

地域情報学の読み解き

——発見のツールとしての時空間表示とテキスト分析

柳澤雅之
高田百合奈
山田太造

はじめに——地域情報学と地域研究

インターネットの普及をはじめ、デジタル化された情報の流通とそれを使うためのツールの開発・普及は、とくに一九九〇年代後半以降、顕著になった。アメリカのカリフォルニア大学バークレー校では、一九九七年、歴史学・考古学等で扱う文化資源のデジタル化とその利用における課題解決を目的とした研究グループ、E C A I (Electric Atlas

Initiative) が結成され、とくに時空間分析とネットワークを通じた資源の共有化を目的とした (Buckland 2008)。文化研究と情報学の融合を目指したもつとも初期の研究グループの一つである E C A I と連携しながら、日本では、地域研究に情報学を導入した、新領域としての「地域情報学」が二〇〇四年に創成された (柴山二〇〇九)。柴山は、一九六四年に書かれたタイ国の広域年間水収支に関する論文を例にして、情報学はその当時からすでに地域研究の論文のなかに組み込まれていたが、情報学の役割は G I S (地理情報システム) やリモートセンシング技術、あるい

は統計分析等の限られた範疇でツールとして利用されたにすぎないという。しかし、その後、情報理論や情報メディア・システム、空間情報学等、情報学における関連分野の進展とともに、情報学の視点に基づいて地域を〈解く〉ことが可能になり、地域情報学が構想されたとする。そして、地域情報学は、地域研究のツール開発やデータ処理という狭義の目的だけでなく、情報とそれに付随する属性とを累積・解析して地域像を説明する「広義」の目的を持つものであると説明する（柴山二〇〇九）。

情報学が進展し、地域情報学が生まれるのと同時に、多くの分野で情報学が取り入れられるようになってきた。地理学や人類学、経済学、歴史学、その他多くの自然科学分野を含む分野で、情報学を融合した研究が進んだ。

しかし、先述した柴山の説明では、情報学を取り入れた多くの分野と地域情報学との手法上の違いが不明瞭である。情報学の視点や技術を取り込んだ点ではいずれも同じであり、地域情報学が、情報学を取り込んだ他の研究分野と異なる唯一の点は地域像の解明を目的とするかどうかにある。柴山が編集責任者となった、地域情報学に関する東南アジア研究誌の特集号でも、柴山が述べるように、個別の論文のどの過程で情報理論が使われたのか、あるいは、

どのように情報システムを駆使したかについては必ずしも判断しがたいものが含まれる（柴山二〇〇九）。

新しく構想された地域情報学の異質性を強調し閉鎖的に学界を形成する必要があるが、他の分野において情報学・情報技術の利用が進む以上、単にツールの提供者ではなく、地域研究にとって、あるいは、地域像の解明にとって情報学がどのような意義を持ち役割を果たすのかを地域情報学は説明する必要がある。そしてそのためには、柴山が特集号で試行したように、多くの具体的事例を積み重ねることが必要であるが、それに加えて、地域研究の実際の研究プロセスのなかでどのように情報学が取り入れられるのかを具体的に検証する必要がある。そのことが地域研究の新しい手法を切り拓く上で重要であると考ええる。

そこで本研究では、情報学を取り込んだ地域研究の新しい研究手法を検討するための最初のステップとして、フィールドワークによって得られる現地の情報を、近年の情報学的手法を用いることによって、いかに発見のツールとして利用することができるかを検討する。地域情報学の研究手法には、問題・仮説の設定、地域情報の収集と集積、分析、成果の発信と共有化というプロセスがあり（柴山・原二〇〇八）、そのいずれの過程でも地域研究と情報

学の融合が認められる。本稿では、一連の研究プロセスの中から、まずは現地での地域情報の収集と課題発見のプロセスに焦点をあてて検討する。それにより、地域情報学の手法のバージョンアップだけでなく、地域研究論のバージョンアップを試みる。地域研究論のバージョンアップが必要なのは、地域研究の対象である世界のバージョンが新しくなっているためであるが（山本二〇二二）、同時に、研究手法のバージョンが新しくなっているからでもある。

現地を得られる地域に関するさまざまな情報（地域情報）から地域像や現地の課題を発見するための情報学的手法として、本稿では、オープンソース・デジタルアースとテキスト分析に焦点をあてる。オープンソース・デジタルアースとは、近年、利用が盛んな、インターネットを介してオープンアクセス可能なデジタル地図のことである。記録者によって収集された地域情報を断片化し、デジタル地図の上に時間と空間を軸にして再配置させることで、個々の地域情報の連関を可視化する。

一方、テキスト分析とは、膨大なデータから知識を発掘するためのデータマイニングのひとつで、主に文字情報（テキスト）を分析対象とする情報学の手法である。地域情報から地域像に関連する用語、および、用語のかたまり

であるトピックを抽出する。そして、トピック間のつながりを統計的に分析し、その結果を時空間上で表示させることにより、テキスト間の連関を可視化する。

そして、両者を組み合わせることにより、断片化された地域情報を、地図上の物理的な空間と記録者の認識上の空間の連関として可視化させ、地域情報からの読み解きのツールとしての可能性を検討する。それにより、従来までの単純なキーワード検索等とは異なり、データの意味内容にまで踏み込んで地域情報を読み解くための手法を考察する。

I 資料と調査方法

本研究では、地域情報として、スマトラを広域に調査した高谷好一によるフィールドノートの記録を利用した。

高谷好一フィールドノートとは、一九六七年～一九九五年に京都大学東南アジア研究センターに在籍した高谷好一・京都大学名誉教授が、東南アジアを中心として、世界各国の地形や植生等の自然環境条件と地域社会の生業体系との関係をテーマとして調査した記録である（高谷二〇一二・二〇一三）。出版された高谷好一フィールドノートの記

た。そして、これらの個別の情報に加えて、テキスト分析の手法で得られた結果を時空間上で表示するシステムを構築し、フィールドノートの記録から読み取れる知見について検討を加えた。

また、筆者の一人の柳澤は、高谷が調査したスマートラの中南部で二〇一五年三月一八日から三月二五日にかけて現地調査し、フィールドノートに記載された情報との比較を行った。本稿はそれらの結果に基づいている。

Ⅱ 地域情報の時空間表示から読み解く

1 時空間表示システムの構築

かつて、現地調査の記録は鉛筆とノートによる手書きの資料として作成された。しかし現在、多くの研究者が現地にパソコンをはじめとする多種多様な電子機器を持ち込み、フィールドワークに活用している。電源が限られているところであっても、太陽光発電のような小規模の発電機やバッテリー能力の向上により、辺境の地でさえ現地調査

時における電子機器の使用は一般的になってきた。

インターネットの普及も同様である。利用可能な容量やスピードに違いはあるものの、多くの地域でインターネットが利用可能になってきた。現地調査時に、インターネットにアクセスし、現地の俯瞰図を現地に居ながらに見ることが可能な時代になってきた。パソコンなどの携帯型端末をインターネットに接続することができるような条件下では、時空間上で表示するシステムを現地調査時に活用することが容易になったのである。

本研究では、フィールドノートの情報を時空間上で表示するために、オープンソース・デジタルアースの一つである Cesium を利用し、時間軸を調整することのできるタイムスライダー機能をつけることにより、地域情報を時空間上で表示可能なシステムを構築した。^{*2} 図2に実際のインタフェースを示した。フィールドノートの個別の情報が、記録された地点にポイントで表示され、そのポイントをクリックすることで情報を閲覧することができる。他に、広域の空間内での情報の位置の特定機能、ズームイン・アウト機能、バックグラウンドの地図表示の変換機能（俯瞰図、地形図等）、地図上での表示と連動したキーワード検索などの機能が備わっている。

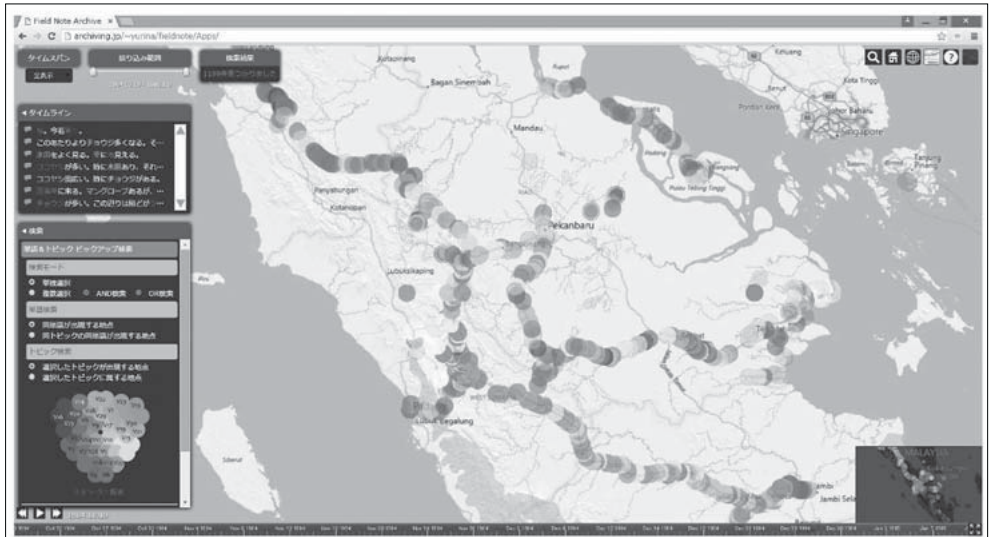


図2 高谷好一フィールドノートの時空間表示(スマトラ中部)

(出所)高田(2015)よりスマトラ中部を表示。

	A	B	C	D	E	F	G
	日付+3ケタの 通し番号	yyyyymmddの形 で4ケタの数字	緯度(10 進法)	経度(10 進法)	本文(FN3-1)	本文に対す る背景	種別(view:風景, note:手書きメモあり hear:聞き取り, loc:地名やその他の点 での行動, sum:まとめ, other:その他)
row	date	lat	long	text	background	class	
13	19900913008	19900913	31.23374	121.4768	本日、上海で帰って、泊。		loc
14	19900914001	19900914	31.23374	121.4768	杭州に行く。		view
15	19900914002	19900914	31.16136	121.5228	9時05分、郊外に出る。野菜を多く作っている。		view
16	19900914003	19900914	31.11896	121.5498	8:04 上海島。工場が多いが、隣には野菜畑が多い。		view
17	19900914004	19900914	31.05036	121.5861	8:19 杭道を越す。		view
18	19900914005	19900914	31.02071	121.5961	8:21 水田が広い。川、小川、田舎、道には雑木が茂り、川がある。ここには稲が多い。		view
19	19900914006	19900914	30.97797	121.6588	8:29 右折。すく、きそな農村になる。		view
20	19900914007	19900914	30.95794	121.404	8:40 水田が広い。そのほとんどは田舎。小さな橋が直立している。所々に幅10mほどの水路がある。それ以外にはメタセコイアがあることが多い。		view
21	19900914008	19900914	30.94099	121.2982	8:54 ここから9kmで60江。このあたりの村には家の周りに農作物が少しある。家は平入りが多い。鬼瓦の所に風車の飾りをつけたものがある。*		view
22	19900914009	19900914	30.93912	121.2221	9:00 黄浦江。幅200m、10mほどの小舟をラグボートが浮いていて、川岸には薪や手網が多い。		view
23	19900914010	19900914	30.93574	121.1838	9:12 川岸にはまきれに家が建ち、水田が広い。稲の中間は時にカラムガムが落ちていて、		view
24	19900914011	19900914	30.93795	121.1588	9:22 大きな川を過ぎるとき。		view
25	19900914012	19900914	30.93496	121.1165	9:34 マコモ田が多い。		view
26	19900914013	19900914	30.93633	121.0918	9:39 幅50mの水路。道は後の木芝木。		view
27	19900914014	19900914	30.93001	121.0728	9:45-10:19 新島。ここで休憩。近の水路には稲が多い。そのうちのいくつかはコンクリート製の。		view
28	19900914015	19900914	30.6889	120.9983	10:56 浙江直に入る。ここはコンクリート製あり。ここから杭州までは122km。(4)		loc
29	19900914016	19900914	30.67156	120.9633	10:59 小さな田舎のある水田。水田がコンクリート田がある。家の多くはハウスキットが少しある。		view
30	19900914017	19900914	30.79767	120.7285	左2kmで鹿野。ここは上海から93km。		loc
31	19900914018	19900914	30.76672	120.6928	95.1 上海島。ここから上海までは122km。(4)		loc

図3 高谷好一フィールドノートをCSV形式でデータベース化した例

(出所)高田2015。

この時空間上での表示システムには、情報学を専門としない研究者の利用を想定しているため、三つの特徴がある。一つめは、情報の可視化に重点を置き、GIS関連のソフトウェアでよく見られるような複雑な分析ツールを付加せず、容易な操作性を実現したことである。

二つめは、ウェブ上で公開されたオープンソース・デジタルアースを用いた点にある。インターネットにつながりさえすれば、世界中のどこでも無料で利用が可能である。また、いったん情報を入力すれば、オフライン上でも利用が可能であり、Googleが無料で提供するシステムに比べても簡易に利用できる点が優れている。

三つめにデータ入力容易な点である。ユーザーは、情報ごとに位置情報を付加したCSV形式のデータをアップロードすることで時空間上に表示させることができる(図3)。フィールドノートの記録を時空間上に表示させるのに、プログラム言語を理解する必要はない。

2 俯瞰視と微視から読み解く

ズームアウトによる地域情報の読み解き

地域情報を取得し、時空間上で表示させれば、自ずから

地域の像が読み取れるわけではない。スマトラを例にして考えると次のようになる。

スマトラでは、島の西部に三〇〇〇メートル級の急峻なバリサン山地が南北に連なり、そこから、西側はすぐ海に達するが、東側は、丘陵・低湿地が百キロ以上にわたって続く地形が形成されている(図4)。マクロに見た時の地形を大地形、ミクロに見た時の地形を微地形と呼び、山脈・丘陵・低湿地というのがスマトラの大地形に相当するが、実際には、それぞれの大地形のなかにもさらに小さな

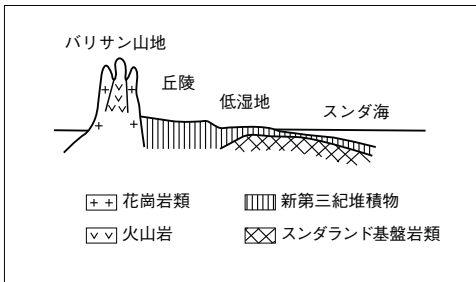


図4 スマトラの横断面図

(出所)古川 1992:4。

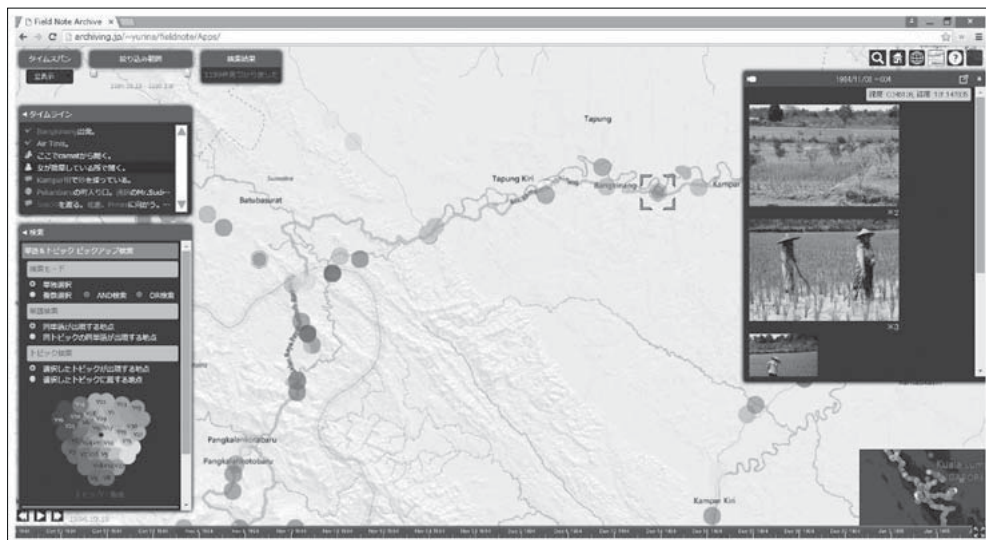


図5 シアク川沿いの自然堤防と後背湿地の複合地形の景観(右上の写真)

(出所)高田(2015)より1984年11月3日の記録を表示。

地形区分が可能であり、もつとも小さい単位を微地形と呼ぶため、大地形から微地形まで、地形の様相は入れ子状態になっていると考えた方がよい。

現地調査で地形を観察する場合、まずはこの微地形に着目する。生業体系や土地利用に関心があるのならば、微地形を、土壌や植生、降水などの他の自然環境要因や、現地の水利用や移住の歴史、開発プロジェクト等と関連させて理解しようとする。

スマトラ中南部の場合、山脈から東に向かい、シアク川やムシ川など、何本もの河川が形成されている。河川が山脈の間の谷を経て丘陵地帯に達すると、河川沿いには自然堤防が形成され、その背後に後背湿地が現れる。たとえば、カンバル市南側を流れるカンバル川流域の場合、カンバル市の南西約二〇キロの地点からバンキナンまでのカンバル川沿いの道は、自然堤防と後背湿地の景観が卓越する。現地で観察可能な微地形は湿地帯や水田、砂質土壌での畑作物栽培、自然堤防上の植生、屋敷地である(図5)。しかし、この図と同じ位置を、時空間表示システムを用いてズームアウトすると図6のようになる。すなわち、図5にある現地で得られた景観を俯瞰視して見れば、カンバル川の流路と三日月湖が入り組んだ景観が川の両側

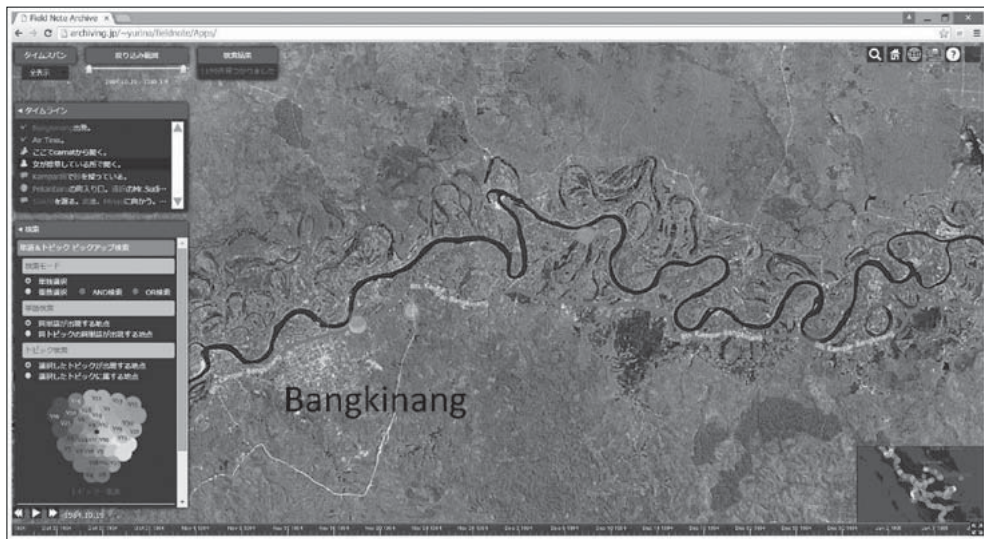


図6 シアク川沿いの自然堤防と後背湿地の複合地形

(出所)高田 2015。

四キロほどの幅をもって連続することがわかる。このことは現地の微視からだけでは読み取ることが難しい。

さらにズームアウトすると、スマトラは急峻な山脈と丘陵、低湿地という大地形から構成されていることがわかる。河川は山脈に源を発し、丘陵地を蛇行したのち、低湿地を経て海につながる。自然堤防と後背湿地の組み合わせは、バンキナンから東側に続き、プカンバルへの別れ道をさらに流下してプカンバルの南東約五〇キロに位置するパシランクリンチまで続いている。そこから先は海の影響を受け、地形的には蛇行する河川が連続し、潮汐によって水位が変動するため洪水の影響を受けにくく、したがって、三日月湖も形成されない。海水の影響を受けるため、河川から供給される真水の量は限定的で、水田のような農地開拓は困難となり、マングローブ林のような、汽水域に特徴的な植生が卓越する。

スマトラの大地形では丘陵と区分される地域のなかに、実際には河川沿いには自然堤防と後背湿地の複合的な地形が形成されることが、地図上でズームイン・アウトを往還することによりわかった。スマトラにおいて河川は重要な交通手段であり、港町に都市が歴史的に形成されてきたことがわかっていく（たとえば、鈴木一九九八）。都市に近

接し、食糧を供給することができた自然環境条件を有する地域は、河川沿いの自然堤防と後背湿地の複合地形に作られた水田地帯であったと考えられる。高谷好一フィールドノートにも「水田はオランダ時代からあるが、二期作になったのは六年前から（一九八四年一月二日、プカンバルからバンキナン方面へ一三キロ地点）」という記述がある。スマトラの大地形区分では丘陵地帯に属するが、河川の中流域にある自然堤防と後背湿地の複合地形に比較的古くからの水田が残っていることがわかる。時空間表示システムを使うことで、さらに、その地形がカンパル川以外の河川沿いにも広くみられることが確認でき、同様の土地利用がスマトラで一般的であることが推測される。このように、現地調査の微視から得られる情報を俯瞰視の情報と組み合わせることで、個別の地域情報の持つ空間的拡がりを確認することができる。

ズームインによる地域情報の読み解き

河川と河川との間の地域には、大地形の区分でいうところの丘陵地帯が卓越する。緩やかに起伏する丘陵が続き、土壌の肥沃度が低く、利用可能な水も限られているため、水田の造成や畑作物の栽培は困難である。丘陵地帯に卓越

する土地利用はゴムとアブラヤシである。

河川沿いの自然堤防・後背湿地の複合地形に卓越する水稲栽培と異なり、ゴムとアブラヤシのいずれが栽培されているのか、あるいは、栽培の主体は小農か企業かといった両者の違いを、自然環境条件で説明することは困難である。したがって、その地理的分布を自然環境条件の観察だけから推測することも難しい。そのためには、むしろ、導入の歴史的経緯を知る必要がある。

しかし、ゴムとアブラヤシが実際に栽培されている土地を、微視と俯瞰視を組み合わせて確認することは可能である。アブラヤシは一株の大きさが直径五メートルほどにもなり、デジタル地図上で確認することができるし、ゴムと異なり、整然と植栽されている。また、企業によるプランテーション経営では面積が広大なモノカルチャーが多いのに対し、小農による経営では、小面積であることに加え、ゴムや果樹園など、他の土地利用がモザイク状に組み合わせられていることが現地調査からわかる。こうしたことを考慮すると、解像度による違いはあるが、スマトラの丘陵地におけるゴムとアブラヤシの分布の概略を確認することができる。図7の道路右側に整然と植えられているのがアブラヤシであり、その周りをゴム園と果樹園が囲んでいる。



図7 Bing Maps Aerial Imagery から見た丘陵地におけるゴムとアブラヤシの分布
ジャンビ州と南スマトラ州の州境からジャンビ方面に20キロ地点

(出所) Bing Maps Aerial Imagery.

このように、現地調査の微視で得られた指標を基に、デジタル地図上から地域情報を読み取ることが可能である。^{*3}

3 歴史から読み解く

地域情報を時空間上で表示することにより、地域研究の現場で、過去の記録を現在の状況と比較することが可能になる。

筆者による現地調査では、バリサン山地から丘陵地帯にかけて、とくに屋敷地やその周辺の菜園・果樹園に多数のカカオが栽培されているのを観察した。そこで、高谷好一フィールドノートから「カカオ」「ココア」をフリーワード検索したところ、ヒット数は「カカオ」○件、「ココア」二件であった。しかし、フィールドノートの記録を読むと、現在のカカオと同じ生育条件下でコーヒーが栽培されていることがわかった。たとえば、西スマトラ州ソロクから東に向かう国道沿いでは次のような記述が見られる。「Minangkabauの家。屋敷地には大きな果樹に交じってコーヒーがある。こうした屋敷地はすべて柵で囲われている。もう暗くて見にくい。(一九八四年一〇月二三日、ソロクの東八三キロ地点)」「左に川、二〇×一・五メートル。そこに灌漑用の水車がか

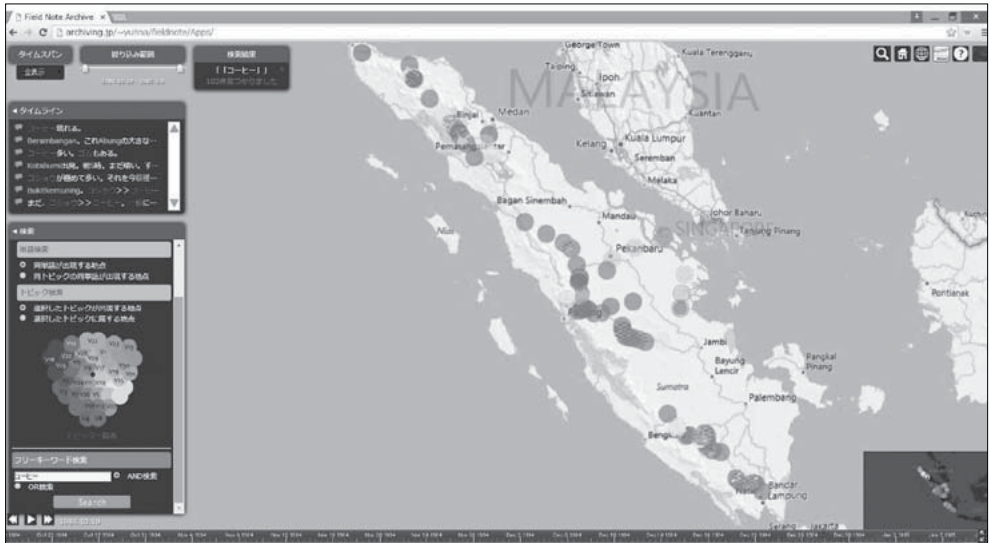


図8 高谷好一フィールドノートで「コーヒー」を検索した結果
(出所)高田 2015。

かっている。周りには二カ月稲あり。斜面にはコーヒーが多い。(一九八四年二月一九日、ソロクの東三七キロ地点)「Silungkangの町。周りにはコーヒーが大変多い。(一九八四年一〇月三日、ソロクの東一九キロ地点)」などである。そこで、コーヒーに関する記述のある地点のみを取り出すと図8のようになった。コーヒーに関する記述は一〇二件がヒットし、スマトラのかなり広域でコーヒーに関する記述が存在することがわかった。筆者による現地調査では、現地ではほとんどコーヒーの木を見ることはなく、上述した高谷の記録にあるソロクから東に向かう国道沿いも同様であった。カカオはアオイ科の常緑樹で、コーヒーと同様、被陰樹の下で栽培されることが多い。筆者の調査時に、高谷の調査時と同様、屋敷地や菜園には多くの果樹が残されていたが、コーヒーはほとんどみられず、代わりに、同様の生育環境で育つかカオに代替されていたのである。

インドネシアは世界第三位のカカオ輸出国であり、国内ではスラウェシでの生産が七五%を占める (International Cocoa Organization 2015; Hasnah et al. 2011)。西スマトラでは、主に二〇〇〇年代以降になって本格的に導入され、ゴムやアブラヤシと異なり、小農による生産が九七%を占めると言う (Hasnah et al. 2011)。筆者の現地調査で観察

されたのは、近年になって拡大したカカオであり、その多くは屋敷地内の果樹園で栽培されており、高谷の調査した一九八〇年代にはコーヒーが植えられている場所であった。地域情報を読み解くには、かつての歴史的な経緯を知ることが重要であり、地域情報の時空間表示を発見のツールとして利用することができる。

地域研究の重要な方法論の一つに臨地調査がある。その一段としての景観観察は高谷の現地調査でも重要な方法論として採用されている。景観観察の手法は、単に観察者のいる道路からの眺め（車窓の景観）だけで地域全体の景観を判断していると批判されることもある。しかし、実際には、地形を見る場合のように、大地形から微地形まで常にズームイン・アウトを頭の中で繰り返している。逆に言えば、その往還がなければ、現地で見ている景観を理解することができない。オープンソース・デジタルアースを用いて地域情報を時空間上で表示することにより、ズームイン・アウトの往還が現場で可能になる。それにより、従来は専門家でなければ困難であった現地での地形分類という作業を、より簡便に行うことができ、現場で見えている景観を地域全体における個別性と普遍性の中で理解することが可能になる。

III テキスト分析から読み解く

1 フィールドノートのテキスト分析

一般に、情報学的なテキスト分析の手法は日進月歩であるが、フィールドノートのような膨大な地域情報から、地域像の抽出や個別事象間の関連性を発見するような、従来は研究者が行っていた思考のプロセスをコンピュータに完全に代替させることは依然として困難である。しかし、現在の情報技術では、ある文書にどの語彙が何回登場するかという単純な量的集計だけでなく、語彙の内容に踏み込んだ質的分析のための手法が発達してきた。そこで本稿では、語彙間の関連性から潜在的トピックを抽出するテキスト分析の手法を用い、統計的に得られた結果が、地域像の解明や地域理解にどのように貢献できるのかを検討する。そのために、トピックモデル LDA (Latent Dirichlet Allocation) の手法を用いて、高谷好一フィールドノートから潜在的トピックの抽出を行った。^{*4} フィールドノートの

記述からLDAを用いてトピックを抽出するためのくわしいプロセスとそこから得られる情報学的な意義、あるいは、抽出された個々の用語についての妥当性は別稿にゆずるとして、本稿では、潜在的トピックが、地域研究者にとつて実際にどのように利用可能かを、時間間上で表示することによって検討する。地域研究の手法という観点からいえば、テキスト分析によって抽出された潜在的トピックは何らかの地域性を示しているのか、示しているとすればどのように地域性の発見に利用できるのかという点が重要である。本稿では、潜在的トピックの可視化を通じて、フィールドノートのテキスト分析に必要な利用者側（地域研究者側）の視点を提供し、情報学的処理を加える際の技術的な改善と、分析結果の有効利用について検討する。

2 潜在的トピックの抽出

本研究ではトピックモデルLDAの手法を用いた潜在的トピックの抽出を行った。テキスト分析は、まずフィールドノートのテキストの構造化、テキストからの用語の抽出、トピックの抽出というプロセスを経る。フィールドノートでは一地点ごとに情報が記載されているため、テキ

ストの構造化では、分析の単位となるのは一地点での情報（場面）である。場面に含まれる文字情報（語彙）を、MeCabとIPADICを利用して形態素解析を行い、名詞と形容詞を対象として用語を抽出した。^{*5}抽出した用語からLDAを用いて、各場面に潜在的にトピックが存在すると仮定し、各場面での用語の共起関係から潜在的トピックを求めた。これにより、場面とトピックの関係、および用語とトピックの関係が出力された。その結果、場面間の類似性やある用語が出現する潜在的トピックはどのようなものか、異なる用語が共起するような場面での潜在的トピックは何かといった分析が可能となった。

分析の結果、高谷好一フィールドノートのスマトラでの記述から、用語の異なり数五六六と、三〇の潜在的トピックを得ることができた（表1）。表1には、V1からV30のトピックと、トピックごとに出現頻度の高い方から上位五つの用語を記載した。

3 用語の妥当性

テキスト分析をするためには、先述したように、テキストをまず構造化した後、用語を抽出するが、その際、テキ

表1 スマトラのフィールドノートの記録から得られた潜在的トピック一覧

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
タイトル	?	?	?	?	?	傾斜畑	地形的特徴	?	池・魚・水田複合	町と市場
1	集落 8	島 10	Loc 8	Bengkalis 8	チガヤ 4	多い 94	松 12	きれいな 6	池 28	町 30
2	乳液 6	多い 8	北進 4	土手 6	Tembilahan 3	オカボ 92	悪い 7	広大 5	水田 25	店 17
3	女 4	周辺 6	島 3	墓 5	急 3	トウモロコシ 86	湖 7	ゴム園 4	魚池 16	市場 9
4	新しい 4	ton 5	苗木 3	オランダヤ 4	松林 3	広い 80	平坦 6	広い 4	魚 13	北 7
5	灌木 4	松林 5	Pekanbaru 2	乾季 4	煉瓦 3	コーヒー 52	所々 6	ヨシ原 3	小池 6	Arsad 5
	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
タイトル	?	ゴム・水田複合	?	中国人・ムラユ	炭	?	サゴ	オランダ時代	ゴム	?
1	Sultan 6	ゴム 59	地区 11	中国人 103	木 16	オカボ畑 5	サゴ 91	オランダ 33	ゴム 142	バガン 19
2	pres 5	水田 44	丸太 7	人 103	炭 10	Buatan 4	工場 59	下 16	広い 90	Tebing Tinggi 6
3	森 5	ゴム園 21	会社 6	自分 82	窯 10	松 4	tual 44	ムラユ 14	ゴム園 41	Tanjung Datuk 5
4	簡単 4	Minangkabau 19	レジ 5	無い 79	炭焼き小屋 8	Amuntai 3	サゴヤシ 35	Raja Kecil 12	タッピング 27	Tanjung Pinang 5
5	Ungku Tugut 3	suku 16	分かれ 5	家 78	直径 7	Loc 3	水 34	間 9	丘 24	核 5
	V21	V22	V23	V24	V25	V26	V27	V28	V29	V30
タイトル	水田	魚、養殖	家集落	屋敷地と菜園	川	民族と水田	水田と除草	耕起	コショウ + ドリアン	ココヤシ
1	水田 108	魚 43	家 77	多い 227	左 4	Banjar 49	草 108	牛 16	コショウ 19	ココヤシ 103
2	広い 84	網 28	多い 59	家 118	下り 3	Bugis 37	鋳 98	長い 16	ドリアン 12	木 75
3	稲 65	長い 26	右 40	コーヒー 95	川口 3	Sapat 37	田 95	クビキ 8	根元 7	自分 58
4	多い 61	inch 23	マングローブ 29	ココヤシ 91	昼食 3	Tembrahan 29	多い 86	土 8	中心 5	良い 42
5	幅 61	エビ 18	集落 24	村 61	湿地 3	稲 27	水田 72	犁先 7	成木 5	泥炭 41

(データ出所)高田 2015。

ストの内容には本来、関係しない用語を排除する必要がある*。このことはテキスト分析を行う際の基本的な情報学的課題であるが、本稿では、とくにフィールドノートの情報を扱う場合に発生する用語の妥当性について検討する。

高谷好一フィールドノートは現場の観察記録である。地形の特徴や農作物の栽培状況などが記載され、実際、「水田」「ゴム」「ココヤシ」「コーヒー」「鋤」「池」といった用語が頻出する。これらの用語は、現地での記録時に、地理的分布や栽培状況と合わせて記録するために、「ゴム園が広い」「ココヤシが多い」といった量的な意味を含む用語を合わせて使うことがしばしばある。その結果、たとえば「広い」と「狭い」では逆の意味になるもの、付随する名詞にあわせていずれの用語も出現頻度が高まる。さらに、フィールドノートを記録する実際の現場では、本来はそのあたりに少ないはず、あるいは、これまでは見かけることが少なかったにもかかわらず、観察中のその場所については「多い」あるいは「少ない」という意味を込めて量的な表現をすることがある。この場合、「多い」か「少ない」かを理解するにはフィールドノートの前後の文脈を知ることがあり、用語のみで多いか少ないかを判断することはできない。同様に、「上下左右」「前後」といった用語も

名詞に付随することが多く、用語として抽出されることがあるが、これらも前後の文脈を考慮する必要がある。

テキスト分析を行う際、扱う資料の性格によって抽出、あるいは、排除されるべき用語は異なってくるであろう。

高谷好一フィールドノートの場合、生業体系や土地利用に関する用語が頻出し、重要な観点が地形と生業の関係であった。そのため、それらに付随して、一見、不要な用語が抽出された。不要な用語を排除すると同時に、文脈を理解した抽出とするためには、相反する言葉が付随する用語の関係を理解し、前後の文脈から意味を相対化して用語を抽出するような工夫が課題となる。

ただし、このことは、情報学だけの課題ではない。「多少」や「広狭」といった量的な表現を現地調査時にフィールドノートに走り書きする場合はよくあるかもしれないが、あいまいな表現であることも事実である。フィールドノートを取る際の工夫としても考慮されるべきであろう。

4 トピックから読み解く

トピックの読み解き

潜在的トピックを抽出した結果、V1からV30で示した

三〇のトピックが得られた(表1)。さらに、トピックに示された用語を、高谷好一フィールドノートの記載や現地調査をふまえて筆者がタイトルをつけた。クエスチョンマークを付したのは、用語だけでは内容が判別しないトピックである。

トピックのタイトルは、主に生業体系や土地利用についてのトピック(V6・7・9・12・15・17・19・21・22・23・24・25・27・28・29・30、以下、生業トピック)と、地域の歴史に関するトピック(V10・14・18・26、以下、歴史トピック)とに分けることができた。^{*7}また、クエスチョンマークを付したものでも、V4・11・20のように、地名や民族の移住の歴史等に触れる用語が見られ、歴史トピックであると分類できるトピックもある。実際、高谷好一フィールドノートを通読してみても、記載されている情報は現地の景観観察によるものばかりではなく、しばしば、町や村で役人や村の幹部、古老から、それぞれの地域の歴史的経緯を聞き取りしていることがわかる。冒頭でも述べたように、現地の生業体系や土地利用を理解するためには、現在の状態だけでなく、そこにいたる歴史的経緯を知る必要がある。歴史的経緯の違いが、それらの多様性にかかわっているからである。

フィールドノートの読み解き

フィールドノートのような膨大な地域情報の集積には、地域に関するさまざまな情報が含まれる。そのため、他人が記録したフィールドノートを、地域情報の源として利用するには、当然フィールドノートをすべて読み込む必要がある。しかし、多数のフィールドノートが集積した場合、記録は膨大になり、すべてを読み込むことは難しい。

フィールドノートに含まれる多様な情報からユーザーが必要な情報を抽出するもつとも簡便で普及している方法はキーワード検索であろう。しかし、フィールドノートの記録のなかで何が重要なのか、言い換えると、現地の地域情報のなかで何に着目すべきかを理解せずして適切なキーワードを得ることはできない。フィールドノートに存在しないものをキーワードにして検索してもヒットしないし、あまりにも普及しているものを検索すると、ヒット数は増大するが、結局多数の記録を読み込まざるをえない。たとえば、スマトラのフィールドノートで「オランウータン」を検索してもヒット数は〇件だが、「水田」では一八二件がヒットし、記録された位置もスマトラ全土に分布する。

しかし、テキスト分析の手法、とくに本研究で利用したLDAによる手法では、関連性が高いと統計的に判断され

た語彙が用語として抽出され、たとえ同一の語彙でも、共起するパターンが異なれば、異なる用語として抽出される原理になっている。表1に示したテキスト分析の結果によれば、トピックのタイトルに「水田」に関するトピックが見出せるのと同時に、用語で見ても「水田」が存在する。異なる複数のトピックに「水田」が見出されることは、テキスト分析の手法により、異なる状態の「水田」が統計的に抽出されたことを意味している。

タイトルに「水田」を含む、あるいは水田に関連するトピックは、V9「池・魚・水田複合」、V12「ゴム・水田複合」、V21「水田」、V26「民族と水田」、V27「水田と除草」、V28「耕起」の六トピックである。V9「池・魚・水田複合」とV12「ゴム・水田複合」は水田が他の土地利用と組み合わせられて存在することを示している。前者は池と水田の組み合わせであり、後者はゴムと水田の組み合わせである。立地環境としては、前者は低湿地が多く、後者は丘陵地が多いであろう。これに対しV21「水田」は、むしろ水田が中心的な話題であり、V26「民族と水田」では、やはり水田が中心ではあるが、それを栽培する民族に関連する記述が多い。また、V27「水田と除草」とV28「耕起」では、水田における農耕技術が中心的な話題

になっていると考えられる。フィールドノートを読まずして、これだけの「水田」を分類することができ、用語の構成や実際の記述からも関連性のあることが示唆される。

語彙間の関係や語彙に付随する言葉の関係から語彙間の関連を統計的に判断するため、テキスト分析の結果は、フィールドノートの記録の全貌を知らずしてトピックや用語の分類を行うことが可能であった。ただし、本研究で統計的に抽出されたさまざまな「水田」が、実際にどのような異なるのかは、現地調査やさらなるトピックや用語の解析が必要になる。統計的な処理により見出された、異なる「水田」が、実際にどのように異なるのかはフィールドノートの記録だけでは判断することができない。テキスト分析によって抽出された用語間の関係性は相関関係であり、因果関係は研究者が検討すべき課題である。そして、抽出されたトピックは統計的に関連がある用語の集合体であり、トピックに現れる頻出用語からトピックのタイトルの考えることは、方法的には作業仮説の提示と言い換えることができる。トピックに現れる用語の因果関係を検討したり、作業仮説を修正したりすることにより、テキスト分析の結果を発見のツールとして使うことができる。

トピックの物語の読み解き

用語の関連性を抽出するテキスト分析の手法は、フィールドノートに頻出し広範囲に存在する「水田」のような用語を統計的に分類することに有用であったが、逆に、特定の地域や対象物を表す用語では、出現頻度が少ないため、キーワード検索とテキスト分析による結果の間に差が生じないと考えられるが、テキスト分析の結果は、必ずしもそうではなかった。

サゴはヤシ科の植物で、通常、海水の影響を受ける、海岸近くの河川沿いの汽水域によく生育する植物である。スマトラでは、サゴの髓に含まれるでんぷんを食用あるいは家畜の餌として利用する。サゴの植物的特性から、スマトラにおけるサゴの生育域は東海岸と西海岸の低湿地帯ではない。実際、フリーワード検索で「サゴ」を検索すると五八件がヒットし、時空間上で表示すると図9の下図のようになった。ヒットした五八件には、「サゴヤシ」や「サゴ工場」など、サゴに関する記述が含まれ、フィールドノートの記録を確認しても、サゴの立地環境から利用まで幅広く情報が記載されていることがわかる。一部、内陸部での分布が見られるが、多くは東海岸沿いに集中し、自然環境要因だけでもサゴの用語の分布をある程度説明するこ

とができた。

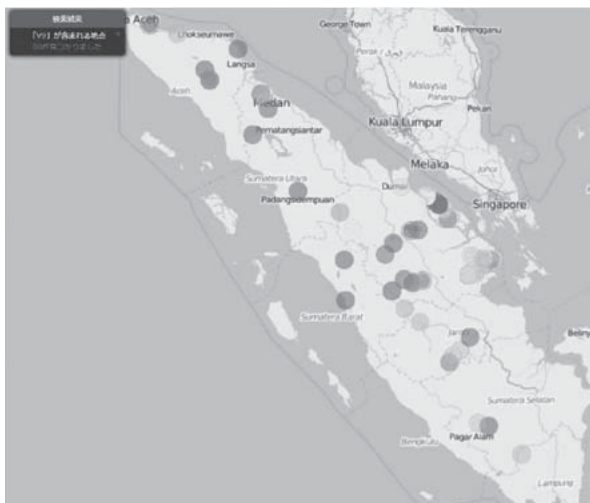
しかし、トピック「サゴ」に分類された用語を検索すると、ヒット数八〇件となり、フリーワード検索の場合よりヒット数は二二件増加した。また、時空間上で表示すると、内陸での分布が多いことが明らかになった(図9上)。

このことは、二二件の情報は「サゴ」を含まないにもかかわらず、「サゴ」という用語と共起する情報として抽出されたことを意味した。トピック「サゴ」に含まれる八〇件のフィールドノートの記録を読むと、「サゴ」の立地環境の話だけでなく、中国からの移民によるサゴの開拓やサゴ工場の建設、あるいはサゴが生育する湿地林の開拓や販売していた現金収入源としての籐(ラタン)について書かれていた。すなわち、トピック「サゴ」の記録には、サゴそのものだけでなく、サゴにまつわる開拓や移民、民族などに関する情報が付随しており、それらの用語が共起すると判断され、その結果、サゴが生育することのほとんどない内陸に位置し、情報としても「サゴ」という語彙を含まないにもかかわらず、「開拓」や「中国」「湿地林」「籐」という用語からなる一部の情報が「サゴ」と共起する情報として抽出されたと考えられる。

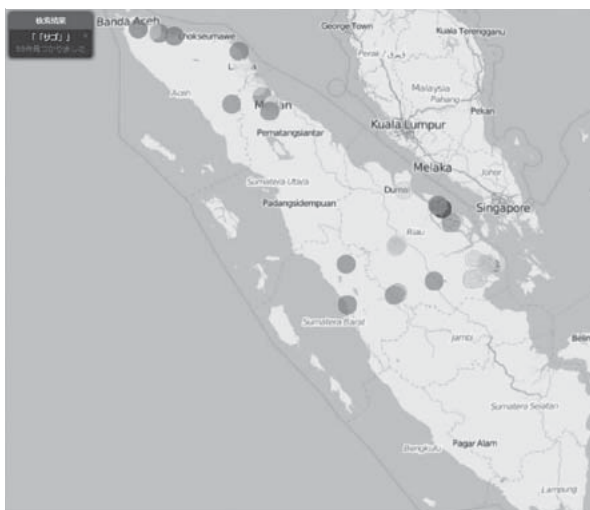
サゴが本来生育しない内陸に「サゴ」に関する情報が抽

出されることは、現場の状況を知らず専門の知識のないユーザーにとっては誤った情報の提供になる。しかし、スマトラ東海岸では、中国からの移民が湿地帯のサゴを開拓し、サゴ工場を建設して経済発展に資するというユニークな歴史的経緯（物語）が存在し、そのことと関連する情報がトピック「サゴ」のなかに見出されたのだと考えれば、

「サゴ」という用語を含まないトピック「サゴ」の二二件の情報をあらためて見直し、トピック「サゴ」に付随する物語を読み解くことで、新たな作業仮説を立て、より具体的な研究課題として追及することが可能になる。



トピック「サゴ」、80件ヒット



フリーワード検索「サゴ」、58件のヒット

図9 「サゴ」に関する情報を可視化した二つの結果

(出所) 高田 2015。

結論

本研究では、情報学を取り込んだ地域研究の新しい研究方法を検討するための最初のステップとして、フィールド

ワークによって得られる現地の情報を、オープンソース・デジタルアースとテキスト分析を用い、いかに発見のツールとして利用することができるとかを検討した。題材としたのは、高谷好一フィールドノートのインドネシア・スマトラにおける記録であった。その結果、スマトラの大地形で丘陵と区分される地域のなかに、自然堤防と後背湿地の複合地形が河川中流域に卓越していることがわかった。また、同じく丘陵と区分される地域のなかで、ゴム園とアブラヤシ園の分布の概略を確認することができた。前者は現地調査の微視だけでは見えない地域情報が、時空間表示システムを用いてズームアウトすることで理解され、後者は逆に、俯瞰視だけでは見えないゴムとアブラヤシの区分をズームインすることで違いの指標を見出し、俯瞰的な図に適用することで、分布の概略を確認することができた。また、現地調査で観察されたカカオが、一九八〇年代の高谷

好一フィールドノートの記録にほとんど見出すことができなかつたものの、当時の記録とカカオの歴史的経緯を踏まえると、かつてのコーヒー栽培がカカオに代替したことがわかった。いずれの場合も時空間表示システムを用いてズームイン・アウトの往還を繰り返すことで、発見の糸口を見出すことができた。

テキスト分析では、LDAの手法を用いて、用語間の関係性を基にして、共起する用語、用語の塊であるトピックとを抽出し、地域研究の手法への適用を検討した。本手法のもつとも重要な特徴は共起する情報をつなぐ点にある。すなわちトピックは、統計的につながっていると判断された情報の塊であり、相関関係の提示であった。本稿では、スマトラで広域に観察される「水田」という用語から、相関関係でつながった用語群からなる六つの異なる「水田」を見出すことができた。相関関係を基に作業仮説を立て、因果関係を検証するというプロセスは研究者による思考プロセスと変わるところはない。人間の思考でも、多くの作業仮説を作っては壊しまた作るのであって、本稿で検討したテキスト分析は、その過程に、コンピュータによる統計的処理という判断で得られた作業仮説を付け加える行為に他ならない。しかも、膨大な情報から、つながりの程度に

応じて重みづけをし、共起する情報を見出すことは人間にとって困難な課題である。その点はコンピュータの得意分野である。テキスト分析を地域研究者の思考プロセスに組み込むことにより、地域情報の記録者やユーザーが明示的に意識していなかったようなトピックの相関関係や、あまりにも膨大な資料群のため検出がこれまでは不可能であったようなトピック間の相関関係を探り出すことに有効であると考えられる。

また本研究では、用語に付随する歴史的経緯のような背景的情報を物語とよび、物語を読み解く必要性を指摘した。数値情報と異なり、テキスト情報の分析にはまだあいまいさがつきまとう。本研究で示したように、「サゴ」という用語が内陸部に分布している結果を前にして、それはテキストのあいまいさに由来する間違った結果なのか、あるいは、何らかの現象を示しているのかの判断は、研究者が現場で作業仮説を立てる時の判断と似ている。地域情報学の場合、地域研究と情報学の視点や手法が異なるため、作業仮説を立てるための判断は共同で行う必要がある。学際研究では異分野の研究者が相手の研究分野に踏み込んでまで議論することが望ましいようだが、双方のかかわりを通じて、地域情報学から新しい手法を見出すことができる。

●謝辞

本研究は、京都大学地域研究統合情報センターの地域情報学プロジェクトから多大な支援を受けた。また、『地域研究』誌の査読者の的確・迅速かつ建設的なコメントにも謝意を表したい。

●注

*1 たとえば（水島・柴山二〇〇九・京都大学フィールド情報学研究会二〇〇九）。また、GISをベースにした歴史分析におけるアジアでの研究体制の整備と研究の進展を図るための学会「アジア歴史地理情報学会（Asian Network for GIS Studies-Japan）」も設立され二〇一二年から活動を開始した（アジア歴史地理情報学会二〇一五）。

*2 以下のURLを参照 <http://heldnote.mapping.jp>。表示システム構築のための情報学的プロセスと課題は（高田ほか二〇一四・Takata et al. 2014）を参照。

*3 もちろん、経営主体はもとより、ゴムとアブラヤシの栽培地の区別についても、正確な境界をみつけ面積を計測するには、画像上でのより正確な解析が必要である。

*4 本稿で用いたテキストマイニングのくわしい手法や情報学的な意義等は、山田（二〇一五a・二〇一五b）による。

*5 地名や数値、助詞、テキストの入力間違い、不適切の分節された専門用語など、抽出される用語として不適切なものを排除した。

*6 LDAでは、共起する用語の関係性が集計されてトピックを区切るため、トピック間の関係性は問わない。そのた

め、本稿でもトピック間の関係性は議論しない。

*7 町の様子や市場の様子を記載したトピックもいくつかの用語は歴史的経緯を反映しているため地域の歴史に関するトピックとして分類した。

*8 トピックの可視化に使用したのは二〇一四年段階の抽出データを用いているが、表1で示したトピックの一覧は用語の抽出時に若干の形容詞を加えた二〇一五年九月時点のデータを採用している。Vの順番に統計的な意味はなく、トピックが表す内容(タイトル)は、トピック内に現れる用語の頻度が高いものを基に判断している。

●参考文献

- アジア歴史地理情報学会 (二〇一五) http://www.iit-tokyo.ac.jp/~angis/about/ANGIS_Japan_jihml (二〇一五年二月一日)
- 古川久雄 (一九九二) 『インドネシアの低湿地』 勁草書房。
- 京都大学フィールド情報学研究科編 (二〇〇九) 『フィールド情報学入門——自然観察、社会参加、イノベーションのため情報学』 共立出版。
- 水島司・柴山守 (二〇〇九) 『地域研究のためのGIS』 古今書院。
- 大阪府立大学学術情報センター (二〇一五) <http://nakaodh.center.osakafu-u.ac.jp/> (二〇一五年二月一日)
- 柴山守 (二〇〇九) 「特集 地域情報学——地域研究と情報学の新たな地平——序論」『東南アジア研究』四六巻四号、四八—一四九頁。
- 柴山守・原正一郎 (二〇〇八) 「地域情報学の目指すところ

——地域研究におけるGISの応用」『アジア遊学 特集地域情報学の創出』一—三、二八—三五頁。

鈴木恒之 (一九九八) 「東南アジアの港市国家」『世界歴史一三 東アジア・東南アジア伝統社会の形成』岩波書店。

高田百合奈 (二〇一五) <http://archiving.jp/yurina/fieldnote/Apps/> (二〇一六年二月二九日)

高田百合奈・渡邊英徳・柳澤雅之・山田太造 (二〇一四) 「位置情報とトピックモデルに基づくフィールドノートのビジュアルイズ手法」情報処理学会人文科学とコンピュータシンポジウム(じんもんこん二〇一四)発表資料、国立情報学研究所／一橋講堂、二〇一四年二月二三日。

高谷好一 (二〇一二) 『地域研究アーカイブズ フィールドノート集成』一—五巻 (CIAS Discussion Paper No. 22)、京都大学東南アジア研究センター・地域研究統合情報センター。

高谷好一 (二〇一三) 『地域研究アーカイブズ フィールドノート集成』六—八巻 (CIAS Discussion Paper No. 36)、京都大学東南アジア研究センター・地域研究統合情報センター。

山田太造 (二〇一五a) 「地域研究資料と対象とした時空間情報に着目したデータの構造化」『人文科学とコンピュータ研究会報告』vol. 2015-CH-105, no.4, pp. 1-6。

山田太造 (二〇一五b) 「地域研究史資料とした時空間的特徴の抽出と場面の構造化」『第一四回情報科学技術フォーラム講演論文集』vol. 14, no. 4, pp. 409-410。

柳澤雅之 (二〇一二) 「自然科学分野の地域研究——地域情報の限定性を克服するために」『地域研究』一二号巻二二号、一

一六一―一三〇頁。

柳澤雅之(二〇一四)「フィールドノート・プロジェクト」

『Seeder』一―号、一四―一二三頁。

山本博之(二〇二二)「新しい地域研究をめざして」『地域研究』二二巻二号、六一―一五頁。

Barkin, G., and G. D. Stone (2004) Field Notes as a Web Site:

Integrating Multimedia into Anthropological Documents.

Field Methods, 16 (2): 203-214.

Bing Maps Aerial Imagery. <http://www.bing.com/maps/> (二〇一六年二月一九日)

Buckland, M. (2008) Cultural Atlases: The Electronic Cultural

Atlas Initiative and Related Studies at Berkeley. 1997-2008.

http://eca.org/about/ECAL_history_Buckland.pdf

Hasnah, F. Euan, A.V. Renato, and P. Ian (2011) *The Potential*

of Cocoa Agribusiness for Poverty Alleviation in West

Sunatra. The 55th National Conference of the Australian

Agricultural and Resource Economics Society, Melbourne,

Australia. Feb. 8-11, 2011. <http://ageconsearch.umn.edu/bitstream/100555/2/Hasnah.pdf> (二〇一六年二月一九日)

International Cocoa Organization (2015) Statistics. <http://www.icco.org/about-us/international-cocoa-agreements/cat-view/30-related-documents/46-statistics-production.html> (二〇一五年一月二日)

Takata, Yurina, Hidenori Watanabe, Masayuki Yanagisawa

and Taiso Yamada (2014) A Visualization Method of Field

Notes based on Locations and Topic Models, Japanese

Association for Digital Humanities Conference 2014,

Tsukuba University, Sep. 20, 2014. (発表資料)

●著者紹介●

- ①氏名……柳澤雅之(やなぎわ・まゆゆき)。
 - ②所属・職名……京都大学地域研究統合情報センター・准教授。
 - ③生年・出身地……一九六七年、奈良県。
 - ④専門分野・地域……農業生態学・ベトナム。
 - ⑤学歴……京都大学大学院農学研究科熱帯農学専攻、農学博士。
 - ⑥職歴……京都大学東南アジア研究センター助手・助教授(六年間)・京都大学地域研究統合情報センター・准教授(二〇年間)。
 - ⑦現地滞在経験……タイ(二三歳、合計一年)、現地調査、三五歳、合計一年半、京都大学東南アジア研究センターバンコク連絡事務所駐在員)、ベトナム(二七歳よりベトナムでの現地調査を開始、二週間〜一カ月程度の調査を年間二〜三回程度を繰り返す、三〇歳、二年間、East-West Center 研究プログラム・コーディネーター)。
 - ⑧研究方法……フィールドワークを主な手法とする。現場の観察と、土地利用履歴や農林業を中心とした生業活動についての農民への聞き取り調査。
 - ⑨所属学会……日本熱帯農業学会、日本熱帯生態学学会、東南アジア学会。
 - ⑩研究上の画期……フィールドワーク中に画期が何度も訪れる。
- ⑪推薦図書……『京大式フィールドワーク入門』京都大学大学院アジア・アフリカ地域研究科・京都大学東南アジア研究所(編)NIT出版、二〇〇六年。

●著者紹介●

- ①氏名……高田百合奈(たかた・ゆりな)。
- ②所属・職名……首都大学東京大学院システムデザイン研究科博士後期課程三年。
- ③生年・出身地……一九八八年、岡山県。
- ④専門分野・地域……情報デザイン。
- ⑤学歴……首都大学東京大学院 システムデザイン研究科、芸術工學修士。
- ⑥職歴……東京大学公共政策大学院学術支援専門職員(平成二四年から一年間)、東京大学公共政策大学院特任研究員(平成二五年から一年間半)、佐賀大学客員研究員(平成二八年から現在に至る)。
- ⑦研究方法……地図上へのデータのビジュアライズ。ソーシャルデータの収集と分析。ユーザの空間把握能力に関する分析。
- ⑧所属学会……日本バーチャルリアリティ学会。電子情報通信学会。
- ⑨研究上の画期……スマートデバイスや、ソーシャルメディアの変遷に應じる。またオープンソース化や各種APIの提供と廃止時期など。
- ⑩推薦図書……『データを紡いで社会につなぐ——デジタルアーカイブのつくり方(渡邊英徳、講談社現代新書、二〇一三年)』。

●著者紹介●

- ①氏名……山田太造(やまだ・たいぞう)。
- ②所属・職名……東京大学史料編纂所・助教。
- ③生年・出身地……一九七七年、愛媛県。
- ④専門分野・地域……データ工学、歴史情報学。
- ⑤学歴……総合研究大学院大学複合科学研究科情報学専攻、博士(情報学)。
- ⑥職歴……国立情報学研究所・特任研究員(二〇〇五～二〇〇七年)、東京大学史料編纂所・特任助教(二〇〇七～二〇一〇年)、人間文化研究機構本部・特任助教(二〇一〇～二〇一三年)、東京大学史料編纂所・助教(二〇一三年)。
- ⑦現地滞在経験……日本とロシアにて前近代日本史に関係する史料を所蔵する機関を中心に短期調査を年に数回。
- ⑧研究方法……データ検索手法の改善、テキストを中心としたデータ分析、史料調査、他分野研究者への聞き取り。
- ⑨所属学会……情報処理学会、日本データベース学会、日本デジタル・ヒューマニティーズ学会。
- ⑩研究上の画期……二〇〇五年の日本史研究者とのふれあい・日本関係史料との出会い、二〇一〇年以降のビッグデータによるパラダイムシフト。
- ⑪推薦図書……『情報検索の基礎』(Christopher D. Manningほか、共立出版、二〇一二年)。