

古事類苑・地部 GIS データの作成

山田 奨治^{†1} 中西 和子^{†1} 尾方 隆幸^{†2}

『古事類苑』は、明治から大正にかけて出版された、本文 1,000 巻、和装本で 350 冊、洋装本で 51 冊の大百科事典である。そこには、前近代の文化概念について、明治以前のあらゆる文献からの引用が掲載されており、人文科学研究を行ううえでたいへん有用な事典として、いまでも利用されている。われわれは『古事類苑』「地部」の情報を地理情報システム (GIS) と連携させる試みを進めている。大正時代の「帝国図」をベースマップに採用して、島嶼部を含めて海岸線と郡境界を抽出し、ポリゴンとしてシェープファイル化したうえで、旧国郡名の属性情報を与えた。また「地部」の見出し語レベルに登場する人工物・自然地名の経緯度情報も作成し、シェープファイル化した。それぞれのシェープから、国際日本文化研究センターの「古事類苑全文データベース」へのリンクも作成した。『古事類苑』本文に記載されている地名を抽出する方法や「地部」の GIS 分析についても、基礎的な検討をはじめている。

Making GIS Data of *Kojiruien Chi-bu*

YAMADA SHOJI,^{†1} NAKANISHI KAZUKO^{†1}
and OGATA TAKAYUKI^{†2}

Kojiruien (The Dictionary of Historical Terms) is a kind of large-volume Japanese encyclopedia with 1,000 volumes in original form, 350 volumes in Japanese-style binding, and 51 volumes in Western-style binding, which was published from 1896 to 1914. It contains numerous words and cultural concepts of pre-modern Japan, citing volumes of books and documents before Meiji restoration (1867), and is frequently used by scholars of the Humanities, even in today. We made basic geographical information for using in the full-text database of *Chi-bu* (Volumes of Geography) of *Kojiruien*, which could be linked to Geographic Information System (GIS). We chose the 1:200,000 *Teikokuzu* map as the base-map for extracting shorelines and administrative areas. In addition, we measured the longitudes and latitudes of artificial/natural place-names. At the same time, a map-to-database query system of *Chi-bu* has been developed. We also conducted preliminary studies of extracting geographical names from the body text of *Kojiruien* and GIS analysis of *Chi-bu*.

1. はじめに

国際日本文化研究センター (日文研)・山田研究室では、国文学研究資料館 (国文研)・相田研究室と協力して、『古事類苑』の全文テキストデータベース化に取り組んでいる¹⁾。そのうち「地部」(洋装本 3 巻)の入力は日文研が担当し、第 1 巻は 2008 年 5 月に正式公開、第 2・3 巻は 2009 年 6 月に試験公開した²⁾。われわれは、「地部」の全文テキストデータベース化と並行して、その情報を GIS と連携させて利用するための、基盤となる情報資源作りを進めている。この報告では、これまで進めてきた作業の概要と今後の課題点について述べる。

2. 「地部」見出し項目への地理情報の付与

2.1 空間属性の分類

まず「地部」の見出し項目をリストアップし、それぞれに経緯度情報を与え、GIS (ArcGIS Ver.9.2) を使ってシェープファイル化した。海岸線と行政界については、歴史性と精度の双方を考慮して、20 万分の 1「帝国図」をベースマップに採用した³⁾。島嶼部については、島名の記述があることを基準に、一定程度の面積を有する島の海岸線を取得した。作業の進行上、データ化する島の選択は、ある程度、作業者の主観的な判断に委ねた。日本列島には海岸線の長さが 100 メートル以上の島は 6,852 あるといわれている。今回シェープファイル化した島の本数は 1,200 である。そのうち最小の島は長崎県の母子島で、ポリゴンでの海岸線長は約 224 メートルある。「帝国図」を選択したことに起因する課題点については後述する。

空間的な広がりを考慮して「地部」の見出し項目をポイント・ライン・ポリゴンの属性に分類した。ポイントデータは、点で代表させることが妥当と考えられる 14 項目 (橋・渡・津・泊・港・山・井・泉・温泉・瀧・岬・野・原・牧) である。これらのうち「井」「泉」「瀧」「岬」はあきらかに点としての性格が強い。「橋」「渡」は厳密には線であるが、その距離は長くないため、中心付近の点で代表させることにした。「津」「泊」「港」は面的な広がりがあるが、広範囲ではないため中心付近の点で代表させた。なお、「津」「泊」「港」の 3 種

^{†1} 人間文化研究機構・国際日本文化研究センター

International Research Center for Japanese Studies, National Institutes for the Humanities

^{†2} 琉球大学・教育学部

Faculty of Education, University of the Ryukyus

と「井」「泉」の2種は、同一のシェイプファイルにまとめた。「山」については、その山域全体が対象とされているとみることができ、山頂の1点で代表させた。複数のピークを持つ山は、最高峰で代表させた。「温泉」については、観光的・歓乐的・保養的な市街地の中心地付近で代表させた。「野」「原」「牧」には面的な広がりがあるが、その範囲を正確に確定することが困難である場合が多い。これらのうち地形的に範囲を設定できるものはポリゴンデータを作成し、できないものは中心点付近のポイントデータとして作成した。

ラインデータは、線的な属性を持つ8項目（道路・河・海・潟・江・磯・濱・浦）である。「道路」および「河」には、あきらかに線の性格がある。「海」「潟」「江」「磯」「浜」「浦」の6項目は、ラインとは必ずしもいえないが、海岸線をトレースしてシェイプファイルを作成した。

ポリゴンデータは、面的な属性を持つ項目のうち、行政界である「国」「郡」と、境界線を明瞭に区切ることができる内水域の「池」「湖」「沼」の3項目、および「野」「原」「牧」のうち地形的に境界線を引くことが可能なものを作成した。

2.2 位置の特定手順

まず「地部」に記述されている本文を熟読し、掲載されている地名のおおよその位置を推定した。参考資料として、平凡社刊行の『日本歴史地名大系』、および角川書店刊行の『角川日本地名大辞典』を使用した。このうち『日本歴史地名大系』は、そのデータベース・サービスを利用した⁴⁾。また、地形の判読が必要になる項目も多いので、東京大学出版会が刊行している『日本の地形シリーズ』（全8巻）も参考にした。

以上の作業によっても場所があきらかにならない項目も、少なからず存在する。そのような項目についてはWeb上の情報を活用して検索したが、こうした情報は必ずしも正しいとは限らず、その信憑性を厳しく評価されなければならない。検索にあたり、公的機関（地方自治体の観光関連部署や教育委員会、大学等の研究教育機関など）が提供している情報は一定の信憑性があるものと判断した。一方、個人の作成した私的なホームページによる情報はできるだけ避け、やむを得ず使用する場合、その信憑性を慎重に検討した。また、資料館や博物館の学芸員への聞き取り調査（電話による聞き取りも含む）も行った。位置について諸説ある場合は、複数の位置情報を作成した。当然のことながら、個々の位置特定で参考にした情報源は、参考情報としてすべて記録してある。どうしても位置を特定できなかった地名については、シェープ作成を保留してある。

上記のようにしておおよその位置を特定したあと、ポイント・ライン・ポリゴンごとに以下のような方法でシェープファイル化した。ポイントで表現できる地名については、国土地



図1 「帝国図」の郡境界
Fig.1 Gun borders on Teikoku-zu.

理院から公開されている「地図閲覧サービス(ウォッチず)」⁵⁾を利用して、経緯度を測定した。このサービスでは、画面に表示された2万5千分の1地形図の任意地点をマウスでクリックすると、その地点の経緯度値が秒単位まで表示される。

ただし、この方法には2つの大きな問題がある。第1に、「地図閲覧サービス」提供側からは、経緯度の数値はあくまで参考値であるとの警告がされている。この簡便な手法で秒単位の精度を求めるのは不可能であり、せいぜいプラスマイナス30秒程度の精度しかないと思われる。第2に、「地図閲覧サービス」で提供されている現代の2万5千分の1地形図と「帝国図」とでは精度に違いがあるため、前者で取得した経緯度を「帝国図」に落とすと位置があきらかにずれることになる。したがって、各ポイントの位置を「帝国図」上で修正することを当初の基本方針にしたが、作業過程でこの方針の妥当性に対する疑問も浮かんだため、現時点では「帝国図」上で位置を修正したポイントと、そうでないポイントが混在したデータになっている。

ラインとポリゴンで表現できる地名については、上記と同様の手順でおおよその位置を特定したあと、「帝国図」をベースマップにして、シェープファイル化した。たとえば「道路」については「帝国図」を用いて旧集落と旧街道の位置を判読し、トレースした。作業にあたり、吉川弘文館から刊行されている『街道の日本史』(全56巻)なども参照した。「河」は「帝国図」上で河道の中心線を結んで、ラインを作成した。ポリゴンのうち、国郡境界は「帝国図」上に表記されている境界をトレースし、同図上の国郡名を属性として付与した(図1)。ただし、島嶼部にかんしては国郡属性を与える作業は完了していない。

3. 「地部」情報のマッピング

以上のようにして作成したシェープファイルを利用して、「地部」をGISといかに結合するかについて検討を進めている。ここでは、3つの方向性について述べることにする。

第1に、「地部」洋装本第3巻にある自然・人工物地名の分布を地図化することを試みた。図2は、14項目のポイントデータ(橋・渡・津・泊・港・山・井・泉・温泉・瀧・岬・野・原・牧)、8項目のラインデータ(道路・河・海・潟・江・磯・濱・浦)、そして6項目のポリゴンデータ(池・湖・沼および野・原・牧の一部)と「帝国図」の海岸線を表示させたものである。一見してわかるとおり、これらの見出し項目は京・大坂・奈良に集中している。『古事類苑』が前近代の事項を収録していることを考えると、これは当然のことである。その京・大坂・奈良地域での自然・人工物地名の分布を詳しくみると、図3になる。

第2に、これら「地部」見出し項目シェープから日文研の「古事類苑全文データベース」

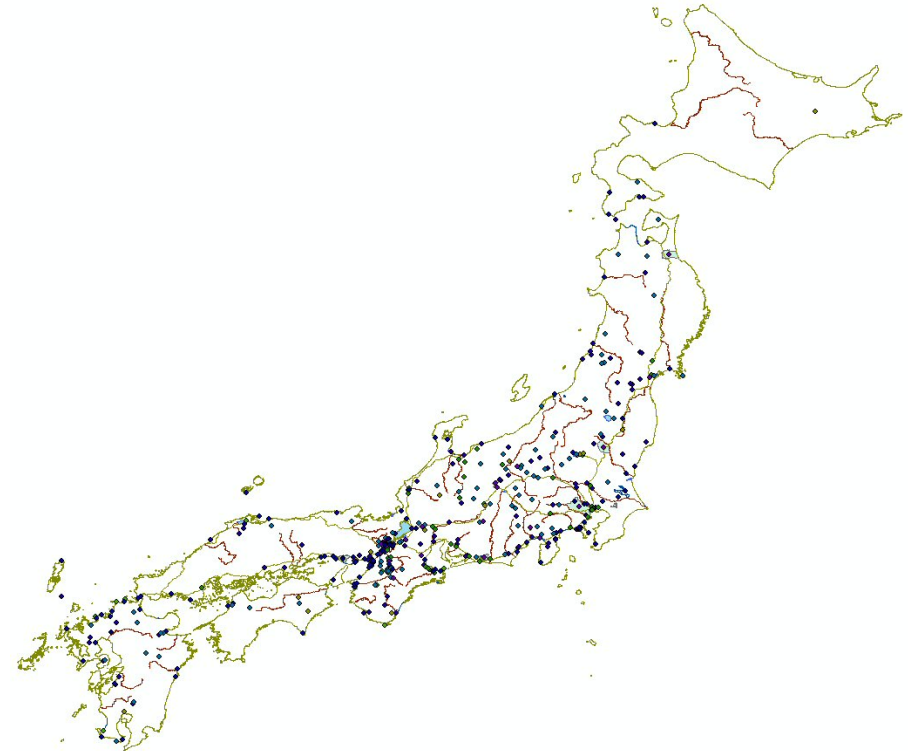


図2 「地部」自然・人工物地名の分布

Fig.2 Plot of natural and artificial geographical names in Chi-bu.

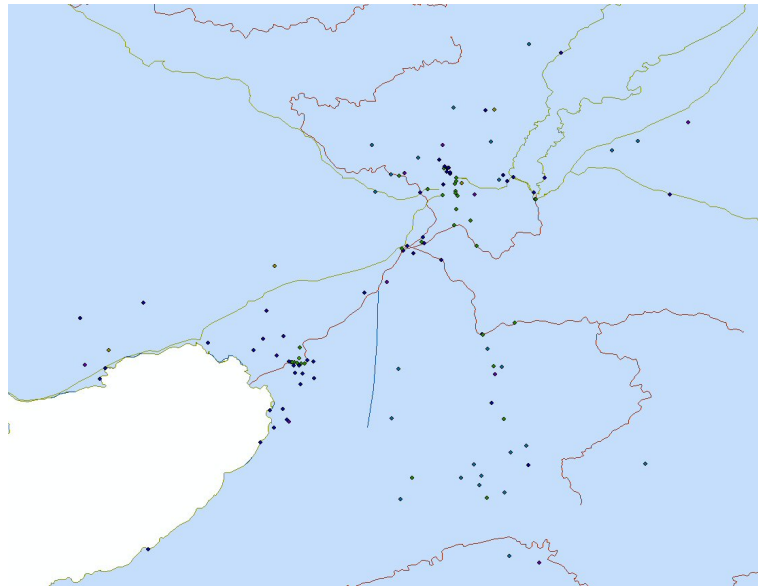


図3 京・大阪・奈良の自然・人工物地名の分布

Fig. 3 Distribution of natural and artificial geographical names in Kyoto, Osaka, and Nara.

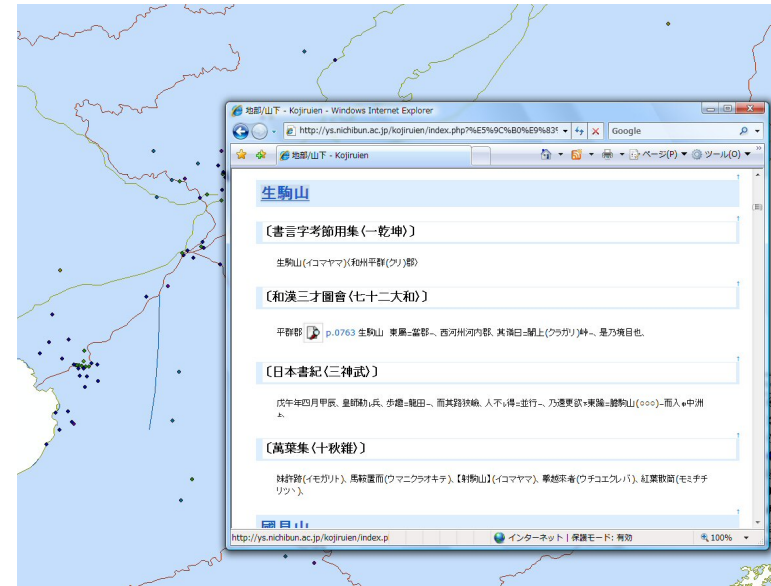


図4 GIS から『古事類苑』本文の表示例

Fig. 4 Link from GIS to *Kojiruien* full text.

へのリンクを作成し、GISの画面から『古事類苑』本文の記述を容易に表示することができるようにした。こういったシステムにより、『古事類苑』に対するより視覚的・直感的で多様な検索を実現することができるだろう(図4)。

第3に、『古事類苑』本文記述から抽出される何らかの統計量を、GISで分析する試みをはじめた。一例として、「地部」洋装本第1,2巻にある国別記述の情報量の多寡を知るために、各国についての資料本文の文字数を算出し、それを「帝国図」の国ポリゴンに属性として与えてグラフを表示させてみた(図5)。ただし資料本文の文字数は、武蔵国と山城国がそれぞれ約14万字、約9万3千字と図抜けて多いのに対して、3番目に多い陸奥国は約6万3千字、もっとも記述が少ない伯耆国は約5,500文字と情報量に開きがあるため、すべての国をグラフ化すると武蔵・山城国以外の情報量の多寡が視覚的に認知しにくくなる。それを避けるために、図5では、武蔵国と山城国を例外扱いして表示していない。また、陸奥国以北は『古事類苑』の陸奥国と「帝国図」のそれとは、領域があきらかに異なる

ので空白とした。島嶼部についても、技術的な問題で省略してある。

図5中の円の大きさは、資料本文の文字数に比例している。これをみると、武蔵国・山城国・陸奥国を別にすれば、どの地域について情報量が多いか、少ないかがひとめでわかる。情報量の偏りの原因をより詳しく分析すれば、『古事類苑』の特質や編纂事業の経緯・手法について、何らかの知見が得られるかもしれない。

4. 『古事類苑』と『大日本地名辞書』のマッチング

以上の作業は、「地部」の見出し語のレベルで位置情報を付与する試みであったが、将来的には資料本文中にあらわれる地名についても、同様の位置情報を作成するのが理想である。とはいえ、それを手作業で行うのは膨大な時間がかかり、現実的ではない。作業を効率化する有効な方法として、位置情報を備えた既存の地名データベースとのマッチングを行うことが考えられる。ここでは、吉田東伍の『大日本地名辞書』を桶谷猪久夫が電子化し経緯度情

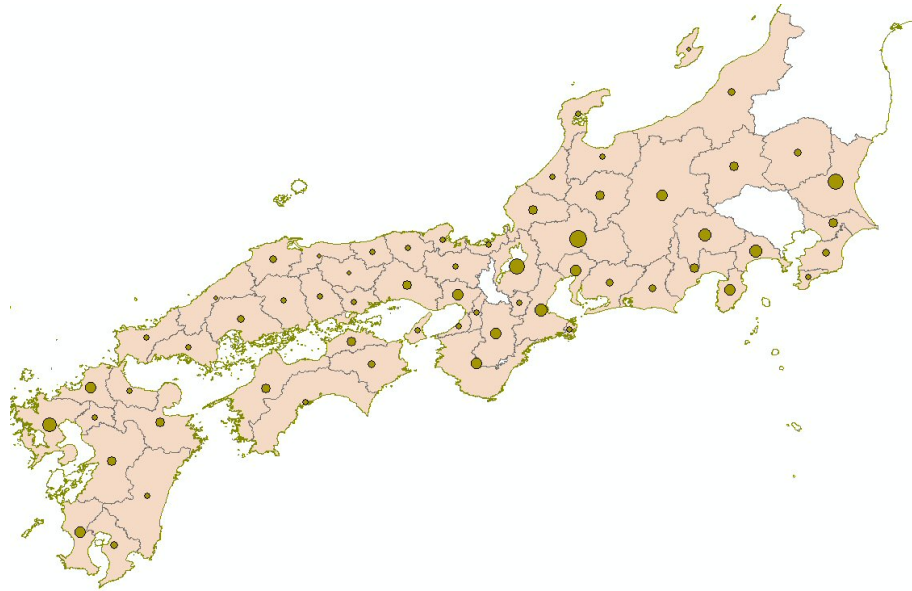


図5 「地部」資料本文の文字数分布（武蔵国・山城国を除外）

Fig. 5 Distribution of the number of characters of the body text in *Chi-bu* (Musashi-koku and Yamashiro-koku are excluded).

報を与えたデータを使用し、それと『古事類苑』「地部」伊勢国の資料本文にあらわれる地名とを、試験的に照合してみた。

地名照合には、ふたつのパターンが考えられる。第1は、『古事類苑』の地名を手作業で同定したうえで、『大日本地名辞書』電子データ上のおなじ地名エントリを発見し、後者に登録された経緯度情報を利用するというパターンである。第2は、『大日本地名辞書』のエントリを使いながら、『古事類苑』の本文記述から地名を探すというパターンである。

第1のパターンから検討を進める。まず、「地部」伊勢国の記述から手作業で地名を抽出したところ、35,334文字のなかから397の地名がみいだされた。それら397地名について、『大日本地名辞書』の電子データと単純な文字列マッチングを行った。ただし、『古事類苑』には「英多」、『大日本地名辞書』には「英多郷」と表記されているなど、後者の記述の末尾が「郷」「村」「郡」などの接尾辞である場合が散見されたため、マッチングにさいしては、『古事類苑』地名の全文字が『大日本地名辞書』の電子データに登録された地名に「先頭一致」した場合を「一致」とみなすことにした。

その結果を表1に示した。397地名のうち、『大日本地名辞書』と一致した地名は253あった。そのうち候補がひとつのものは75（うち正解があるのは32）で、複数の候補を返したものは178（うち正解があるのは99）あった。もっとも多い候補が見つかったのが「山田」で59の候補があった。『大日本地名辞書』に正解が見つからなかったもののうち、候補数が最大だったのは「大島」の29候補だった。かりに、複数のマッチング結果が得られた場合、手作業で個別の精査が可能な候補数を5以下とするならば、1~5件の候補が見つかった地名のうち正解を含むものが、実用的に両者の対応が取れる地名になり、その数は397地名のうち91ということになる。

表1 『古事類苑』「地部」伊勢国資料本文中の地名から『大日本地名辞書』への文字列マッチング結果
Table 1 The result of character matching from the body text of Ise-koku in *Kojiruien*, *Chibu* to *Dainihon Chimei Jisho*.

一致数	地名数（うち正解がある数）
6以上	77 (40)
5	13 (5)
4	16 (9)
3	35 (19)
2	37 (26)
1	75 (32)
0	144 (17)
合計	397 (148)

つづいて、第2のパターンである『大日本地名辞書』から『古事類苑』へのマッチングについて検討する。まず前者の地名エントリをリスト化した。そのさい「国」「郡」「郷」の接尾辞がつく地名については、接尾辞を取ったエントリも用意した。各エントリについて、「地部」の伊勢国の資料本文と文字列が一致した数をカウントした。その結果は、『大日本地名辞書』の地名49,557件のうち172件だった。ちなみに伊勢国に地域を限ると、『大日本地名辞書』には1,074の地名が収録されている。この1,074件に絞って「地部」伊勢国の資料本文とマッチングを取ると、143の地名が一致した。先述した第1のパターンでのマッチングで一致した地名は148であったので、数のうえからいえば、第1のパターンも第2のパターンも一致する地名の数は同程度であることがわかる。ただし、第2のパターンについては、これらの文字列一致した地名が、実質的にも正解か否かの検証はまだ行っていない。

これらの結果を総括するならば、『古事類苑』の資料本文中から手作業で地名を抽出してそれらに『大日本地名辞書』電子データを半自動でリンクするにせよ、あるいは『大日本地名辞書』の地名を頼りに『古事類苑』資料本文中から地名を半自動で発見するにせよ、完全な手作業でやる場合とくらべて40%以下しかカバーできないことになる。このカバー率は、異体字に対応した文字列マッチングを行えば多少は上昇すると見込まれる。しかし以上のデータからは、『古事類苑』と『大日本地名辞書』の半自動リンクツールを作ったとしても、その有効性に過大に期待することは難しいといえる。

5. 今後の課題

5.1 『古事類苑』と「帝国図」のタイムラグ

『古事類苑』は、「本書二記載スル所八、神代二始マリ、慶應三年二終ル、但シ事ノ同年以降二関連スルモノハ此限りニ在ラズ」（編纂條例第四條）で謳うとおり、基本的には明治以前の文献をもとにし、項目ごとに列挙する「類書」のスタイルをとる。したがって、おなじ見出し語であっても、本文中では典拠文献ごとに示す場所そのものが異なる場合、もしくは示す領域が異なる場合も少なくない。わけても行政区画の場合は顕著である。

「地部」においては、洋装本第1・2巻の国別地誌編にかんしては、編纂條例第四條の「此限りニ在ラズ」が適用されており、近代資料も採録されている。「国」の項目立ては、「延喜式」に則ってなされている。しかし、見出し語項目である「郡」にかんしては、表記・区分とともに、最終典拠資料となる明治14(1881)年内務省作成の『郡名異同一覧』（同資料内の情報は明治11(1878)年の太政官布告による郡区町村編成を反映したものが最終情報である）中の地誌提要にしたがって構成されている。

以上を踏まえたうえで、見出し語項目のレベルで地図化作業をより完璧にするためには、少なくとも郡区町村編成以前の時点におけるシェイプ作成を行うのが妥当であろう。「帝国図」では、郡の再編および市制施行後の境界が記されているため、同図をそのままシェイプファイル化した現在の状態では、『古事類苑』の典拠資料の時代とくらべてポリゴンの形・数とも不整合が生じている（表2）。

以上のことから、郡ポリゴンの編集が今後必要なのであるが、その作業はまた困難が予想される。「帝国図」の郡境界を『古事類苑』の時代に近づけるために必要なポリゴン編集作業を、伊勢国の場合についてまとめたのが表3である。容易に出来る編集もあれば、地籍図が必要なもの、あるいは員弁郡・旧中野村のようにもはや境界画定が困難なものもみられる。したがって、郡ポリゴンの編集を行うにしても、ある部分で妥協を強いられることになるであろう。

5.2 「帝国図」の補正と精度の向上

先行文献3)で示したとおり、精度と歴史性の両方の観点から、われわれは「帝国図」をベースマップに選択して、海岸線や国郡境界は「帝国図」のものをシェイプファイル化した。しかしながら、作業効率の観点から自然物・人工物地名の多くは、まず現代図上での経緯度情報を取得することになった。そのうち多くの地名については、「帝国図」上にマッピングしたときに位置がずれないように修正したのだが、現時点では作業結果に統一が取れていない。

今後は、「帝国図」に忠実にシェイプファイル化する部分と、現代図から得られるより正確な経緯度情報を重視する部分とにわけて作業を進める必要性を感じている。すなわち、海岸線と行政界は「帝国図」にしたがった現行のシェイプを尊重し、「帝国図」のシェイプファイルとして公開する。それと同時に、各時代の地理情報を重ねて分析できるというGISの特徴を活かすためには、「帝国図」の海岸線・行政界を現代図と重ねられるように補正したバージョンも作成する必要がある。

しかしながら、経緯度情報の作成にさいして国土地理院の「地図閲覧サービス」に頼っている現状では、精度は分単位でしか確保されていないと考えるべきである。それ以上の精度でデータを作成することは、資料の制約上、不可能である。また、「原」や「野」のように、実際には面的な属性を持つ項目でもポイントで代表させたものが多く、経緯度値や地図上の地点に厳密性は求められない。

いうまでもないが、『古事類苑』はさまざまな時代のさまざまな史料をとりまとめたものであり、項目ごとに対象としている時代は異なるし、おなじ項目であっても複数の時代・場

所を対象としていることもしばしばある。すべての項目について『日本歴史地名大系』『角川日本地名大辞典』などの同一の地誌資料集を参照していないことも、現時点での不十分な点である。データ精度の均一性を保つためには、すべてをおなじ資料集で調べ直し、その結果に基づいてシェイプファイルを修正すべきであろう。そういった作業を進めていくうえで、歴史地理学・地形学・情報学などの多分野からの意見や協力が必要である。

謝辞 この報告は、平成 19～22 年度日本学術振興会科学研究費補助金・基盤研究 A「近代日本の歴史的時空間データマイニングのための基盤整備」(課題番号:19200019, 研究代表者:山田奨治)の補助を受けた研究成果の一部である。『大日本地名辞書』の電子データをご提供いただいた大阪国際大学の桶谷猪久夫氏、『古事類苑』全文テキスト入力にあたってご助言・ご協力いただいた国文学研究資料館の相田満氏、シェーブファイル作成にご助力いただいた岡屋純子、駒野恭子のおふたりに感謝の意を表す。

参 考 文 献

- 1) 山田奨治, 早川聞多, 相田満: 古事類苑(天部・地部)の全文入力と Wiki 版の試行 - 近代の文化概念の情報資源化 -, 情報処理学会研究報告, Vol.2006, No.112, pp.39-46 (2006).
- 2) <http://ys.nichibun.ac.jp/kojiruuen/>
- 3) 山田奨治, 尾方隆幸: 古事類苑・地部データへの経緯度情報付与と地図化, 情報処理学会研究報告, Vol.2008, No.47, pp.85-92 (2008).
- 4) <http://rekishi.jkn21.com/>
- 5) <http://watchizu.gsi.go.jp/>

表 2 『古事類苑』編纂・地図作成・郡界変更の経緯
Table 2 A brief history of *Kojiruien*, maps, and municipal mergers.

年	古事類苑編纂	地図作成	郡界変更
M04 (1871)			廃藩置県(1888年12月まで幾度か改編)
M05 (1872)		『日本地誌提要』刊行開始	大区・小区制施行
M10 (1877)		『日本地誌提要』刊行終了	
M11 (1878)			郡区町村編制(太政官布告) 改称・分割多数
M12 (1879)	文部省により編纂開始		
M13 (1880)			
M14 (1881)	「地部」最終採録記事	『郡名異同一覧』	釧路国網尻郡 北見国網走郡に編入 渡島国津軽郡・福島郡 合併して松前郡
M15 (1882)			
M16 (1883)			
M17 (1884)		「輯製図」作成開始	
M21 (1888)			市制・町村制公布(県域最終確定)
M22 (1889)			栃木県寒川郡 下都賀郡に編入
M23 (1890)	皇典講究所の管轄となる		府県制・郡制公布
M24 (1891)			
M25 (1892)			
M26 (1893)		「輯製図」完成	
M27 (1894)			
M28 (1895)	神宮司庁の管轄となる		
M29 (1896)	刊行開始		編入・合併・分割合併多数
M30 (1897)			郡界変更(岐阜・愛媛)
M31 (1898)			郡界変更(広島)
M32 (1899)			郡界変更(香川)
M33 (1900)			郡界変更(岡山)
M34 (1901)		「帝国図」作成開始	
M40 (1907)	編纂完了		
T01 (1912)	「地部」(1)刊行		
T02 (1913)	「地部」(2)刊行		
T03 (1914)	刊行完了		
T04 (1915)			
T05 (1916)			
T06 (1917)		「帝国図」完成	
T12 (1923)			郡制廃止(郡を越えた合併開始)

表 3 「帝国図」郡境界統合・分割作業（伊勢国の場合）

Table 3 Division and integration of *gun* borders required to *Teikoku-zu* (in case of Ise-koku).

作業種別	作業内容
分割	三重郡 三重・朝明に分割（村界使用可能） 飯南郡 飯高・飯野に分割（村界使用可能） 河芸郡 河曲・奄芸に分割（村界使用可能）
分割・統合	多気郡から旧向粥見村を分割して飯高郡に統合（地籍図使用で可能か？） 津市 旧大部田村を分割して奄芸郡に統合（地籍図使用で可能か？） 旧太田部村以外を安濃郡に統合（可）
統合のみ	員弁郡・旧中野村部分（南石加村内）を分割して朝明郡に統合（不可能か？） 宇治山田市 度会郡に統合（可） 四日市市 三重郡に統合（可）