

国立国語研究所学術情報リポジトリ

First year of appearance of standard Japanese forms and gravity center by railway distance : East and west clusters in 3D distribution graphs of railway distance, usage rate and first year of appearance

メタデータ	言語: jpn 出版者: 公開日: 2019-03-25 キーワード (Ja): キーワード (En): 作成者: 井上, 史雄, INOUE, Fumio メールアドレス: 所属:
URL	https://doi.org/10.15084/00002130

標準語形初出年と鉄道距離重心

——鉄道距離・使用率・初出年の3D散布図と東西クラスター——

井上 史雄

(東京外国語大学)

キーワード

言語地理学, 標準語形, 計量方言学, 鉄道距離重心法, 初出年

要 旨

この論文では、従来分析を進めてきた標準語形使用データについて、二つの単純化を適用した。地理的位置を鉄道距離によって表現したことと、語形の地理的分布を1点の重心で示したことである。本稿では、まず「河西データ」の県別使用率のグラフにより、標準語形の中でも古代初出語の一部が辺境残存分布を示すことを確認した。つぎに「河西データ」の行にあたる各語形について、鉄道距離重心を計算し、各語形の全国使用率、初出年との対応をみた。2要素ずつを組み合わせた2次元のグラフを考察し、また3要素の関係を示す3次元のグラフを考察した。

さらに古代・近代2時代への区分と東西2クラスターへの区分を組み合わせて82語を4区分して検討した。古代初出の東部クラスターの語は、初出年との相関を見せない。しかし他の三つの区分では、初出年がかなりの相関を見せ、しかも近似直線の数値が似ていて、1000年につき31~36%の減少を示す。これは普及年速1キロ(弱)という仮説と矛盾しない。文化的中心地から新しく出た語は、最初は全国使用率が低いが、年数が経つと古く出現した語と同じ過程をたどって全国に広がって、全国使用率が高まると考えられる。また古代に出た語は、その後の新形に侵食されて、文化的中心地を明け渡すことがある。

1. 本研究の位置付け

1.1. 研究の出発点

「河西データ」は、『日本言語地図』(LAJ)のうちの82項目について、県ごとの標準語形¹使用率を計算した行列である²(河西 1981)。河西データの総まとめにあたる県別の標準語形使用率は、河西(1981)以外に井上(2001)などに掲げてある。82項目総合地図として、4段階または7段階に分けた大まかなものだった。その後約20段階に分けた地図(Kawaguchi & Inoue 2002)により、標準語形使用率が東京中心にはほぼ同心円状に分布することが、読み取れた。ただし北陸の低さなどは、東京からの直線距離では説明できず、交通路が日本アルプスなどによって遠回りになることを考慮に入れる必要が感じられた。

本稿の関心は、河西データの文献初出年と他の属性との関係である。初出年が、地理的分布パターンと関連することは、井上(2001:第8章)で確かめた。また因子分析・クラスター分析の適用結果により、地理的分布パターンが全国使用率と関連することも分かった(井上 2003a)。し

かし初出年と全国使用率との間には、わずかな関連性しか見えなかった。本稿では、初出年を二つの時期に分け、かつ地理的分布パターンも2種に分けることにより、両者の関連性が薄くなった理由を確かめることができた。1000年以上の歴史の中で、かつて京都から広がった語形のみに、すでに中央を他の新形に譲り渡して衰退期に入ったものが混じったためである。

1.2. 河西データと鉄道距離（県の分析）

「河西データ」に鉄道距離を組み合わせた分析については、いわば連作の形で発表している。紙数制限による分割のため、また分析手法の段階的進化のためである。以下では本稿に必要な部分のみ、これまでの研究をたどる。方言区画の研究史については、井上(2001, 2003a)などを参照されたい。データの性格と研究史については、Inoue(2004, in print, forthcoming), 井上(2004a, 2004c, 予定)などを参照されたい。

地域間のコミュニケーションにおいては、直線距離よりは交通路が基本的影響を与える。交通路の手がかりとして、鉄道距離は便利な指標である。前代の街道などの交通路をなぞる形で発達したので、近代以前の交通をも反映することによる³。また『日本言語地図』のインフォーマントの時代には鉄道が主要な運輸手段だったこともある。ここでは1980年代の国鉄時刻表により、東京駅から都道府県庁所在地の駅への距離を計算した⁴。

先行論文では、列にあたる県の分析を進めた。各県について鉄道距離が計算してあったが、本格的な分析はなかった。2003年に各県の標準語形使用率平均値と鉄道距離とを散布図に描いたところ、きれいな相関関係をみせ、鉄道距離が日本語標準語形使用の地域差をよく説明することが分かった(井上 2004b, 2004c)。言語現象の分布・伝播が交通(路)に影響されることは、古くから言語地理学で言われてきたことであるが、多量のデータに基づいて数量的に解明された。京都および東京からの鉄道距離グラフを比較したところ、それぞれに異なった中心性を示したので、かつての「京都周囲分布」と近代の「東京周囲分布」という、標準語の歴史の2段階を考えた。また現代の中学生の標準語形使用では鉄道距離がほとんど意味を持たないので、これを第3段階「全国一律分布」とした(井上 2004a, 2004b; Inoue in print)。

これをもとに、『日本言語地図』の標準語普及の2段階は、以前に適用した因子分析・クラスター分析による語形の分類に反映されるのではないかと考え、いくつかのクラスターに分けて標準語形使用率平均値と鉄道距離とを散布図に描いた。予想どおり西部クラスターは第1段階の京都周囲分布を示し、東部クラスターは第2段階の東京周囲分布を示した(井上 2004a)。その後クラスターをさらに細分し、東部クラスターのうちでも東京クラスターの20語が、東京中心に東日本に急速に普及するパターンを示すことが分かった(井上 予定)。以上は県についての分析である。

1.3. 河西データと初出年（語形の分析）

行にあたる各語形についても、①平均鉄道距離（重心）が計算できるので、本稿では、これを扱う。各語形の②全国使用率との対応が考察可能である⁵。また各語形の③文献初出年との対応

もみた。文献初出年については、井上（2001:第8章）で地図化し、初出の古い語は京都中心に分布して、西日本に偏ること、ときに近畿地方が空洞のようなことになること、近代初出の語は東京付近に分布し、東日本に偏ることなどを、地図に表示できた。その後『日本国語大辞典』第二版によって初出年を調べなおし、日本の文化的中心地の移動との関連をみた（井上 2003a）。

1.4. 古代近代・東部西部別県別使用率

まず県別の計算に基づく結論部分を、先取りして本稿冒頭で示す。図1では、以前に適用したクラスター分析（井上 2001:図2-1）に従って82語を東部クラスター43語と西部クラスター39語に分け、さらに初出14世紀を境に古代初出34語・近代初出48語に分けた（後掲表2参照）⁶。縦軸に県ごとの平均使用率を示した。横軸の県の配列は旧自治省の番号によるもので、鉄道距離は考慮に入っていない。下端東京・京都の位置には目立つように▲印を付けた。この図では、東部・西部のクラスターの差がきれいに出た。それにくらべると、古代・近代の違いは小さい。しかし細かくみると、伝播に関して示唆に富む現象が読み取れる。

まず古代初出東部クラスターの語（■）が特徴的で、標準語としての東京からの新たな普及でなく、他の語形に追われての辺境残存の傾向を示す。栃木はじめ北関東が最大使用率で、東京が低い。また山梨・長野も高いが、中部地方は古形をとどめやすいという指摘がある（西岡 1956; 山口 1982）。西日本は一般に低いが、熊本・高知がやや高い。ここも古形を残す辺境にあたる。新潟・富山・石川の低さは、実は東北、西日本の低さと同格である。つまりは、関東・中部（の

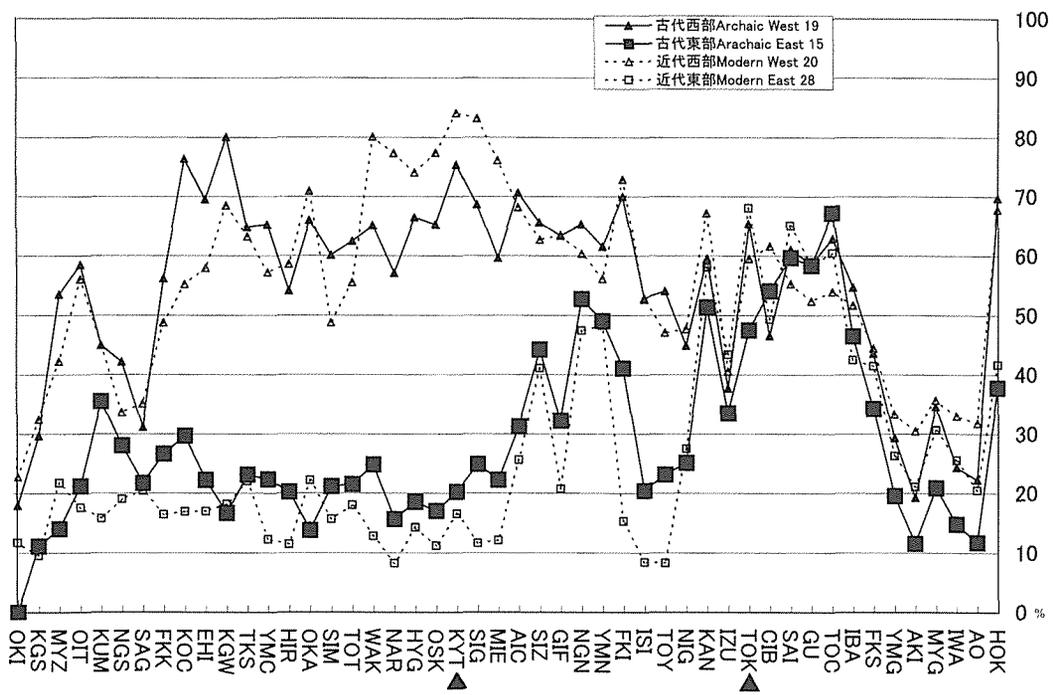


図1 古代近代・東部西部別県別使用率

周辺部) だけが高いという東部クラスターの特性を示している。個々の分布図で古代東部クラスターの語の地理的分布をみると⁷、「つむじかぜ、におい、とかげ、すてる、(塩味が) うすい」などで、確かに東京よりも北関東の使用率が高い。

近代初出東部クラスターの語(□)の平均使用率は、東京ついで埼玉がとび抜けて高い。東京からの鉄道距離を横軸にとり使用率を縦軸にとって散布図を描いたところ、かなりきれいな反比例関係が得られた(本稿では略す)。明らかに東京から鉄道距離に応じる形で広がりつつある。他の県は古代東部クラスターの使用率より下が多い(東北地方は別)。古代から広がり続け、今は衰退の道をたどる語も含む古代初出の語を、近代初出の語は、まだ追い抜けないわけである。

西部クラスターのうち、古代初出西部クラスターの語(▲)では、四国の香川・高知(と愛媛)が固まってピークをなす。京都・滋賀・福井・愛知が続く。東北地方と(大分・宮崎を除く)九州地方では使用率が際立って低い。近代標準語としての新たな普及でなく、関西文化圏の周辺部での残存傾向を示す。古代西部クラスター19語の具体的分布をみると、「いえ、しあさって、やる、からい、けむり、あぜ、こおる、こおり」などで確かに四国地方の使用率が近畿より高い。その原因の多くは、のちに京都から別の(標準語の地位に上昇しそこねた)語形が広がったためである⁸。

近代初出西部クラスターの語(△)は、京都・滋賀がピークで、近畿地方一帯が高い。新潟・富山・石川の低さは、東西双方から新形を受け入れないことを示す。大分を除く九州は低い。東北も低い。しかし北海道は、(古代初出・近代初出ともに)西部の言い方を受容している。これは内陸部の近代の開拓者のおかげであろう。京都からの鉄道距離を横軸にとり使用率を縦軸にとって(北海道・沖縄を除く)散布図を描いたところ、きれいな反比例関係が得られた(本稿では略す)。

まとめると、東西クラスターともに近代初出語は、文化的中心地からの鉄道距離とみごとな対応を見せる。これに対し古代初出の語は、東西それぞれの中心地から離れたところ(辺境)で多く使われる傾向があり、京都や東京の近くでは使われなくなる。つまり古代初出語は、標準語と一致する語形であっても、地理的には辺境残存分布を示し、退縮過程を表す⁹。1000年近く経つと、国土に広がりきって、領域の一部を他の語形に奪われるわけで、栄枯盛衰、勢力交代の過程を示す。

以上のように、地理的分布パターンのクラスター分析によって語を東部・西部に分け、かつ初出年によって古代・近代初出に分けて、4グループに分けると、有意な傾向が見られた¹⁰。しかし後掲図2、図3にみられるように、初出年は、クラスター分析の結果や全国使用率と全体としてつきあわせると、対応関係がないようにみえた。図1の4分類に至るには試行錯誤が必要だった。以下錯誤の部分を除いて、試行の過程をたどる。

1.5. 鉄道距離重心の計算法

以下では、河西データ(河西 1981)のうち県(つまり列)ごとに測った鉄道距離から、語形(つまり行)ごとの鉄道距離重心を計算する手法について、述べる。鉄道距離の語形ごとの平均

値は、人口の地理的重心を測ると同様の数式（奥野 1977:113f）で計算できる。

$$\frac{w^1 \times x^1 + w^2 \times x^2 + w^3 \times x^3 + \dots + w^n \times x^n}{w^1 + w^2 + w^3 + \dots + w^n} = \text{Distance Center}$$

ここで $w^1 \dots w^n$ は weight（この場合は県の標準語形使用率）を示す。

また $x^1 \dots x^n$ は地理的位置（この場合は県の鉄道距離）を示す。

各項目ごとに各県の使用率と鉄道距離を掛け算して行列ができる。その総計を各県の使用率合計で割る。これで82項目それぞれの鉄道距離重心が求まる¹¹。人口重心は、人口に応じた重りを地図に載せ、地図に穴を開けて糸で吊るすと釣り合うような点である。ここでは人口の代わりに標準語形使用率に応じた重りを、各県に載せたと考えていい。地理的位置は、鉄道距離によるので、いわば鉄道距離に従って地図をデフォルメして、鉄道で遠回りになる北陸や大分・宮崎が実際より遠くに位置するように図化したようなものである。その鉄道距離重心を、方角を無視して、東京（または京都）からの距離だけで表す。そのために1個の数値で位置を表現できる¹²。また人口重心の年ごとの移動を示すなら、数個、数十個の数値計算の結果を地図上にプロットすることになる。ここでは82の語について計算し、グラフにプロットした。なお井上(2004b, 予定), Inoue(forthcoming) などでは、各県の鉄道距離と標準語形使用率の散布図を描いた。

本稿では東京からの鉄道距離のみを扱う。京都からの鉄道距離についても対応するグラフを出力し、原稿も整えた。両方を照合しつつ考察すると、深い洞察が得られるが、紙数の関係で京都関係部分を省いた。

2. 鉄道距離重心・使用率・初出年：2次元グラフ

2.1. 鉄道距離重心と全国使用率

過去においては鉄道距離重心と使用率と初出年のうち、2要素ずつを組み合わせた2次元のグラフ（散布図）を作り、残りの要素を記号で示すなどの手法で考察してきたが、隔靴搔痒の感を免れなかった。第3節の、鉄道距離重心と使用率と初出年との関係を示す3次元のグラフが、これまでの分析の総合、集大成になる。ただ3次元グラフは、全体の相互関係をとらえるにはいいが、82項目におよぶ本データについては、複雑すぎて読み取りが困難な点がある。また個々の項目への配慮が及ばないことがある。第2節で、予備的分析として2要素を2次元ずつを組み合わせて考察する。

まず図は略すが、東京基準の鉄道距離重心と全国使用率との関係をみると、両者には密接な関係が見られた¹³。

2.2. 初出年と全国使用率

82の標準語形は、京都続いて東京からの普及の過程を多様な形で、連続的に示すものと考えられた。そのための歴史的手がかり、文献初出年が、鉄道距離や全国使用率とどんな関係にあるかを見よう。

まず、初出年と全国使用率の関係を確かめよう。図2は、横軸が文献初出年、縦軸が全国使用率である。個々の項目名を示し（注6参照）、また東西のクラスター境界を線で示した。使用率の多い語の集中する時期については二つのピークが見られる。700年代奈良時代は、多量の日本語文献が現れ、多くの語が記録されたためである。また1600年代は、キリシタン宣教師の日葡辞書や江戸時代以降の豊富な文献によって、多くの日常語が記録されたためである。14世紀初出の語がないのは、そのころの文献資料が少ないためでもある¹⁴。文献初出年は、その語が口頭で使われはじめた時期を忠実に示すものではない。そのほかにも文体差・階層差なども考慮すべきである。つまり実際にはいずれかの地の民衆のことばとして、文献初出よりはるかに早く使われていた可能性がある¹⁵。

全国使用率は西部クラスターで高く、東部クラスターで低い傾向があるので（井上 2003a）、図2でもほぼ横の線で区切られる。横軸の初出年との関係をみると、上の西部クラスターも下の東部クラスターも、14世紀を境に二つに分かれる。西部クラスターの語は全時代に多いが、ことに16世紀以降に多い。東部クラスターの語は16世紀以降ことに19世紀に多く固まる。

近似直線（最小2乗法による直線）ではゆるい右下がりになり、

$$y = -0.0102x + 51.06$$

である。つまり（xに年を代入すると分かるが）西暦0年初出という仮定の語形は全国平均使用率が51%という計算になる。800年初出の語では約8%低く約43%になり、2000年初出の語では約20%低く約31%になる。初出が100年遅い語は約1%ずつ使用率が低いという傾向を示す。古

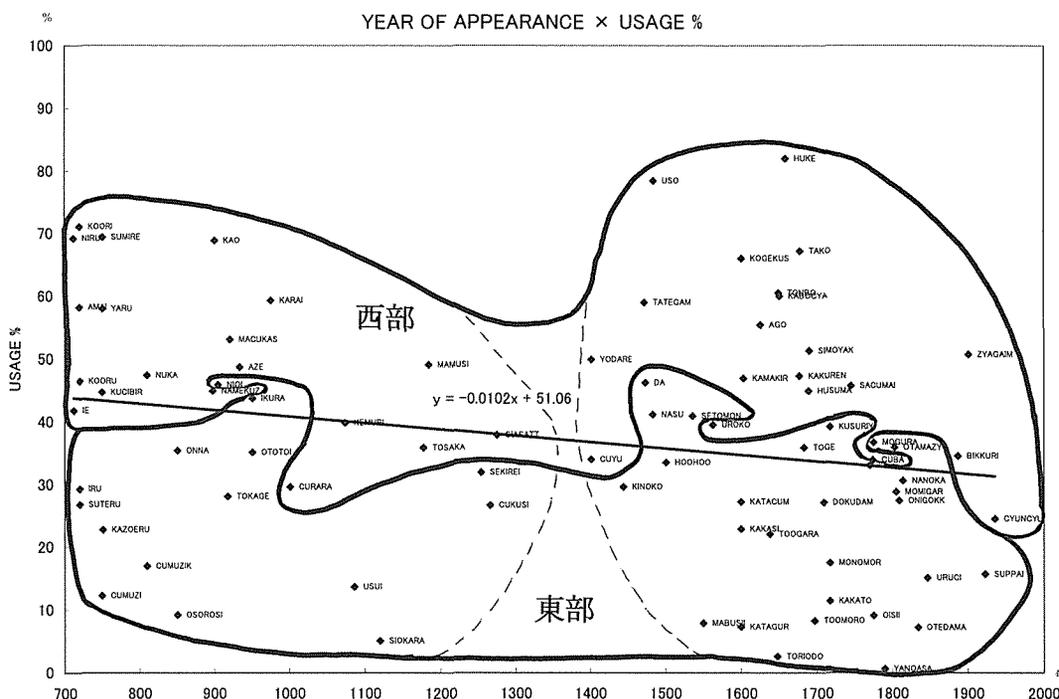


図2 初出年と全国使用率（82項目）

くから使われていた語は全国に普及するが、その後出た語はまだ領域拡大の途上である、と読みとれる。図で二つのピークがあることを説明するためには、もっと近似式の次元を多くするほうがよさそうで、4次、5次方程式の方が、図2のパターンをよりよく再現しうる。つまり8世紀あたりをピークにゆるやかに下がる線と、16世紀あたりをピークに急に下がる線である（後掲図8参照）。

語形の普及の典型が年速約1km（以下）というまとめがあるが（徳川 1993, 1996; 井上 2003b)¹⁶、図2の右上の「うそ、ふけ」のように、15世紀、17世紀初出の語でも70%から80%の使用率に達することをみると、標準語形の普及年速は、近代にはもっと速まった例もありそうである。これを手がかりにすると、全国普及に必要な年数は一応考察可能なはずである（図8で後述）。この図は、井上(2003a)で論じたように、中世・近世以降京都（および各地）で別の言い方が生じて、かつての標準語の領域がむしばまれる過程（近畿空洞分布）をもとらえている。

2.3. 初出年と鉄道距離重心

以上の関係を別の面から見るために、鉄道距離重心と初出年との散布図を作った。図3では、縦軸が初出年を、横軸が東京からの鉄道距離重心を示す。

図3では、右端を（地図上の配置に従って）東京の位置0kmにした。図の左に西部クラスター、右に東部クラスターが分布する。クラスターはここでは東西の位置に対応し、鉄道距離重心の信頼性・便利さが示された。ただし図2にくらべると、クラスターの境界は複雑で、左下に飛

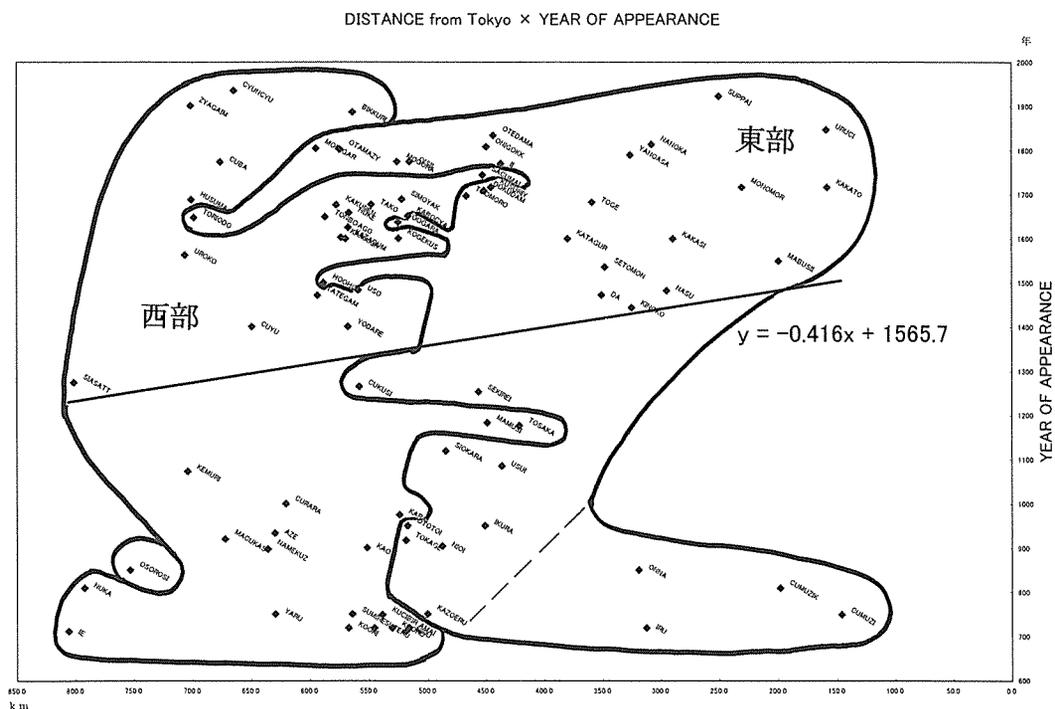


図3 初出年と鉄道距離重心（82項目）

地もある。初出年と鉄道距離重心の関係が単純でないためである。

東部クラスターの多くは、右の東京付近から450キロあたりにかけて位置する。日本の人口重心（岐阜）の距離に相当する。右上つまり初出が江戸時代以降の語が多い。左側にも一部伸びる。例外が右下の「つむじ、つむじかぜ、おんな、いる」の4語である。古代初出語の関東での復活であり、関西でその後生まれた語形（ぎり類、まき類、おなご、おる）が標準語に上昇しなかったことによる（佐藤 2002）。

西部クラスターは、左の京都（500km 付近）以遠の地に重心がまとまる。古代初出の語は左（東京の遠く）に位置し、近代初出の語は真中（東京のやや近く）に位置する傾向がある。東京からの伝播も働いたのだろう。

縦軸の初出年を手がかりに全体をみよう。東京付近に重心があるのに例外的に初出が早い右下の4語「つむじ、つむじかぜ、おんな、いる」を除くと、全体傾向がはっきりする。全体として左下から右上にプロットされる。初出の古い語は左下、つまり京都以遠に重心があり、初出の新しい語は右上、つまり東京から近い東日本に固まる。一方1700年前後初出の語も500km 付近、つまり京都付近に多く固まる。

図3の近似直線は、ゆるい右上がりになる。

$$y = -0.416x + 1565.7$$

とは、(x に距離を代入すると) 東京（つまり0km）の地点に分布する語の平均初出年がおおよそ1566年であり、京都付近(500km)の地点に分布する語の平均初出年がおおよそ1358年であることを示す。換言すると東京から100km 離れるごとに平均約42年古い語になる。近代初出の語は東京付近に重心があり、中世以前初出の語は遠いところに重心があることを示す。これは後世江戸・東京からの新形が標準語形として普及したことの反映である。

方言圏論によれば、古く出た語が遠くまで伝播し、古語は辺境に残ることになり、初出年の古い語の鉄道距離重心は遠いはずである。図3の全体傾向は、圏論に合致する。逆にいうと、近代になって東京が勢力を増して、東京から広がった語が多いことを示している。ただし各語の使用率までは示されないので、この図は単独で読みとれるわけではない。図4の3次元グラフの考察が必要になる。

なお図8で示すように、初出年で14世紀を境に古代・近代に区切ると意味のある傾向が出てくる。図3の近似直線はほぼ14世紀の境界線と一致するので、これを利用すると、語形を古代・近代と西部・東部の組み合わせで4分類できる（左端の「しあさって」のみが古代西部クラスターなのに近代に入るように見える）。

3. 鉄道距離重心・使用率・初出年：3次元グラフ

3.1. 鉄道距離重心・使用率・初出年の3次元グラフ

以上、語形についての鉄道距離重心、全国使用率、初出年を、二つずつ組み合わせる形で考察した。しかしこれら3要素は相互にからみあう関係にある。これを見るために、3次元グラフを作成し、かつ近似値を計算してみよう。まず82項目全体として示す。統計ソフト STATISTICA

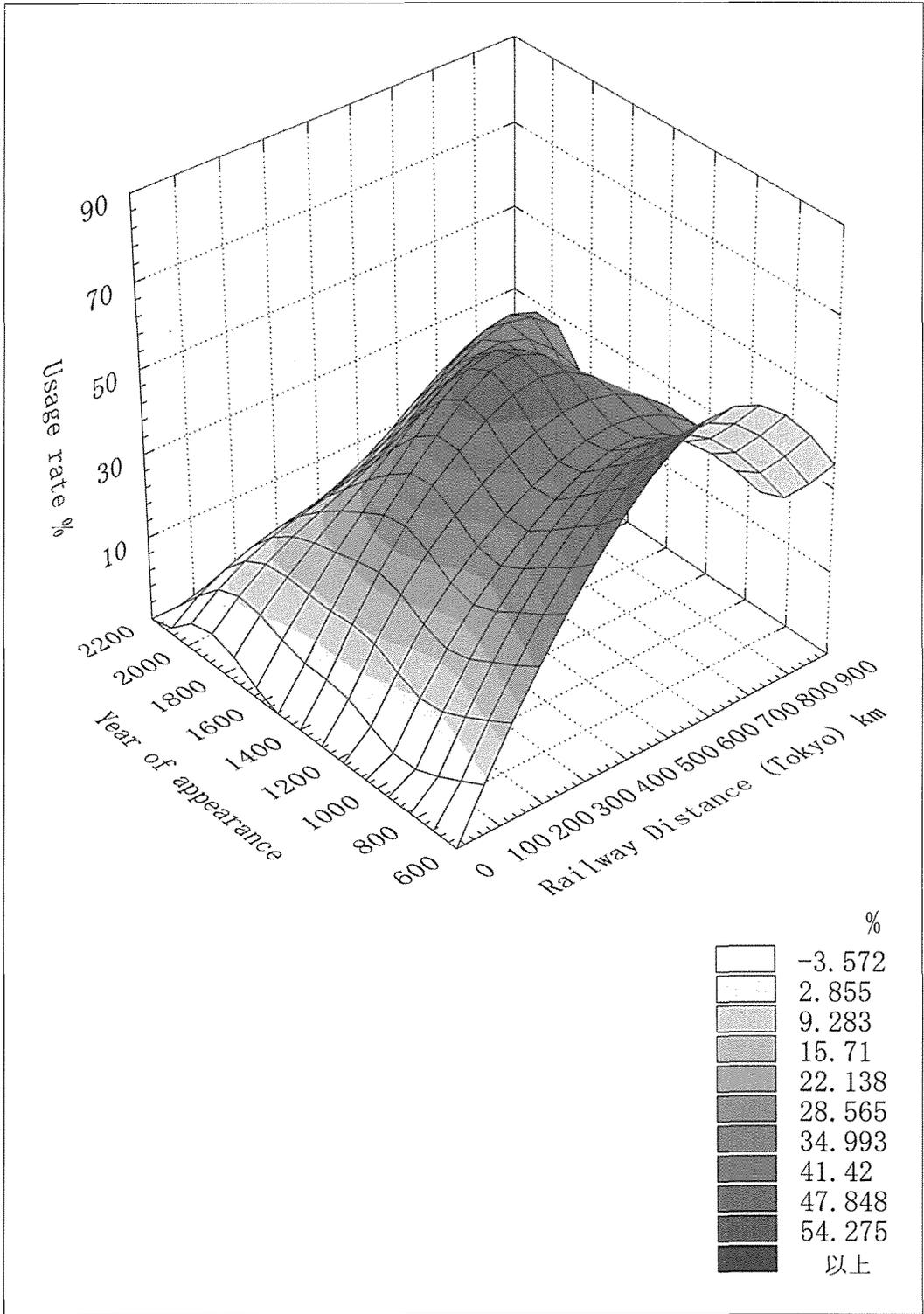


図4 82項目 3D鳥瞰図 (重み付き最小2乗近似)

では様々な技法が適用可能なので、多種類のグラフを出力したが、複雑なパターンを忠実に示す技法を主に提示する。これまでの直観的把握とほぼ一致し、さらに精密に示すことができた。

図4～6の手前から右に向けての軸は東京からの鉄道距離で¹⁷、0キロから900キロである。手前から左奥に向けての軸は初出年である。600年から2200年までの目盛があるが、実際には700年代から1900年代までである。上下の軸は使用率で、90%まで目盛があるが、実際には最大で82%である。なお右端凡例の数値は標準語形使用率（縦軸）の切れ目を示す。全体を把握するためにいくつかの近似値計算法を適用したが、図4以下では、重み付き最小2乗近似によるものを示した。個々の語形を示した3次元グラフ¹⁸と対照すると、細かい起伏を忠実に示す。

図4によると、鉄道距離重心と使用率については、全体として東京に近い左側で低く、途中高くなり、また下がるという傾向が見えた。図3と照合すると、西部クラスターが600キロ付近に集中すること、および東京付近に使用率の低い語形が分布して、なだらかな傾斜を描くことが影響している。この傾斜を反映して、山が遠方で平面に付くのは、グラフの端の900kmより向こうである。図3では実際には800キロ以上で位置する語はない。東京からかつて伝播して、東京自身のさらに新しい言い方によって周囲（残存）分布をなすような現象は、実際にはないことを示す。つまり東京のことばの新しさを示す。

図4で初出年と使用率については、古代と中世の二頂点のある山をなして、初出年が古いほど使用率が高いという傾向が見られる。これは図2の二つのピークを反映している。また奥（距離が遠くて初出の新しい語）に中程度の峰が見えるのは、図3で初出の新しい語が東京付近に多いのと印象と違う。使用率を計算に入れると、初出の新しい語は東京付近に多いが、使用率が低いのである。

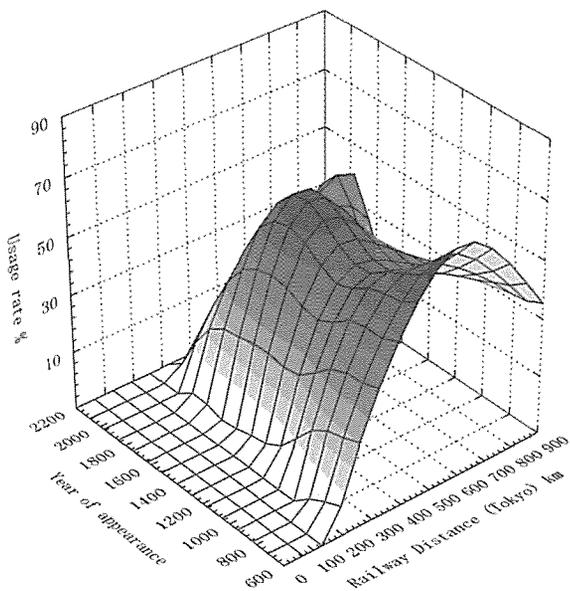
この図4は、これまで2要素ずつ組み合わせて考察した3要素の相互関係全体像を忠実に表し、それ以上の関係も示している。後述のように初出年を基準に古代・近代に分ける手がかりも得られる。以下では全体を2分割して考察する。

3.2. 東西クラスター別3次元グラフ

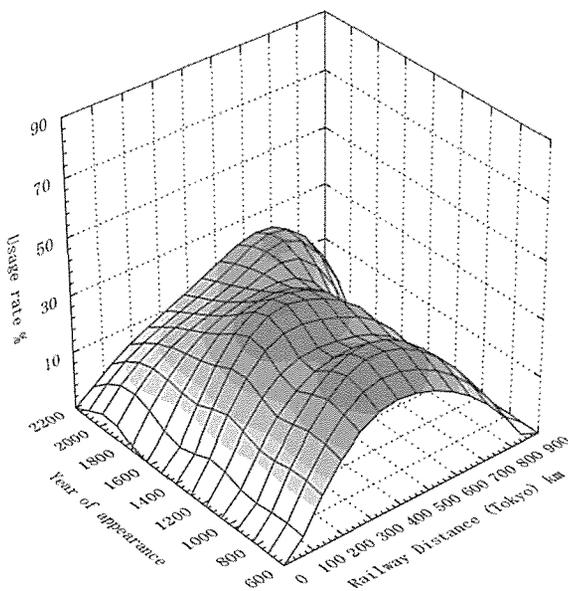
次に、これまで図2～3に線で補入した東西2クラスターとの関連をみるために、全体を東部・西部クラスターに分割して示す。重み付き最小2乗近似によると、図5のように、上の西部クラスターは、（東京からの鉄道距離の遠いところを中心に）図4の全国の様子をよく再現しているし、東部クラスターも、（鉄道距離が近く）使用率は低いながら、図4の全国のパターンに似ている。どちらのクラスターでも古代と中世の二つのピークが見られる。

3.3. 古代近代別3次元グラフ

文献初出年と東西2クラスターは、前述図2のように、きれいな対応関係を見せず、東部西部クラスターが初出年と交差する傾向を見せる。しかし図4、図5で2頂点が現われたのを手がかりに、図2をもう一度見ると、8世紀のピークから下がり、また16世紀前後のピークから下がるという、よく似た二つの下降パターンが読み取れる。14世紀初出の語がないので、ここを境に



西部



東部

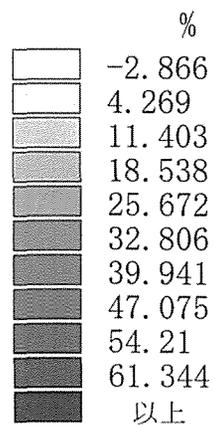
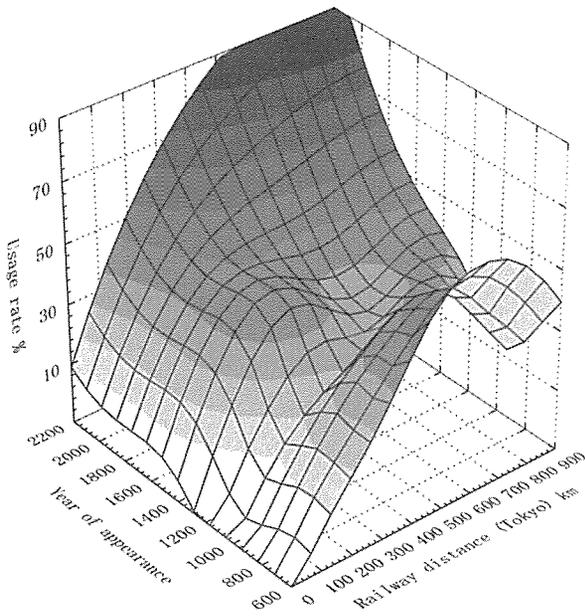
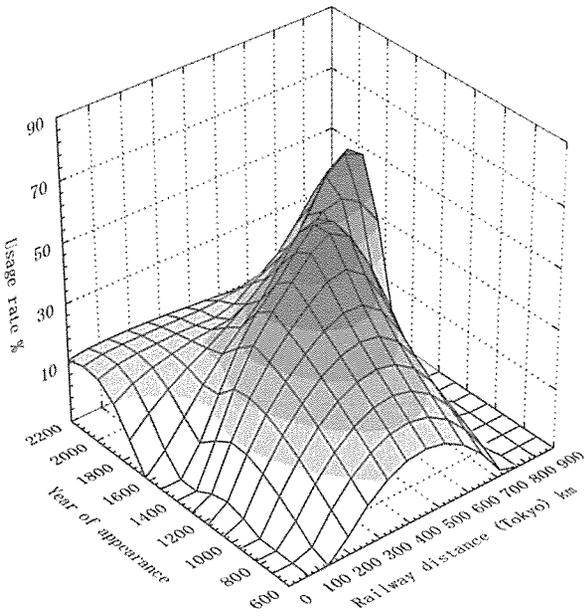


図5 東西クラスター別82項目 3D 鳥瞰図 (重み付き最小2乗近似)



古代



近代

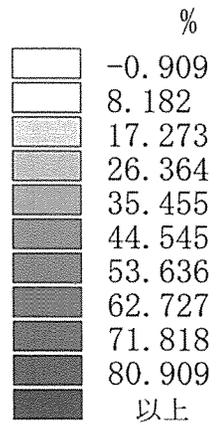


図6 古代近代別82項目 3D鳥瞰図 (重み付き最小2乗近似)

「古代・近代」の二つの時代に分割して、図6を描いてみた。

図6ではやや違うパターンが浮かぶ。重み付き最小2乗近似を適用したために、上の古代初出語のグラフでは14世紀以降の語形（図の左奥）が実際にはないにもかかわらず、近代の遠方に向けて使用率が増えるかのように描かれた。12, 13世紀の語の使用率傾向を延長したためである。しかし古代の山は残っている。また下の近代初出語の図でも、右手前のデータがないにも関わらず、中世ころの山が再現されている。これらデータ不在にもかかわらず表示されている山を除いて見れば、標準語形使用傾向に、古代のピークと中世のピークの二つがあることは、確実である。

3.4. 古代近代別・東西クラスター別3次元グラフ

それでは、東西と古今二つの要素を組み合わせて分割するとどうなるか、図7で、単純化を進めて平面で近似した3D鳥瞰図をみよう¹⁹。

左下の古代東部クラスターだけが、例外的に下面に対してほぼ水平で、鉄道距離も初出年も使用率に影響を与えない。これにくらべると他のクラスターは傾斜のパターンが似ていて、いずれも手前、つまり古代初出で東京への距離が近いところで使用率が高い。左上の古代西部クラスターの近似面は、偶然3Dグラフの傾斜と一致したので、線のように見える。つまり、古代初出、近代初出ともに、古くから出ていた語は使用率が高い。また西部クラスター（上の二つ）については、東京方向に重心のある語は使用率が高い。ことに右の壁面に投影される線は角度が似ている。これをさらに詳しくみるために、再び3Dグラフを離れて、初出年を考察する。

4. 初出年と東西2クラスター・2時代

4.1. 古代・近代別初出年と東西クラスターの使用率

前出の図2では、14世紀を境に二つの使用率ピークが似たパターンを繰り返すように見えた。後述のように年速約1キロ（以下）で伝播が行われれば、15, 16世紀初出の語は21世紀までに国土の中央部には普及する。また8世紀初出の語で地域差を示すものは、むしろ後世中心領域をむしろまかれた可能性がある。つまり古代と近代で標準語形の位置付けが異なる可能性がある。河西データでは14世紀初出の語がないので、ここを境にして、図8を描いた。原図は図2と同じで、語形を省いたものである。「古代・近代」の二つの時代に分けて、また「東部・西部」クラスターに分けた結果を示す。別々に計算した線形近似の線を4本記入し、数式を表1に示した²⁰。

表1 2時代2クラスターの近似直線

古代東部クラスター	$y = 0.0003x + 25.295$
古代西部クラスター	$y = -0.0364x + 83.776$
近代東部クラスター	$y = -0.0319x + 77.499$
近代西部クラスター	$y = -0.0351x + 109.63$

図8左下、古代初出の東部クラスターの語は、図7でみたように、初出年との対応関係を見せない。図3右下（および注6）で具体例が分かるが、多くは例外的な分布で、京都から広がって関

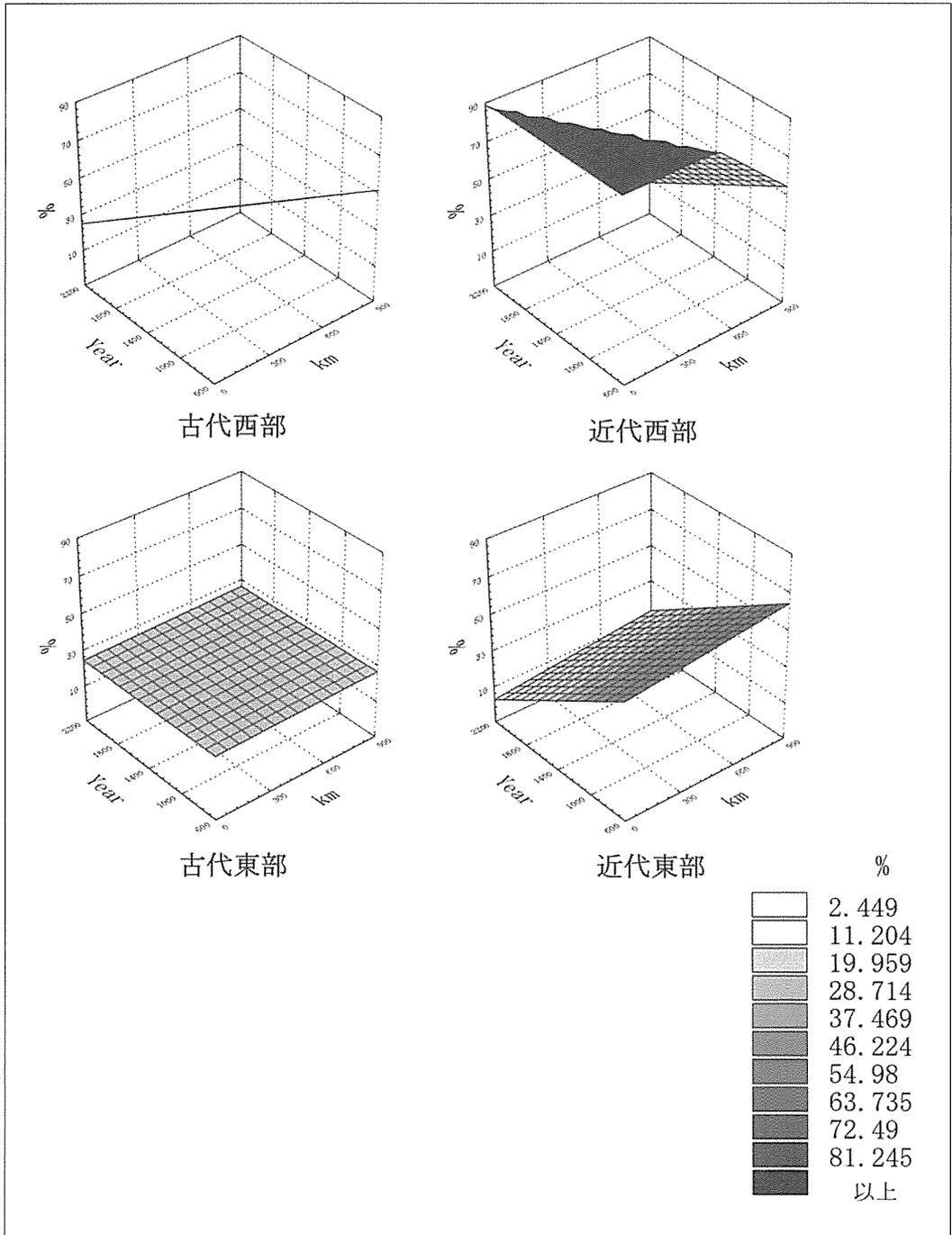


図7 古代近代・東部西部別82項目 3D 鳥瞰図

東に残った古形が、のちに東京中心の標準語形として復活したものである。

しかし他の三つ（古代西部、近代東部、近代西部）の組み合わせでは、初出年と使用率がかな

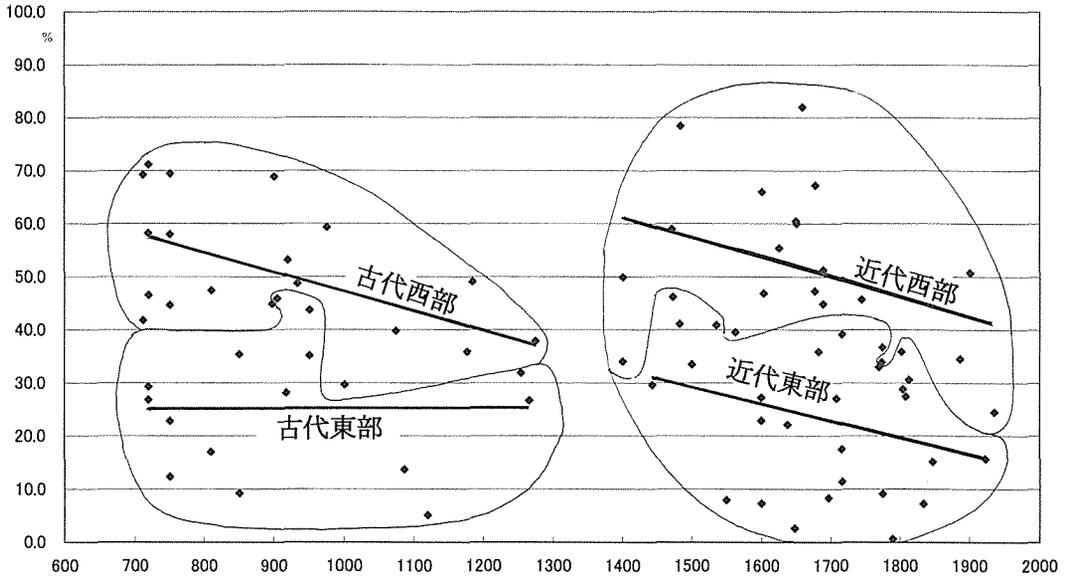


図8 初出年と全国使用率の近似直線（古代・近代+東西）

りの相関を見せ、しかも近似直線の傾斜の共通性が目立つ。表1でも数値が似ていて、1000年につき、32~36%の減少にあたる。換言すれば、100%が0%になるのに3000年かかるような変化・伝播である。日本全土でこの82項目で非標準語形が使われていたと仮定すると、全土が標準語形に置き換わるのに約3000年という計算になる²¹。

図8によれば、古代西部クラスターの線は、そのまま延長すると近代東部クラスターの線に重なる。県別の使用率を鉄道距離重心と対比した結果（井上 2004c, 予定）、西部クラスターは、京都中心にゆっくりと広がるパターンで伝播したと読み取れた。東部クラスターは、江戸・東京から急速に伝播したとみられた。両者の線がつながるのは、日本全体の基本的な動きを反映すると、考えていい。つまり4本の線のうち、図の左上から右下への2本の線が基本的な動き（京都周囲分布と東京周囲分布）で、左下の傾斜のない線は（近畿空洞分布から）復活した古語、右上の線は近代の全国的普及の勢力を、示すと見られる。

いずれにしろ、図8で3本の線が右下への下降を示しているのは、歴史的伝播過程を示す。つまり古く出現した語は時間が経てば遠方に伝播して全国使用率が高まる、新しく出た語はまだ使用率が低い、年数が経てば古く出現した語と同じ過程をたどって広がる、と考えられる。

4.2. 普及の年速と初出年

以上、古代・近代の二つに分けたところ、初出年と鉄道距離重心の関係が浮き彫りにされた。ここで、時代を二分する正当性を検討しよう。図8は、標準語形が一定の普及速度を中心にして全国に波及したことを思わせる。かつての標準語形の伝播は「年速1キロ（以下）」という大まかな捉え方がある（徳川 1993, 1996; 井上 2003b）。京都から鹿児島・青森は約1000キロだから、

1000年（以上）たてば京都の言い方が国土の両端にたどりつくことになる。紀元1000年つまり平安時代の仮名文学隆盛のころの言い方は全国に使われるし、それ以前の奈良時代初出の語はさらに本土の先（沖縄県）まで行き着くことになる。江戸時代初頭1600年代出現の語は、400年間で400キロ程度動くとしたら、京都から長野・広島・高知あたりに達する（その範囲内では100%になる）はずである²²。

個々の分布地図を見ると、8～10世紀初出の標準語形は、その後関西に広がった新しい語形に追われて、関東・中部や中四国などに分かれて、辺境残存分布（近畿空洞分布、周囲論的分布の古形）を示すことが多い。つまり京都からみれば古めかしい言い方が、近代に東京付近で使われていたために標準語に採用されたというパターンである。典型が以前にあげた「つむじ、つむじかぜ」である。

これに対して、おおよそ11世紀以降初出の語は、それ以前の語形（古語辞典に載るような語形）に入れ替わって、今なお全国制覇の過程にあるものが多い。ただし近代初出の語形でも、京都でまたは各地で新たに方言形が生まれたりして、内部が虫食い状態になり、全国制覇に至らない例もある。これらによって、平均の年速は遅くなる。また図8でみたように全国普及に一見3000年かかるかのような数値が得られる。

4.3. 辺境残存分布と近畿空洞分布

以上を踏まえて、各分布図に戻り、また佐藤(2002)によって他の方言語形との張り合い関係を見、他の語形の文献初出年などを検討して、河西データにさらに情報を追加した。地理的分布パターンからみて辺境残存分布を示すと思われる標準語形をマークすることができた。このデータにソートを繰り返す手法で、さまざまな要因の相互関係を探ることができた。

その結果を表2に示した。河西データで初出のない14世紀を区切りにすると、辺境残存分布を示す標準語形の割合に差があり、古代初出語の半分以上になる。古代初出の(19+15=)34語中では半数以上の(10+8=)18語が、(関西で別語形が生まれたために)関東や中四国に辺境残存分布を示す。これに対し、近代初出の(20+28=)48語中で、辺境残存分布を示すのは、(3+4=)7語(1割台)に過ぎない。

表2 辺境残存分布を示す標準語形の割合
(古代近代・東部西部別)

	辺境残存分布	総語数	%
古代西部クラスター	10	19	52.6
古代東部クラスター	8	15	53.3
近代西部クラスター	3	20	15.0
近代東部クラスター	4	28	14.3
計	25	82	30.5

なお表2で分かるように、西部クラスター(19+20=)39語のうち古代初出語は19語で約半分を占めるが、東部クラスター(15+28=)43語のうち古代初出語は15語で3分の1程度にすぎない。

82語中28語、つまり3分の1が近代東部クラスターであることが、響いている。これは河西データでは東日本に古代初出語が少ないことを物語る。

具体的な地理的分布をみると、西部クラスターに属する語が図2で古代と近世の二つのピークを作ったことは、必然である。かつては京都中心に伝播し、奈良時代初出の古い語ほど全国に行き渡るようなパターンだった。近世以前初出の語で全国制覇したものは、地域差がないので、LAJでは調査対象にならなかった。取り上げられたのは、一度各地に普及したあと別の新語形が近畿や各地に生まれて、分布地域を狭めはじめた語である。一部は近畿空洞分布である。関東に圏分布の辺境残存の形で残った語形が、文化的中心地の移動のために、標準語としての地位を得て再浮上したという例もある。これらが古代のピークである。

その後江戸・東京中心の伝播パターンが有力になり、京都から広がった語形を江戸東京で採用して、全国への普及が速まった。近代に伝播速度が速くなり、かつ二重の力が働いたため、中世以降初出の語でも全国に行き渡るようになった。これが近代のピークである。

先にクラスター分析・因子分析の結果と全国使用率のグラフによって、標準語の地理的伝播（普及と衰退）の4つのルートを提示した（井上 2003a）。古代の京都からの普及と衰退、近世・近代の江戸・東京からの普及と衰退が、別のルートを通ることを示した。今回の分析でも同様の結果が得られたことになる。伝播のメカニズムは一つだが、日本語では時代を違えて文化的中心地が移動し、「二都物語」が成立したために、初出年と使用率に関して、複雑な関係ができたものと考えられる。これは、普及の年速約1キロ（以下）、1000年で1000キロ（以下）という仮説と矛盾しない。

5. 結論：鉄道距離の有効性

本稿執筆以前には、東西クラスター別に、クラスターの数を変えながら、考察を続けてきた。図1でも東西クラスターの違いは再現された。しかし、初出年別のグラフでは、地域差がはっきり出なかった（図は略した）。東西クラスターを一緒に集計していたためである。図1のように初出年別をさらに東西に分けることによって、意味のある読み取りができた。

初出年で古代・近代に分けると、意味のある傾向が現れるのではないかという予測は、図2の考察から生まれた。図8にみるように、一定の傾向は出たが、2時代に分ける理論的正当性は、当初なかった。図1で、古代初出の語は（一部の語のために）辺境に残存する傾向を示し、進出一方の近代初出語と区別されることをみたので、4分する根拠が得られたことになる²³。また本稿では省いたが、東京からの鉄道距離重心と使用率との散布図により、古代初出語は近代初出の語にくらべてばらつきが大きいことが分かった。合わせて、普及年速についても、傍証が得られた。

以上2次元グラフと3次元グラフを見た。82項目全体で見て、鉄道距離が使用率と非常に強い関係を持ち、また初出年とも複雑な関連を見せることが、読み取れた。かつての京都からの伝播の過程は、古代西部クラスターで典型的に示された。その後の東京からの伝播は、近代東部クラスターで示された。

いずれにしろ、3次元グラフで近似値を表示することによって、正確な考察ができた。データをいかに処理するかで読み取りが違ってくることの典型といえる。大まかな傾向は分かっていたが、多変量解析法の適用により、具体的な数値をもとに、視覚化できた。

標準語の使用に地理的距離が影響を及ぼすことは見当がついていたが、鉄道距離とその語形ごとの重心が、全国使用率をよく説明することが分かった。先に県ごとの鉄道距離と標準語形平均使用率との関係を考察して、全体としても、クラスター分析の結果を組み合わせても、日本語標準語の歴史的伝播パターンの変遷を読み取ることができた。本稿では、語形ごとの一般傾向を読み取ったことになる。

注

- 1 「日本言語地図」段階の「標準語形」は、近代になって標準語として採用された語形である。伝播の当初には他の方言形と同じ資格で、競い合う形で京都などから広がったのだろう。近代の標準語と同じ語形が偶然各地の方言で使われていた、と考えるべきである。以下で言及する徳川(1993, 1996)の京都から広がった語形と重なるものもある。各地の中心地から広がった方言形と、伝播パターンが違うだろうことは井上(2003b)で指摘したが、本質的な違いはないと思われる。しかし、近代・現代のマスコミを通じて広がる新語・流行語の類とは明らかに異なっている。
- 2 82行(82語)×48列(47都道府県+伊豆諸島)の行列の形をなしており、各セルには標準語形使用率が示してある。例えば京都府36地点中24地点で標準語形を答えていれば使用率は66.7%である。なお併用として答えられても1地点と数えてある。
- 3 井上(2004b, 2004c, 予定)で論じた。またその後の変化も少ない。『日本言語地図』段階のデータ解釈には新幹線による変化は無関係である。
- 4 計算法・問題点などについては井上(2004c)を参照。ここでは東京からの鉄道距離のみを扱った。京都からの鉄道距離は井上(2004c)、Inoue(in print)などで扱ったが、古い時代の伝播の様相をよりよく示す傾向がある。
- 5 基本部分はInoue(2002)で口頭発表した。Inoue(2004)はその増補改定版である。
- 6 図2、図3の語形との照合の便のために、全82標準語形を掲げる。東西クラスターに分け、かつ古代・近代に分けて、初出年の古い順から配列する。他に井上(2001)、河西(1981)、河西・真田(1982)などに様々な順に並べたリストがある。古代初出西部クラスター：にる、いえ、あまい、こおる、こおり、やる、くちびる、すみれ、ぬか、なめくじ、かお、まつかさ、あぜ、からい、つらら、けむり、とさか、まむし、しあさって。近代初出西部クラスター：よだれ、つゆ、たてがみ、うそ、うろこ、こげくさい、かまきり、あご、かほちゃ、とんぼ、ふけ、かくれんぼ、たこ、ふすま、しもやけ、さつまいも、つば、びっくりする、じゃがいも、ちゅんちゅん。古代初出東部クラスター：いる、すてる、かぞえる、つむじ、つむじかせ、おそろしい、おんな、におい、とかげ、いくら、おととい、〈塩味が〉うすい、しおからい、せきれい、つくし。近代初出東部クラスター：きのこ、だ、なす、ほうほう、せともの、まぶしい、かたぐるま、かかし、かたつむり、とうがらし、とりおどし、とげ、とうもろこし、どくだみ、ものもらい、くすりゆび、かかと、いい、もぐら、おいしい、やのあさって、おたまじゃくし、もみがら、おにごっこ、なのか、おてだま、うるち、すっぱい。

- 7 因子分析結果のグラフに縮小した地理的分布図を82枚貼りこんだ図でも分かる (Inoue 2004: Fig5)。
- 8 中国地方がピークになる標準語形は、以前の分析で鉄道距離との散布図を描いたときにも得られた。「近畿空洞分布」にあたる。徳川(1993)で見られるような近畿発の非標準語形が地理的に進出したためである。井上(2003a)図3で全国分布パターンからの退縮ルートを描いたが、ルートを追加すべきといえる。
- 9 これは、古代初出の語は東西クラスターともに昔はもっと使用率が高かったことを示唆する。図8左側の2本の近似線はかつてはもっと上に位置し、右の2本の近似線の延長の位置にあったと考えることもできる。(古代東部の線の傾斜はその後ほぼ水平になった。)
- 10 東部西部と古代近代の分類を組み合わせることによって、有意義な結果がえられた。ある社会で、男女で分けても年齢で分けても髪の長さの傾向をつかめなかったが、両方を組み合わせることによって、若い男性と老いた女性に長髪が多いことが分かった、という分析に似ている。
- 11 プラハの国際言語学者会議では、この数値を用いて発表した(Inoue in print)。なお人口重心の位置としては都道府県庁所在地を用いた (井上 2004c)。そこに各都道府県のインフォーマントが集合したと仮定して、標準語使用率を計算したことになる。地理学で人口重心を計算するときにも、市町村役場所在地を計算基準にするが、自治体の実際の人口分布は偏っていることがある。なおインフォーマント数に基づく標準語形使用率に都道府県の人口をかけて計算しなおした結果は、語形ごとに井上(2002)に提示した。人口を考慮しない分布図よりは、全体が東に寄る傾向が見られたが、全体の分布パターンはくずれない。標準語形を弁別し、かつ各都道府県を性格づけようという本研究にとっては、インフォーマントの数や人口を考慮してまで計算を複雑にする意義は認められない。
- 12 東京発と京都発の鉄道距離を別に計算したので、2個の数値になった。ただし本稿では東京からの距離のみを提示する。
- 13 東京基準の鉄道距離重心と全国使用率のグラフは、井上(2004b), Inoue(in print) で考察した。
- 14 宮島(2002)によれば、『日本国語大辞典』(第二版)における基本語1000語の初出年は、半数近くが8世紀以前である。その後はほとんど変動がないが、14世紀前後には初出語がやや減る。また『国語学大辞典』『国語年表』をもとに世紀別文献数グラフを作ったところ(未発表)、世紀を追って国語文献数が増えるが14世紀ころに一時減る。中世に資料が少なくなったのは確からしい。
- 15 なお河西データの文献初出年は、『日本国語大辞典』(初版)により1980年代に入力してあった。『日本国語大辞典』(第二版)により2002年に調べなおしたところ、初出年が早くなった語が19語あった。大部分は1, 2世紀の違いで、数世紀もさかのぼる語はまれである。逆に初出の(わずかに)遅くなる語は8語あった。計27語。82語中3割強で変動があったことになる。また個々の研究文献でさらに早い初出が指摘されてもいる(本稿には反映させられなかった)。従って、以下の初出年の扱いには、用心・保留が必要である。ただし、後述のように古代・近代に二分するが、この区分が(近代から古代に)変わった語は2語(イクラとツララ)だけである。
- 16 徳川の2回の研究手法は違い、数値も違う。徳川(1993)では0.93km/y、徳川(1996)では0.62km/yだった。その理由については、井上(2003b)で論じた。
- 17 前述のように、本稿では京都からの鉄道距離は扱わない。
- 18 本稿では省略した。Inoue(in print) 参照。

- 19 重み付き最小2乗近似を適用すると、古代・近代について実在しない山まで極端なパターンを形成するので、ここには提示しない。
- 20 図8で線に囲まれた4個の領域に属する語形について、別々に最小2乗近似によって計算した。
- 21 ただし、日本祖語の状態で標準語形以外を100%使っていたとは考えられない。またある地域の使用率が0%にならなくとも、「空からばらまいた」ような分布だと、そこに標準語が使われていると、ふつうとらえる。標準語形は1県を完全に覆ったあと次の県に伝播するわけではない。某県で100%になった段階で、離れた県に数%であれ伝播していることもある。さらに様々な伝播パターンの語を一緒に扱ったために、近似線の傾斜がゆるやかになったという事情も考えられる。したがって、標準語形が全土に伝わるのに3000年かかるわけではなく、もっと速く(年速1キロに近い速度で)伝わりと考えられる。
- 22 平山(1992~94)をもとに、標準語形の全国での使用状況と初出年の関係をみたところ、初出が15世紀以降でも全国的に使われる語があることが分かった。この辞典はコンピュータ可読データになっているので、編者たちによる調査が待たれる。年速の速い語では1000年かからずに全国に普及する。1000年以上経った語ではむしろ別の語形に置き換わる過程が問題になる。これが図2、図8で古代初出語と近代初出語とが一見繰り返しのパターンを示した理由である。
- 23 古代初出の語が京都付近に少なく、近畿空洞分布を示すことは、井上(2001:第8章)でも見られた現象である。

引用文献

- 井上史雄(2001)『計量的方言区画』明治書院。
- 井上史雄(2002)「方言区画論の再生」日本方言研究会編『21世紀の方言学』73-86, 国書刊行会。
- 井上史雄(2003a)「文献初出世紀と方言分布重心にみる標準語の歴史」『国語学』54(2), 59-74, 国語学会。
- 井上史雄(2003b)『日本語は年速1キロで動く』講談社。
- 井上史雄(2004a)「併用現象と言語変化の中間段階—河西データ3クラスターの普及過程—」『語研論集』9, 1-20, 東京外国語大学。
- 井上史雄(2004b)「『隣のことは』の近接効果—社会言語学における距離—」『月刊言語』33(9), 24-30, 大修館書店。
- 井上史雄(2004c)「標準語使用率と鉄道距離にみるコミュニケーションの地理的要因」『社会言語科学』7(1)(方言特集号), 1-11, 社会言語科学会。
- 井上史雄(予定)「標準語形普及の3段階—鉄道距離と4クラスター別標準語形使用率—」『言語研究』, 日本言語学会。
- 奥野隆史(1977)『計量地理学の基礎』大明堂。
- 河西秀早子(1981)「標準語形の全国的分布」『言語生活』354, 52-55, 筑摩書房。
- 河西秀早子・真田信治(1982)「『日本言語地図』による標準語形の地理的分布」『日本語研究』5, 34-47, 東京都立大学。
- 国立国語研究所(1966~1974)『日本言語地図1~6』大蔵省印刷局。
- 佐藤亮一(2002)『お国ことばを知る 方言の地図帳』小学館。
- 徳川宗賢(1993)『方言地理学の展開』ひつじ書房。
- 徳川宗賢(1996)「語の地理的伝播速度」『言語学林1995-1996』893-909, 三省堂。

- 西岡秀雄(1956)『人文地理学 下』桜書院.
- 平山輝男編(1992～94)『現代日本語方言大辞典』明治書院.
- 宮島達夫(2002)『『日本国語大辞典』(第二版)における初出文献の改訂』『近代語研究』11, 638-619, 武蔵野書院.
- 山口幸洋(1982)『方言から見た東海道』秋山書店.
- Inoue, Fumio (2002) Multivariate analysis, geographical gravity center and the history of the standard Japanese, the Symposium of “*Measuring lexical variation and change*”, Leuven.
- Inoue, Fumio (2004) Multivariate analysis, geographical gravity centers and the history of the standard Japanese forms, *Area and Culture Studies* 68, 15-36, 東京外国語大学.
- Inoue, Fumio (in print) Geographical distance center and multivariate analysis of the standard Japanese, *Proceedings for ICL 17 Prague (International Congress of Linguists)*, CD-ROM.
- Inoue, Fumio (forthcoming) Geographical distance center and rate of diffusion of standard Japanese, *Proceedings of ICDG 4 Riga (International Committee of Dialectologists and Geolinguists)*.
- Kawaguchi, Yuji & Fumio Inoue (2002) Dialectology in Japan in historical perspectives: PART II Historical characteristics and geographical distribution of standard Japanese forms, *Revue Belge de Philologie et d'Histoire* 80, 816-829.

初校に際して：図4～7はもともとカラーであったが，編集委員会の要請で，白黒となった。また（日英語別の）文献配列，その他も，編集委員会の方針に従った。注が多くなったのは，編集委員諸氏の細かい指摘に応えたためである。

(投稿受理日：2004年3月15日)

(改稿受理日：2004年8月26日)

井上 史雄 (いのうえ ふみお)
東京外国語大学外国語学部
183-8534 東京都府中市朝日町3-11-1
innowayf@nifty.com

**First year of appearance of standard
Japanese forms and gravity center
by railway distance:
East and west clusters in 3D distribution graphs of railway
distance, usage rate and first year of appearance**

INOUE Fumio

(Tokyo University of Foreign Studies)

Keywords

linguistic geography, standard Japanese, computational dialectology,
railway distance gravity center method, first year of appearance

Abstract

In this paper two techniques of simplification are attempted to represent geographical distribution patterns of standard Japanese. The first one is a representation of two-dimensional geographical distribution patterns by one dimension. The second one is plotting the geographical locations making use of railway distances from cultural centers. By simplifying the two-dimensional geographical distribution into one by railway distance, another dimension can show the average percentage of usage of standard Japanese forms for each prefecture. “Kasai data”, or numerical data of the “Linguistic Atlas of Japan”, was utilized, and columns of the word-form were analyzed. Gravity centers of railway distance were calculated for each standard Japanese form, and correspondence with the nationwide rate of usage and the first year of appearance in historical documents was considered. Two-dimensional graphs (scattergrams) of the two factors were considered, and also three-dimensional graphs were analyzed.

Division of the data into Archaic and Modern words by means of first year of appearance and also division into eastern and western clusters were attempted. Archaic Eastern words did not show any correspondence with first year of appearance. However, the other three categories showed fairly good correspondence with first year of appearance. This shows that the newly appearing words from a cultural center have lower percentage of usage at first, but increase its nationwide usage rate by taking the same steps with words which had spread before. However, words which had appeared early in Archaic era may later be eroded by newer dialectal forms.