

「歴代宝案」データベースと検索システム

柴山 守：大阪市立大学・学術情報総合センター・教授

琉球国における『歴代宝案』は、明・清朝や李氏朝鮮、暹羅国など東南アジア諸国との間で通商貿易等のために交わされた外交文書の集成であり、第一集から第三集、別集、目録の計270巻として編集されている。この宝案は、「関係史料が十分でない琉球国の歴史、特にその対外関係史の基本史料として重要である」と共に四百数十年に渡る「前近代の東アジアの国際関係のあり方、いわゆる冊封体制・朝貢貿易を知る上で貴重な文献である」とされ、内外の研究者に注目されると同時に現在も数多くの研究が進められている。また、この宝案文書は、当時の公用文である漢文文書であると共に特殊な様式と構成をもつ公文書であることから、「難解な宝案文書の解読には、集団的な英知を結集する必要があり、正確な解読のためにはその様式と構成の特徴を把握しておくことが不可欠の前提である」と指摘されている。

本重点領域研究「沖縄の歴史情報研究」では、この『歴代宝案』文書のデータベース化において

- (1) 県立図書館歴代宝案編集室で入力された第一集、第二集の校正作業（星野、柴山）、
- (2) 『歴代宝案』台湾大学本の画像データベース化（星野）
- (3) 『歴代宝案』テキストのための外字ファイルの生成（柴山）

を行い、同時に

- (4) 本文テキストに対し、読み下し文をベースにした構文解析の試み（柴山）、
- (5) 語彙検索のための検索システム（DOS版、Windows3.1版、Windows95版、WWW版）の開発（柴山）

を進めた。以下に、これらの各々の内容について報告する。

(1) 『歴代宝案』第一集、第二集の校正作業

沖縄県立図書館歴代宝案編集室（平成7年当時）から提供された『歴代宝案』第一集、第二集の本文テキストに足して、『歴代宝案』校訂本を底本として校正を進めた。平成9年12月現在、第一集については、後述する外字作成を残してほぼ完了した。第二集は、約40%の進捗状況である。

(2) 『歴代宝案』台湾大学本の画像データベース化

『歴代宝案』台湾大学本約8,900頁のすべてのページについてスキャナー入力を行い、CD-ROM版『歴代宝案』画像データベース（Mac版）を作成した。その後、すべての画像データに対してファイル変換を行い、ISO-9660フォーマットによるWindows版を作成、Netscape等のWWWブラウザを介して参照することができる。この画像データの例を図1に示す。

また、本『歴代宝案』台湾大学本に対応したページ情報を『歴代宝案』テキストに付加した検索システムを開発した。本システムを利用することによって、検索結果に表示されるページ数から当該の検索語を含む画像データを参照することが出来る。

自永樂二十二年
至萬曆四十八年

歷代寶案

卷第一
一集

歷代寶案 一集 目錄

詔勅

卷之一

起永樂二十二年
至康熙九年

卷之二

起弘治七年
至崇禎十七年

卷之三

起順治四年
至康熙三十二年

禮部咨文

卷之四

起宣德二年
至崇禎十三年

卷之五

起順治六年
至康熙八年

卷之六

起順治十一年
至康熙十年

卷之七

起康熙十二年
至康熙五十四年

福建布政使司等咨

卷之八

起成化十五年
至天啓七年

卷之九

起崇禎元年
至崇禎十二年

卷之十

起康熙十一年
至康熙三十四年

卷之十一

起康熙三十五年
至康熙三十六年

表奏

図1 『歷代寶案』(台湾大学本)画像データベースの表示例

(3) 『歴代宝案』外字ファイルの作成

『歴代宝案』本文テキストのデータベース化に際して、JIS第1水準、第2水準に含まれない多くの外字を作成しなければならない。第一集の校正作業では、約500字が必要になった。この外字ファイルを作成する手順は、つぎのとおりである。なお、詳細については「41.04『歴代宝案』の情報化と漢字処理」で述べる。

(a) IRIZ漢字ベースを用いて、JIS第1水準、第2水準に含まれないフォントを抽出する。

(b) 抽出したフォント(.bmpファイル)をWindows3.1版 - 外字フォントファイルに登録する。

(c) Windows3.1版 - 外字フォントファイル "USERFONT.FON" をフロッピーディスクに複写する。 (d) 上記の "USERFONT.FON" ファイルを配布し、Windows95日本語版上で標準提供されている外字エディタを使ってロードする。 以上の操作で、校正された本文テキストの外字は、Windows95版 のアプリケーション上で表示される。

(4) 『歴代宝案』構文解析の試み

『歴代宝案』第一集(文書数 1,037通)に対して、那覇市史の資料編として編纂された第一巻四「歴代宝案第一集抄」に、原文と共に読み下し文、語句注等 423通が掲載されている。この文献をベースに原文から読み下し文を自動生成することを目指して、まず原文を語単位に分かち書きする形態素解析を行うこと、また読み下し文から自動的に辞書を作成することを試みた。対象は、前述の那覇市史資料編第一巻四「歴代宝案第一集抄」から計算機への入力完了した暹羅国との文書7通に対して辞書を作成し、その内の2原文についての形態素解析である。

構文解析を行うための読み下し文には、那覇市史資料編第一巻四「歴代宝案第一集抄」から暹羅国との間で交わされた文書20項から、表1に示す7つの咨文を用いる。ここで、「咨文」とは同級の官庁または同品の官吏が相互に移送する公文書である。

表1 構文解析の辞書生成に用いた宝案文書の咨文

書No.	文 書 内 容	文
14	中山王(尚巴志)より暹羅国あて、磁器・蘇木等の自由な売買の許可についての咨文(1425年)	
15	中山王(尚巴志)より暹羅国あて、進貢品の収買についての咨文(1425年)	
22	中山王(尚巴志)より暹羅国あて、返礼の使者派遣についての咨文(1426年)	
37	(暹羅国王)より(琉球国王)あて、礼物の奉献についての咨文(1430年)	
53	中山王(尚巴志)より暹羅国あて、自由な売買の許可を乞う咨文(1431年)	
59	中山王(尚巴志)より暹羅国あて、貿易等についての咨文(1432年)	
60	中山王(尚巴志)より暹羅国あて、貿易についての咨文(1432年)	

表1に示す咨文の中から、文書No.15について原文及び読み下し文を図2、及び図3に示す。

まず、図3に示した読み下し文から、辞書の自動生成を行う。このために読み下し文から、見出し語、付属語、送りがな等を自動的に抽出しなければならない。これは、読み下し文を単なる漢字かな混じりから成る記号列と考える。

15 琉球国中山王為
 進貢事切照本国稀少貢物為此今遣正使
 浮那姑是等坐駕仁字号海船裝載磁器
 前往
 貴国出産地面収買胡椒蘇木等貨回国
 以備
 進貢
 大明御前仍備礼物
 詣前奉獻少伸遠意幸希收納仍煩聽今
 差去人員及早打發*趁風迅回国庶使
 四海一家永通盟好今將奉獻礼物数目
 開坐于後須至咨者
 今開
 織金段五匹 素段二十匹
 腰刀五柄 摺紙扇三十柄
 硫黄五千斤今報二千五百斤正
 大青盤二十箇 小青盤四百箇
 小青碗二千箇
 右 咨
 暹羅国
 洪熙元年 月 日
 咨

15 琉球国中山王、進貢の事の為にす。
 切に照らすに、本国は貢物稀少なり。
 此が為、今、正使浮那姑是等を遣わし、
 仁字号海船に坐駕して、磁器を装載して、
 前みて貴国の出産地面に往き、胡椒・
 蘇木等の貨を収買して回国し、以て大明
 御前への進貢に備えんとす。仍、礼物を
 備えて詣り前みて奉獻し、少しく遠意を
 伸ぶ。幸希わくは收納せられよ。なお、
 煩わくば、今、差去わす人員は早きに及
 びて打發し、風迅に*趁して回国するを
 聽さば、四海一家永く盟好を通ぜしむる
 に庶からん。いま、奉獻の礼物数目を將
 て後に開坐す。須く咨に至るべき者なり。
 今、開す
 織金段五匹 素段二十匹
 腰刀五柄 摺紙扇三十柄
 硫黄五千斤 今報じたるは二千五百斤正
 大青盤二十箇 小青盤四百箇
 小青碗二千箇
 右、暹羅国に咨す
 洪熙元年 月 日
 咨す

図2 文書No.15 の原文

図3 文書No.15 の読み下し文

実行した結果を、表2に示す。暹羅国に関する7咨文の総単語数は、889語である。

表2 読み下し文からの辞書の自動生成結果

単語No.	見出し語	単語	付属語	句読点
		[略]		
227,	琉球国中山王,	琉球国中山王,	,	,
228,	進貢,	進貢の,	の,	,
229,	事,	事のためにす,	のためにす,	。
230,	切,	切に,	に,	,
231,	照,	照らすに,	らすに,	,
232,	本国,	本国は,	は,	,
233,	貢物稀少,	貢物稀少なり,	なり,	。
		[略]		

暹羅国との7咨文、総単語数 889語
 一意語数 532語

表2の結果から一意な語を抽出し、分類した結果の単語数は532語である。つぎに前述の辞書を用いて、原文に対する形態素解析を行う。形態素解析は原文と辞書の見出し語を比較参照しながら、最長一致法(1

ongest match) を用いる。形態素解析において、入力した文書No.15 の原文に対する分かち書きの結果を図4 (一部) に示す。

```

*Total : 539
Text Length : 193
W:0001 [琉球国中山王]-[琉球国中山王]-【琉球国中山王】 L:06
W:0002 [為]-【為】 L:01
W:0003 [進貢]-[進貢]-【進貢】 L:02
W:0004 [事]-[事]-[事]-【事】 L:01
    - 略 -

W:0037 [奉献]-[奉献]-[奉]-[奉献]-[奉献]-[奉献]-[奉献]-【奉献】 L:02
W:0038 [少]-【少】 L:01
W:0039 [伸]-[伸]-【伸】 L:01
W:0040 [遠意]-【遠意】 L:02
W:0041 [幸希]-[幸希]-【幸希】 L:02
W:0042 [収納]-[収]-【収納】 L:02
W:0043 [仍]-【仍】 L:01
    - 略 -

```

図4 分かち書きの結果 (一部)

文書No.15 の形態素解析の結果は、入力文字数 193字、辞書の見出し語参照回数 539回で、分かち書きによって抽出された単語数は88語であった。また、図4において、記号 [] で囲まれた字は、辞書参照によって出現した見出し語を示す。記号【】は単語として認識した結果の字を示している。

(5) 『歴代宝案』テキスト検索システムの開発

本文テキストから語彙を検索する『歴代宝案』テキスト検索システムの開発を行った。本検索システムの特徴は、(a) 検索したい語に対してすべて部分一致で検索する。(b) 表示結果は、KWIC形式である。(c) 表示結果の【】は原書上での所在、は第一集抄での文書番号を示す。なお、本文テキストには上記の所在情報、文書番号をタグとして、予め挿入しておくが必要である。

DOS版検索システムの例 (図5) を以下に示す。

* 『歴代宝案』テキスト検索システム *

- 検索語：字号船

検索結果 [字号船]

【卷16-4】	13	部洪熙元年十二月十七日仁	字号船	進貢馬式拾疋硫黄壹万
【卷16-4】	13	伍阡斤大正使者浮那姑是盤	字号船	進馬式拾疋硫黄壹万斤
【卷16-10】	27	也馬二十五疋硫黄五千斤天	字号船	使漫泰束結制馬十五疋
【卷16-17】	47	起三隻正事林喜齋表右咨荒	字号船	硫黄一万斤馬五疋永字
【卷16-17】	47	字号船硫黄一万斤馬五疋永	字号船	馬二十五疋礼部義字号
【卷16-17】	47	永字号船馬二十五疋礼部義	字号船	馬二十五疋宣德六年九

[略]

* 『歴代宝案』テキスト検索システム *

- 検索語：字号？船

検索結果 [字号？船]

【巻16-4】	13	齋表箋文各壱通及坐駕仁盤	字号 海*式隻裝載馬肆拾疋硫
【巻16-4】	13	須至咨者一件番貨事所有各	船 附搭蘇木等物煩為乞賜免抽
【巻40-3】	22	部洪熙元年十二月十七日仁	字号 船進貢馬式拾疋硫黄壱万
【巻40-6】	31	熙元年十二月十七日仁	船 進貢馬式拾疋硫黄壱万斤小
			[略]
【巻40-3】	22	特遣正使南者結制等坐駕義	字号 海船一隻前詣貴国奉獻少
【巻40-6】	31	使南者結制等坐駕義	船 一隻前詣貴国奉獻少伸酬謝
		此今遣正使実達魯等坐駕天	字号 海船壱隻裝載磁器等物前
		正使実達魯等坐駕天	船 壱隻裝載磁器等物前往貴国
			[以下、略]

図5 『歴代宝案』テキスト検索システムの実行例

本検索システムにおいて「字号?船」における検索は、『仁』字号船、『恭』字号船、『天』字号船などの字号船の直前に現れる字、及び「字号海船」の検索にも有用である。また、「両平収買」による検索と「両平?収買」による検索結果に違いがある。

つぎに、Windows95版検索システムの例(図6)を掲げる。

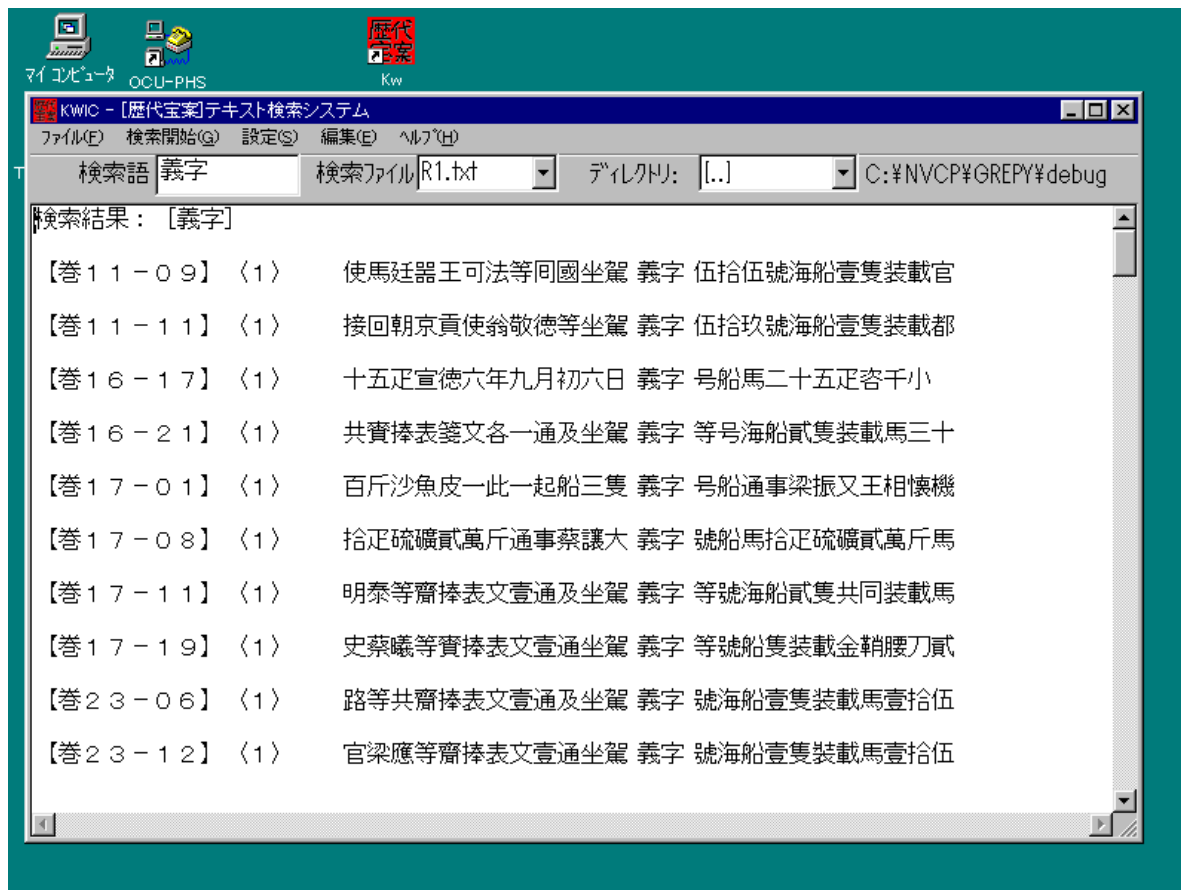


図6 Windows95版『歴代宝案』テキスト検索システム

図6の検索システムでは、ウインドウ上部の「ディレクトリ」にテキストが存在するディレクトリを指定する。「検索ファイル」では、検索対象となるテキストのファイル名を指定する。「検索語」には、検

索する語を指定し、例では”義字”を指定している。[検索開始(G)]をクリックすると検索が開始され、結果が表示される。また、デスクトップ上のアイコン[歴代宝案]は、プログラム起動用アイコンである。