

ビッグデータからみた地域の諸文化

—方言と食文化を事例に—

桐 村 喬*

I. はじめに

1. 背景

例えば、「ある地域では納豆を毎日のように食べる」、「ある地域ではゴミを捨てることを「ゴミを投げる」と言う」というような、食物や料理、習慣や方言など、一般的な意味での文化に関する地域的差異は、我々の日常会話においてたびたび話題にされる。このような文化についての地域的差異を主なテーマとして取り上げるテレビ番組も多く、様々な現象に存在する地域的差異のなかで、文化の地域的差異は、現代日本における一般社会の関心が比較的高いものと考えられる。

文化の地域的差異のうち、食物への関心の高さは、観光対象としての「食」にも結び付いており、近年は、比較的安価な名物料理の観光資源化など、食をテーマとした地域振興やまちづくりが行われるようになってきている¹⁾。宇都宮市による「餃子」の活用はその代表例であり、観光客の94.1%が宇都宮市を「餃子のまち」と認識し、観光客の60.2%が餃子を食べることを目的として来訪するようになっている²⁾。餃子については浜松市も同様に注目されており、2015年1月に公表さ

れた家計調査の結果によって、2014年の浜松市の1世帯当たりの年間支出金額が宇都宮市のそれを上回ったことがニュースとして全国に報道されるなど³⁾、食物についての地域的差異への関心の高さと観光資源としての価値は、地域間での競争も引き起こしている。

また、方言に関しても、テレビ番組の企画をもとに、「アホ」や「バカ」に関する分布を日本全国で調査した結果が『全国アホ・バカ分布考』⁴⁾としてまとめられ、方言研究者だけでなく、一般社会においても同書への関心が寄せられたのは記憶に新しい。また、東京女子大学の篠崎ゼミでは、インターネット上で辞書・事典情報を提供するジャパンナレッジ⁵⁾と共同で、「出身地鑑定!! 方言チャート」⁶⁾というサービスを開発・提供しており、新聞⁷⁾やインターネットメディア⁸⁾で紹介されるなど、インターネットを通じて方言とその地域的差異についての関心が高まりつつある。

2. 既存の調査手法の課題

納豆や餃子などの食物や料理、習慣や方言など、文化に関する地域的差異は、どのような手法によって学術的に把握されてきたのであろうか。文化に関する地域的差異は、文

* 東京大学空間情報科学研究センター

キーワード : SNS、Twitter、位置情報、新方言、家計調査

Key words : SNS, Twitter, Geotag, New Dialect, Survey of Household Economy

化地理学や言語地理学、民俗学、文化人類学などの研究対象となっており、多くの研究においては、フィールドワークを通じた観察や聞き取り調査、アンケート調査、資料調査などによって、文化に関する地域的差異が把握されてきた⁹⁾。このような調査手法は、文化の詳細な実態や形成要因を明らかにすることができる一方で、必要となる労力は膨大であり、多くの場合、分析対象にできる空間的範囲は限定されてしまう。

一方、食物や料理、物品などの消費量については、多くの場合、総務省統計局が公表している家計調査などの統計資料から把握されている。例えば山下は、1963年と1990年の家計調査結果を利用して、多変量解析によって日本の食文化の地域性とその変容を検討している¹⁰⁾。家計調査結果を利用すれば、1963年以降の日本全国の都道府県庁所在都市について時系列的に分析できる。しかし、家計調査結果として品目別の年間支出金額が表章されている地域は、都道府県庁所在都市と政令指定都市のみであり、文化に関する地域的差異を分析するには空間単位が粗く、分析対象から漏れる地域が多くなる。例えば前述の浜松市は、2007年4月1日に政令指定都市に移行し、2008年の家計調査結果から年間支出金額が表章されるようになったため、それ以前については家計調査結果から判断することはできなかった。もし、今後、品目別の年間支出金額が表章される地域が拡大していくことになれば、餃子への年間支出金額が最も多い地域は、浜松市や宇都宮市のような現時点で上位に位置する地域とは全く異なる別の地域になる可能性もある。

このように、研究者自身による聞き取り調査やアンケート調査では空間単位を詳細に

することができるものの、分析対象の空間的範囲は限定的になる。また、既存の統計資料の利用によって、分析対象の空間的・時間的な拡張は可能であるが、空間単位は粗くなり、主要都市以外の地域に局所的に存在する文化については把握できない。もちろん、研究者の欲求をすべて満たしうる手法や資料は存在しないと考えられるものの、双方の中間的な特徴をもつような、まったく別の調査手法ないしは資料は存在しないのだろうか。

3. ビッグデータの利用可能性と研究目的

文化に関する地域的差異についての調査上の課題を解決しうるものとして、近年注目されているビッグデータの1つである、SNS（ソーシャル・ネットワーキング・サービス）のログデータがある。SNSの利用者は、SNS上で、文章や写真、動画などを利用して、様々な人々との会話や情報共有を図ることができる。SNSのログデータはこのような交流に用いられた文章などが蓄積された総体としてのデータであり、多くの場合、位置情報を付与することができるため、SNSログデータを用いた空間分析も可能である。

社会を構成する人間によって習得・共有・伝達されてきた行動・生活様式である文化は¹¹⁾、SNSの利用者間の交流や情報共有に用いられたログデータにも含まれるはずである。例えば、海外においては、ParkらがSNSの1つであるTwitterの投稿データを利用し、使用される顔文字の種類に関する異文化間の比較を行っており、個人主義的な文化の人々は口が主体の顔文字を、集団主義的な文化の人々は目が主体の顔文字をそれぞれ好む傾向にあることを示した¹²⁾。また、様々なSNSログデータについての空間分析を行い、インターネットを通じて情報発信を行っ

ている研究グループ「Floatingsheep」¹³⁾を率いる Zook and Poorthuis は、ビールや関連する用語について言及している Twitter の投稿データを抽出し、“ビールに関する空間”の可視化を試みている¹⁴⁾。日本ではまだ、文化に関する研究事例はそれほど多くないが、高は、日本語の Twitter ユーザによる中国人や韓国人についての言説の計量的分析を行っている¹⁵⁾。顔文字やビールに関する海外の既往研究と同様に、SNS ログデータからは日本に存在する様々な方言の使用状況を把握できると考えられ、昼食で何を食べたのかのような情報からも、よく消費される食物や料理など日本の食に関する文化も把握できると考えられる。分析対象にできる時間的スパンは家計調査などと比べて短いものの、利用者が全国に分布するような SNS であれば、空間単位の粗さや分析対象の空間的範囲の大きさを、聞き取り・アンケート調査や家計調査結果のように制限されることは少なく、大都市とその中小都市、村落との比較や、全国の市区町村単位の分析なども可能になる。

そこで本稿では、文化に関する地域的差異を把握するための資料としての、位置情報が付与された SNS ログデータの利用可能性を検討する。検討の対象とする SNS は Twitter であり、2012 年 2 月から 2014 年 10 月までに、日本国内に所在する携帯電話とタブレット PC から投稿され、一般に公開されていた約 1 億 8 千万件の SNS ログデータを利用する。

文化の地域的差異の分析対象としては、方言と食文化を取り上げる。方言については、まず「ダカラ」と「ヤカラ」を SNS ログデータから抽出する。「ダカラ」は、断定の助動詞である「ダ」と順接の助詞「カラ」からなるか、それ自体で接続詞となる語であり、「ヤ

カラ」は、断定の助動詞「ダ」に対する方言形式としての「ヤ」と順接の助詞「カラ」からなる語である¹⁶⁾。抽出された SNS ログデータの分布と、国立国語研究所編『方言文法全国地図』¹⁷⁾に収録されている言語地図とを比較し、SNS ログデータの妥当性を検討する。続いて、SNS ログデータから把握できる、新方言などについての地域的差異を示す。食文化については、麺類に注目し、家計調査結果との比較を考慮して「ラーメン」についての分布を検討する。これらの分析結果を通じて、文化の地域的差異を把握するための資料としての SNS ログデータの利用可能性を議論する。

II. ビッグデータを利用した方言の分析

1. 分析の対象

SNS ログデータには、SNS 利用者間の会話や独り言の文章が含まれている。これらの文章は、必ずしもすべてが共通語で記述されているとは限らず、方言で記述されているものもあるはずである。

方言の地域的差異の分析事例として、「東西対立」が顕著である方言形式に注目する。東西対立とは、日本の東西において語法が顕著に異なることを示す日本語学の用語である。断定の助動詞「ダ」は東西対立が顕著である語の代表例であり、東日本では「ダ」、西日本では「ジャ」・「ヤ」が用いられることが知られる¹⁸⁾。そこで、SNS ログデータから助動詞「ダ」および西日本で用いられる方言形式「ヤ」の使用例を抽出する。ただし、SNS ログデータ中の文章は品詞分解されていないため、助動詞としての「ダ」や「ヤ」のみを抽出することができない。そこで、順接の

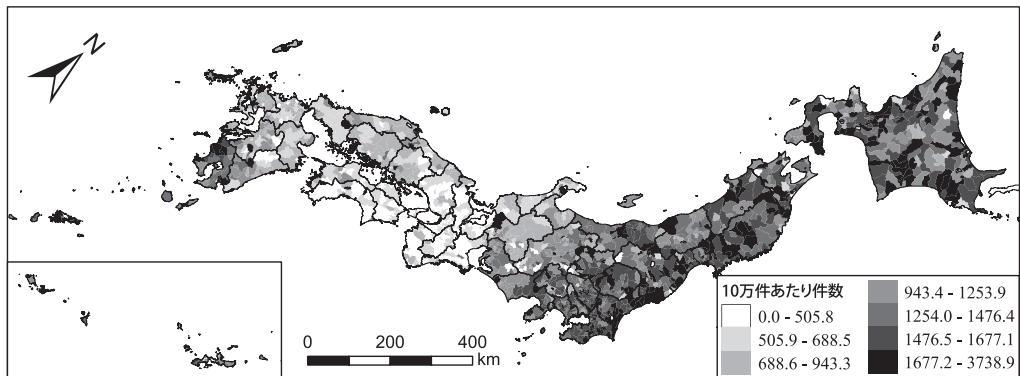
助詞「カラ」と組み合わせた「ダカラ」と「ヤカラ」を抽出することにする。これによって接続詞の「ダカラ」も同時に検出されることになるものの、『方言文法全国地図』によれば、接続詞「ダカラ」と断定の助動詞「ダ」および順接の助詞「カラ」を使用する地域はおおむね一致しており¹⁹⁾、分析上の誤謬は少ないものと思われる。

2. SNS ログデータからみた「ダカラ」・「ヤカラ」の地理的分布とその妥当性

日本国内の位置情報が付与された SNS ログデータのうち、「ダカラ」を含むものは 202 万 1,219 件、「ヤカラ」を含むものは 44

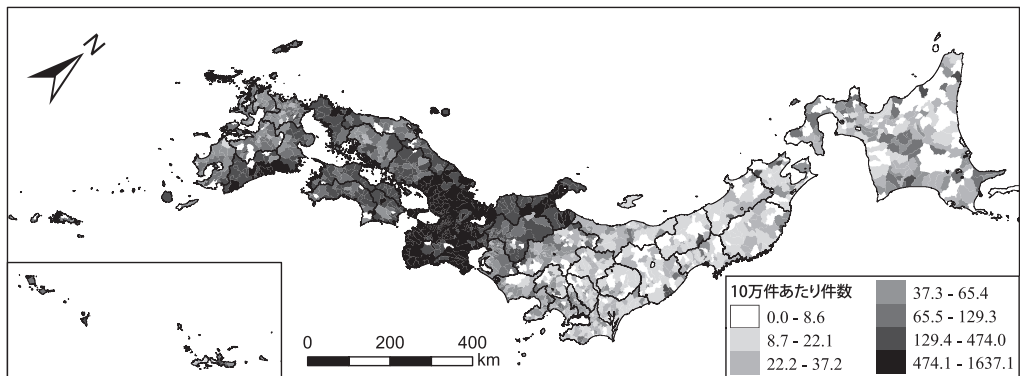
万 5,974 件であった。SNS ログデータの件数は通常、大都市ほど大きくなるため、何らかの母数となる値で基準化する必要がある。そこで、2010 年 10 月 1 日現在の市区町村単位で、位置情報が付与されたすべての SNS ログデータ 10 万件あたりの「ダカラ」・「ヤカラ」を含む件数を求めた（第 1 図・第 2 図）²⁰⁾。

SNS ログデータ 10 万件あたりの「ダカラ」の件数が多いのは、東日本と鹿児島県、沖縄県であり、「ヤカラ」については西日本で件数が多くなっている。明治期の国語調査委員会が調査し、編纂した『口語法分布図』に基づいて徳川が示した、中部地方における断定の



第 1 図 「ダカラ」を含む SNS ログデータの地理的分布

出典：Twitter 投稿データ



第 2 図 「ヤカラ」を含む SNS ログデータの地理的分布

出典：Twitter 投稿データ

助動詞「ダ」と「ジャ」・「ヤ」の東西の境界線は、おおむね愛知県、静岡県、長野県、新潟県と、富山県、岐阜県、三重県との境界に引かれている²¹⁾。SNS ログデータにおいては、前者を含む東日本で「ダカラ」が多く、後者を含む西日本で「ヤカラ」が多くなっており、明治期に存在した東西対立の境界線とおおむね対応していると判断できる。この東西対立の境界線は、1979年から1982年に調査が行われた『方言文法全国地図』でも大きくは変移しておらず²²⁾、SNS ログデータから、2010年代までその状況が続いていることが確認できる。

一方で、徳川は、西日本にも「ダ」を使用する地域が存在することを指摘しており、具体的には山陰地方と熊本県としている²³⁾。しかし、SNS ログデータからは、これらの地域において「ダカラ」がそれほど多く分布しているようには見えない。また、『方言文法全国地図』などとの相違点として、鹿児島県や沖縄県に「だから」が多く分布していることが指摘できる。このうち鹿児島県に関しては、『方言文法全国地図』では指摘されていない、鹿児島方言である「ダカラヨ」の存在²⁴⁾が、「ダカラ」のSNS ログデータの件数を押し上げているものと予想される。また、沖縄県においては、相槌としての「ソウ」の意味での「ダカラ」の使用が確認されており²⁵⁾、SNS ログデータにおいて「ダカラ」を含む件数が多いことについての一定の説明になりうる。

このように、既往研究・調査の結果とSNS ログデータに基づく分布図を比較すると、若干の差異が認められた。ただし、『方言文法全国地図』に関する調査は、比較的古い言語を採集することを目指しており²⁶⁾、方言の

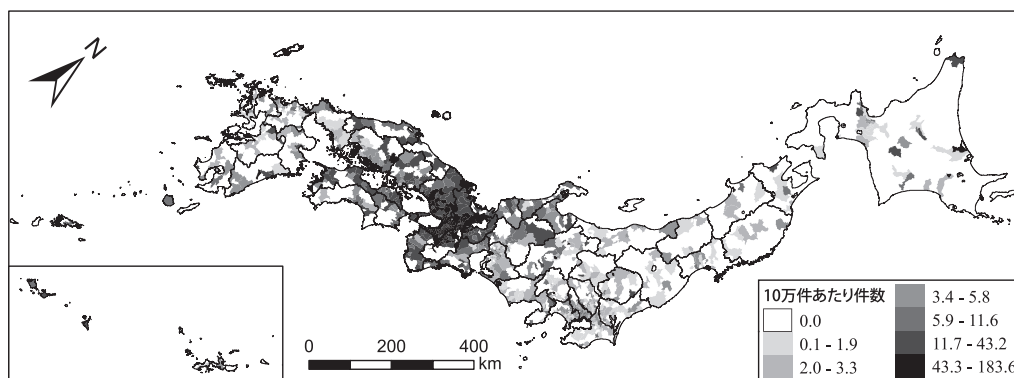
リアルタイムな状況についての情報を含んだSNS ログデータとは、資料的な性格が異なっている。そのため、「ダカラ」に関して確認された、既往研究・調査の結果とSNS ログデータとの相違は、方言の分析資料としてのSNS ログデータの妥当性を否定するものではなく、2010年代までの方言の変化の結果として生じているものと考えられる。鹿児島県や沖縄県での「ダカラ」を含むSNS ログデータの件数の多さは、新方言などがSNS ログデータに含まれていることを示唆しており、『方言文法全国地図』以降に発生した「新方言」などの新しい方言の分析のための資料としての妥当性を認めることができる。

3. SNS ログデータにみる新方言

前節で検討したように、SNS ログデータは「新方言」のような、現代においてよく使用される方言を分析することに適している。そこで、代表的な新方言について、若干の事例を地図化する。

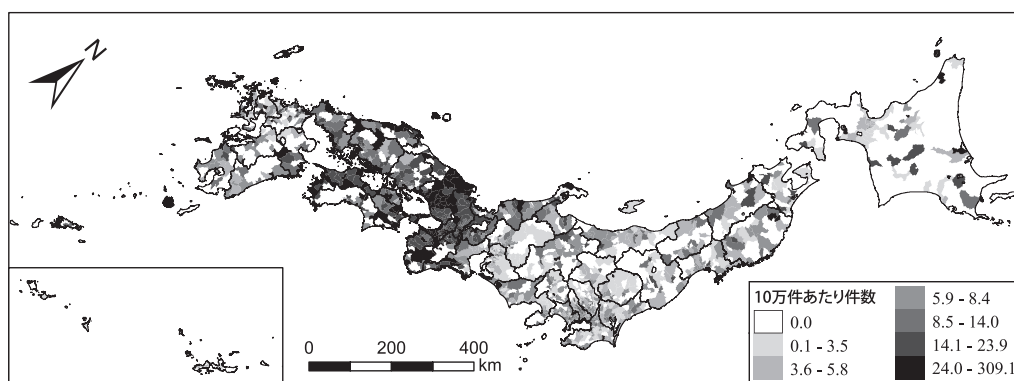
外食産業大手の日本マクドナルドが運営する「マクドナルド」は、日本全国にその店舗網を広げており、2014年末現在の店舗数は3,164である²⁷⁾。井上によれば、顧客側からのマクドナルドの略称は、全国的には「マック」が多く、近畿地方では「マクド」が多いとされる²⁸⁾。SNS ログデータから、「マクド」を含むデータを抽出すると、第3図のようになる²⁹⁾。近畿地方、特に京阪神大都市圏においてSNS ログデータの件数が多い。近畿地方全体ではなく、京阪神大都市圏により集中した分布を示しているのは、マクドナルドの店舗が大都市圏に集中しているためと考えられる。

もう1つの事例として、理容所の呼称についても検討する。日本全国で約7万人が加入



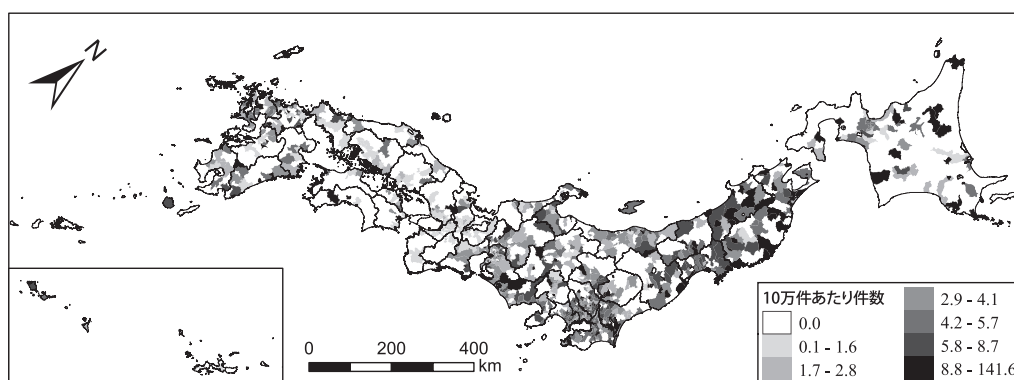
第3図 「マクド」を含む SNS ログデータの地理的分布

出典：Twitter 投稿データ



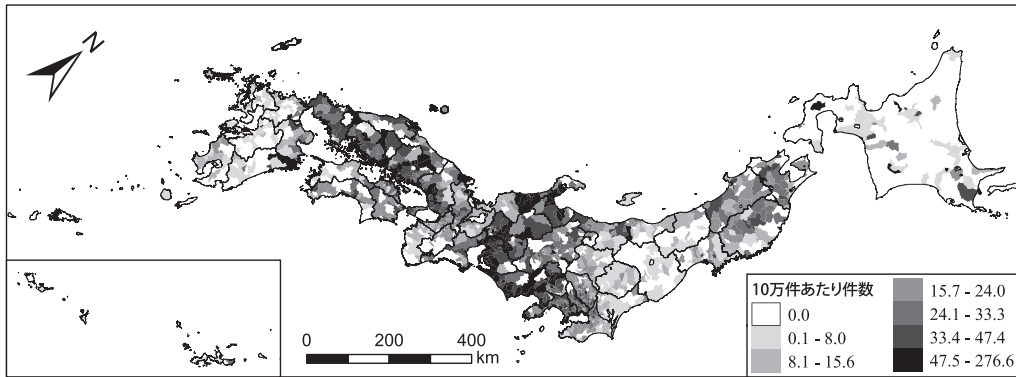
第4図 「散髪」を含む SNS ログデータの地理的分布

出典：Twitter 投稿データ



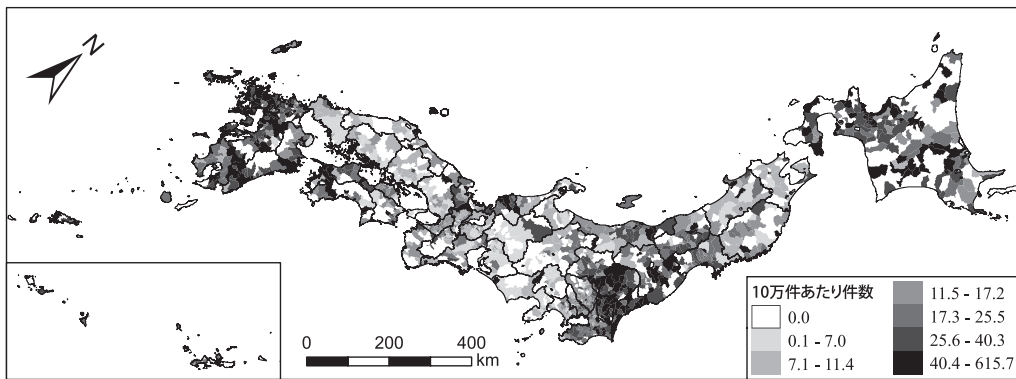
第5図 「床屋」を含む SNS ログデータの地理的分布

出典：Twitter 投稿データ



第6図 「美容院」を含む SNS ログデータの地理的分布

出典：Twitter 投稿データ



第7図 「美容室」を含む SNS ログデータの地理的分布

出典：Twitter 投稿データ

する全国理容生活衛生同業組合連合会（全理連）が調査した結果によれば、理容所を「散髪屋」または「散髪」と呼ぶ例が多いのは近畿地方、中国地方、四国地方と、大分県、沖縄県であり、「床屋」と呼ぶ場合が多いのは東日本を中心として、佐賀県や長崎県、熊本県にも広がっている³⁰⁾。そこで、SNS ログデータにおいて、「散髪」と「床屋」、関連して「美容室」と「美容院」を含むデータ件数を求めた。「散髪」を含む SNS ログデータの件数は、近畿、中国、四国地方と、大分県の南部で若干多くなっており、おおむね全理連の調査結果と一致している（第4図）。一方で、「床屋」

に関しては、関東地方および東北地方で多いと考えられるものの、佐賀県や長崎県、熊本県ではそれほど多くはなく、全理連の調査結果との整合性は「散髪」ほど高くはないものと考えられる（第5図）。次に、「美容院」については、関東地方のうち東京大都市圏を中心とする南部と、福井県を除く中部地方、兵庫県から山口県にかけての地域で多くなっている（第6図）。最後に、「美容室」については、残る関東地方や北海道、九州地方などで多くなっている（第7図）。このように、理容所や美容所の呼称についても関する一定の地理的パターンを認めることができる。

以上のことから、SNS ログデータを方言分析のための資料として用いることで、新方言をはじめとした、リアルタイムに使用されている方言の地域的差異を検討できるものと考えられる。

III. ビッグデータを利用した食文化の分析

1. 分析の対象

SNS ログデータには、方言だけでなく、食物や料理の名称についての情報も含まれるはずである。例えば、昼食や夕食、好きな料理などについての会話では、料理などの名称を含んだ投稿が行われる。SNS ログデータを利用することで、食文化の地域的差異を検討することができるものと思われる。

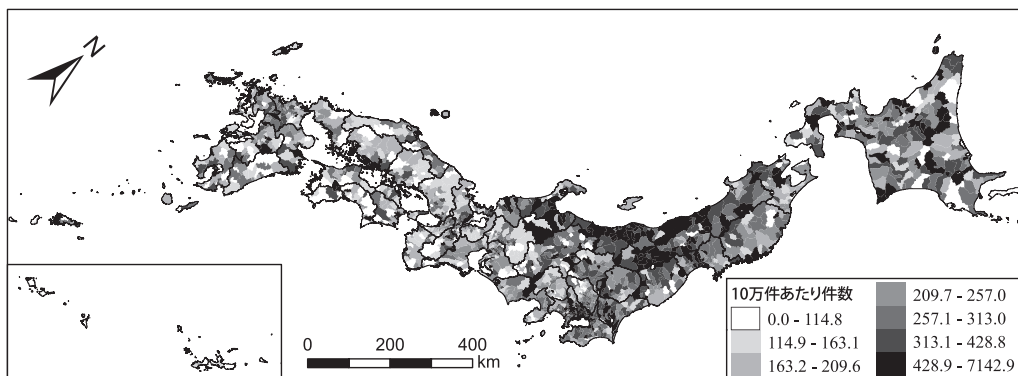
分析の対象として取り上げるのは麺類の名称である。麺類のうち、「うどん」や「そば」は全国的に消費される代表的なものと考えられ、特に香川県は「うどん」を主軸においた観光キャンペーンを行っていることはよく知られている³¹⁾。しかし、家計調査の収支項目分類では、食材としての「生うどん・そば」および「乾うどん・そば」と、外食の食事代のうちの「日本そば・うどん」に分類

されており、「そば」と統合されている³²⁾。そのため、「うどん」と「そば」は、SNS ログデータと同時期の家計調査の結果を比較するための分析対象には不適切である。そこで、家計調査結果との比較を考慮して、「ラーメン（中華そば・中華麺）」に注目する。家計調査の収支項目分類におけるラーメンは、食材については「中華麺」として、外食の食事代については「中華そば」として単独で表章されており、SNS ログデータとの比較が容易である。

2. SNS ログデータからみた「ラーメン」の地理的分布とその妥当性

まず、方言の場合と同様に、「ラーメン」³³⁾が含まれる SNS ログデータを抽出し、市区町村別に 10 万件あたりの件数を求めた（第 8 図）。「ラーメン」を含む SNS ログデータの件数は、東日本、特に日本海側において多い傾向にあり、東京大都市圏、福岡県などでも多くなっている。

この分布と、総務省統計局が実施した家計調査結果に基づく、2014 年の都道府県庁所在都市別の 1 世帯あたり年間支出金額³⁴⁾と比較する（第 1 表）。食材である「中華麺」への支出が最も多いのは青森市であり、宇都



第 8 図 「ラーメン」を含む SNS ログデータの地理的分布

出典：Twitter 投稿データ

第1表 家計調査による中華麺・中華そばの1世帯あたり年間支出金額（2014年）

地域	中華麺		中華そば	
	支出金額 (単位：円)	全国を100とする 指数	支出金額 (単位：円)	全国を100とする 指数
全国	3,089	100.0	5,929	100.0
札幌市	2,653	85.9	5,549	93.6
青森市	4,004	129.6	6,662	112.4
盛岡市	3,420	110.7	8,515	143.6
仙台市	2,633	85.2	15,194	256.3
秋田市	3,409	110.4	7,886	133.0
山形市	3,270	105.9	11,910	200.9
福島市	3,010	97.4	9,003	151.8
水戸市	2,916	94.4	9,704	163.7
宇都宮市	3,746	121.3	10,616	179.1
前橋市	3,233	104.7	5,485	92.5
さいたま市	3,044	98.5	5,235	88.3
千葉市	2,403	77.8	8,570	144.5
東京都区部	2,942	95.2	7,414	125.0
横浜市	3,369	109.1	4,664	78.7
新潟市	2,587	83.7	8,962	151.2
富山市	3,258	105.5	7,729	130.4
金沢市	3,093	100.1	6,272	105.8
福井市	3,149	101.9	6,168	104.0
甲府市	3,220	104.2	7,050	118.9
長野市	3,342	108.2	7,368	124.3
岐阜市	3,193	103.4	5,457	92.0
静岡市	3,369	109.1	4,031	68.0
名古屋市	3,474	112.5	8,227	138.8
津市	3,150	102.0	2,884	48.6
大津市	3,314	107.3	5,204	87.8
京都市	3,187	103.2	4,805	81.0
大阪市	2,837	91.8	4,812	81.2
神戸市	2,590	83.8	5,622	94.8
奈良市	2,484	80.4	3,643	61.4
和歌山市	2,614	84.6	3,153	53.2
鳥取市	2,455	79.5	3,265	55.1
松江市	2,904	94.0	4,179	70.5
岡山市	3,054	98.9	6,687	112.8
広島市	3,031	98.1	4,745	80.0
山口市	1,843	59.7	3,040	51.3
徳島市	2,131	69.0	3,782	63.8
高松市	2,776	89.9	5,000	84.3
松山市	3,054	98.9	3,151	53.1
高知市	2,511	81.3	6,357	107.2
福岡市	2,681	86.8	7,885	133.0
佐賀市	1,677	54.3	5,353	90.3
長崎市	2,581	83.6	3,651	61.6
熊本市	1,805	58.4	4,619	77.9
大分市	2,296	74.3	5,225	88.1
宮崎市	1,899	61.5	4,045	68.2
鹿児島市	2,855	92.4	4,845	81.7
那覇市	3,207	103.8	3,474	58.6

出典：総務省統計局『家計調査（家計収支編）』総世帯に関する詳細結果表（2014年の年次数値）
<http://www.e-stat.go.jp/SG1/estat/List.do?lid=000001129456>

宮市、名古屋市、盛岡市、秋田市と続いている。SNS ログデータの件数が多い山形県や福島県、新潟県、富山県の各県庁所在都市での「中華麺」の年間支出金額は、全国で3,089円であることを考えれば、それほど支出金額が大きいとはいえない。一方、外食の食事代としての「中華そば」の年間支出金額は、仙台市が最も多く、15,194円となっている。全国の年間支出金額は5,929円であり、東北や北陸の各県庁所在都市や、千葉市、東京都区部、名古屋市、福岡市などで多くなっている。SNS ログデータとの対応関係については、「中華麺」よりは「中華そば」の年間支出金額のほうが明瞭と考えられる。

「ラーメン」を含む SNS ログデータの地理的分布は、少なくとも外食の食事代である「中華そば」の年間支出金額の分布とは、一定程度対応していることが示された。食材である「中華麺」との対応関係は明瞭ではないものの、SNS ログデータは、食物や料理、食文化についての地域的差異を分析できる資料であるといえる。ただし、「中華麺」よりも「中華そば」との対応関係が相対的に強いことを考慮すれば、SNS ログデータには、食材の名称よりも外食の際の料理名が含ま

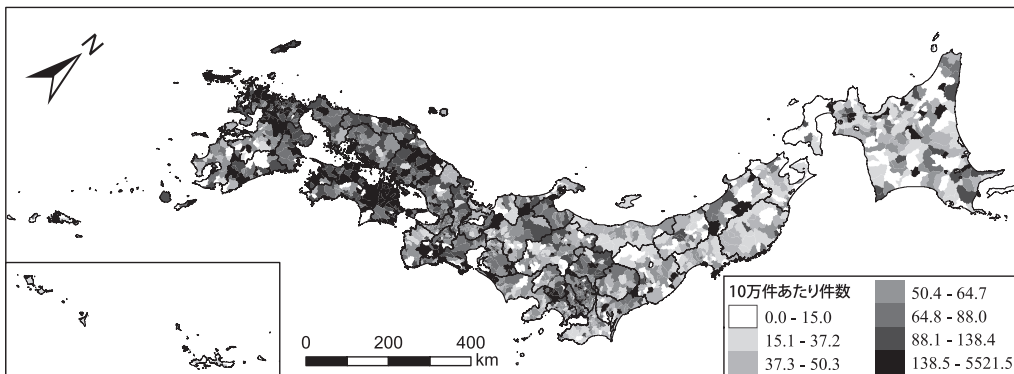
れていることが多いものと思われる。

3. SNS ログデータからみた「うどん」の分布の時空間的特性

ここまで、食文化の地域的差異についての分析資料としての SNS ログデータの妥当性が一定程度示された。そこで、家計調査では個別に把握することのできない、「うどん」を事例として、「うどん」を含む SNS ログデータの分布と、その時空間的特性について若干の検討を行う。

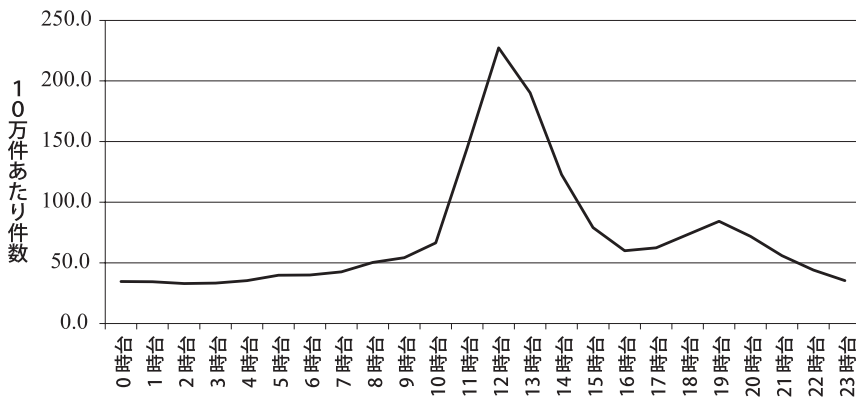
これまでと同様の手順で「うどん」を含む SNS ログデータを抽出し、その地理的分布を地図に示した（第9図）³⁵⁾。「うどん」を含む SNS ログデータの件数の多い地域は、主に西日本であり、特に、香川県や岡山県、愛媛県、福岡県などで多くなっている。また、関東地方の西半分や秋田県などでも多く、それぞれの地域の食文化としての様々なうどん食の様式が存在が示唆される。

さて、SNS ログデータには、そのデータが投稿された日時の情報が含まれている。例えば、「うどん」を含む SNS ログデータを、1時間ごとの時間帯別に集計し、1時間ごとの日本全国すべての SNS ログデータ 10 万件に対する「うどん」を含む件数を求めると第



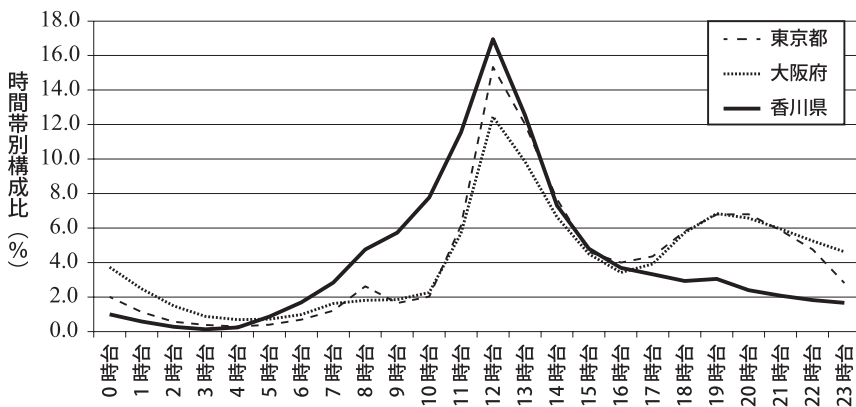
第9図 「うどん」を含む SNS ログデータの地理的分布

出典：Twitter 投稿データ



第10図 「うどん」を含むSNS ログデータの時間帯別件数

出典：Twitter 投稿データ



第11図 「うどん」を含むSNS ログデータの時間帯別構成比 (東京都・大阪府・香川県)

出典：Twitter 投稿データ

10図のようになる。このグラフから明らかのように、「うどん」を含むSNS ログデータの件数には、12時台および13時台という正午前後の時間帯にピークが存在し、夜間の19時台にも第2のピークも認められる。つまり、昼食の時間帯に「うどん」についての言及が多くなり、また、夕食の時間帯にも一定程度「うどん」への言及が行われている。このような時間による「うどん」への言及の差は、地域的な差異を伴うものなのであろうか。この点の検証のために、「うどん」を含むSNS ログデータを、都道府県別に、1時間ごとの

時間帯別に集計し、都道府県ごとに時間帯別の構成比を求めた。「うどん」を含むSNS ログデータの件数が多い、東京都と大阪府、香川県、それぞれの時間帯別構成比を比較したものが第11図である。東京都と大阪府は、全国と同様に、昼食の時間帯に最大のピークを、夕食の時間帯に第2のピークをそれぞれもつパターンを示している。一方で、香川県については、ピークは昼食時間帯のみであり、夕食時間帯の構成比はそれほど高くはない。また、8時台や9時台など、朝食時間帯の構成比が東京都や大阪府と比較して高く、

香川県における「うどん」への言及の時空間的特性は、朝食・昼食時間帯に多くなる点にあるといえる。

このように、SNS ログデータから、食文化についての地域差を分析することができ、かつ、SNS ログデータがもつ日時の情報をもとに、その時空間的特性についても検討できる。

IV. ビッグデータを活用した地域文化研究の可能性と課題

ここでは、SNS ログデータの利用上の課題と、利用の方向性に注目しながら、SNS ログデータの地域文化研究への利用可能性について若干の検討を加える。まず、これまでの方言に関する分析において確認されたように、SNS のようなデジタル機器や文字を通した他者との交流のためのメディアにおいても、伝統的に存在する東西対立を伴いながら方言が使用されていることが示された。また、新方言のような現代においてよく使用されるようになった方言も検出することができた。

しかし、SNS ログデータは、生の音声ではなく、デジタル機器を通して文字として記録されるため、SNS ログデータから把握できる方言には、主として入力のためのデジタル機器のバイアスがかかることになる。例えば、漢字とともにひらがなで方言を入力しようとした場合、誤って方言が漢字に変換されてしまうこともある。このように、方言を意図した形式で入力するためには、一定の注意を払う必要があり、そのような配慮が SNS において方言を利用する際の障壁にもなりうる。また、方言は、相手によって使い分け

られ、親しい関係においては方言を利用し、そうではない関係においては、共通語を利用することもある。SNS においても同様の使い分けが考えられ、どのような相手との会話で用いられているのかを把握する必要がある。

食文化についての分析では、家計調査から把握した特定の食物や料理への年間支出金額の地域性に対応する形で、SNS ログデータにおいても、当該の食物や料理への言及が行われていることが示された。また、SNS ログデータに含まれる投稿日時の情報を利用することで、食物などへの言及の時空間的特性についても検討でき、SNS ログデータから、食文化に関する様々な空間分析が展開できるものと思われる。ただし、SNS ログデータに含まれるのは、食物や料理の名称のみである点に注意が必要である。例えば、うどんのだしの味の地域的差異は、ほとんどの場合、SNS ログデータには表れないものと予想され³⁶⁾、料理を構成する食材についての地域的差異の分析には、SNS ログデータはあまり適していないものと思われる。

本稿では方言と食文化を事例として取り上げたものの、これら以外の様々な文化的現象に関する空間分析の展開も可能と考えられる。例えば、SNS ログデータの件数は非常に膨大であり、これまで十分に分析されてこなかった、限られた地域に存在する文化に対する言及も SNS ログデータから検出することができるかもしれない。また、新たに生まれた方言やテレビなどに紹介されることで知られるようになった名物料理や、いわゆる「ゆるキャラ」のようなものについてのリアルタイムな分析も可能である。SNS ログデータに含まれる会話の相手に関する情報から、リアルタイムに生じた現象に関する分析を

より詳細に検討すれば、SNS 上での会話やテレビなどのメディアを通じた現象の空間的な伝播に関する研究の展開も期待できる。加えて、方言がより親密な相手との会話に利用されることを考慮すれば、方言の使用の有無を、相手との親密度の指標の1つとして用いることで、伝播におけるその差異を検討できよう。最後に、SNS ログデータ、特にTwitterの投稿データに特有の課題について指摘する。Twitterのユーザ登録の段階では、メールアドレス以外の利用者の個人情報を入力する必要がない。そのため、利用者自身が公表している場合を除いて、分析者を含めた他者が利用者の属性を把握することは難しい。このような問題に対して、実名で利用している利用者の氏名の情報に基づいて性別、年齢、民族を推定することで、Twitterの投稿データに見られる時空間的な人口特性を明らかにした事例もある³⁷⁾。日本においても同様の手法を援用できるかどうかは不明であるものの、何らかの方法によって、限られた情報から利用者の属性を推定することができれば、文化に関する分析資料としてのTwitterの投稿データの価値を高めることができる。ただし、そのような推定を高精度に行うことで、利用者の居住地などの情報が特定される可能性もあり、個人情報の保護や取扱いには細心の注意を払う必要がある。

V. おわりに

本稿では、ビッグデータの1つであるSNS ログデータに注目し、文化に関する地域的差異を把握するための資料としての、位置情報が付与されたSNS ログデータの利用可能性を検討してきた。文化の地域的差異を把握す

るため事例として、方言と食文化に注目し、既往研究の成果や既存の統計調査の結果と、SNS ログデータから得られた地理的分布とを比較し、それらの分布について、一定の妥当性が認められることを示してきた。

SNS ログデータに含まれる文章の情報は、入力のためのデジタル機器による技術的なバイアスがかかった情報であり、その読み方や語句の意味までを把握できるものではない。また、Twitterの投稿データの場合、ユーザ自身の情報については分析者が入手することができないという問題もある。これらの課題が残されているとはいえ、これまでの方言および食文化に関する分析事例から、ビッグデータの1つであるSNS ログデータは、文化に関する地域的差異を分析する資料として大きな価値を持っていると考えられ、様々な空間分析の展開可能性が期待できる。

数千万件、数億件のような規模をもつビッグデータの1つであるSNS ログデータは、その件数の膨大さゆえにハンドリングが非常に難しいものの、文化だけでなく社会における様々な現象についての地理学的分析の展開可能性を広げるものでもある。地理学において、地域に関する情報を含んだ、様々なビッグデータの活用が進み、地理学自体の更なる発展と、社会における地理学のプレゼンスがより向上することを期待する。

〔付記〕本稿の骨子は、2014年度第26回立命館地理学会大会（2014年12月6日）において発表した。また、本稿の作成にあたっては、JSPS 科研費 15K12886 の助成を受けた。

注

- 1) 安田亘宏「日本のフードツーリズムの変遷についての考察」、日本国際観光学会論文集 19、

- 2012、103-109 頁。
- 2) 宇都宮市経済部観光交流課『平成 25 年宇都宮市観光動態調査報告書』http://www.city.utsunomiya.tochigi.jp/dbps_data/_material/_files/000/000/017/794/H25dotai.pdf 2015 年 5 月 17 日参照。
- 3) 朝日新聞デジタル 2015 年 1 月 31 日付 <http://www.asahi.com/articles/ASH1Z451HH1ZUTPB00K.html> 2015 年 5 月 17 日参照。
- 4) 松本 修『全国アホ・バカ分布考—はるかなる言葉の旅路』、太田出版、1993、446 頁。
- 5) <http://japanknowledge.com/personal/> 2015 年 7 月 13 日参照。運営会社は、株式会社小学館が出資する株式会社ネットアドバンスである。
- 6) <http://ssl.japanknowledge.jp/hougen/hougen100/index.php> 2015 年 7 月 13 日参照。現在は「出身地鑑定!! 方言チャート 100」として改良版が公開されている。
- 7) 例えば、日本経済新聞 電子版 2013 年 8 月 31 日付 http://www.nikkei.com/article/DGXNASDG2703Q_Q3A830C1CR0000/ 2015 年 7 月 13 日参照。
- 8) 例えば、週刊アスキー 2015 年 7 月 12 日付 <http://weekly.ascii.jp/eleme/000/000/355/355339/> 2015 年 7 月 13 日参照。
- 9) 例えば、藤原与一『方言研究法』、東京堂出版、1964、28-191 頁。など。
- 10) 山下宗利「わが国における食文化の地域性とその変容」、研究論文集 39-2、1992、115-133 頁。
- 11) 松村 明・三省堂編修所編『大辞林 第三版』、2008（電子版）。
- 12) Park, J., Baek, Y. M. and Cha, M.: Cross-Cultural Comparison of Nonverbal Cues in Emoticons on Twitter: Evidence from Big Data Analysis, *Journal of Communication* 64, 2014, pp. 333-354.
- 13) <http://www.floatingsheep.org/> 2015 年 7 月 11 日参照。
- 14) Zook, M. and Poorthuis, A.: Offline Brews and Online Views: Exploring the Geography of Beer Tweets, in Patterson, M. and Pullen, N. H. eds.: *The Geography of Beer, Regions, Environment, and Societies*, Springer, 2014, pp. 201-209.
- 15) 高 史明「日本語 Twitter ユーザーの中国人についての言説の計量的分析—コリアンについての言説との比較—」、人文学研究所報 53、2015、73-86 頁。
- 16) 接続詞としての「ダカラ」に対応するのは「ソヤカラ」である。ただし、SNS ログデータから「ヤカラ」を含むものを抽出することで、「ソヤカラ」も同時に抽出されることになる。
- 17) 国立国語研究所編『方言文法全国地図』は、1979 年から 1982 年にかけて調査が実施され、1989 年から 2006 年にかけて、全 6 集が刊行された。特筆しない限り、『方言文法全国地図』とした場合は、全 6 集全体を指す。
- 18) ①徳川宗賢『言葉・西と東（日本語の世界 8）』、中央公論社、1981、345-359 頁。②牛山初男「語法上より見たる東西方言の境界線について」、国語学 12、1953、59-63 頁。
- 19) ①国立国語研究所編『方言文法全国地図 第 1 集 助詞編』、大蔵省印刷局、1989。②国立国語研究所編『方言文法全国地図 第 3 集 活用編 2』、大蔵省印刷局、1993。
- 20) 分析対象とする語を明示するために、本文中ではカタカナによる表記を行っているが、実際に SNS ログデータから抽出したのは、ひらがなの「だから」または「やから」を含むデータである。
- 21) 前掲 18) ① 347 頁。
- 22) 前掲 19) ②。
- 23) 前掲 18) ① 348 頁。
- 24) 二階堂整「福岡の大学生の携帯メールにおける方言使用」、山口国文 32、2009、17-26 頁。
- 25) 井上史雄「新方言辞典稿・インターネット版（1996 年版）」<http://triaez.kaisei.org/~yari/Newdialect/> 2015 年 5 月 17 日参照。
- 26) 国立国語研究所編『方言文法全国地図解説 1』、大蔵省印刷局、1989、20 頁。
- 27) 日本マクドナルド株式会社『マクドナルド CSR レポート 2014』http://www.mcdonalds.co.jp/company/csr/pdf/csr_report2014.pdf 2015 年 5 月 17 日参照。
- 28) 井上史雄「新方言辞典稿」、東京外国語大学論集 47、1993、185-212 頁。
- 29) SNS ログデータから、「マクド」にカタカナが続かない言葉を含むものを抽出した。そのため、「マクドナルド」を含むものはこの抽出結果に含まれていない。
- 30) 全国理容生活衛生同業組合連合会「史料室：こぼれ話篇 1」http://www.riyo.or.jp/library/etc_kobore_01.html 2015 年 5 月 17 日参照。ただし、調査の期日や規模、対象などの詳細は明示されていない。
- 31) 香川県の公式ウェブサイト上では、香川県の名称の横に「うどん県 それだけじゃない 香川県」というキャッチコピーが表示されている。<http://www.pref.kagawa.lg.jp/index.shtml> 2015 年 5 月 17 日参照。
- 32) 食材としての「うどん」についての単独の項目が設定されていたのは、1969 年の改訂までであり、食事代としての「そば」についての単独の項目「日本そば」が設定されていたのは 1995 年の改訂までである。
- 33) 「ラーメン」だけでなく、ひらがな表記の

- 「ラーメン」を含むものも抽出した。
- 34) <http://www.e-stat.go.jp/SG1/estat/List.do?lid=000001129456> 2015 年 7 月 11 日参照。
- 35) SNS ログデータから、「うどん」、「ウドン」、「鰻 鮎」のいずれかを含むものを抽出した。
- 36) だしの地域的差異に関する言及が行われることもあるが、「うどん」そのものへの言及などと比較すれば、相当少ないことは明らかである。
- 37) Longley, P. A., Adnan, M. and Lansley, G.: The geotemporal demographics of Twitter usage, *Environment and Planning A* 47, 2015, pp. 465–484.