

## 〈論文〉

## 日本語世界進出のグーグル言語地理学

—グーグルインサイトにもみる外行語総合分布—

井上 史 雄

キーワード：グーグルインサイト 言語地理学 外行語 外来語 借用語

## 1. グーグルインサイト世界言語地図

本稿では、グーグルインサイトによる新種の言語地図を作製し、総合的結果を考察する。以下に掲げるのは、これまで目にできなかった世界言語地図である。日本語の単語が海外で（世界で）どう使われているかを示すもので、色の濃い国では多く使われる。以下では「外行語」という用語を使う（三輪1977）。「外来語」の対として作られた用語で、日本語を中心にした借用語 loanword の両方向（borrowing, lending）を区別して扱うには便利である。

井上2001では世界約30カ国の世論調査の結果を地図にした。それまで日本語の単語がアメリカ（英語）に入り、アジア諸国や移民の多い南米で使われることについて、個別の語や国家の報告や調査はあったが、多数の国家についての同一条件での報告はなかった。多くの国で日本語の単語が使われていることを提示できた。ただし、初出年や意味分野などによって分類して傾向を見出すには不十分な語数だった。元のデータを集めるのに、大変な経費と年数がかかった（新プロ「日本語」総括班編1999）。

それに対し、このグーグルインサイト地図の作製は、簡単である。慣れれば、1分もかからない。以下に掲げた世界地図は、貴重な情報であり、新鮮な印象を与える。グーグルマップの利用法については井上（2011.2）で述べた。以下では、地図の出力の技法について説明する。また高度な技法として、数値データをダウンロードして、多数の地図を総合的に分析する手法についても論じる。

以下の4枚の地図は、かつて公表された世界地図と対比できる（井上2001, p.67）。

図1-1. kimonoは東南アジアと欧米とブラジルに目立つ。井上2001と一致する。

図1-2. sakuraは東南アジア付近と中南米に多い。井上2001では中国・韓国にも多いが、図1-2には反映されていない。後述のように、アルファベットつづりでGoogle検索をしたためである。

図1-3. shogunは東南アジアと欧米・ロシアとブラジルなどに目立つ。井上2001ではブラジルで出なかった。その後約10年の変化か。映画『シヨーゲン』またはゲーム・アニメなど登場による可



図 1-1. kimono



図 1-3. shogun

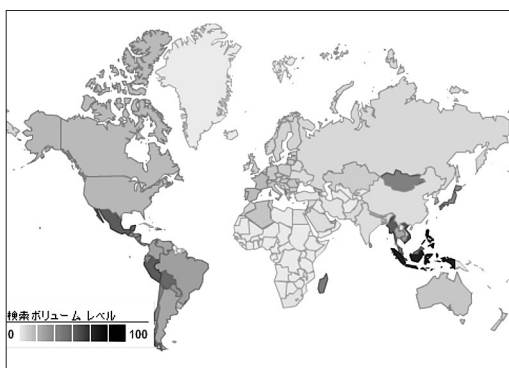


図 1-2. sakura

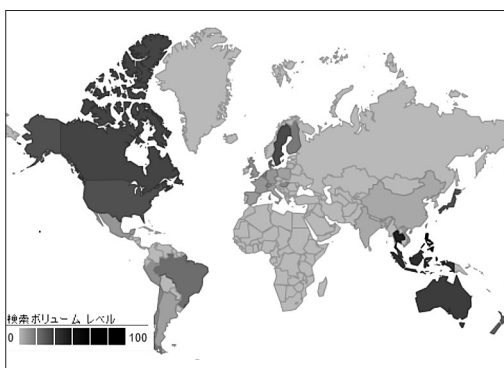


図 1-4. sukiyaki

能性がある。

図 1-4. sukiyaki は東南アジアと欧米とブラジルなどに目立つ。井上 2001 とほぼ一致する。

## 2. グーグルインサイト外行語の総合処理

### 2.1. 外行語総合表の処理

外行語については、原口 (1998)、早川 (2006) などを元に、多数の語についてグーグルマップを作っていた (井上 2011.2)。そのリストを元にグーグルインサイトの表・地図を作製し、エクセルの表をダウンロードした。多数のエクセルファイルを一つのファイルに統合し、扱ったすべての語を横に並べた大きな表を作った。縦に「anime」の使用率の順によって国家が並び、横に検索した数十語が並ぶ。同じ語を重複して検索した可能性もあるので、単語をアルファベット順にソートして重複語を削除した。また「anime」で登場しなかった (小) 国家が他の語で登場することもあるので、それらを手で配列した<sup>1</sup>。以下では 81 語、124 国家 (地域 regions) を扱う<sup>2</sup>。

まず行と列の数値の合計を出した。次に全体の傾向をみるための便利な (必須) 手段として、行と列の合計数値を基準にしてソートした。表 2-1a, b に全体の結果を上位語と下位語に 2 分割して示

す。見やすくするために、ゼロを空白に置き換えた。また数値100に色を付けた。行と列でソートしたので当然のことながら、左上には高い数値が並び、右下には空白（ゼロ）が並ぶ。左下は限られた国でよく使われる語が並び、右上は限られた語をよく使う国が並ぶ。アメリカやアジア諸国の数値が多いことが読み取れた。また「karate」がトップで、自動車やコンピューターゲーム関係の語が上位になることが分かった。

この傾向からはずれた分布パターンの違いが読み取れる。縦の列に並ぶ単語ごとの違いとして、下まで伸びる単語が散在する。左の表2-1aの上位語ではkaraoke, nissan, nintendo, samurai, toyota, zen, manga, pokemon, anime, tsunami, ninja, karaokeが全世界に散らばる。現代日本の技術や文化に関する語が多い。右の表2-1bの下位語ではnippon, geisha, sumo, teriが散在する。やや古い日本語が混じる。なお右端の語を見ると、100とゼロだけの単純な語がある。たった一つの国家だけで使われた語である。最下位のshinkanseï（ミスタイプ）とhinoshi（ハワイ日本語のアイロン）は、使用率がどの国でもゼロだったので、後の表、グラフでは削除した。

横に並ぶ国家ごとの違いとして、左まで伸びる国の違いを見ると、違いが目立たない。現代日本文化は、地理的距離や経済的交流と関わりなく、ほぼ全世界に広がったとみられる。

## 2.2. 外行語の順位

外行語と国家をもっと具体的に示すために、行と列の合計数値を他のシートに（必要に応じて行と列を入れ換え、値として）コピーして、折れ線グラフを作成した。その結果に従って、国家と語形で目立たせたいものに●を付加した。

図2-1に外行語の単語の順位を示す。上位10語ほどは比較的使用率が高い<sup>3</sup>。あとは連続的に低くなる。井上（予定）で調査語を増やした後の20位までは、以下の順番である。Sony, honda, toyota, pokemon, anime, nissan, manga, nintendo, suzuki, ninja, sushi, karaoke, soy, tsunami, zen, tycoon, karate, samurai, sakura, origami.

いくつかの基準で語形に色を塗って、合計数との関連を探った。20世紀末期に世界進出した日本語が上位になる。図の見出しに●を付けた。19世紀以前または江戸時代以前に海外で使われた語は、今はあまり使われない。これについては二つの仮説が成立しうる。まず早くから使われる語は、多くの国に普及するまでに十分な時間があるので全体の数値が大きくなり、最近の語はまだ普及しないので小さい（右下がり）。一方昔使われた語は縁遠くなって数値が下がり、最近の語は身近なので数値が上がる。またコミュニケーション手段の変化（インターネットなどの普及）により、最近出た語ほど世界中でよく使われる（右上がり）。具体例をみると、右上がりと考えられる。OEDなどの初出年の基準を使うと、連続的な数値として、散布図で考察できる。

次に意味論的な性格をみると、ほとんどの語は日本に特有のものである。「特有語」つまり某文化特有の現象をさすことばである。これらは多くの文化で外行語になる。基準は別の言語で言いかえ・代替alternativeがあるかどうかで、新語論における「新物新語」（旧物新語の対義語）に相当する。







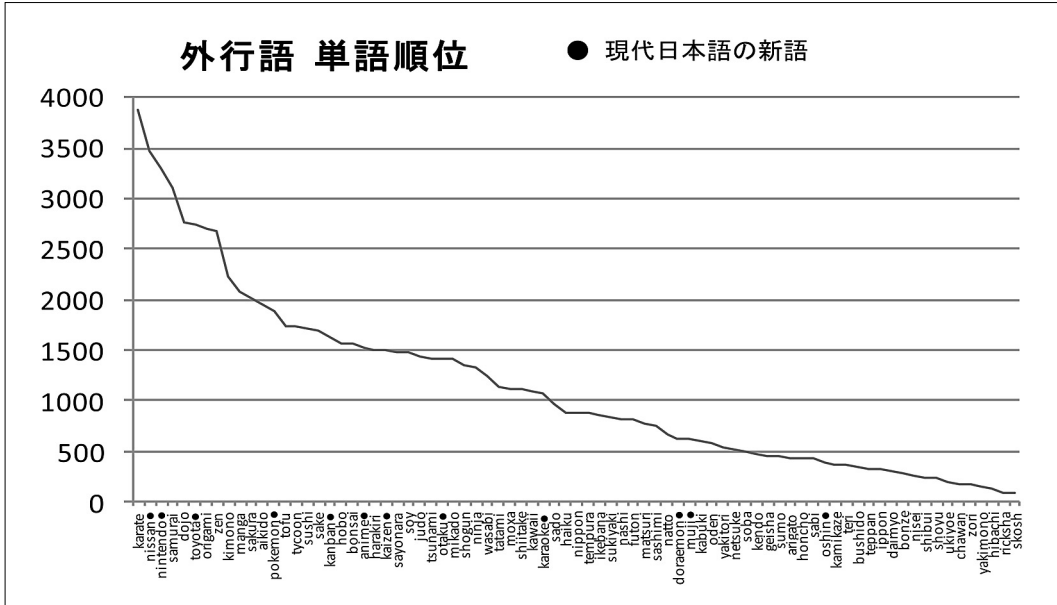


図 2-1 外行語の順位

英語の単語が現在世界各地の現地語の単語を置き換える形で普及しているのと、性格が違う。日本語には和語・漢語の代わりに（かっこよさから）英語からの外来語（洋語）が広がっている。

この観点からいうと、日本語と英語の借用語には大きな違いがある。日本語からの外行語は、新物新語、かつ特有語が大部分で、知的要因により、語彙の拡大として進出する。これに対し日本語への外来語は、旧物新語として、従前の和語・漢語を置き換える形で、情的要因によって流入する。

### 2.3. 外行語受容国の順位

図 2-2a, b に外行語使用国の順位を示す。国家・地域は数が多いので、上位と下位の二つのグラフに分割した。アメリカをトップに、東南アジア諸国が並び、南アメリカ、(西)ヨーロッパ諸国が続く。単純な地理的近接効果とは読み取れない。日本との経済交流の多さが関わると考えられる。もう少し下位の国家までみると、多様な要因がからむようである。様々な色を塗って考察した。外行語はまずアメリカ英語に入って他国に普及するとも考えられるので、英語を公用語にする国の末尾に●を付けた。確かに最上位の国と一致する。しかしアフリカの英語を公用語とする国は下位になるので、もっと別の要因を考えるべきである。

ただしここで言語的制約要因を考える必要がある。まず外来語受容に関する傾向（または言語政策）である。中国語では外来語を避けて翻訳語を使う傾向がある<sup>4</sup>。また文字も影響する。アルファベットで英語のグーグルインサイトで検索したデータだから、表意文字（漢字）・表音文字（音節文字（かな）・音素文字（アルファベット・タイ語などの文字）による検索をする国のデータは反映されない（ただしグーグル検索が民族文字でできない言語では、アルファベット検索を使う）。つまり

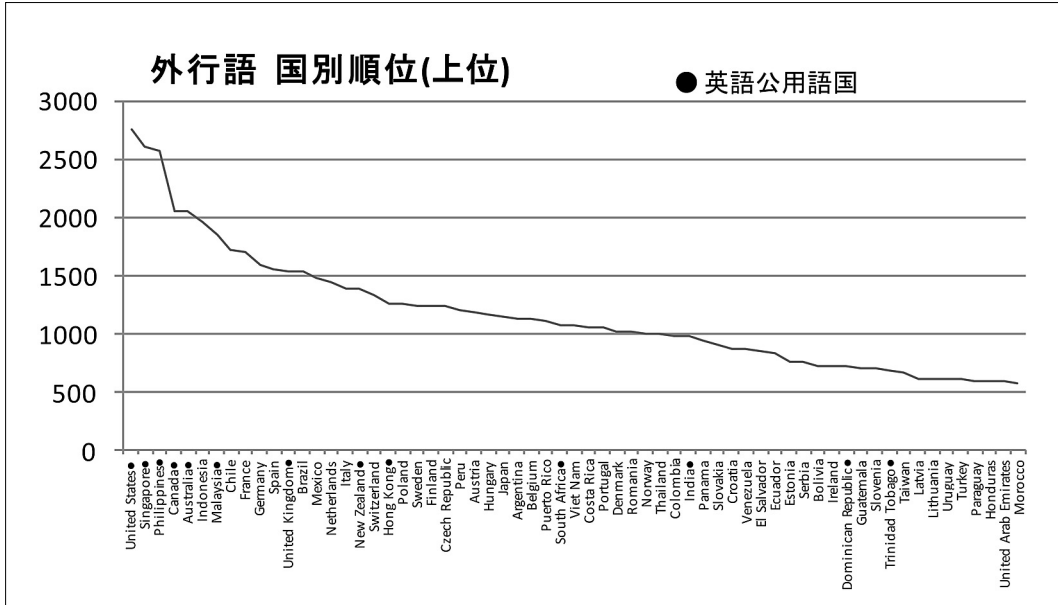


図 2-2a 外行語の国別順位 (上位)

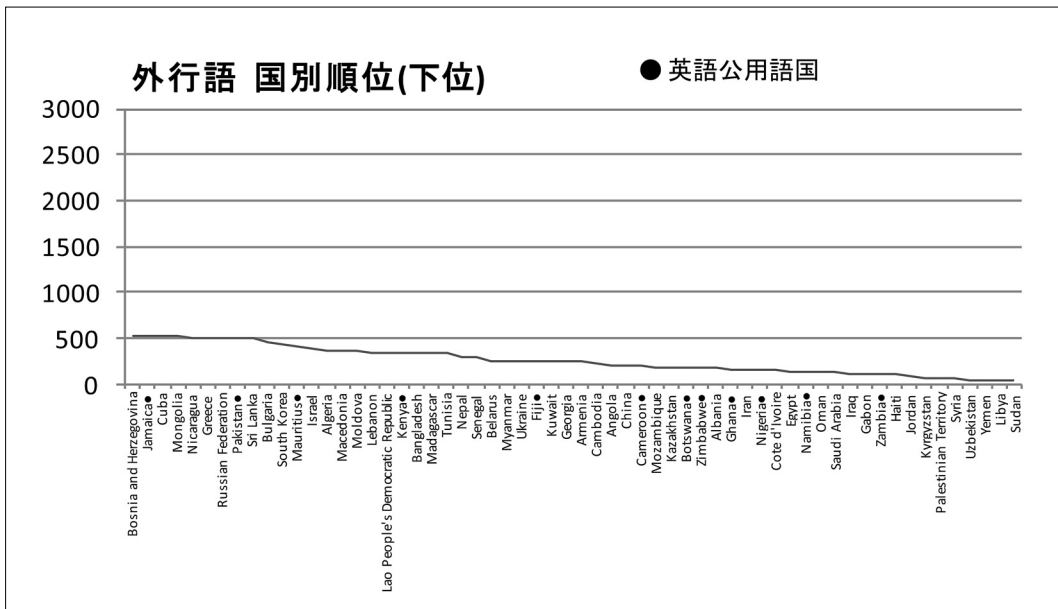


図 2-2b 外行語の国別順位 (下位)

アジア大陸の多くの国家には、制約がかかる。

大陸ごとにも違いがあり、アフリカでは外行語が少ない。非言語的要素を使って、これらの相関を調べることも考えられる。地理的距離（航空距離）直行便の有無や乗り換えによる時間、週ごとの便数などで国別にしたときには、ヨーロッパ諸国で差が開き過ぎる。貿易量などとの関連、日本

への親近感などとも関連が考えられる。日本語観国際センサスのデータは、多様な統計資料と照合して分析できた。本稿のグーグルインサイトデータも、同様に多種の資料と照合できるだろう。今後の課題にする。

#### 2.4. 使用率上位下位の対比

表2-1a,bで、使用率の大小によって、単語を左右に2分して示した。また国家も上下に2分できる。世界中で使われているか、日本語交流の盛んな（英語）国に限って使われているかを、判別できそうである。そのくらいの傾向性なら因子分析にかけなくとも、エクセルの処理で判別できる。表を上下左右に4分割し、それぞれの行列の合計を出した。その多数語と少数語、多用国家と稀用国家の数値の比を計算する形で、単語と国家を配列しなおした。その結果の順番は、かなり食い違うが、理由が説明できる。

この結果を組み合わせると、表2-1の縦横の順番を変えた表ができあがる（紙数の関係で提示できない）。左上は多数語を使う国家、左下は少数語をも使う国家である。右上は多数語も少数語も使う国家、右下は多数語も少数語も使わない国家である。ここでは上位の一部を論じる。

国家でみると、多数語と少数語の比が小さいのは日本付近のアジア国家・アメリカなどで、アフリカ諸国などは比が大きい。単語で言うと、多用国家と稀用国家の双方で使われる語は、以下の順番である。ippon, teri, toyota, nissan, tsunami, karate, judo, ninja, oden, kamikaze, karaoke, manga, anime, sumo。

#### 2.5. 研究史 言語間方言学 借用語論

以上、結果の提示を先に行った。以下でこの分野の研究史を振り返る。言語地図は多様なものがあるが、個々の単語について、世界全体を（または国境を超えて）見渡す研究も多様である。外来植物などは、生物学者や地理学者により実物の流入とからめて名称が調べられている。言語学者・方言学者の研究も多いが、具体的文献は他稿に譲る（井上 2000, 2001, 2010.2）。

インターネットでの言語現象分布の研究はグーグルマップの活用で急展開した。井上 2010.2などで紹介した。グーグルインサイトの活用によって、さらに研究の可能性が広がった（井上 2011.9, Inoue 2012.1）。なおグーグルブックスの200年間の驚異的な語数の調査を参照されたい（Michel 2010）。

### 3. グーグルインサイト地図の作成

#### 3.1. グーグルインサイト地図作成の技法

グーグルマップの検索方法、およびグーグルマップの画面の保存については、井上 2010.2で解説した。掲載の「明海日本語16」はインターネットでも見られるので参照していただきたい。グーグルインサイト地図作成の技法も基本的には同様である。最初にどんな検索画面でもいいから、「グー



グーグルインサイト」または「Google insights」と入力して呼び出す。図3-1のようなトップページが出る。「検索クエリ」の欄に単語を入力して、「検索」をクリックすれば、グラフや地図が出る。そのあとは知りたい単語を「検索クエリの追加」欄に入力して「検索」をクリックすれば最大5語までグラフや地図が出る。

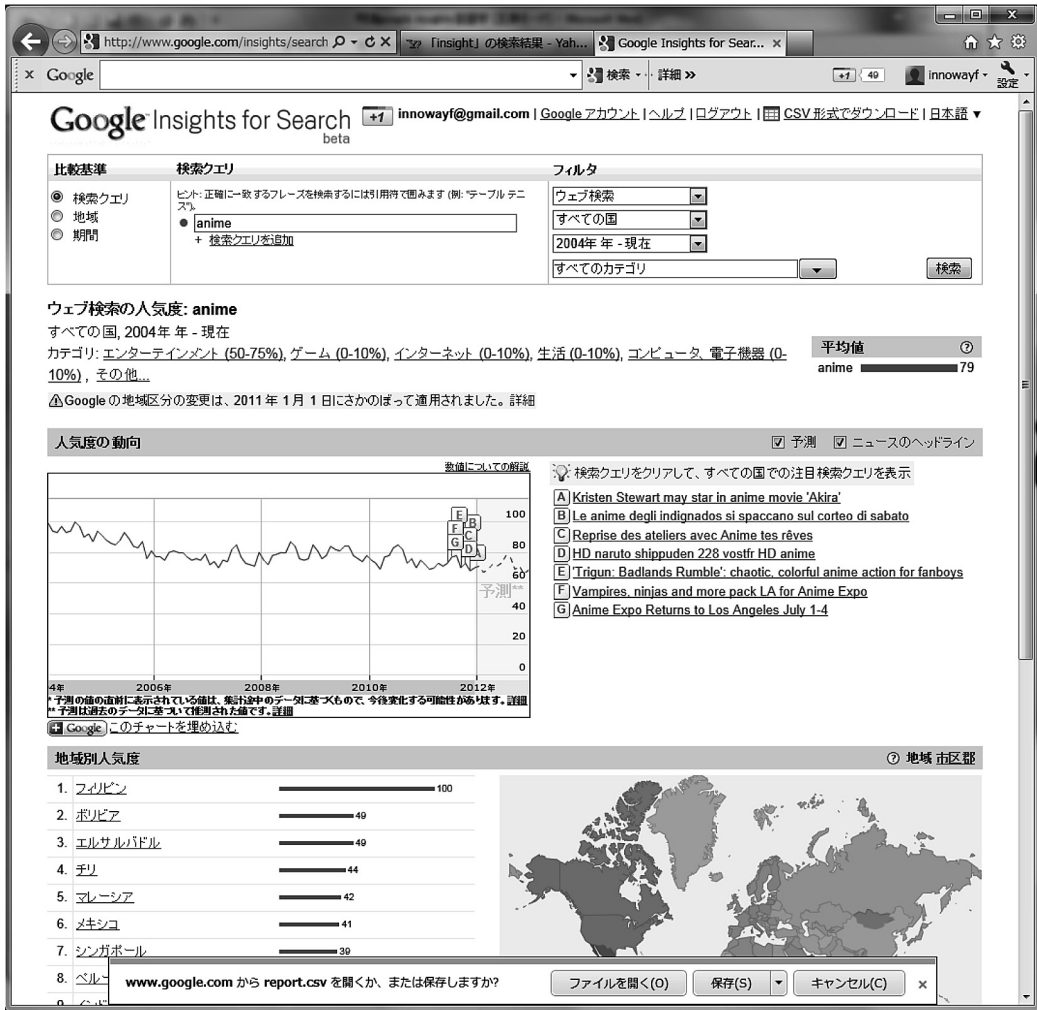


図3-1. Google insights 日本語版トップページ anime入力結果

### 3.2. 発展 数値データの集計

グーグルインサイトでは、グラフや地図の基礎になった数値を、エクセルの表の数値としてダウンロードできるので、さらに加工して研究できる。Gmailに加入して（無料）、グーグルインサイト検索の前に電子メールを使える状態にする。図3-1の画面右上の「CSV形式でダウンロード」をクリックすると、パソコンの別画面にエクセルファイルが表れるので、名前を付けて保存する。

### 3.3. 手順 ダウンロード

まずGoogle insightsにアクセスする前に、Gmailのメールを使える状態にする（自分の他のメールアドレスに空メールを出す）。次にGoogle insightsにアクセスする。グラフや地図などが画面に出る。画面右上の「CSV形式でダウンロード」をクリックする。「Report.csvを開くか」を聞いてくるので、「ファイルを開く」または「保存」をクリックする。

このエクセルファイルを開くと、文字化けが起こることがある。日本語をUNICODEで表記するかを選択で避けられるが、アルファベットだと文字の問題が起こらないので、画面左上の「English」をクリックして使う。図3-2にGoogle insights 英語版のanime, tsunami他の検索結果を示す。



図3-2. Google insights 英語版 tsunamiの検索結果

5語検索してファイルを開くと、**図3-3**のように語が左右に並んだ形で数表が出る。

このエクセルの表を「名前を付けて保存」する。すべてにReportという名前が与えられているので、区別できる名を与える。検索した単語を名前に書いておくとあとで整理に楽である。

数表には、英語版だと、まずグラフ用の数値が出る (Interest over time, Week)。その下に地図用の国家・地域別数値が出る (Top regions for anime)。さらにその下に都市別などの情報も入っている。

The screenshot shows an Excel spreadsheet with the following data:

1	Web Search Interest: anime; kimono; shogun; sukiyaki; teppan					
2	Worldwide; 2004 - present					
3						
4	Interest over time					
5	Week	anime	kimono	shogun	sukiyaiki	teppan
6	2004-01-0	92	2	4	0	0
7	2004-01-1	92	2	4	0	0
8	2004-01-1	94	2	4	0	0
9	2004-01-2	90	2	4	0	0
10	2004-02-0	91	2	4	0	0
11	2004-02-0	86	2	4	0	0
12	2004-02-1	89	2	3	0	0
13	2004-02-2	89	2	3	0	0
14	2004-02-2	89	2	3	0	0
15	2004-03-0	100	2	3	0	0
16	2004-03-1	90	2	3	0	0
17	2004-03-2	95	2	3	0	0
18	2004-03-2	86	2	3	0	0
19	2004-04-0	93	2	3	0	0
20	2004-04-1	94	1	3	0	0
21	2004-04-1	90	2	3	0	0
22	2004-04-2	90	2	3	0	0
23	2004-05-0	90	2	3	1	0
24	2004-05-0	83	2	3	0	0
25	2004-05-1	90	2	3	0	0
26	2004-05-2	86	2	3	0	0
27	2004-05-3	88	2	3	0	0
28	2004-06-0	89	2	3	0	0
29	2004-06-1	92	2	3	0	0
30	2004-06-2	95	2	3	0	0
31	2004-06-2	98	2	3	0	0
32	2004-07-0	95	2	3	0	0
33	2004-07-1	94	2	3	0	0
34	2004-07-1	94	2	3	0	0
35	2004-07-2	93	2	3	0	0
36	2004-08-0	93	2	3	0	0
37	2004-08-0	91	2	3	0	0
38	2004-08-1	91	2	3	0	0
39	2004-08-2	91	2	4	0	0
40	2004-08-2	93	2	3	0	0
41	2004-09-0	89	2	3	0	0
42	2004-09-1	84	2	3	0	0
43	2004-09-1	89	2	3	0	0
44	2004-09-2	85	2	3	0	0
45	2004-10-0	88	2	2	0	0

図3-3. Google insights 英語版 animeのエクセル集計表

以下では国家・地域別にしぼる。国家名でソートして、そろえれば、多くの語について集計できる。本稿のデータは、2011/12/7に出力した。データは1週ごとに更新されるので、数値は変化する。パソコンの画面を記録した日付は、右下に出る。

### 3.4. 手順 エクセルデータ総合処理

多数の語を統一して扱えば、総合的な状況、一般的傾向が知られる。グーグルインサイトでは5語までを同時に処理できるが、もし国家の順に並べようとする、各語により順番も登場する国家も違うので、多数の語を一つの表にまとめるのは、手数がかかる。ふと思いついたのは、世界中の多くの国家で使われる単語を、いつも検索語（の最後）に入れておく手法である。配列の統一指標になる。以下のデータに関しては「anime」をいつも入れた。4語ずつ検索すると、エクセルの表にはいつも「anime」の使用順によってソートされた形で使用率を示した数表が出る。

単語ごとに0から100までに数値化されている。順番ではなく、比率であり、グーグルインサイトのヘルプの説明では、以下のような数値である。

「グラフ上の数値は、Googleで検索された総検索ボリュームに対する、特定の検索クエリの検索ボリュームを相対的に表したものです。実際の検索ボリュームではなく、データを正規化して0～100のレベルで表したものです。グラフ上の各ポイントは最高値（100）で割った値です。十分なデータがない場合は0が表示されます。グラフ上部の検索クエリの横には概数または合計数が表示されます。」多数の単語の単純な合計であるが、初期的な処理を試みる<sup>5</sup>。

## 4. まとめと今後の課題

### 4.1. 地図からの出発と回帰

グーグルマップとグーグルインサイトは、地図の形で語の使用状況を示す。前者は使用地点を点としてプロットし、後者は国家・地域ごとに使用率を面として段階表示する。グーグルインサイトの面表示は、分かりやすいが、ミスリーディングでもあり、アメリカ、ブラジル、ロシアのような広い地域の状況に読み取りを支配され、香港やシンガポールなどは目立たない。読み取りが主観的になる恐れがある。これを避けるために、本稿で扱ったように、数表として分析する手法が役立つ。また単語については、少数の語だけを扱うと、個別の事情に攪乱される。例えば国際センサスで扱った4語のうち、shogunは映画としての上映と関係がある。また語によっては外行語と偶然一致する語形が某言語に存在することがありうる。本稿では扱わなかったが、Fujiはスペイン語では「逃げる」の意味になる。この危険も、多数の外行語を集合として扱うことによって、影響力を弱めることができる。この手法は、他のジャンルの単語にも広げる価値がある。

### 4.2. 外行語総合表の細分析

言語現象を地図やグラフにしてみると、様々な法則性が浮かび上がる。「単語」で方言差のありそ



うな語を地図で見してみる。本稿では外行語に絞った。ある程度広く分布している外行語と、特定の国で分布している外行語がある。これまで外行語については、英語でまず指摘されていた。グーグルインサイトを使うことにより、英語国以外に世界中に外行語が普及していることが分かった。一方井上2001などにより、北アメリカ、西ヨーロッパ以外に、南アメリカ、オセアニアと東アジアでも外行語が多いことが推測できたが、語数が少なく、偏っていたので、一般化が困難だった。本稿の大量データの分析により、これまでの分析で得られた傾向を再確認できた。新古双方のデータと分析手法の信頼性を相互に補強しあう成果と言える。要するに日本との交流の多い国家に外行語が多く分布するのである。

今回のデータに関しては、因子分析にかけ、相関係数の表を分析する手法でさらに細かいパターンに分けられるだろう。さらにこのデータは、単語の順番と国家の順番を説明できる変数を求める作業が、残っている。グーグルインサイトでは単語ごとに全体使用量（ボリューム）が出るので、この数値を掛け算できるが、エクセル集計表には出力されない<sup>6</sup>。この手法は、語彙統計の異なり語数に対する延べ語数の集計に相当する。（ネット上の検索で）多く使われる単語に重みづけすることに相当する。今回の分析で上位になった語は現代の若者が使いそうな単語なので、単語ごとの差がさらに開くだろう。また国家によってインターネット普及率が違うので、国家間の差も明らかになると期待される。

#### 4.3. 外行語総合表の拡大

単語も再検討の可能性がある。外行語はもっと多し、最近さらに増えているので、新たに探索する必要がある。一方かつて指摘されていた外行語（英語辞書に載っている全語）をグーグルインサイトで検索することも考えられる。今回の試みの最下位のhinoshiのように、使用率ゼロの語は、集計に含めることができないが、その数値自体に意味がある。単なるグーグル検索では（ホームページ内の語として）使われている可能性がある。

また韓国語、タイ語、中国語、アラビア語、ロシア語などを使い、アルファベットをあまり用いない国家については、翻訳ソフトで相当する表記を探し、コピーアンドペーストで貼り付けてグーグルインサイトで検索すれば、世界全体の外行語がさらに詳しく分かる。

調査対象の語は、今後英語（やドイツ語、フランス語、スペイン語）の世界的な違いにも広げたい。外行語に相当する語は、多くの言語に存在する。術語としての一般性を高めるためには、「輸出語」「輸入語」がふさわしいだろう。多くの言語の例を探れば、借用語を手がかりにして、世界の文化の流れが分かるだろう。借用語は水のように、高いところから低いところへ流れると言われるからである。

さらに国家内部の県別、州別等の使用率の差も分かるので、方言差も分かる。日本国内だけでなく中国・韓国やアメリカ、カナダなどの方言差も観察できる。交流の多さと関係すると思われる。以上ではグーグルインサイトの地図部分のデータ分析を論じた。エクセルデータではさらに最近7年間の推移（トレンド）も分かる。どんな外行語が最近増えているか、減っているか、全体として

どうかも確認できる（井上予定）。

従前にくらべて画期的な研究手法で、短時間で多くの現象を知ることができる。学問研究におけるコストパフォーマンスの高い研究方法で、将来性の大きい新分野である。卒業論文に役立つ。

#### 〈注〉

- (1) 語数が少なく、国家数が増えすぎるので、今回の分析では表から省いて処理した。
- (2) グーグルインサイトではregionsと呼んでおり、国家と独立国家でない地域の両方を含むが、本稿では便宜上「国家」と呼ぶ。
- (3) この順番は各単語の使用率が考慮に入っていないので、仮のものである。井上（予定）では各語の使用率比率を計算できたので、そちらを重視すべきである。
- (4) 漢字などのアルファベット以外の民族文字の検索も可能である。世界中のどの地域（中国語圏、華僑の多い地域とか）で中国語がその綴りで使用されているかなども分かる。
- (5) 各語の日常での使用率が分からないので、各言語（国家）での多数の辞書の記載数を扱っているような分析である。場面や人によって使用が違うだろうが、考慮に入らない。数値は、順番を合計して並べるよりましで、オリンピックでメダル数を単純に数えるのに似る。金メダルか銀・銅メダルかを考慮に入れないし、団体競技で複数を大勢が手にするのを考慮するのに似る。比喩としては、プロ野球も高校野球も草野球も一緒にして、様々な野球選手の打率を並べるようなものである。
- (6) 毎週の使用状況を示すトレンドグラフの数値を掛け算すると、検索ボリュームを考慮した集計ができる。

#### 参考文献（五十音順）

- 井上史雄（2000）『日本語の値段』（大修館）
- （2001）『日本語は生き残れるか——経済言語学の視点から』PHP新書
- （2010.11）「Google マップで見る関西弁の世界進出」『地域語の経済と社会』第127回 三省堂 word wise web
- （2010.12）「「体操服」のGoogle マップ」〈ことばの散歩道〉151 日本語学 29-15
- （2011.1a）「Google マップによる「モータープール」の世界分布」『地域語の経済と社会』第132回 三省堂 word wise web
- （2011.1b）「消える雪」〈ことばの散歩道〉152 日本語学 30-1
- （2011.2）「Google 言語地理学入門」明海日本語 16 pp.43-52
- （2011.9.10）「グーグルインサイトによる「キャンディーバー」と「チョコレートバー」」三省堂ホームページ地域語の経済と社会第167回（<http://dictionary.sanseido-publ.co.jp/wp/2011/09/10/>）
- （予定）「日本語世界進出の動向」
- Inoue, Fumio (2012.1) “Improvements in the sociolinguistic status of dialects as observed through linguistic landscapes” *Dialectologia* 8  
www.publicacions.ub.edu/revistes/ejecuta\_descarga.asp?codigo=713
- 新プロ「日本語」総括班編（1999）『日本語観国際センサス 単純集計表（暫定速報版）』（国立国語研究所）
- 早川勇（2006）『英語になった日本語』春風社
- 原口庄輔・原口友子編訳（1998）『新「国際日本語」講座』洋販出版
- 三輪卓爾（1977）「外行語の昨日と今日——海を渡った日本語——」言語生活 312
- Michel, Jean-Baptiste et al. (2010) “Quantitative Analysis of Culture Using Millions of Digitized Books” / [www.sciencexpress.org /](http://www.sciencexpress.org/)