タイのデジタル地名辞書基盤の構築

永田　好克（大阪市立大学 大学院創造都市研究科）

地名はさまざまな記述や知識を連結する情報の一つである。そのため、冊子体の地名辞書は古くから編纂されてきた。これらをデジタル化することは古書であった地名辞書の活用価値がよみがえるだけでなく、人文社会科学における基盤情報として永続的に寄与できるものである。さらに現在の地名に関連付ければ、地名履歴データとして時空間情報データベースに発展が可能となる。本稿ではタイ国東北部を主な対象としたデジタル地名辞書構築への取り組みについて報告する。地名辞書だけでなく、約 100 年前に発行された地図も地名データ収集の対象とした。位置情報を頼りにした地名履歴の復元には課題が多いが、一方で現在の地図にある地名表現では判然としない地域社会の多様性を顕在化するデータセットとなることが期待できる。

Development of a spatio-temporal digital gazetteer of Thailand

Yoshikatsu Nagata (Graduate School for Creative Cities, Osaka City University)

Place name is an indispensable key to link information among various descriptions and knowledges. To digitize analog old gazetteers is not only to revive their utility value but also to contribute to enrich fundamental information for social sciences and humanities. Furthermore, if place names in old days can be linked to their corresponding current names, a digital gazetteer can be expanded to a spatio-temporal gazetteer. In this article, interim status of development of a spatio-temporal digital gazetteer mainly covering Northeast Thailand is described. Major source materials are old gazetteers, and also maps published in almost a century ago are referred to collect place names. There are many difficulties to link place names over time relying on their data on geographic coordinate; however, local ethnic diversity which are not clearly shown in place names on present day maps may become apparent on such a spatio-temporal digital gazetteer.

１．はじめに

地名はさまざまな記述や知識を連結し、また他者との共有を円滑にする情報の一つである。そのため、書物としての地名辞書は古くから編纂されてきた。近年は、これら冊子体の地名辞書の情報に経緯度の位置情報を加えたデジタル地名辞書の構築も取り組みが進められてきている。過去の地名を収録したデジタル地名辞書は人文社会科学における基盤情報として今後永続的に貴重なものとなるばかりでなく、デジタル化によって古書であった地名辞書の活用価値がよみがえるものである。地名辞書のデジタル化は、まず、その地名辞書が編纂された時代の地名空間を再現することから始まるものであるが、経緯度を頼りとして現在の地名に関連付けることができれば、地名履歴データとして時空間情報データベースに発展が可能となる。

一方で、地名辞書のデジタル化において、もともと経緯度が記載されていないものに経緯度の値を付加することは容易ではない。特に、過去の地名で現在の地名とは異なる場合には、該当する現在の地名との関連を知る手がかりが別途必要となる。手がかりとして有用な資料のひとつは同時代に発行された地図である。もうひとつは、地名の履歴や別名が記された資料である。いずれにせよ経緯度の値を付加しながらのデジタル化は、地道な照合作業が必須であり、明治期に発刊された日本の地名辞書をデジタル化した桶谷の長年にわたる取り組みと成果[1]はその儀範である。

本稿ではタイとくにタイ国東北部を対象にしたデジタル地名辞書構築への取り組みについての報告を行う。タイ国東北部を対象としているのは、筆者が長年研究対象の地とする中で必然的に地名に接してきたからである。タイに関連する地名辞書やタイで発行された地形図だけでなく、外邦図などの戦前期の地図に多数接するなかで、地名の変遷や表記のゆれ、場合によっては誤記と思われる別名まで記録されていることに注目した。これらを統合的にデータベース化することは、今後さまざまな資史料に記述された地名を読み解くにあたっての長寿命で発展的な基礎資料となることが期待できる。

また、村落レベルの地名において現在と過去の名称が異なるものの中には、過去の地名においてタイ語とは異なる言語を起源とする名称が散見されることから、現在の地図に見える地名の由来を含めた地域の民族多様性を顕在化できる可能性がある。

©2017 Information Processing Society of Japan

２．地名辞書と地図

本稿での報告に活用したタイに関する冊子体の地名辞書は、以下である。

(a) 大南洋地名辞典 [2]

(b) Gazetteer to Maps of Thailand [3]

(c) Thailand: Official Standard Names Gazetteer No.97 [4]

このうち(a)(b)は第二次世界大戦期に刊行されたものであり、また(c)はベトナム戦争期の刊行であることから、いずれも平穏な時代に編纂されたものではないものの、当時の緊迫した時勢下における戦略的かつ網羅的な記録である。さらに、学術的視点で編纂された大日本地名辞書と異なり、地理情報重視の(a)(b)(c)には精度の精粗はあるものの、経緯度が記されている。

(a)は単に地名と経緯度だけでなく、周辺地名との関係や地理上の位置づけなどの記述を含み、掲載項目として選定された背景が垣間見えるものである。地名はローマ字表記を見出しとして一部にカタカナが併記されている。地名に関する記述がある分、掲載数はそれほど多くなく 3,047 地名にとどまる。

(b)は(a)と同時期に米国で刊行されたものであり、ローマ字表記の地名と経緯度の一覧である。(a)とは異なり地名に関する記述は一切ないが、地名ごとに典拠の地図が判別できるようになっている。前書きによればタイ国地図局発行の 20 万分の 1 図を主要な典拠地図としている。約 84,000 地名が記載されているが、別称の記載を除くと約 66,000 地名となる。

(c)も米国で 1966 年に刊行されたものであり、ローマ字表記の地名と経緯度の一覧である。後述の(e)である 5 万分の 1 図を主要な典拠地図としてはいるものの、地名の選択にあたっては 20 万分の 1 図に拠ると前書きに記載されており、掲載地点数は(b)よりも少ない。約 45,000 地名が記載されているが、別称の記載が多いため、これを除くと約 24,000 地点の記載である。

これらの網羅的な地名辞書のほかに用いたのが以下の地図である。

(d) 外邦図

(e) L708 シリーズ地形図

(f) L7017 シリーズ地形図

(d)は旧日本陸軍参謀本部陸地測量部が発行した日本の領土以外の地図の総称である。タイに関する外邦図のうち、広範囲に村落レベルの地名を収集できるものには、インド測量局が発行した地図に基づき 25 万分の 1 に複製したものとタイ国地理局が発行した 20 万分の 1 図を複製したものがある。前者は原図の地名がローマ字表記であるため図画外の凡例表記を日本語化したにとどまる。後者は原図の地名表記がタイ語であるため、すべての地名をカタカナ表記にしたものであり、1920 年前後に発行された原図をもとに 1940 年から 1941 年にかけて複製印刷されたものである。本稿では後者のうち、タイ全土の約 3 分の 1 の面積を占めるタイ国東北部を対象とする図葉を用いた。外邦図として発行された図と原図の描写内容、すなわち等高線、道路、河川、村落、三角点などが同一であることは、タイ国地理局発行の原図の写しとの対照により確認済みである。

(e)は第二次世界大戦後に AMS(U.S. Army Map Service)がタイ国地図局とともに作成した 5 万分の 1 地形図で、1950 年代から 1960 年代にかけて発行された。版によって AMS 作成とするものとタイ国地図局作成とするものが混在する。地名はタイ語とローマ字の併記である。(d)およびその原図では、タイ国東北部のほとんどの地域で記載位置に大小の齟齬がみられたが、(e)では大幅に位置精度が改善された。測地座標系は Indian Datum と凡例に示されている。

(f)はタイ国地図局作成で 1980 年代に発行された(e)の更新版である。凡例には測地座標系を Indian Datum と示しているが、厳密には(e)と同じではない。しかしながら、5 万分の 1 という縮尺であるため、実用上は同一であるとして差し支えない。

３．デジタルデータ化

タイのデジタル地名辞書基盤の構築にあたって、まず前述の地名辞書(a)(b)(c)のデータベース化を行った。筆者の関心はタイ国東北部にあるが、地名辞書の地名一覧から該当する地名だけを抜粋することはかえって煩雑であることから全地名を対象とした。これらには経緯度も記載されていることから、各地名の名称、区分、経緯度、典拠図について入力を進めた。ここでの区分とは、地名の属性のことであり、村、町、駅、山、岬、港などである。

それぞれの地名辞書に掲載された地名をデータベースとして入力したのち、KML でデータを出力した。KML ファイルを Google Earth 上に展開することで地名の分布や位置精度について概略的に考察を行った。

経緯度は位置を表記する一般的な方法であるが、準拠する測地座標系は測量が精緻になるとともに変遷してきた[5]。従って経緯度の値だけでなく測地座標系の情報を得ておくことが望ましい。Official Standard Names Gazetteer は概ね出典が戦後に発行された地形図に拠っているため、経緯度の基準となる測地座標系は Indian Datum に相当する。

これより発行が古い大南洋地名辞典、Gazetteer to Maps of Thailand、および外邦図とその原図については基準となる測地座標系が明確ではない。現在の全球の標準の測地座標系に変換する都合上、これらについても測地座標系は Indian Datum であると仮定している。

なお、ローカルな測地座標系である Indian Datum と Google Earth などで採用する世界測地系

©2017 Information Processing Society of Japan

WGS-84 との差異は東西方向約 14 秒、南北方向約 5 秒、あるいは距離にして約 450 メートルである。

3.1 大南洋地名辞典

大南洋地名辞典には、掲載された 3,047 地名のうち 2,900 地名に経緯度が記載されている。経緯度の記載単位は統一されておらず、1 秒、10 秒、1 分のものが混在する。同書の前書きで複数の典拠資料について述べているものの、個々の地名とその経緯度の典拠は明らかではない。経緯度の 1 秒は約 30 メートルに相当するところであるが、ほぼ同時期の刊行である Gazetteer to Maps of Thailand での最小単位が 1 分であることから、記載単位の有効性については別途検証する必要がある。

掲載地点の分布には地域による濃淡があり、バンコク周辺や沿岸部には多く分布するが、内陸部には掲載地名の分布が非常に疎である地域もある。秋田県に相当する面積を占めるルーイ県には掲載地点が一箇所もない。タイ全土の約 3 分の 1 の面積を占めるタイ国東北部から掲載された地名数は約 6 分の 1 である。

3.2 Gazetteer to Maps of Thailand

Gazetteer to Maps of Thailand には別称として重複する地点を除くと約 66,000 地名掲載されている。うち区分として最も多いのが村落であり約 25,700 地点が掲載されている。次に多いのが沼の約 11,300 地点であることは興味深い。経緯度の記載精度には、1 分、5 分、10 分のものが混在するが、これは典拠とする地図の縮尺がそれぞれ 5 万分の 1、20 万分の 1、100 万分の 1 であることに由来する。

掲載地点の分布の濃淡は地域による差異よりも典拠とした地図の縮尺に由来するようであり、当時入手可能であった地図から満遍なく地名を収集したものと考えられる。

3.3 Official Standard Names Gazetteer

Official Standard Names Gazetteer には別称として重複する地点を除くと約 24,000 地点が掲載されている。区分として最も多いのが村落であり約 13,600 地点が掲載されている。次に多いのが河川の約 3,800 件であり、Gazetteer to Maps of Thailand で多く掲載されていた沼は約 800 地点にとどまる。経緯度の記載単位は 1 分である。

村落数の掲載地点は Gazetteer to Maps of Thailand の半分弱にとどまることや、沼の掲載数が大幅に減っている。Official Standard Names Gazetteer の編集にあたっては既刊の Gazetteer to Maps of Thailand の再利用は行わずに新たに当時入手可能な地図資料に基づいて地名を収集したのであろう。掲載地点の分布にはやや不自然な濃淡が見られる。第二次世界大戦以前の地図では位置が相当に不正確な地域が多数あり、戦後の測量成果を待って慎重に経緯度を定めていく必要があったことから地名の収集が後回しになったのが理由であろうと考える。

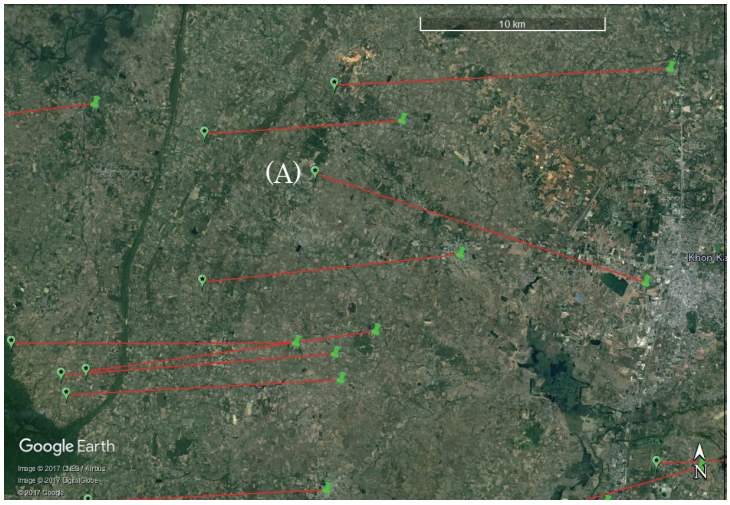


図 1 外邦図における記載位置のずれ
Figure 1 Discrepancy of location in *Gaihozu*

3.4 外邦図

外邦図からの地名データの収集は、図面からの手作業での読み取りで行った。その位置は図面上の緯線経線からの換算で計算した。当面はタイ国東北部に限って作業を進めており、現時点でタイ全土の約 5 分の 1 に相当する領域から約 2,500 地名を収集ずみである。

前三点の地名辞書とは異なり地図からの収集であることから、一般的には地図からの読み取り精度の範囲内で精緻な位置情報を得ることが期待できるのであるが、読み取った結果を前三点と同様に Google Earth 上に展開した結果、さまざまな課題が明らかになった。

ひとつは記載位置の大幅な齟齬である。当時の州庁所在地であるコーンケン近郊のシーターン村は実際の所在地よりも西に約 19 キロメートル離れた位置に記載されており、この村にとどまらず周辺の村もほぼ同様に大幅にずれた位置に記載されている(図 1 の(A))。図 1 においてバルーンが外邦図上での記載位置であり、ピンが実際のその村の位置である。外邦図やその原図の縮尺が 20 万分の 1 であることから、図面上のずれは 10 センチメートル近い。図 1 にみるように、単に特定の村だけでなく近隣の村落を含めて全体の位置がずれている。

これをさらに範囲をひろげてタイ国東北部全体で比較すると一定の範囲の地名が総じてずれている地域が多数あることが判明した。図 2 はタイ国東北部の外邦図での記載位置と実際の位置

©2017 Information Processing Society of Japan

とのずれを、近隣の地名でグループ化して平均値を求め、ずれの距離と方向を示したものである。黒丸が外邦図上での位置で、黒丸から伸びる線がずれの距離と方向を示す。当時の測量技術が近代化途上にあって、数少ない基準点から遠くはなれたタイ国東北部では精緻な測量と作図がまだ及ばず、概略図を越えないものであったことを如実に物語っている。

データ化するにあたっての別の問題点は、カタカナ表記された地名のデジタルデータとしての再現性である。日本語ではない言語で書かれた地名をカタカナ化したための苦心の結果であろうが、通常の日本語では用いない小書きのカタカナが多数登場する。

例えば Gazetteer to Maps of Thailand でも多数収録されている沼を「ノーング」として「グ」を小書きにしている。「グ」の小書きは Unicode 実装の原則に従えば「ク」の小書きと濁点の合成文字とすることでデータ上取り扱うことは可能である。ただし現状では合成した結果を不自然ではない組版で再現できるソフトウェアは一般的ではない。

「コヲーンケェン」や「ラーヲ村」などの表記でしばしば用いられている小書きの「ヲ」は現在の Unicode では未収録であるため、当面は大書きの「ヲ」か小書きの「オ」で代用せざるを得ない。この「コヲーンケェン」とは前述で位置の齟齬を例示したコーンケンのことであり、コーンケンを含む図葉のインデックス名にもなっている(図 3)ことから、書誌情報としての正確性を確保する上でも Unicode への収録を検討すべき課題である。

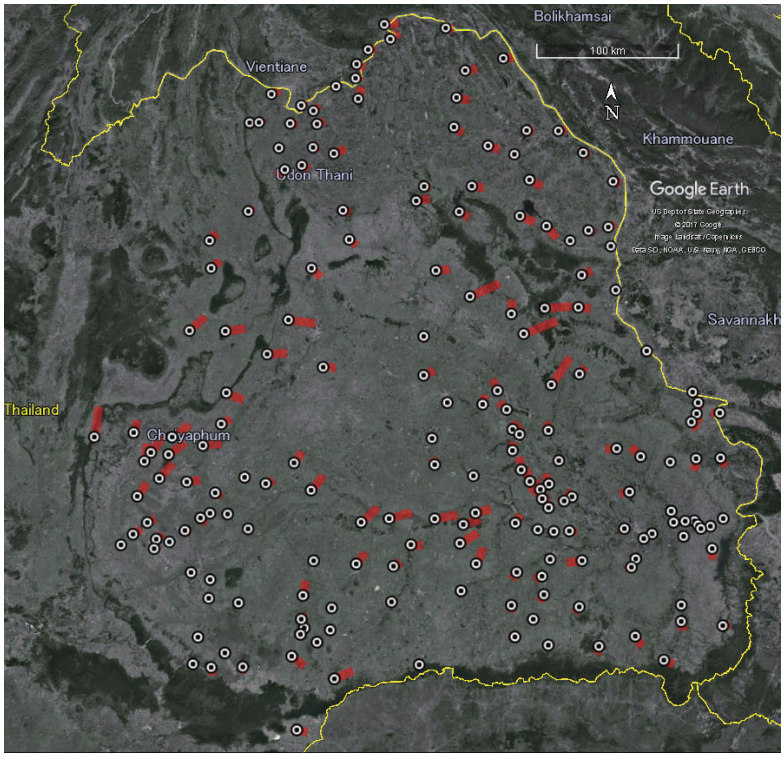


図 2 タイ国東北部の外邦図における記載位置のずれ
Figure 2 Tendency of locational discrepancy in Northeast Thailand in *Gaihozu*

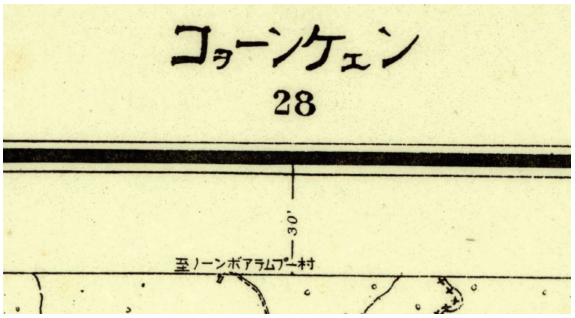


図 3 外邦図に使われている小書きカタカナの例
Figure 3 Small katakana used in *Gaihozu*

ほかに、外邦図から読み取った小書きの「マ」が現時点で Unicode に未収録である。

4. 時空間情報への連携

異なる時代の地名データセットを、位置を基準にして連携することによって、地名履歴データ、すなわち時空間を加えた地名情報データベースを構築することができる。これはどの地名データセットも測地座標系が明らかでかつ隣接の同じ属性の地名と峻別可能な充分な位置精度により構成されていれば容易なことである。しかし、本稿で取り扱う地名データセットは測地座標系の確認が困難なものがあるだけでなく、位置精度も近隣の異なる村を区別できるほどではない。それればかりか、近代的な測量の成果ではなく概略図として記載されてきた地域が広範囲に含まれている。従って、位置情報を信頼できるキーとして連携することは困難である。

時系列上に地名データセットを連携するもうひとつのキーが地名そのものである。まず、異なるデータセット間での照合が容易な著名な地名の連携から開始した。著名な地名とは、小縮尺の地図でも掲載されている地名、あるいは、県庁所在地などの地域を代表する地名である。

前章で述べたデータセットはいずれも原語表記ではなく翻字された表記であるため、地名としての照合は読まなければ進まない。例えば著名な地名のひとつであるコーンケンは以下のように

©2017 Information Processing Society of Japan

表記されている。(a)から(d)は 2 章で述べた資料に対応する。

- Ban Khon Khen (a)
- Moeang Khon Khen (a)
- MUANG KHON KAEN (b)
- Khon Khen (b)
- KHON KAEN (c)
- AMPHOE MUANG KHON KAEN (c)
- AMPHOE PHRALAP (c)
- AMPHOE PRALAP (c)
- CHANGWAT KHON KAEN (c)
- CHANGWAT MUANG KHON KAEN (c)
- KAWN KEN (c)
- KHOHN KEN (c)
- KHON KAN (c)
- KHON KHEN (c)
- KHONKEAN (c)
- KHONKHAEN (c)
- KONKAHEN (c)
- MUANG KAUN GEAN (c)
- MUANG KHAUNKAEN (c)
- MUANG KHON KAEN (c)
- ムァンコヲーンケェン (d) (ヲは小書き)

(c)の Official Standard Names Gazetteer には 2 章で述べたように別称の記載が多く、特にコーンケンのような著名な地名では顕著である。別称が多数掲載されていることについて、当初は翻字の揺れに由来してデータ入力上煩雑なだけと考えていた。しかし、単に翻字の揺れに留まらない真の別称あるいは旧名と確認できる情報が含まれることがあり、類似性が乏しい読みの地名を連携する際に活用できる場面が多かった。コーンケンの例では Phralap が郡名の旧称として用いられていたことが確認できる。

地名を手がかりにした連携に加えて必要なのが所在地の経緯度である。位置情報を連携するためのキーとして用いることが難しいと前述したように、(a)から(d)それぞれのデータから得る経緯度の情報は必ずしも同一の値ではない。例えばコーンケンに関しての値はそれぞれ以下である。

- 北緯 16 度 28 分 東経 102 度 50 分 (a)
- 北緯 16 度 30 分 東経 102 度 50 分 (b)
- 北緯 16 度 26 分 東経 102 度 50 分 (c)
- 北緯 16 度 29 分 東経 102 度 39 分 (d)

地名データを時空間情報として現在につなげるためには戦後の近代的な測量によって編纂された地図と連携しなければならない。そのために用いたのが 2 章で述べた L708 シリーズ地形図と L7017 シリーズ地形図である。例えばコーンケンの郡役所の位置を L708 から読み取ると「北緯 16 度 25 分 44 秒 東経 102 度 50 分 17 秒」であり、分単位に丸めれば(c)の値は信頼できることが確認できる。(c)の Official Standard Names Gazetteer は座標の典拠として L708 シリーズ地形図を用いているので当然の帰結ではあるが、他のデータセットに収録されていても(c)には収録されていない地名もあることから、正確な経緯度の値を得るためには L708 または L7017 シリーズの地形図を参照する必要がある。

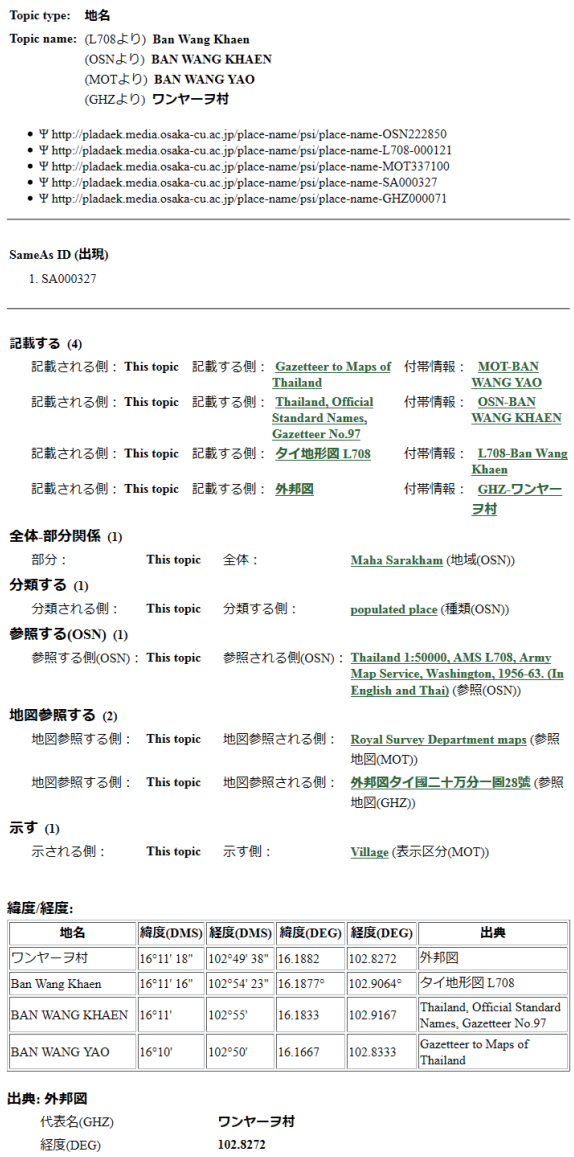
図 4 地名履歴データの見出し例

Figure 4 Index of digital gazetteer of Thailand

このような過程で地名データセット間の連携を行った後、著名な地名は周辺の村落や沼などのより小さな単位の地名を照合するためのキーとなる。経緯度の値を頼りにすることができないため、相対的な位置関係や当時の主要な交通路であった河川に沿った並び順を考慮して照合を進めてきた。また古い資料に記録された位置精度を検証するために L708 または L7017 の地形図との照合も行ってきた。現在複数のデータセットを連携した地名履歴データは約 2,000 地点分である。

現時点までの地名履歴データ収集の成果は、ナレッジ・シナジー株式会社内藤求氏の協力のもとでトピック

©2017 Information Processing Society of Japan



Topic type: 地名

Topic name: (L708より) Ban Wang Khaen
(OSNより) BAN WANG KHAEN
(MOTより) BAN WANG YAO
(GHZより) ワンヤーヲ村

- Ψ http://pladaek.media.osaka-cu.ac.jp/place-name/psi/place-name-OSN222850
- Ψ http://pladaek.media.osaka-cu.ac.jp/place-name/psi/place-name-L708-000121
- Ψ http://pladaek.media.osaka-cu.ac.jp/place-name/psi/place-name-MOT337100
- Ψ http://pladaek.media.osaka-cu.ac.jp/place-name/psi/place-name-SA000327
- Ψ http://pladaek.media.osaka-cu.ac.jp/place-name/psi/place-name-GHZ000071

SameAs ID (出現)

1. SA000327

記載する (4)

記載される側： This topic 記載する側： Gazetteer to Maps of Thailand 付帯情報： MOT-BAN WANG YAO

記載される側： This topic 記載する側： Thailand, Official Standard Names, Gazetteer No.97 付帯情報： OSN-BAN WANG KHAEN

記載される側： This topic 記載する側： タイ地形図 L708 付帯情報： L708-Ban Wang Khaen

記載される側： This topic 記載する側： 外邦図 付帯情報： GHZ-ワンヤーヲ村

全体-部分関係 (1)

部分： This topic 全体： Maha Sarakham (地域(OSN))

分類する (1)

分類される側： This topic 分類する側： populated place (種類(OSN))

参照する(OSN) (1)

参照する側(OSN)： This topic 参照される側(OSN)： Thailand 1:50000, AMS L708, Army Map Service, Washington, 1956-63. (In English and Thai) (参照(OSN))

地図参照する (2)

地図参照する側： This topic 地図参照される側： Royal Survey Department maps (参照地図(MOT))

地図参照する側： This topic 地図参照される側： 外邦図タイ国二十万分一図28號 (参照地図(GHZ))

示す (1)

示される側： This topic 示す側： Village (表示区分(MOT))

緯度/経度:

地名	緯度(DMS)	経度(DMS)	緯度(DEG)	経度(DEG)	出典
ワンヤーヲ村	16°11' 18"	102°49' 38"	16.1882	102.8272	外邦図
Ban Wang Khaen	16°11' 16"	102°54' 23"	16.1877°	102.9064°	タイ地形図 L708
BAN WANG KHAEN	16°11'	102°55'	16.1833	102.9167	Thailand, Official Standard Names, Gazetteer No.97
BAN WANG YAO	16°10'	102°50'	16.1667	102.8333	Gazetteer to Maps of Thailand

出典: 外邦図

代表名(GHZ) ワンヤーヲ村

経度(DEG) 102.8272

図 5 地名履歴データの詳細例
Figure 5 Spatio-temporal data of place name

マップ形式のデータベースにもしている。図 4 は地名インデックスの例であり、図 5 は構築した地名履歴データを詳細に表示した例である。

5．課題と展望

本研究の大きな課題のひとつは、古い時代のアナログな地図から地名と位置を抽出してデジタルデータとする過程であり、この作業に時間を要することである。さらに、時系列上の関係を定めるためには、単に地図から地名と位置を抽出するだけでなく、別のデータセットとの照合が必要であるが、位置情報の精度が低いことや地名表記にゆれや変更があることから、情報処理として自動化することは現時点では相当に困難である。

一方、位置情報に頼った自動化が困難であることの裏返しとして、20 世紀初頭のタイにおける地図作成の精度を地域別に提示できるデータを蓄積できる可能性がある。現在、外邦図あるいはその原図となったタイ国地理局作成の地図上の地名について、小地域に分けてその記載位置の計量的な誤差や傾向を分析しつつある。

また、古い地図ほど地域の伝統的な呼称を用いた表現で地名が掲載されている。典型的な例は、タイ国東北部の南部にあたるカンボジアとの国境に近い地域に多数みられるクメール語系の地名表現である。これらの地名は現代の地図における表記では、クメール語起源であることがわかりにくくなっている。このことから、古い地名まで連携した地名履歴データは地域の根源的な民族多様性を現代に照射する手がかりとなる情報を提供するものとなる。

本研究への取り組みからは、Unicode 文字セットに未収録の小書きカタカナが複数存在することが判明した。今後データの持続性と再現性に齟齬が生じないような形で Unicode に収録されることを待望する。

謝辞

外邦図はその多くを東北大学が公開する外邦図デジタルアーカイブによった。また、JSPS 科研費 24500312 および 26300010 の成果の一部である。

小書きカタカナの取り扱いとその課題については、国立国語研究所の高田智和准教授から多くの知見を得た。

参考文献

1) 桶谷猪久夫 : 地図・地名データベース，情報処理学会研究報告， Vol.2009-CH-83 No.3， pp.1-8 (2009).
2) 南洋経済研究所：大南洋地名辞典第四巻泰國及佛領印度支那(1943).
3) War Department, Army Map Service, U. S. Army : Gazetteer to Maps of Thailand (1944).
4) United States Board on Geographic Names : Thailand, Official Standard Names Gazetteer No.97 (1966).
5) Clifford J. Mugnier : Grids & Datums – Kingdom of Thailand, “Photogrammetric Engineering & Remote Sensing”, Vol.77 No.2, pp.109-111 (2011).

©2017 Information Processing Society of Japan