

「近世琉球語彙索引」作成の試み

柴山 守：大阪市立大学学術総合情報センター

「琉球産業制度資料」語彙索引の作成

「琉球産業制度資料」データベース化に際して、コンピュータ利用による『琉球産業制度資料』計算機索引、すなわち語彙索引の作成を試みた。手順は、つぎに示すとおりである。

- (1) 「琉球産業制度資料」は、「桐」TBL形式で格納されている。これをASCII形式ファイルに変換する。
 - (2) 変換された原文テキストから、漢字から特殊記号、漢字から平仮名、漢字から数字、特殊記号から漢字、平仮名から漢字、特殊記号から数字等への変移点をコンピュータ・プログラムにより見つけて自動セグメンテーションを行う。
 - (3) セグメントされた語彙・文節は適切な単位ではなく、後に人手によるセグメント作業が必要である。そのために、コンピュータを使って分類・排列（ソート）を行う。
 - (4) 分類・排列された語彙・文節は同一語彙や文節が複数個出現するために、一意な語彙・文節にプログラムによりまとめる。これで、「語彙リスト」ができあがる。
 - (5) ただし、本「語彙リスト」は適切な単位にセグメントされていないために、人手によるセグメント作業を行う。この作業の後、再度(4)の作業を繰り返す。
 - (6) これまでの作業で完成した「語彙リスト」を用いてテキスト検索を行い、語彙索引を作成する。
- 『琉球産業制度資料』計算機索引は、以上のような手順にて作成する。平成9年12月現在、上記(5)の段階の作業を進めている。以下に、これらの主な例を挙げる。

(1) 「琉球産業制度資料」原文テキスト

「琉球産業制度資料」原文テキストの一部を図1に示す。本テキストは、「桐」TBL形式からDOSテキスト（ASCII）形式に変換したものである。

村々地与合之儀、田畑耕方其外諸上納物調兼候節致補助肝要之事にて、去亥年表御方御案内之上、地与合相立置候処、至当分は与々強弱出来候由候間、此節無親疎致与直、其首尾可申出事（咸豊三〔癸〕丑年（嘉永六年）十一月十七日恩納間切御手入日記）／琉球産業制度史料、#第一巻土地制度（近世地方経済史料・沖縄県・\$第九巻）／4／
村々地割之儀、田畑共上中下差分け、与々鬮取を以て致割符〔府〕於与々も右同断鬮取を以家内之厚薄人居之多少見合、分々に応し持地可相授事。附、本地山野之内より面々はんた所持いたし候田畑地割之砌、此中持口通致割符〔府〕候ては以後不締之体可相立候間、此節之儀は都て此中之持口取替致地割候様丑十一月十七日／琉球産業制度史料、#第一巻土地制度（近世地方経済史料・沖縄県・第九巻）／4／

図1 「琉球産業制度資料」原文テキスト

(2) プログラムによる自動セグメンテーション

自動セグメンテーションの結果を図2に示す。セグメンテーションは、漢字、平仮名、記号、数字等の状

態変移点でセグメントする。

- 1, 村々地与合之儀, 村々地与合之儀, , 、
- 2, 田畑耕方其外諸上納物調兼候節致補助肝要之事, 田畑耕方其外諸上納物調兼候節致補助肝要之事にて, にて, 、
- 3, 去亥年表御方御案内之上, 去亥年表御方御案内之上, , 、
- 4, 地与合相立置候処, 地与合相立置候処, , 、
- 5, 至当分, 至当分は, は,
- 6, 与々強弱出来候由候間, 与々強弱出来候由候間, , 、
- 7, 此節無親疎致与直, 此節無親疎致与直, , 、
- 8, 其首尾可申出事, 其首尾可申出事,
- 9, 咸豊三〔癸〕丑年, 咸豊三〔癸〕丑年, , 。
- 10, 嘉永六年, 嘉永六年, , 。
- 11, 十一月十七日恩納間切御手入日記, 十一月十七日恩納間切御手入日記, , 。
- 12, 琉球産業制度史料, 琉球産業制度史料, , 、
- 13, 第一巻土地制度, 第一巻土地制度, , 。
- 14, 近世地方経済史料, 近世地方経済史料, , 。
- 15, 沖縄県, 沖縄県, , 。
- 16, 第九巻, 第九巻, , 。
- 17, 4, 4, , 。

図2 自動セグメンテーションの結果

図2において、左端から順に出現順序番号、見出し語、語彙、送りかな、句読点である。

(3) 語彙の分類・排列（ソート）

セグメントされた語彙から一意な語彙を得るために、分類・排列（ソート）を行う。この結果を図3に示す。

- 81652, 亜宝山, 亜宝山, , 。
- 44917, 阿嘉親雲上, 阿嘉親雲上, , 、
- 45051, 阿嘉親雲上, 阿嘉親雲上, , 、
- 28872, 阿嘉村構宇佐嶺山二百五十坪全焼失候不届,
阿嘉村構宇佐嶺山二百五十坪全焼失候不届に, に,
- 28878, 阿嘉村山当山真境名, 阿嘉村山当山真境名, ,
- 28868, 阿嘉村赤頭赤田仁屋妻, 阿嘉村赤頭赤田仁屋妻, ,
- 45583, 阿波根親雲上, 阿波根親雲上, ,
- 59959, 阿波根里之子親雲上, 阿波根里之子親雲上, ,
- 74596, 阿波茶仲間真栄田三, 阿波茶仲間真栄田三ヶ, ヶ,
- 54922, 逢, 逢, ,
- 82760, 逢難船, 逢難船に, に,
- 86570, 逢難船致打荷大島迄取着候付, 逢難船致打荷大島迄取着候付, , 、
- 85771, 逢不申無是非間, 逢不申無是非間も, も,
- 69304, 悪鬼納并諸島高八万九千八拾六石之内,
悪鬼納并諸島高八万九千八拾六石之内, , 、
- 14624, 悪砂糖煎込有之候間, 悪砂糖煎込有之候間, , 、
- 14736, 悪砂糖煎込有之候間, 悪砂糖煎込有之候間, , 、
- 14927, 悪砂糖入交候儀無之儀, 悪砂糖入交候儀無之儀と, と,
- 14877, 悪砂糖入交候者, 悪砂糖入交候者も, も,

14652, 悪砂糖不交入様検見, 悪砂糖不交入様検見させ, させ,
 34563, 悪水土不流入様可仕事, 悪水土不流入様可仕事, , 。
 86590, 灘状申請直, 灘状申請直に, に,

図3 セグメントされた語彙の排列(ソート)結果

図3において、同一の語彙が複数出現していることが判る。

(4) 一意な語彙の抽出

前述のセグメントされ、ソートされた語彙は、一意でない。したがって、これを入力ファイルとしたプログラムにより、一意な語彙とその頻度を求める。これを図4に示す。

000001	5	々
000002	1	々々へ
000003	1	々より
000004	1	々右帳立添本文同断可為致次渡候
000005	1	々科律并罪科禁日法義込る
000006	1	々休或調方
000007	1	々上納にて
000008	1	々申渡置処
000009	1	々道具帳に
000010	1	々二度づゝ
		[略]
002587	1	伊野波村居住若狭町村次男
002588	1	伊野波村崎原筑親雲上イク
002589	1	伊野波村百姓カナ
002590	2	伊野波里之子
002591	1	伊良皆村
002592	1	伊良波
002593	1	伊良部
002594	1	伊礼親雲上
		[略]
051261	1	饒辺名村屋我地松敷五敷
051262	2	饒辺名二ヶ
051263	1	騙し
051264	1	鬪を
051265	1	鬪取様は
051266	2	鬪先の
051267	1	鬪当候面付相定申候
051268	1	鬪入を
051269	1	麿相有之代銀高直に
051270	1	麿末之拳動共於有之は
051271	3	・手形・
051272	1	・手形・地頭代認可の

図4 一意な語彙リストと出現頻度

図4において、左端から順に順序番号、出現頻度、語彙を示す。この語彙リストから、入力テキストに出現する語彙数は、51,272語であることが判る。

(5)、(6)については、省略する。

. WWW による「琉球産業制度資料」検索システムの実現（柴山）

「琉球産業制度資料」の語彙が、インターネットの WWW を介して検索できれば、研究情報の発信と共に資料の共有の視点からも重要であり、沖縄・琉球研究の進展に大いに貢献する。

「琉球産業制度資料」本文テキストを WWW サーバ上に格納しておき、インターネット上の WWW ブラウザから検索のための語彙が指定されるとサーバ上で検索を行い、検索結果を KWIC 形式で WWW ブラウザ上に表示する。また、KWIC 形式で表示された結果に付加して表示される「位置」のリンク情報をクリックすると当該の本文テキストが表示される。

ホームページを図 5 に示す。

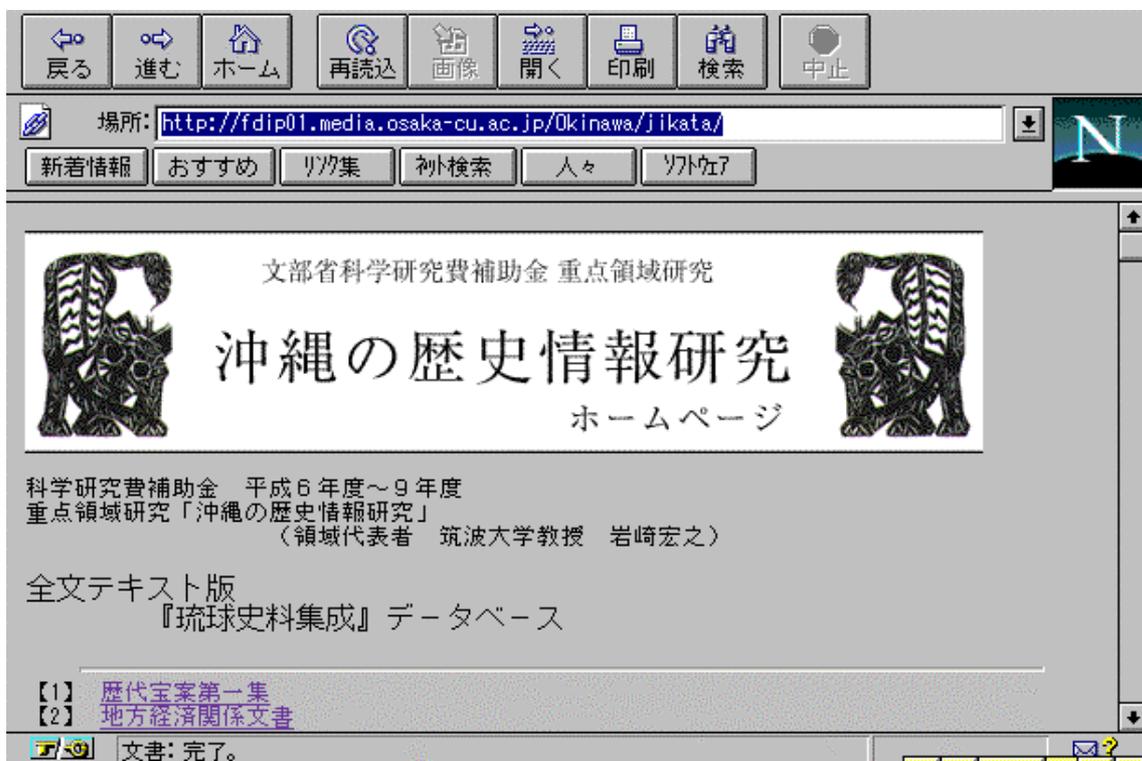


図 5 「琉球産業制度資料」検索システムのホームページ

図 5 において、「【2】地方経済関係文書」をクリックすると「琉球産業制度資料」テキスト検索ページが表示される。本検索ページでは、検索する語彙を入力し、[検索]ボタンをクリックする。検索結果が表示される。これを図 6 に示す。

図 7 は、図 6 に表示されている KWIC 形式の検索結果にリンクされた本文テキストの結果を表示している。例では、[*67*]の箇所をクリックした結果である。

場所: <http://fdip01.media.osaka-cu.ac.jp/Okinawa/jikata/jikata00.html>

検索結果 [砂糖]

[#20*](#) #08
一 砂糖 七十二斤八合二勺六才

[#67*](#) #08 年焼過 砂糖 多御払直成相下り及御不

[#67*](#) #08 申談候処、諸間切并伊江島 砂糖 之儀、去亥年御相用候所

[#67*](#) #08 相細め候はゞ、自から焼過 砂糖 出来少御払宜成相進可申

[#67*](#) #08 之、御元上納より相成候 砂糖 之儀、間切間切多少又は

[#67*](#) #08 之、親疎相見得申候間、右 砂糖 七十五万斤之分は、以前

《地方経済》データベース 検索ページ

<検索例>
 (1) キーワード指定欄にキーワードを入力します。
 (2) 検索ボタンを押下します。
 (3) つぎのページで、以下に示すような検索結果が表示されます。

* 《地方経済》テキスト検索システム [V1.01] *
 * 検索結果 [琉球] *

[フレーム1へ戻る](#)

文書: 完了。

図6 検索ページと語彙「砂糖」の検索結果

場所: <http://fdip01.media.osaka-cu.ac.jp/Okinawa/jikata/jikata00.html>

検索結果 [砂糖]

[#20*](#) #08
一 砂糖 七十二斤八合二勺六才

[#67*](#) #08 年焼過 砂糖 多御払直成相下り及御不

[#67*](#) #08 申談候処、諸間切并伊江島 砂糖 之儀、去亥年御相用候所

[#67*](#) #08 相細め候はゞ、自から焼過 砂糖 出来少御払宜成相進可申

[#67*](#) #08 之、御元上納より相成候 砂糖 之儀、間切間切多少又は

*Specified Record : 67
 【67】
 #00 0067
 #01 甘蔗敷地制限の事
 #02 1860.1
 #03
 #04
 #05
 #06 湧川里之子親雲上、他
 #07 各間切島下知役検者、他
 #08 近年焼過砂糖多御払直成相下り及御不勝手に候間、例相宜敷地減少方吟味を以可申上旨被仰渡、役人中へも吟味させ私共申談候処、諸間切并伊江島砂糖之儀、去亥年御相用候所も有之、又は去申年御例被仰付置候処も有之候処、亥年御例之儀、申年御例より格別相減しに付、一統亥年御例に被仰付敷地相細め候はゞ、自から焼過砂糖出来少御払宜成相進可申と奉存候、尤其通被仰付事候はゞ、依

《地方経済》データベース 検索ページ

<検索例>
 (1) キーワード指定欄にキーワードを入力します。
 (2) 検索ボタンを押下します。
 (3) つぎのページで、以下に示すような検索結果が表示されます。

* 《地方経済》テキスト検索システム [V1.01] *
 * 検索結果 [琉球] *

[フレーム1へ戻る](#)

文書: 完了。

図7 リンクされた本文テキストの検索結果