意味


日本語意味読み


： ここより下は、日本語意味 と 日本語意味読み の繰り返し

この形式の例を次に示す。

例 1：入力されたファイル MDAllSort.jmm の先頭部分

a

 n.
 282/302
 陽 [2. 時令部・時令 1]。
 ヨウ
 陰陽の陽。
 ヨウ
 太極が動いて生成するもの。
 タイキョクガウゴイテセイセイシルモノ

a

 n.
 2867/3088
 陽 [7. 文書部・書 5]。
 ヨウ
 陽。奇の爻。
 ヨウ

a

 n.
 2952/3179
 字牙 [7. 文書部・書 8]。
 ジガ
 滿蒙字の劃 (かく)。
 マンモウジノカク

[形式Ⅱ] 清文總彙全卷順.jmm, 12 本文順.jmm, 折奏成語順番.jmm

ローマ字表記

中国語意味 [出典情報]

[形式Ⅲ] MD190.jmm, MD220.jmm, MDXXX.jmm

ローマ字表記

中国語意味 [出典情報]

中国語意味読み

[形式Ⅳ] MD200.jmm, MD210.jmm, MD230.jmm, MD240.jmm, 雍正.jmm
ローマ字表記

例文（ローマ字表記例文¶例文日本語訳）

ただし、MD200.jmm, MD230.jmm, MD240.jmm には、
「ローマ字表記 → 参照ローマ字表記」の行のみのものあり。

なお、これらのファイル作成時に参考にしたものは次のようなものであることである。

MDAllSort.jmm	清文鑑
清文總彙全卷.jmm	清文総彙
12 本文.jmm	大清全書
折奏成語.jm	摺奏成語, 折奏成語
MD190.jmm	清文備考, 同文彙書
MD200.jmm	禮科史書, 滿文太宗実録, 宗人府史書
MD210.jmm	内国史院档, 崇徳 2 年分
MD220.jmm	同文彙書
MD230.jmm	滿文老檔太祖紀
MD240.jmm	滿文老檔太祖紀
MDXXX.jmm	清文備考
雍正.jmm	宮中档雍正朝奏摺
彙	清文彙書

Ⅲ. 辞書作成までの処理手順

処理 1. 形式的なチェック

以下のようなチェックによる警告出力により、かなり多くの形式的な間違いを見つけて修正することができた。見つかったのは各項目の記述の間違いだけでなく、項目の記述位置が間違っているものもあった。また、グループ行数のチェックにより、本来なら二つのグループであるものが、空行で区切られずに、

一つのグループとなっていたものが見つかった。

- ファイルの各グループの行数のチェック。
- ローマ字表記部：英小文字と満族語文字のローマ字記述用に追加したコード F070～F074（16 進数）以外の文字があれば，警告。
- 満族語文字部：満族語文字以外の文字があれば，警告。
- 品詞部：英小文字とピリオド（.）以外があれば，警告。
- 中国語意味：漢字と出典記述を囲む半角鍵括弧（[]），数字，数字の区切りのピリオド以外があれば，警告。最後が，句読点の丸（。）でなければ警告。
- 中国語意味読み：全角カタカナ以外の文字があれば，警告。
- AA 番号：数字，およびスラッシュ（/）以外の文字があれば，警告。
- 日本語意味：漢字，ひらがな，アルファベット，コンマ（,）ピリオド（.）以外の文字があれば，警告。最後が，句読点の丸（。）でなければ警告。
- 日本語意味読み：全角カタカナ以外の文字があれば，警告。
- 例文（¶ ローマ字表記部：日本語意味部）例文の最初は記号¶であり，ローマ字表記部と日本語意味部の区切りは全角コロン（:）。
- ・ ローマ字表記部：ローマ字（F070～F074 のコードの文字を含む）とコンマ以外があれば警告。
- ・ 日本語意味部：日本語文字，出典記述を囲む半角鍵括弧（[]），出典記述中のピリオド，数字，英大文字 A, B, C 以外のものは警告。

処理 2. タグ付け

数字あるいは英文字を半角の大小記号（〈 〉）で囲ったタグを，すべてのファイルの各項目の先頭に付けた。タグを付けた理由は，主に次の 2 つである。

理由 1：処理のある段階で全部のファイルを一つのファイルとして、見出し語を元に辞書式順序に並べ替えて（ソートして）、さらに同じ見出し語（ローマ字表記部）のグループが複数あったときには、一つの見出し語の下にそれらのグループの見出し語以外の部分を纏めて、さらに辞書としての形式に整える。この纏める時に最初は何のファイルから、次にはどのファイルからというように順序を指定されていた。そこで、纏めた直後に意図したファイルからの順になっているかを確認するためである。

理由 2：一つのファイルにした後の処理で問題が生じたとき、あるいは辞書データとして間違いが見つかったときに、その間違いの部分はどのファイルからのものであるか、さらには、最終的な辞書の形式での校正の段階で、修正すべき箇所がどのファイルからのものであるかが分かるようにするためである。辞書データの修正などは、ここで述べている処理により作成されるファイルに対して行うのではなく、「Ⅱ. 辞書データファイルとその形式」で述べた「MDAllSort.jmm」などの辞書作成者が入力している辞書データファイルに対して行うことにしているので、どのファイルを修正すべきかのタグ情報は非常に役立つものである。

もちろん、最終的な辞書作成の時には、これらのタグは除かれる。数字のタグは「MDAllSort.jmm」からの部分であることを示している。英大文字はファイルを特定できるものとし、英小文字のタグはその項目の内容を表すことにしている。そのために、英小文字のタグは複数のファイルで重複していることもある。最初は、全項目に異なる英大文字でタグを付けようとしたが、辞書形式になったときに、各部分がどのファイルからのものであるかが分かればいいので、このようなタグの付け方とした。なお、最終的には各グループの先頭のローマ字表記に付けた数字 1 のタグと英大文字のタグだけを用いた。タグ付けを考えてプログラムを作成する段階では、全部のタグを用いる必要があるかと思っていたが、以後のプログラムを作成しているときに、各グループ先頭のタ

グだけでよいことが分かったが、遑ってタグ付けのプログラムから変更することはしなかった。

(1) MDAllSort.jmm

このファイルに関しては、タグ付けだけでなく、グループ内の先頭の番号〔形式Ⅰ〕参照〕の削除と、グループ内の項目の順序を次のように入れ替えて、タグ付けを行った。

タグ付加後	タグ付加前の順序
番号（この項目はタグ付加のときに削除）	[1]
〈1〉ローマ字表記	[3]
〈2〉満族語文字表記	[2]
〈3〉品詞	[4]
〈4〉AA 番号	[7]
〈5〉中国語意味〔出典情報〕	[5]
〈6〉中国語意味読み	[6]
〈7〉日本語意味	[8]
〈8〉日本語意味読み	[9]
〈7〉日本語意味 ここより下は、〈7〉～〈8〉の繰り返し	[10]
〈8〉日本語意味読み	[11]

(2) 清文總彙全卷順.jmm

- 〈A〉ローマ字表記
- 〈a〉中国語意味〔出典情報〕

(3) 12 本文順.jmm

- 〈B〉ローマ字表記
- 〈b〉中国語意味〔出典情報〕

(4) 折奏成語順番.jmm

- 〈C〉ローマ字表記
- 〈c〉中国語意味〔出典情報〕

(5) MD190.jmm

- 〈D〉ローマ字表記
- 〈d〉中国語意味〔出典情報〕
- 〈6〉中国語意味読み

(6) MD220.jmm

〈E〉 ローマ字表記

〈e〉 中国語意味 [出典情報]

〈6〉 中国語意味読み

(7) MDXXX.jmm

〈F〉 ローマ字表記

〈f〉 中国語意味 [出典情報]

〈6〉 中国語意味読み

(8) MD200.jmm

〈G〉 ローマ字表記

〈9〉 例文 (¶ ローマ字表記例文：例文日本語訳)

ただし, 「ローマ字表記 → 参照ローマ字表記」の行のみのグループあり。つまり, 〈G〉のみのグループがあり得る。

(9) MD210.jmm

〈H〉 ローマ字表記

〈9〉 例文 (¶ ローマ字表記例文：例文日本語訳)

(10) MD230.jmm

〈I〉 ローマ字表記

〈9〉 例文 (¶ ローマ字表記例文：例文日本語訳)

ただし, 「ローマ字表記 → 参照ローマ字表記」の行のみのグループあり。つまり, 〈I〉のみのグループがあり得る。

(11) MD 240.jmm

〈J〉 ローマ字表記

〈9〉 例文 (¶ ローマ字表記例文：例文日本語訳)

ただし, 「ローマ字表記 → 参照ローマ字表記」の行のみのグループあり。つまり, 〈J〉のみのグループがあり得る。

(12) 雍正.jmm

〈K〉 ローマ字表記

〈9〉例文（㊦ローマ字表記例文：例文日本語訳）

例2：タグ付けされた例

<1>a
 <2>㊦
 <3>n.
 <4>282/302
 <5>陽 [2. 時令部・時令 1]。
 <6>ヨウ
 <7>陰陽の陽。
 <8>ヨウ
 <7>太極が動いて生成するもの。
 <8>タイキョクガウゴイテセイセイスルモノ

<1>a
 <2>㊦
 <3>n.
 <4>2867/3088
 <5>陽 [7. 文學部・書 5]。
 <6>ヨウ
 <7>陽。奇の爻。
 <8>ヨウ

<1>a
 <2>㊦
 <3>n.
 <4>2952/3179
 <5>字牙 [7. 文學部・書 8]。
 <6>ジガ
 <7>滿蒙字の劃 (かく)。
 <8>マンモウジノカク

処理3. 全部のファイルの統合

「Ⅱ. 辞書データファイルとその形式」で述べたように、四つの形式のファイルがあるが、これらすべてを、一つのファイルに纏めるが、上記処理2で記したファイルの順番で取り込んだ。つまり、最初に「MDAllSort.jmm」、次に「清文總彙全卷順.jmm」、以下「12本文順.jmm」、「MD190.jmm」、などの順である。これは、これら複数のファイルから辞書としての内容を取りまとめて作成するときに、同じ見出し語（ローマ字表記の部分）の複数のグループがあった場合に、作成する辞書では、一つの見出し語の後に、それら複数のグループ

の見出し語を除いた2番目以降から取り出した部分を並べるが、そのときに、一番最初に持ってくるのは、「MDAllSort.jmm」、次には「清文總彙全卷順.jmm」というように、順序を要請されたことに対応して、次の処理4のソートで、扱い易いようにするためである。なお、一つに纏められたファイルは複数の形式のグループからなり、それらの間は空行である。

処理4. 統合したファイルのソートとファイルの分割

処理3で作成したファイルは複数の形式のファイルの一つに纏めたものであるので、当然その纏めたファイルには複数の形式のグループがあるが、どのグループも先頭は見出し語となるローマ字表記であり、そのローマ字表記をもとに、辞書式順序でのソートを行った。ソートの方法としては処理速度の速いクイックソートではなくバブルソートを用いた。これは、同じ見出し語（ローマ字表記）のグループが複数あったときに、バブルソートでは元の順序を保った（つまり、ソート前にファイルの先頭側にあったものはソート後も先頭側に位置する）ソートが可能であるためである。バブルソートでは、処理に4分36秒かかった。ちなみに、元の順序は保たないが、処理時間の短いクイックソートでは、1秒未満であった（CPUはIntel Core i7-3770 3.4 GHz）。なお、このソートと処理8のLaTeXでの処理以外での各処理時間は1秒未満である。

なお、各グループはどのファイルからのものであるかはタグを見れば分かるので、タグを利用してクイックソートで処理することも考えたが、ソートはそれほど頻繁に行うものではないので、バブルソートを採用した。ソート後に同じ見出し語の元に纏められた複数のグループがあったときに、要請されたような元のファイルの順になっていることの確認はタグを見ることにより行った。

ソートの後に、グループの見出し語の先頭文字（A, B, C など）により、例えば先頭文字がAのものはファイル「MDicAA.jmm」に、Bのものは「MDicBB.jmm」というように、ファイルに分けて出力し、LaTeXで処理するためのファイル（MJDic.tex）中で、それらのファイルをLaTeXのコマンド「\input」を用

いて読み込むことにしている。これは、次の理由による。

最終的な辞書として仕上げるときに、たとえば、Aで始まる見出し語の最後の部分と、Bで始まる最初の見出し語の間で、大きな文字でBを入れ、かつその上下をいくら空けるかなどに対して、各ファイルの「¥input」の間で記述して、対応しやすくするためである。現時点では各ページに爪を付けることを考えているが、爪の文字と爪位置の縦方向の変更指定も「¥input」文の間で行う。

ファイル「MJDic.tex」には、出力用紙の大きさ、用いる英文字フォント、二段組みにする指定、利用するLaTeXのスタイルファイル、辞書用に作成したマクロの定義なども入れている。現時点の辞書の形式としては、見出し語の先頭文字が代わるときに、3行程度の空行、大きな英文字、さらに半行程度の空行を入れることにしているが、それらは、ファイル読み込みの「¥input」コマンドの間に入れることにより辞書内容のファイルを作成するプログラムでは扱わなくてよくなる。つまり、「MDicAA.jmm」などのファイルとしては、辞書内容に関することだけを含み、辞書の形式などに関することを含めてプログラムで含めなくてよいようにしている。また、現時点では辞典に「爪」をつけることにしているが、これについても、「MJDic.tex」中で指定することにより、「MDicAA.jmm」などの辞書内容のファイル中で指定しなくてよいようにしている。

処理5. 見出し語が同じグループを一つの見出し語に纏める

同じ見出し語（ローマ字表記）のものが複数あったときに、それらを一つの見出し語の元に纏める。つまり、見出し語一つのあとに、各グループの2番目以降の部分を順に纏める。なお、この時点で「日本語意味読み」、「中国語意味読み」の部分は纏めたものに入れない。例2のものを纏めたものが次の例3である。MDAllSort.jmmからのグループについては、一つのグループ中に複数の日本語意味がある場合は、それらは一行の日本語意味に纏められる。纏めたも

のを例3に示す(例3には例2に記したものの以外のもも含まれている)。例2の最初のaでは日本語意味が7行目の「陰陽の陽。」と9行目の「太極が動いて生成するもの。」の二つがあるが、纏められた例3ではこれは5行目の一行に纏められている。

複数のファイルを一つに纏めて、さらに同じ見出し語の元に纏めたが、これには一種類から四種類の形式の異なるグループからのものが含まれていることがある。ここでの処理は纏められた見出し語からの先頭部分がMDAllSort.jmmのファイルからの場合と、そうでない場合で異なる。

○MDAllSort.jmm からの部分がある場合：

原則として次のような項目の順となる。

見出し語 (ローマ字)

満族語文字での表記

品詞 ←(A)

AA 番号 ←(B)

日本語意味 ・ 一つのグループに複数の日本語意味がある場合は一行にまとめる。

中国語意味 ←(C)

： ・ (B) から (C) までの繰り返し。

： ただし、繰り返しのないこともある。

： ←(D)

品詞 ・ 再度 (A) から (D) までの繰り返し

AA 番号 繰り返しのないこともある。

日本語意味

中国語意味

：

同じ見出し語で、形式Ⅱの中国語意味、形式Ⅲの中国語意味、形式

IVの例文が各一行で続く

なお、最後の「同じ見出し語で、形式Ⅱ…で続く」は、例3の最後の2行のように、各ファイルからのものごとにそれぞれ一行となる。

例3：同じ見出し語のものを一つの見出し語のもとに纏めたもの

```
<1>a
<2>ㄣ
<3>n.
<4>282/302
<7> 陰陽の陽。太極が動いて生成するもの。
<5> 陽 [2. 時令部・時令 1]。
<4>2867/3088
<7> 陽。奇の文。
<5> 陽 [7. 文學部・書 5]。
<4>2952/3179
<7> 滿蒙字の劃 (かく)。
<5> 字牙 [7. 文學部・書 8]。
<3>int.
<4>5864/6272
<7> あー。ふーん。人の話すのに応ずる声。
<5> 慢應聲 [12. 人部 3・問答 2]。
<4>7103/7588
<7> あっ。驚きの聲。
<5> 驚懼聲 [14. 人部 5・聲響 1]。
<a> < A >開口驚訝詞○寫滿字一曲之曲○陰陽之陽 [總彙.1-1.a3]。
<b> < B >開口驚訝之詞 [全 0101a3]。
```

○MDAllSort.jmm からの部分がない場合：

この場合は、[形式Ⅱ]の中国語意味、[形式Ⅲ]の中国語意味、[形式Ⅳ]の例文を、例4のように、一つの見出し語のもとに纏める。この場合、通常の辞書のような品詞、訳などの部分はなく、見出し語、中国語意味、あるいは見出し語、例文などの形のものであるが、辞書作成者の意向では、このような形のものも辞書に入れるとのことである。

例4：一般の辞書形式でないもの（品詞などを含まない）

（この場合は、形式Ⅱのファイル「清文總彙全卷.jmm」と

「12 本文.jmm」からのものを纏めている)

<A>abka ehenehe

<a> < A >天候が悪くなった。天變了 [總彙.1-28.b3]。

 < B >天變了 [全 0128a4]。

ここでの処理では、見出し語が同じものは統合されるので、グループ数と行数が減少する。つまり、2 番目以降のグループの見出し語は除かれるし、日本語意味読み、中国語意味読みの部分も除かれる。また「MDAllSort.jmm」の一つのグループ内に複数の日本語意味がある場合も、それらは一行にされるので、行数が減少する。このように、グループ数、行数が減少するときにその数を数えておき、プログラムの最後に、

処理前の グループ数, 減ったグループ数, 処理後のグループ数

処理前の 行数, 減った行数, 処理後の行数

を出力し確認することにより、プログラムのエラーによりデータが消失することを防ぐための確認を行った。

実は、処理 1～処理 4、処理 6、処理 7 の処理でも、データの消失が生じていないことの確認は、それぞれのファイルについて、処理前と後のグループ数と行数を表示し、同じ値であることにより確認はしている。

処理 6. 最終的な辞書の形の LaTeX 用のコマンドを含んだ形式への変換

処理の説明の前に、まず、処理結果を例 5 に示す。変換結果には、筆者が定義した LaTeX の環境 DicWord, コマンド ¥MIDASHI, ¥MANMOJI, ¥HINSHI, ¥AANO が含まれているが、これらの定義は処理 6 の最後に記す。

例5：LaTeX用のコマンドを含んだ形式へ変換されたもの

```

¥begin{DicWord}
  ¥item[ ]¥MIDASHI{a}{a}
    ¥MANMOJI{ㄣ}
    ¥HINSHI{n.}
    {¥bf 1.¥,}¥AANO{282¥,./¥,302}
    陰陽の陽。太極が動いて生成するもの。
    陽 [2. 時令部・時令 1]。
    {¥bf 2.¥,}¥AANO{2867¥,./¥,3088}
    陽。奇の爻。
    陽 [7. 文學部・書 5]。
    {¥bf 3.¥,}¥AANO{2952¥,./¥,3179}
    滿蒙字の劃 (かく)。
    字牙 [7. 文學部・書 8]。
    ¥HINSHI{int.}
    {¥bf 1.¥,}¥AANO{5864¥,./¥,6272}
    あー。ふーん。人の話すのに応ずる声。
    慢應聲 [12. 人部 3・問答 2]。
    {¥bf 2.¥,}¥AANO{7103¥,./¥,7588}
    あっ。驚きの聲。
    驚懼聲 [14. 人部 5・聲響 1]。
    / < A > 開口驚訝詞○寫滿字一曲之曲○陰陽之陽 [總彙.1-1.a3]。
    < B > 開口驚訝之詞 [全 0101a3]。
¥end{DicWord}

```

(6-1) 処理5で纏められた各グループを、定義した DicWord 環境に入れたものとして出力する（例5の先頭行の ¥begin {DicWord} から最後の ¥end {DicWord} で囲ったもの）。

(6-2) 見出し語部分を、定義した LaTeX コマンド ¥MIDASHI を用いて、

```
¥item[ ]¥MIDASHI{見出し語を加工したもの}{見出し語}
```

のようにして出力する。なお、見出し語を加工したものとは、次の (a)、(b) の処理をしたものである。

- (a) 見出し語中に半角小括弧 (パレンセシス ()), あるいは半角大括弧 (ブラケット []) で囲まれた部分がある場合に、その囲まれた部分を除く。
- (b) 見出し語が単語間の空白を含めて 36 文字以上の場合、36 文字より短いところまでの単語を出力し、その後は「…」とする。

- (6-3) 満族語文字部分を定義した LaTeX コマンド $\text{\textbackslash MANMOJI}$ の引数に入れる。
- (6-4) 品詞部分を定義した LaTeX コマンド $\text{\textbackslash HINSHI}$ の引数に入れて出力する。
- (6-5) AA 番号部分を定義した LaTeX コマンド $\text{\textbackslash AANO}$ の引数に入れて出力する。そのあとの、日本語意味と中国語意味は、そのまま AA 番号に続けて出力する。
- (6-6) MDAllSort.jmm 以外のファイルからの中国語意味、例文などがあれば、全角スラッシュ (／) に続けてそのまま出力する。

(6-2) での「見出し語を加工したもの」は、その見出し語がページの最初あるいは最後の見出し語となったときに、それぞれページ上部の左と右に表示されるようにするためである (例 7 のヘッダ部参照。左側にこのページの最初の見出し語「acafi beiderede baitalara hoošan, …」, 真ん中にページ番号, 右側に最後の見出し語「acambi」が表示されている)。そのときに、注釈的に書かれている括弧内の記述は入れないようにするということである。また、36 文字までとしたのは、見出し語の中には長いものがあり、そのまま表示するとそれだけで中央のページ番号あるいは右側の見出し語に重なってしまうことを避けるためである (例 7 のヘッダ部の左側の見出し語は横線より下の辞書内容の最初の見出し語よりも短くされ、省略された部分は…となっている)。なお、このページ上部の左右に見出し語を表示するには、スタイルファイル「fancyhdr.sty」と「fixltx2e.sty」を用いた。

(6-5) の AA 番号部分の出力については、同じ品詞中に複数の AA 番号が複数あれば、順に番号を付け、一つだけの AA 番号の場合は番号を付けない。例 5 は番号を付けたものであり、 $\text{\textbackslash bf 1.\textyen,}$, $\text{\textbackslash bf 2.\textyen,}$, $\text{\textbackslash bf 3.\textyen,}$ などが順に番号を付けたものである。なお、 \textbackslash はその前のピリオドと次のものの間は少し空けるといいう LaTeX のコマンドである。

環境 DicWord, コマンド ¥MIDASHI, ¥MANMOJI, ¥HINSHI, ¥AANO は次のように定義している。なお, 見出し語用に ¥MIDASHI, 満族語文字用に ¥MANMOJI, 品詞用に ¥HINSHI, AA 番号用に ¥AANO を定義したのは, 見出し語, 品詞などの書体や大きさを処理プログラムを変更することなく定義を変更するだけで簡単に変えることができるようにするためである。

```
%%***** DicWord 環境 定義 の開始*****
¥newenvironment {DicWord} [0] {}
¥def¥stepno##1 {}
¥def¥makelabel {¥hspace* {1zw} ( ¥bf □ ) ¥hfill ¥item} %
¥mybptmsep=-0mm%
¥begin {list} {} {¥settowidth {¥myitmzsz} {a} %
    ¥setlength {¥topsep} {-0.7zh} %
    ¥setlength {¥leftmargin} {-0zw}%
    ¥addtolength {¥leftmargin} {¥labelsep} %
    ¥setlength {¥rightmargin} {¥myitmzsz} %
    ¥setlength {¥labelwidth} {¥myitmzsz} %
    ¥setlength {¥baselineskip} {15pt} %
    ¥setlength {¥itemsep} {-2mm} %
    ¥setlength {¥parsep} {3mm} %
    ¥addtolength {¥labelwidth} {-¥labelsep} %
    ¥renewcommand {¥makelabel} {¥usefont {T1} {txr} {b} {n} ¥large} %
} {¥end {list} ¥baselineskip=¥mybslnskpsv ¥vspace {¥mybptmsep} }
%%***** DicWord 環境 定義 の終了*****
```

%%見出し語の形式：見出し語は左に全角で 0.7 文字分, さらに, 下に
%%全角 0.1 文字分ずらした位置から出力し, ボールド体, 文字サイズ
%%¥large とする。

```
¥def¥MIDASHI#1#2 {¥markboth {¥bf #1} {} ¥noindent¥hspace {-0.7 zw} %
```

$\backslash\mathrm{vspace*}\{.1\mathrm{zh}\}\backslash\mathrm{bf}\backslash\mathrm{large}\#2\}\backslash\mathrm{¥}$

%%満族語文字列の形式：

$\backslash\mathrm{def}\backslash\mathrm{MANMOJI}\#1\{\#1\}$

%%品詞の形式：品詞はイタリックとする

$\backslash\mathrm{def}\backslash\mathrm{HINSHI}\#1\{\backslash\mathrm{it}\#1\mathrm{¥},\}$

%%AA 番号の形式：

$\backslash\mathrm{def}\backslash\mathrm{AANO}\#1\{[\#1]\}$

処理 7. 満族語文字および文字鏡文字を LaTeX で処理できるように変換

編集システムでは、満族語文字、文字鏡文字がその字形でパソコンの画面に表示して扱えるようにしているが、そのままでは、LaTeX では正しく扱われない。そこで、正しく扱えるように変換する。例 6 - (a) は、変換前の満族語文字列とその変換結果、例 6 - (b) は変換前の文字鏡文字列とその変換結果である。

例 6 - (a) : LaTeX で処理できるようなものへの変換 (満族語文字)

変換前： $\backslash\mathrm{MANMOJI}\{\backslash\mathrm{ᡤᡠᡳᡳᡳᡳᡳᡳ}\backslash\mathrm{ᡳᡳᡳᡳ}\}$

変換後：(以下の三行は実際には改行と先頭の空白なしの長い一行である。)

$\backslash\mathrm{MANMOJI}\{\backslash\mathrm{Mann}\{\backslash\mathrm{¥}\mathrm{symbol}\{"81\}\backslash\mathrm{¥}\mathrm{symbol}\{"1E\}\backslash\mathrm{¥}\mathrm{symbol}\{"17\}\backslash\mathrm{¥}\mathrm{symbol}\{"14\}$
 $\backslash\mathrm{¥}\mathrm{symbol}\{"90\}\backslash\mathrm{¥}\mathrm{symbol}\{"0E\}\backslash\mathrm{¥}\mathrm{symbol}\{"24\}\backslash\mathrm{¥}\mathrm{symbol}\{"81\}\backslash\mathrm{¥}\mathrm{symbol}\{"2D\}$
 $\backslash\mathrm{¥}\mathrm{symbol}\{"0C\}\backslash\mathrm{¥}\mathrm{symbol}\{"84\}\backslash\mathrm{¥}\mathrm{symbol}\{"24\}}\}$

例 6 - (b) : LaTeX で処理できるようなものへの変換 (文字鏡文字)

変換前：美 崩 机 平 映 彤 恸 看 咆 控 結 涵 愔 谿 紆 飢

変換後： $\backslash\mathrm{¥TMO}\{072229\}\backslash\mathrm{¥TMO}\{029346\}\backslash\mathrm{¥TMO}\{014547\}\backslash\mathrm{¥TMO}\{028215\}$
 $\backslash\mathrm{¥TMO}\{086489\}\backslash\mathrm{¥TMO}\{030363\}\backslash\mathrm{¥TMO}\{085253\}\backslash\mathrm{¥TMO}\{066439\}$
 $\backslash\mathrm{¥TMO}\{023242\}\backslash\mathrm{¥TMO}\{060587\}\backslash\mathrm{¥TMO}\{079140\}\backslash\mathrm{¥TMO}\{083693\}$
 $\backslash\mathrm{¥TMO}\{083618\}\backslash\mathrm{¥TMO}\{020548\}\backslash\mathrm{¥TMO}\{079059\}\backslash\mathrm{¥TMO}\{044033\}$

¥MLMann は満族語文字用のフォントを切り替えるものであり、次のように定義している。

```
¥def¥MlxMann {¥fontencoding {T1} ¥fontfamily {ManRP} ¥selectfont}
¥DeclareTextFontCommand {¥MlMann} {¥MlxMann}
```

また、¥TMO は Tomoaki Honda 氏が作成した LaTeX で文字鏡文字を扱うためのスタイルファイル Mojikyo.sty の中で定義されているものである。

処理 8. LaTeX での処理により辞書の印刷イメージの dvi ファイルの作成

処理 7 までで、LaTeX で処理できる形式になっているので、最後に LaTeX で処理して DVI ファイルを作成する。作成した一部を例 7 に示す。なお、ここで示した例 7 ではたとえば、3 行目に中国語意味の前に<F>、のタグが残っているが、これは、上記でも述べたが、チェックで誤りが見つかったときに、どのファイルからのものであるかが分かるよう残したものであり、もちろん最終的には、このようなタグを除いたものとする。なお、この例は一段組みであるが最終的には例 8 のような二段組みとする予定である。校正のためには、見やすさを考えて一段組みとした。

最後に、二段組みとして、かつ元のファイルを示す<A>、などのタグも除いた最終的な形のものを例 8 に示す。

例8 二段組みでの最終的な形に近いもの

a	1	abalanambi
A		
a 𐄂 n. [282 / 302] 陰陽の陽。太陽が動いて生成するもの。陽 [2. 時令部・時令 1]。2. [2867 / 3088] 陽。奇の文。陽 [7. 文学部・書 5]。3. [2952 / 3179] 滿蒙字の劃 (かく)。字牙 [7. 文学部・書 8]。int. 1. [5864 / 6272] あー。ふーん。人の話すのに応ずる声。慢應聲 [12. 人部 3・問答 2]。2. [7103 / 7588] あっ。驚きの聲。驚懼聲 [14. 人部 5・聲響 1]。／開口驚訝詞○寫滿字一曲之曲○陰陽之陽 [總彙。1-1. a3]。開口驚訝之詞 [全。0101a3]。		
a a 𐄂 int. [5865 / 6273] あ。あん。かりそめに応対する声。隨便答應聲 [12. 人部 3・問答 2]。		
a fa sere ongolo 𐄂 𐄂 𐄂 ph. [9698 / 10343] どうこうという前に。どんなだか分からない前に。誰よりも先に。未見怎麼樣 [18. 人部 9・散語 1]。		
a i bukdan 𐄂 𐄂 n. [3077 / 3310] 陽扣。紙の表を表にして疊んだもの。おもてだたみ。おもて折り。陽扣 [7. 文学部・文学什物 1]。／陽扣凡接扣摺子紙向外摺口——[總彙。1-1. a3]。		
a jijun i acangga 𐄂 𐄂 c 𐄂 n. [1038 / 1111] 銅で造った陽字の合札 (あいふだ)。夜間城門の通行用に携帯するもので、内府に存貯してある札→ e jijun i acangga。陽文合符 [3. 論旨部・論旨]。／陽文合符銅鑄陽文字以備夜間開城之鍵内廷存貯 [總彙。1-13. a5]。		
a si 𐄂 int. [7112 / 7597] あっしっしっ。鶏などを嚇し逐う聲。逐鶏鳥聲 [14. 人部 5・聲響 1]。		
a si seme 趕逐鶏雀聲 [總彙。1-1. a3]。		
a ta 𐄂 onom. [1889 / 2035] くんずほぐれつ (亂闘のさま)。わいわい。がやがや。亂闘狀 [5. 政部・争闘 1]。		
a ta seme 亂闘狀 [總彙。1-1. a4]。		
aba 𐄂 𐄂 𐄂 [3763 / 4043] 卷狩。卷狩の列。攷獵 [9. 武功部 2・攷獵 1]。ad. [5827 / 6233] 何處に。在何處 [12. 人部 3・問答 1]。／打圍○田獵○何在○那裡 [總彙。1-3. b1]。何在。那裡。田獵。問人口氣也。yargiyan i emu jalan i baturu bihe kai, te aba 此固一世之雄也而今安在哉 [全。0107b3]。𐄂 tereci emu abade, emu nirui niyalma juwe ilan jergi fere de bahafi dosime yabuha : それから一く狩獵>に一 niru の者は二三度、圍底に入り行くことができるようになった [老。太祖。4. 30. 萬曆。43. 12]。𐄂 abaci tucike manggi, jai amcane arcame feksifi gabta : <狩りの囲み>から獣が出たら、また追いかけ先回りしてさえぎり、馳せて射よ [老。太祖。4. 32. 萬曆。43. 12]。𐄂 alin be aba sindafi genere de : 山でく狩りの囲みを>配置して行くとき [老。太祖。4. 45. 萬曆。43. 12]。		
		aba abalambi 𐄂 tefi muse giyahūn maktame aba abalame yabuki : 我等が住んで鷹を放ちく狩猟しに>行く [老。太祖。7. 22. 天命。3. 9]。
		aba bargiyambi 𐄂 𐄂 v. [3869 / 4152] 巻き狩りを収める。巻き狩りを終わる。巻き狩りの陣の両翼端が離れて、後にもどる。收圍 [9. 武功部 2・攷獵 3]。／收圍乃尾獵兩頭散了 [總彙。1-3. b4]。
		aba i ba muse baita akū bai bisire anggala, aba i bade genefi buthašame yabuki : 我々は仕事もなくぶらぶらしているくらいなら、<圍場に>行って狩りをしよう [雍正。佛格。233C]。
		aba saha 𐄂 𐄂 𐄂。巻き狩り。田獵○打圍○與aba 同 [總彙。1-3. b1]。
		aba sarambi 𐄂 𐄂 v. [3800 / 4082] 巻き狩りの陣を展開する。囲みの陣を立てるため、左右兩端の勢子がそれぞれの方向に進む。撤圍 [9. 武功部 2・攷獵 2]。／圍場兩邊分開行 [總彙。1-3. b2]。
		aba sindambi 𐄂 𐄂 v. [3799 / 4081] (巻き狩りの陣立てをすると) 勢子を繰り出す。中央の旗から一人一人勢子を繰り出して左右兩翼を作って行く。放圍 [9. 武功部 2・攷獵 2]。／放圍乃從中獵兩邊一個個放去也 [總彙。1-3. b1]。𐄂 tere dabori aba sindafi : その夜、<狩り (をするように) 囲みを配置し> [老。太祖。8. 17. 天命。4. 3]。𐄂 aba sindambi : 狩りの囲みを展開する。𐄂 emu alin be aba sindame kafi : 一山を<狩りの囲みを展開する>ように囲み [老。太祖。8. 28. 天命。4. 3]。𐄂 aba sindame adafi gemu waha : <巻き狩りのように囲みを置き>列んでみな殺した [老。太祖。10. 11. 天命。4. 6]。
		abai ba abai babe giyarire janggan guwanyembooo : <圍場>を巡察する章京 親音保 [雍正。佛格。550B]。
		abai ebubun 𐄂 𐄂 n. [17341 / 18573] 歸妹。易卦の名。兌の上に巽の重なったもの。歸妹 [補編卷 1・書 2]。／歸妹易卦名○兌上震曰——[總彙。1-3. b2]。
		abalabumbi 𐄂 𐄂 v. [3766 / 4046] 卷狩りをさせる。狩りをやらせる。使打圍 [9. 武功部 2・攷獵 1]。／使打圍○使打獵 [總彙。1-3. b3]。
		abalahahai 正行打圍 [全。0107b4]。
		abalambi 𐄂 𐄂 v. [3765 / 4045] 卷狩りをする。狩獵をする。打圍 [9. 武功部 2・攷獵 1]。／打圍○打獵 [總彙。1-3. b2]。打獵 [全。0107b4]。
		abalame genehe 巡幸 [清備。兵部。7b]。
		abalame meljembi 獵較 [全。0107b4]。
		abalame yabure 遊畋 [清備。兵部。7b]。
		abalanambi 𐄂 𐄂 v. [3767 / 4047] 卷狩りにいく。狩獵に出かける。去打圍 [9. 武功部 2・攷獵 1]。／打圍去 [總彙。1-3. b3]。

Ⅳ. さ い ご に

満族語・日本語辞書が完成したあとは、さらに同じデータから日本語・満族語辞書の作成を行う予定である。ただし、別形式のデータについては、例えば、

満族語単語のローマ字表現

その単語を用いたローマ字表現での例文

の形式のものがあり、日本語意味の部分およびカタカナでの日本語読みの部分がないために、そのままでは日本語・満族語辞書に利用することはできない。カタカナでの日本語読みは、日本語見出しを辞書式順序に並べるために必要なものである。また、

満族語単語のローマ字表現

その単語の中国語訳。日本語訳。

のように、中国語意味と日本語意味が区別できない形で入力されており、カタカナでの日本語読みもないファイルについても日本語訳を適切に取り出せないもので、このようなものも日本語・満族語辞書に利用することはできない。

このことはそのような別形式で作成したファイルも満族語・日本語辞書用のデータであるとお聞きしたときに、そのままの形式では日本語・満族語辞書に取り込むための適切な処理ができないということをお伝えしたが、まず、今のままの形で進めて満族語・日本語辞書を完成させることを優先するということであった。

したがって、日本語・満族語辞書のためには、辞書データの変更が必要であり、完成にはもう少し時間がかかりそうである。

参考文献

- [本田・山田 1990] 本田道夫, 山田勇「言語学研究へのパーソナルコンピュータの応用」香川大学経済論叢第 63 巻第 2 号
- [本田・吉岡・山田 1991] 本田道夫, 吉岡珠実, 山田勇「スラブ系・ラテン系の言語研究のための基礎システム」香川大学経済論叢第 64 巻第 2・3 号
- [本田・今井 1995] 本田道夫, 今井慈郎「日本語・満州語の辞書作成のためのシステム (I)」香川大学経済論叢第 67 巻第 3・4 号
- [本田 1998] 本田道夫「日本語・満州語の辞書作成のためのシステム (II)」香川大学経済論叢第 71 巻第 3 号
- [本田 2003] 本田道夫「日本語・満州語の辞書作成のためのシステム (III)」香川大学経済論叢第 76 巻第 1 号
- [本田 2006] 本田道夫「日本語・満州語の辞書作成のためのシステム (IV)」香川大学経済論叢第 78 巻第 4 号
- [本田 2006] 本田道夫「日本語・満族語の辞書作成のためのシステム (V)」香川大学経済学部研究年報 46
- [本田 2008] 本田道夫「日本語・満族語の辞書作成のためのシステム (VI)」香川大学経済学部研究年報 48
- [本田 2011] 本田道夫「日本語・満族語の辞書作成のための補助システム (VII)」香川大学経済学部研究年報 51