

国立国語研究所学術情報リポジトリ

『日本言語地図』のデータベース化と計量的分析：
併用現象，標準語形の分布と交通網，方言類似度の
観察

メタデータ	言語: jpn 出版者: 公開日: 2020-03-18 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://doi.org/10.15084/00002695

『日本言語地図』のデータベース化と計量的分析

—併用現象、標準語形の分布と交通網、方言類似度の観察—

熊谷 康雄

(国立国語研究所)

1. はじめに

『日本言語地図』データベース（以下、LAJDB）は、1999年頃より本格的にデータベースの構築に取り組み始めたもので、この3年間は国立国語研究所共同研究プロジェクト「大規模方言データの多角的研究」の柱の一つとして、データベースの整備と分析を進めた。本稿では、まず、『日本言語地図』のデータベース化の概要を簡単にまとめてLAJDBの特徴を説明する。次に、このLAJDBのデータの一部を利用し、LAJDBを活用した『日本言語地図』の計量的な探索、分析の試みを報告し、LAJDBの可能性の一端を示したい。

国立国語研究所編『日本言語地図』（全6巻、収録地図300枚、以下LAJ）は、方言研究の基盤的な資料のひとつであり、広く研究に使われてきている。全国規模の日本初の本格的な言語地理学的調査による言語地図であり、1966年から1974年にかけて刊行された。調査項目は語彙を中心とし、調査は1957年から1965年にかけて臨地調査により行われた。調査項目数285（地図化された項目240項目）、調査地点数2400、昭和30年代の日本全国の方言の地理的分布を捉えた方言研究における基礎資料である。LAJの調査開始から数えて今年（2013年）で57年目になる。LAJの原資料であるカード（約54万枚）は国立国語研究所に保管されている。

LAJの刊行期間はコンピュータの利用が普及する以前の1966年から1974年であり、すべて紙と手で行われた。これまで、LAJを用いた計量的な分析は、多くの場合、刊行された地図に基づいて、個々の研究者が、個別に必要な部分を地図から数えて直接集計したり、必要な範囲のデータを作成して集計・分析したりしたものがほとんどであった。利用可能なデータの制約は、研究の方法にも影響する。例えば、計量的な分析に関しては、全国的な分析では、県を集計の単位としたものが多いことや、地点を単位として集計したものは地域が限定されていることなどにも影響していると思われる。全国規模の等質的な方言分布データとしてLAJの持つ可能性を発揮させるには、LAJの全体に渡る一貫したデータ化が必要である。

2. 『日本言語地図』データベース（LAJDB）

LAJDBは、資料の保存と利用の高度化を目指すもので、1999年頃より本格的にデータベースの構築に取り組み始め、データベース科研の補助を受けて構築を進めた（熊谷2007）。前述のように、この3年間は、この国立国語研究所共同研究プロジェクト「大規模方言データの多角的研究」の柱の一つとして、LAJDBの整備と分析を進めてきた。LAJDBの構築では、原カードの画像のスキャンを工程の第一段階としており、これまでに全体の9割近くまでカードの画像データ化を進めたが、ここ3年間では、特に、既に原カードの画像化の済んでいる項目をデータベースとして使えるようにするための作業の部分に注力して、

利用可能な項目を増やすことに努力した。LAJとして地図化されている240項目のほぼ半数である119項目を公開できる段階に来たところである(表1, 表2)。

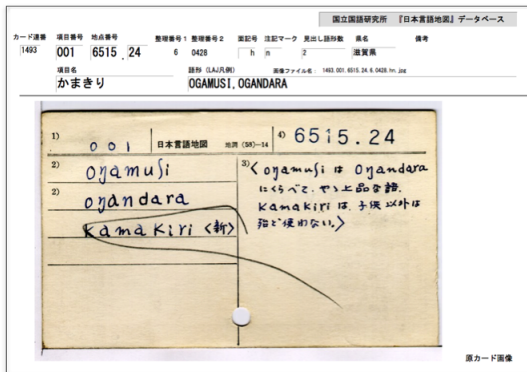


図1 LAJDB 画像データベース：画面例

The screenshot shows an Excel spreadsheet with columns labeled A through L. The data includes item numbers, codes, and names. The first few rows are:

1	001	日本語地図	6515.24										
2	001	Ojamura											
3	001	ogandara											
4	001	Kama Kiri (新)											

 The spreadsheet continues with many more rows of similar data.

図2 LAJDB 文字データ (Excel 形式2)

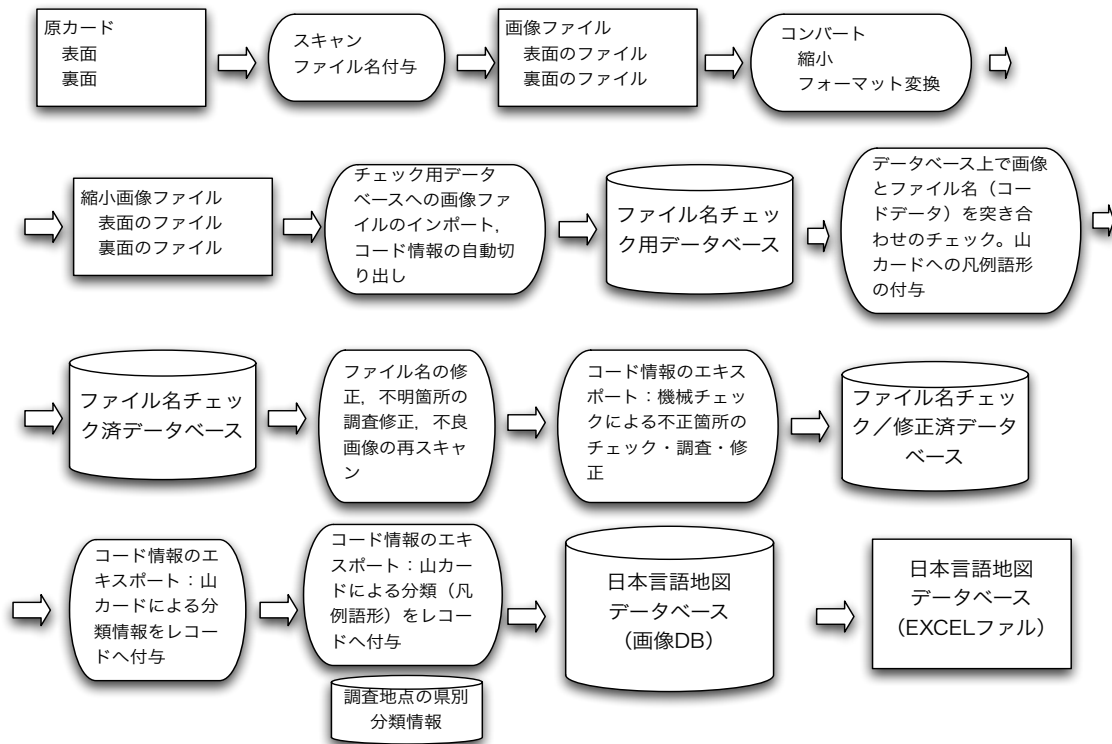
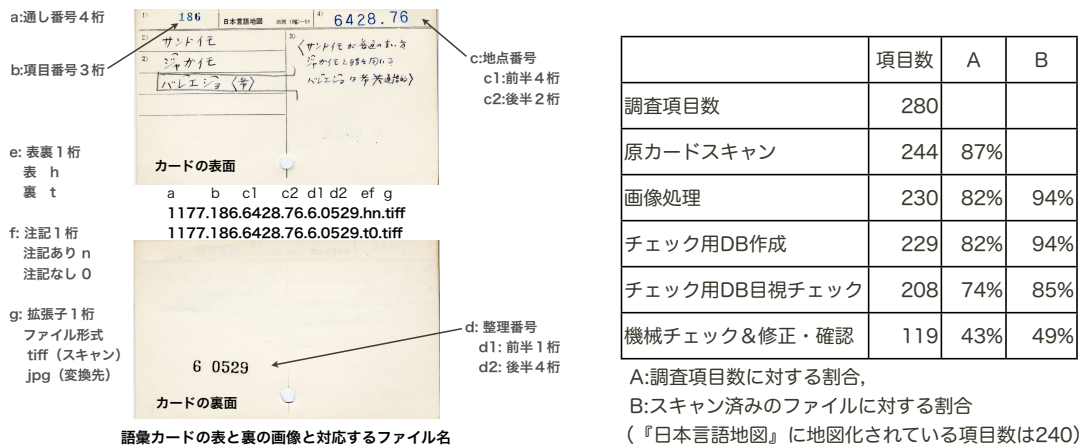


図3 LAJDB：電子化・データベース化の工程 (概念図)

LAJDBの構築では、個々の地点の項目毎の回答を記した原資料のカードを画像データ化し、また、刊行されたLAJの地図上に示された語形の地理的分布情報は文字データとしてデータ化している。LAJDBの画像データベース(図1)はこれらを統合して構築した。文字データは画像データベースとは別途に独立したデータ(図2)としても利用できるようにエクセルファイルも作成してある(データ項目は、図2の形式2[複数回答の場合には回答語形毎に1レコードを作成した形式]では(1)通し番号、(2)項目番号、(3)地点

番号, (4) 整理番号, (5) 項目名, (6) 県名, (7) 凡例の語形, (8) 併用語形数, (9) 語形の併用パターンである)。

表 1 『日本語地図』 データベースの構築状況



	項目数	A	B
調査項目数	280		
原カードスキャン	244	87%	
画像処理	230	82%	94%
チェック用DB作成	229	82%	94%
チェック用DB目視チェック	208	74%	85%
機械チェック&修正・確認	119	43%	49%

A:調査項目数に対する割合,
B:スキャン済みのファイルに対する割合
(『日本語地図』に地図化されている項目数は240)

図 4 原カード情報のデータベース化

表 2 LAJDB の処理済み項目一覧

LAJ_Q001: かまきり	LAJ_Q083: たけうま	LAJ_Q165: まわた	LAJ_Q233: くるぶし (課)
LAJ_Q005: かたつむり	LAJ_Q089: かくれんぼ	LAJ_Q166: いと	LAJ_Q235: すてる
LAJ_Q006: なめくじ	LAJ_Q090: おかね (貨幣)	LAJ_Q167: きぬいと	LAJ_Q237: おそろしい (恐ろしい)
LAJ_Q007: おたまじやくし	LAJ_Q091: おつり (釣銭)	LAJ_Q169: はたいと (機糸)	LAJ_Q238: なのか (七日)
LAJ_Q012: とかげ (蛭蝸)	LAJ_Q092: かぞえる (お金を)	LAJ_Q173: こめ (米)	LAJ_Q239: ここのか (九日)
LAJ_Q031: あたま	LAJ_Q095: やる	LAJ_Q174: うるち	LAJ_Q240: ひまご
LAJ_Q032: つむじ	LAJ_Q103: きとう	LAJ_Q176: はんまい (飯米)	LAJ_Q241: やしやご
LAJ_Q036: ものもらい	LAJ_Q104: おととい	LAJ_Q179: ぬか (糠)	LAJ_Q244: いえ (家)
LAJ_Q037: はな (鼻)	LAJ_Q105: さきおととい	LAJ_Q182: あぜ	LAJ_Q248: ふすま
LAJ_Q038: におい (芳香)	LAJ_Q108: あした	LAJ_Q184: とりおどし	LAJ_Q250: きれいだ
LAJ_Q039: におい (悪臭)	LAJ_Q109: あさって	LAJ_Q185: かかし	LAJ_Q252: とうがらし
LAJ_Q042-7: においをかぐ (前部分)	LAJ_Q110: しあさって	LAJ_Q186: じゃがいも	LAJ_Q253: おいしい
LAJ_Q042-8: においをかぐ (後部分)	LAJ_Q111: やのあさって	LAJ_Q187: さといも	LAJ_Q261: おんぶする (幼児を負う)
LAJ_Q047: くちひる	LAJ_Q113: あしたのばん	LAJ_Q189: さつまいも	LAJ_Q264: かつぐ (材木)
LAJ_Q048: した	LAJ_Q116: つき	LAJ_Q190: とうもろこし	LAJ_Q265: かつぐ (天秤棒)
LAJ_Q051: うすい	LAJ_Q117: あめ (雨)	LAJ_Q191: かぼちゃ	LAJ_Q266: かつぐ (二人で)
LAJ_Q052: あまい	LAJ_Q118: つゆ (梅雨)	LAJ_Q192: すみれ	LAJ_Q268: いる
LAJ_Q056: ほほ	LAJ_Q119: ゆうだち	LAJ_Q194: つくし	LAJ_Q280: カガミ (鏡) の-Gの音
LAJ_Q057: かお	LAJ_Q122: いなずま	LAJ_Q195: すぎな	LAJ_Q282: なす
LAJ_Q059: あざになる	LAJ_Q124: にじ	LAJ_Q200: とげ (いばら等の)	LAJ_Q284: とんぼ (蜻蛉)
LAJ_Q060: ほくら (小)	LAJ_Q125: ゆき	LAJ_Q214: おうま	
LAJ_Q061: ほくら (大)	LAJ_Q127: こおる: 水が凍る	LAJ_Q215: めうま	
LAJ_Q063: おやゆび	LAJ_Q129: つらら	LAJ_Q216: こうま	
LAJ_Q064: ひとさしゆび	LAJ_Q131: ごみ (目にはいるもの一塵)	LAJ_Q217: たてがみ	
LAJ_Q066: くすりゆび	LAJ_Q132: ごみ (掃除の対象一塵芥)	LAJ_Q218: うし	
LAJ_Q067: こゆび	LAJ_Q134: ごみ (川のごみ一塵芥)	LAJ_Q219: おうし	
LAJ_Q068: しもやけ	LAJ_Q135: じしん	LAJ_Q220: めうし	
LAJ_Q072: すわる (座る)	LAJ_Q148: たく	LAJ_Q221: こうし	
LAJ_Q074: あか (垢)	LAJ_Q149: いる	LAJ_Q222: もうもう	
LAJ_Q075: ふけ	LAJ_Q153: ゆけ (飯の場合)	LAJ_Q223: もぐら	
LAJ_Q076: うろこ	LAJ_Q155: すりばち	LAJ_Q228: すずめ	
LAJ_Q080: おとこ	LAJ_Q156: すりこぎ	LAJ_Q229: ちゅんちゅん	
LAJ_Q081: おんな	LAJ_Q157: せともの	LAJ_Q231: はげる (禿げる)	

LAJDB では, 言語地図の電子データと原カードの画像データの両者があることにより, 凡例の形に統合される前の個々の語形の前表記, 注記などの地図化されなかった情報や併用処理の記録などにも簡単に触れることができる。LAJとして分類整理された情報の原情報に戻ることや, LAJには載っていない調査の情報に容易に触れることができる。LAJDBの工程の概念図を図3に示す。この他, 付随データ, 関連データの整備も行っている。

LAJDB構築にあたっての基本的な考え方は熊谷(2007)に述べた。現在は, 最終的な処理の工程や機械チェックの方法, 文字データの形式など, 当時と変更点もあり, また, 最終的な機械チェックと修正・確認の過程では, 作業の中で, 新たに手順を整理した点多々あるが, 根本的な考え方は熊谷(2007)に示してあるので, 本報告書に再録した(p.165)。

LAJDB の詳しい解説は、「『日本言語地図』データベース解説」として、作成中である。

3. 『日本言語地図』データベースの計量的分析

LAJ の計量的分析としては、LAJ の県別の標準語形使用率を集計した河西データ（河西 1981）に多変量解析や計数的な手法などによる分析を加えた井上史雄による一連の多角的な研究（井上 2001 など）がある。現在の LAJDB の 119 項目段階では、河西データ 82 項目と一致するのは 52 項目であり、河西データを用いた井上の分析と十分な比較をするには、さらにデータの増補が必要な段階である。LAJDB のデータが揃えば、河西データでは対象としなかった項目も含めた全体的な位置付けを考えることもできる。今回の分析は、暫定的、探索的な性格のものである。

本発表では、LAJDB を利用して、LAJ における語形の併用現象、併用処理、標準語形の分布、語形の共通度から見た調査地点間の類似度による領域などを取り上げ、相互関係や交通網（道路網）などとの関係を見ることによって、LAJDB を探索的に扱った。

4. 『日本言語地図』における併用現象の地点別集計の地理的分布

語の併用は言語の接触、伝播、変化に関わる現象である。言語地理学な分析においても、併用は、重要な情報のひとつであり、併用回答における語の新古や使い分けに関するインフォーマントの判断の情報は分析の重要な手がかりの一つであり（柴田 1969）、また、そのような判断の分布の様相自体が言語の動態を考える上で興味深い研究の対象と言える。

LAJ における併用について、稲垣（1980）、井上（2004）がある。稲垣（1980）はいくつかの項目の LAJ の地図で言語接触と併用の在り方を具体的に考察した。井上（2004）では、併用を言語変化の中間段階とし、この併用自体を考察する意味を主張し、標準語の普及過程に併用を位置づけ考察している。また、LAJ における「併用処理」についても言及している。回答地点が 1 地点しかない語形は「孤例」（徳川 1973、澤木 1988 など）と呼ばれ、併用との関係でも興味深い。徳川（1973）は『日本言語地図』における孤例を扱う中で、その関係において併用に触れている。

LAJ では「併用処理」が行われている。「併用処理」とは「同一地点から二個以上の回答があつて、一方の標準語と一致する回答に、共通語的である・新しい・上品である・改まった場合に使う・まれにしか使わない、およびこれに準ずる説明がある限り、原則としてその回答を地図に記載しなかった」（LAJ 第 1 集解説書 方法 p.33）というものである。

稲垣（1980）の中で、井上史雄氏は、LAJ の併用処理や、調査員の差などの論点を上げ、編集当時、いわゆる標準語形、共通語形について、「共新上希の注のあるものは、ある県とか地方にかたまつて、それ自体面白いということが分かっている、それは載せなかったはず」と、編集時に併用に関して、その分布、地域的な違いが意識されていたことに触れている。また、同時に調査員による標準語形、共通語形の調査における扱いの違い、個人差の可能性についても指摘している。佐藤（1986）は、共通語形の俚言形地域への伝播、侵入の分析との関係で併用処理を取り上げ、「スッパイ」を事例として、併用処理された地点を復活させると、地図上で共通語形の伝搬の傾向がより明瞭になったとしている。

LAJ の併用を多数の項目について集めてみるような分析は、これまでほとんどないようである。小西（2007）では方言文法全国地図（GAJ）に関して回答数の分析をしている。

小西（2007）では、GAJにおける回答数の地理的な分布には、調査者との関係が強く、地理的に有意なパターンは見いだせなかったとしている。

LAJDB のデータを使い、いくつかの項目について個別に併用の起きている地点の分布を見たところ、交通網との関係などで意味のある分布が期待できる観察がされた。具体的な内容に踏み込む前に、併用回答の地点の現れ方に地理的な分布の傾向が現れるのかどうかを探った。LAJ では調査期間中に項目の削除や追加があり、調査地点数が異なる項目がある。データの集計作業時点での LAJDB から調査地点数がほぼ 2400 あった 55 項目（表 3）で暫定的に集計し、地点毎の語形が併用になっている項目の数の分布を観察した。

併用の回答のあった項目の数を各地点について集計し、地図に示したものが図 5 である。全国的にばらつくようなことはなく、地理的な分布を見ることができる。ほぼ県単位で分担されている調査員の分布についても地図化し、図 5 と比較したが、併用現象の分布には調査の分担区域を超えたひろがりでも連続する分布がある様子が観察され、また、同一調査員による調査地域内でも、後述する道路網などとの関係をはじめ、有意な関係の中で分布を読み取ることができ、単なる調査員の差ではないことが確認できる。

5. 『日本語地図』における併用処理項目数の地点別集計の地理的分布

LAJ には上で述べた併用処理により除かれている併用がある。この〈共新上希〉の併用はどのように分布しているか、これらは、図 5 の分布に対して、どのような分布を示すか。

LAJDB では画像を参照して併用処理を確認することができるが、併用処理のデータ化には、国立国語研究所に保管されている『日本語地図注記一覧』を利用した。これには、「A：普通注記」、「B：除いた共通語」、「C：除いた特殊語」に分けて（B,C を区別していないものもあるが）記録が残されており、編集作業時の判断の記録として扱える。この情報をデータ化し、LAJ の併用と同様に集計して地図上に描いた。B と C を区別せずに地点毎に除かれた併用の項目の数を集計し地図にしたものが図 6 である。結果は、図 5 に類似し、パターンの骨格的な部分に領域を狭めながら図 5 の分布の中にはまるような分布を見せる。LAJ の併用語形が削除される処理の前の併用の分布を見るとすれば、この図 5 と図 6 を合わせたものが対応することになる。

さらに標準語形に関する「併用処理」のみの分布をみたものが図 7 である。上の「B,C」を区別していないものについて、注記一覧および必要に応じて LAJDB の原カード画像を参照し、「B」、「C」を切り分けたデータを作成した。この図 7 では、LAJ において併称処理との関係で認定されている標準語形がひとつになっている項目を選択して取り上げた。項目によっては、複数の標準語形を認定している場合があり、併用処理がなされていない項目もある。55 項目（表 3）から、標準語形が LAJ の解説で複数認められている項目（表 3 の○印）と標準語形が LAJ の解説では特定されず、併用処理も行われていない項目（表 3 の□）を除き、標準語形がひとつ認定されている項目数は 42 項目ある。ここから、注記一覧がないものや併用処理がなされていない項目を除いた 37 項目に関して、各地点において標準語形が併用処理された項目数の分布を地図化したのが図 7 である。なお、個々の項目の標準語形と併用処理の分布も有意な分布を見せているが、ここでは全体をまとめて傾向を扱う。（なお、作成した個々の項目毎の標準語形と併用処理の分布図の一部は、本報告書所載の小林・熊谷の報告に掲載してある。）

表3 『日本語地図』の併用地点数の項目別集計 (B/A 順) 55項目

	質問項目	(A)調査地点数	(B)併用地点数	B/A*100 [%]		質問項目	(A)調査地点数	(B)併用地点数	B/A*100 [%]
○	Q186: じゃがいも	2400	724	30.2		Q124: にじ	2399	235	9.8
	Q038: におい (芳香)	2400	546	22.8		Q223: もぐら	2399	229	9.5
○	Q122: いなずま	2400	502	20.9		Q129: つらら	2399	227	9.5
	Q118: つゆ (梅雨)	2400	492	20.5		Q066: くすりゆび	2400	216	9.0
	Q048: した	2399	490	20.4	○	Q165: まわた	2400	216	9.0
	Q057: かお	2400	484	20.2		Q076: うろこ	2400	212	8.8
○	Q005: かたつむり	2399	428	17.8		Q179: むか (糠)	2400	211	8.8
	Q191: かぼちゃ	2399	426	17.8		Q149: にる	2400	209	8.7
	Q219: おうし	2400	422	17.6	○	Q001: かまきり	2400	196	8.2
○	Q056: ほほ	2400	412	17.2	○	Q012: とかげ (蜥蜴)	2400	196	8.2
○	Q187: さといも	2399	408	17.0	○	Q188: さつまいも	2400	195	8.1
	Q116: つき	2400	406	16.9		Q200: とげ (いばら等)	2400	189	7.9
	Q119: ゆうだち	2399	384	16.0	○	Q127: こおる: 水が凍る	2400	186	7.8
□	Q104: おととい	2400	360	15.0		Q194: つくし	2400	186	7.8
	Q007: おたまじゃくし	2400	356	14.8	□	Q105: さきおととい	2400	183	7.6
	Q221: こうし	2400	354	14.8		Q006: なめくじ	2400	171	7.1
	Q220: めうし	2400	344	14.3		Q032: つむじ	2400	171	7.1
	Q185: かかし	2400	333	13.9		Q060: ほくろ (小)	2400	169	7.0
	Q051: うすい	2400	325	13.5		Q052: あまい	2400	146	6.1
	Q036: ものもらい	2400	312	13.0		Q083: たけうま	2400	139	5.8
	Q072: すわる (座る)	2399	311	13.0		Q148: たく	2400	132	5.5
	Q059: あざになる	2400	307	12.8		Q089: かくれんぼ	2399	130	5.4
	Q039: におい (悪臭)	2399	292	12.2		Q067: こゆび	2400	130	5.4
	Q214: おうま	2400	291	12.1		Q063: おやゆび	2401	101	4.2
	Q215: めうま	2400	282	11.8		Q064: ひとさしゆび	2400	86	3.6
	Q174: うるち	2401	270	11.2		Q110: しあさって	2400	63	2.6
	Q103: きのう	2400	266	11.1		Q111: やのあさって	2400	13	0.5
○	Q216: こうま	2400	264	11.0		合計	131991	15328	11.6

併用回答の割合 (B/A×100) [%]

質問項目の枠の下の罫線が太い項目は河西データ82項目と一致する項目

○印: 標準語形が LAJ の解説で複数認められている項目

□印: 標準語形が LAJ の解説では特定されず、併用処理も行われていない項目

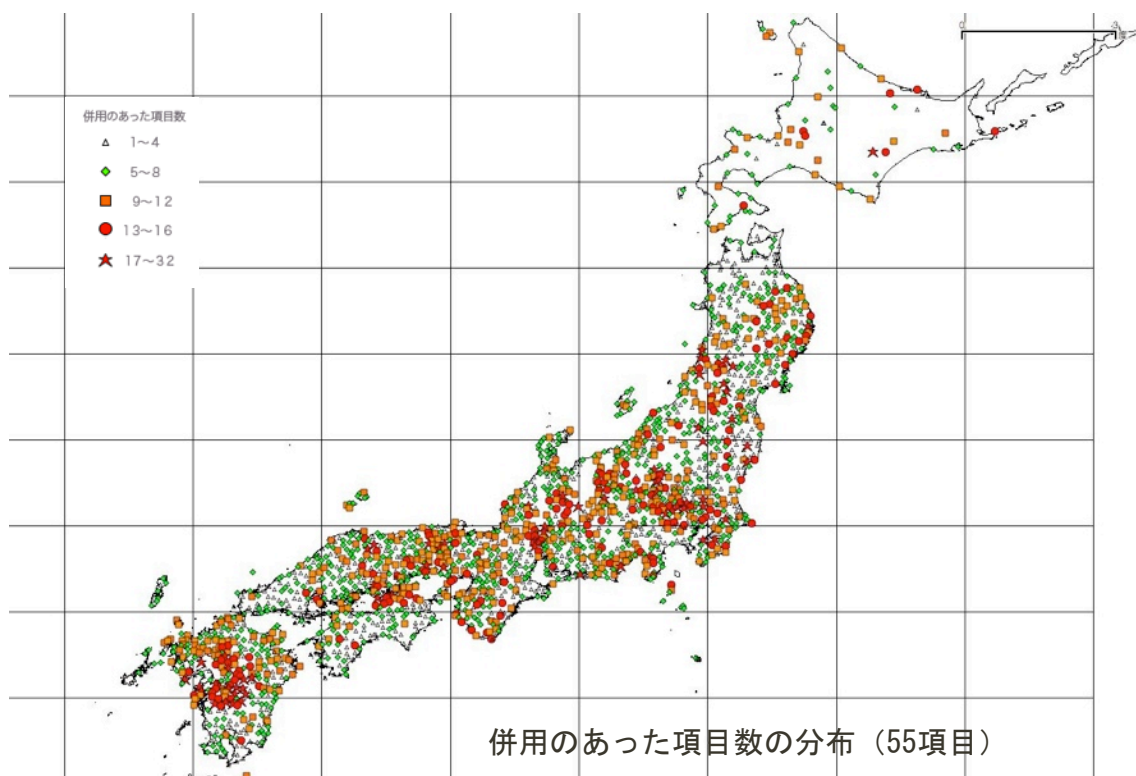


図5 併用のあった項目数の分布

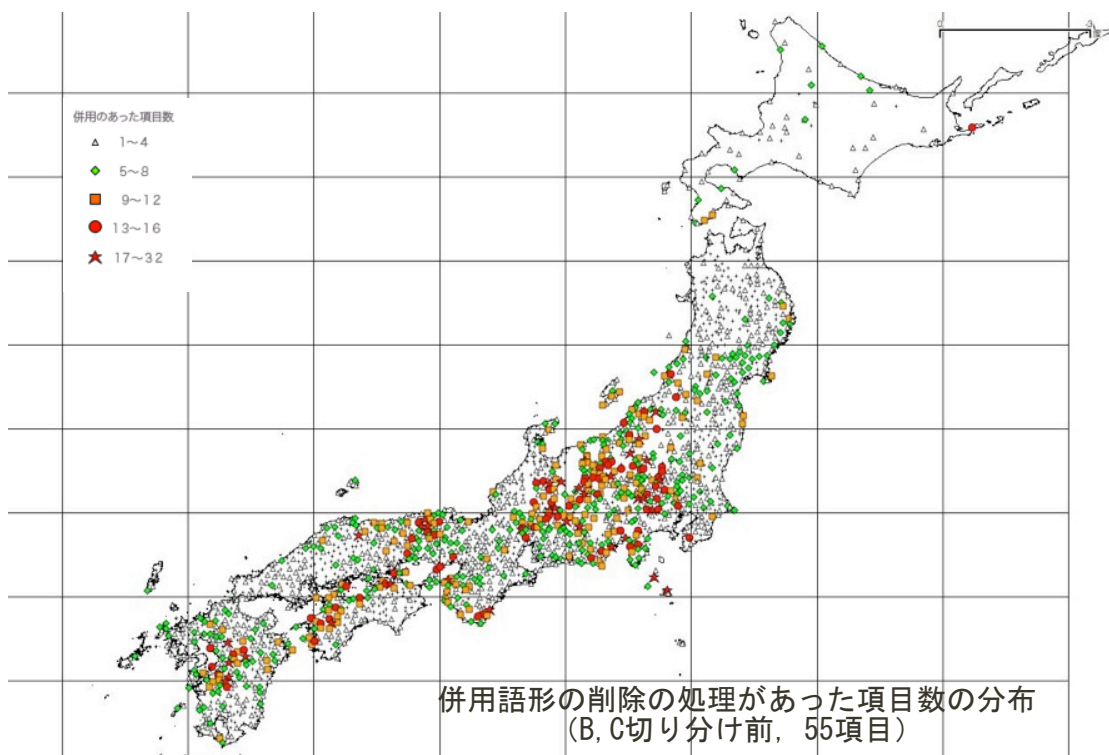


図6 併用語形の削除のあった項目数の分布 (標準語形以外の場合を含む)

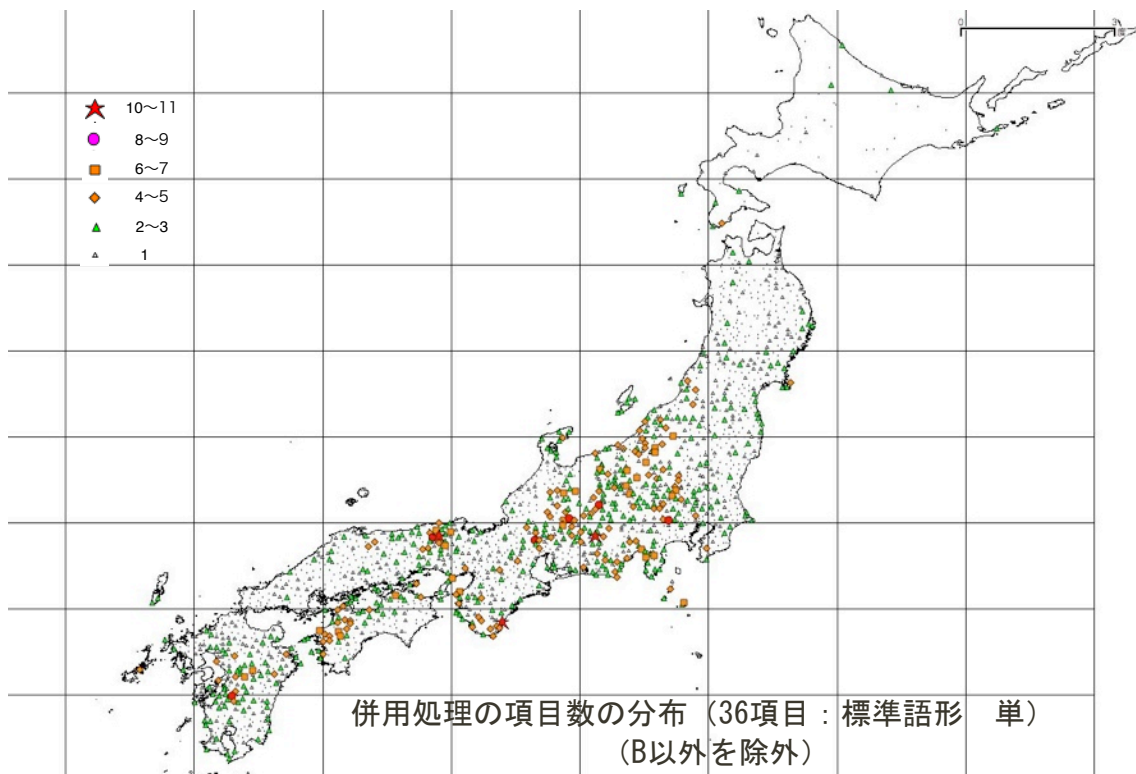


図7 併用処理の項目数の分布 (標準語形の併用処理のみ)

6. 『日本言語地図』における標準語形数の地点別集計の地理的分布

井上 (2004) は LAJ の県別標準語形使用率を分析しているが、データは標準語形ではあるが、方言体系の中に根をおろした語形である。〈共新上希〉の併用は、語形としてはおなじ標準語形でだが外されたものであるが、図7の分布からは地図化された方言形と連続した位置にあり、地理的な要因に関係なく空からばらまかれたような分布ではない。LAJ 上の併用の出現の仕方に、LAJDB の部分的なデータによるものではあるが、併用処理の場合も含め、LAJ 上で併用現象全体に地理的な分布が観察される。これは、井上 (2004) の主張を補う。

LAJ の計量的分析としては、LAJ の県別の標準語形の使用率のデータ (河西データ) を分析した井上史雄による一連の研究がある。河西データに多変量解析や計数的な手法による様々な分析を加えている。今回のデータは LAJDB から 55 項目を選んだものである。項目は LAJ 6 巻全体にある程度ばらついており、見通しを得ることはできる。

図8に上の42項目の標準語形の地点毎の度数を地図化したものを示す。河西 (1981) の表にある標準語形の県別使用率のデータを地図化したものが図9(左)である。河西 (1981) の地図では、段階化して表示されているが、ここでは、連続的な濃淡の変化として表示してある。図8の LAJDB の42項目について、同様に、県別に集計して県別使用率のデータを地図化したものが図9(右)である。河西データと今回の LAJDB のデータを用いた集計は全体的には非常によく似たパターンが得られていることが分かる。図10に河西データと LAJDB の42項目との県別の標準語平均使用率による県の順位を比較したグラフを示す。

大局的によく似たパターンが得られている。差異を伴いつつも、大きくは類似している様子を示しているが、この集計はあくまでも暫定的性格であるので、河西データと同じ項目に関するデータの整備が整った段階で、さらに詳細な検討に入っていくこととしたい。

LAJ の 2400 地点のデータを地点別に集計すると県別の集計では見られなかった細かさで分布の様子が捉えられている。

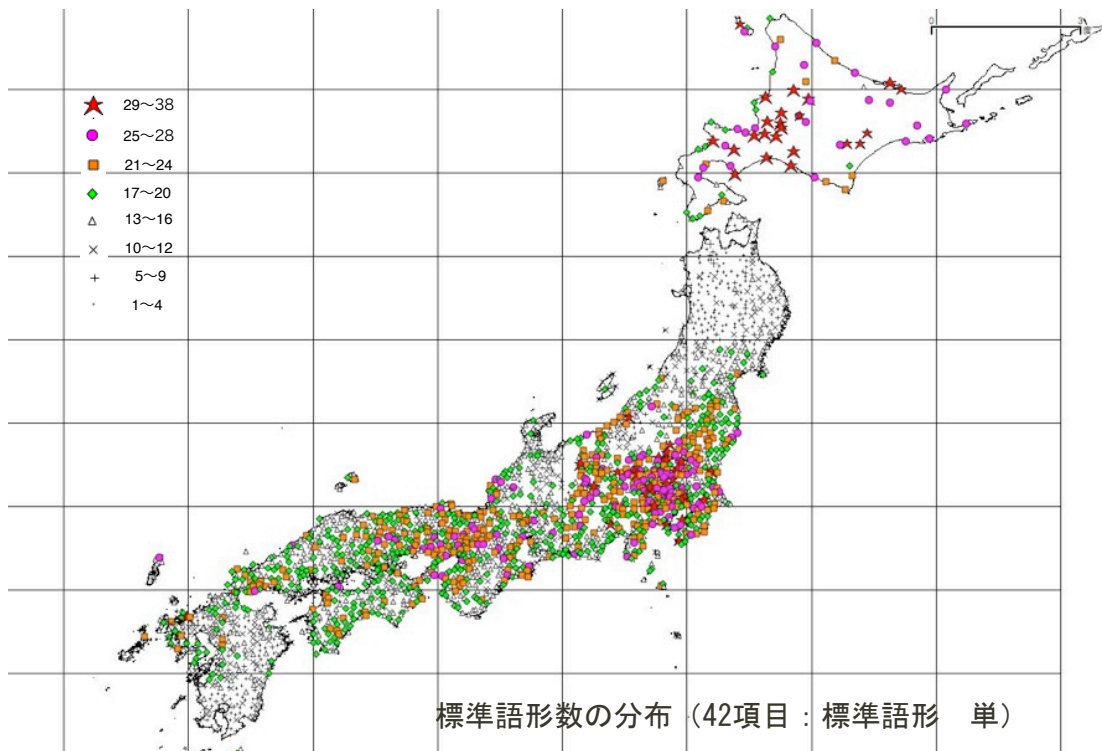


図8 標準語形数の分布

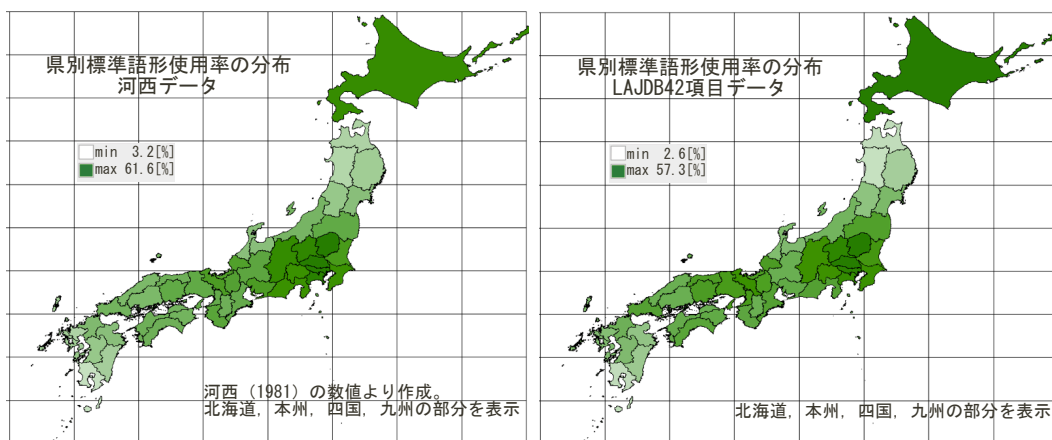


図9 標準語形県別使用率の分布 (河西データと LAJDB 42 項目)

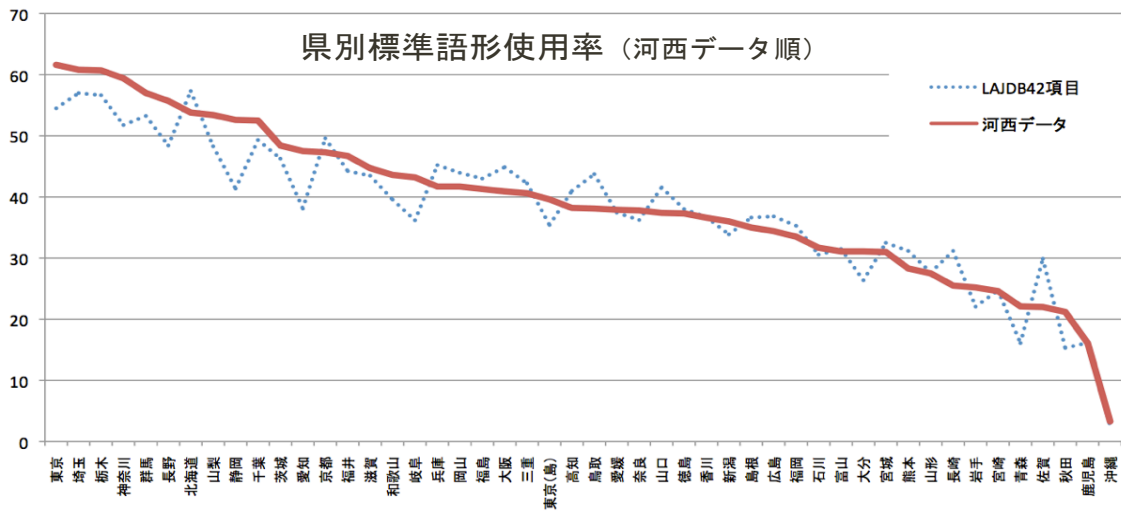


図10 県別の標準語形使用率 河西データ順

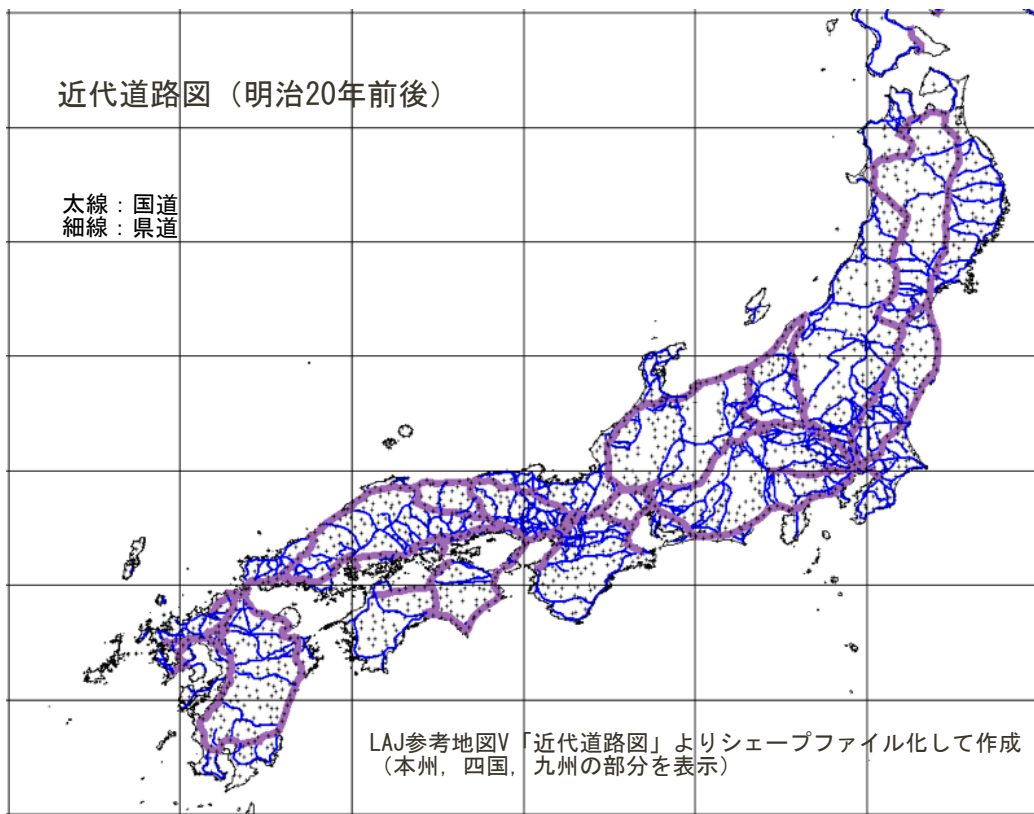


図11 近代道路図（LAI参考地図Vより作成）

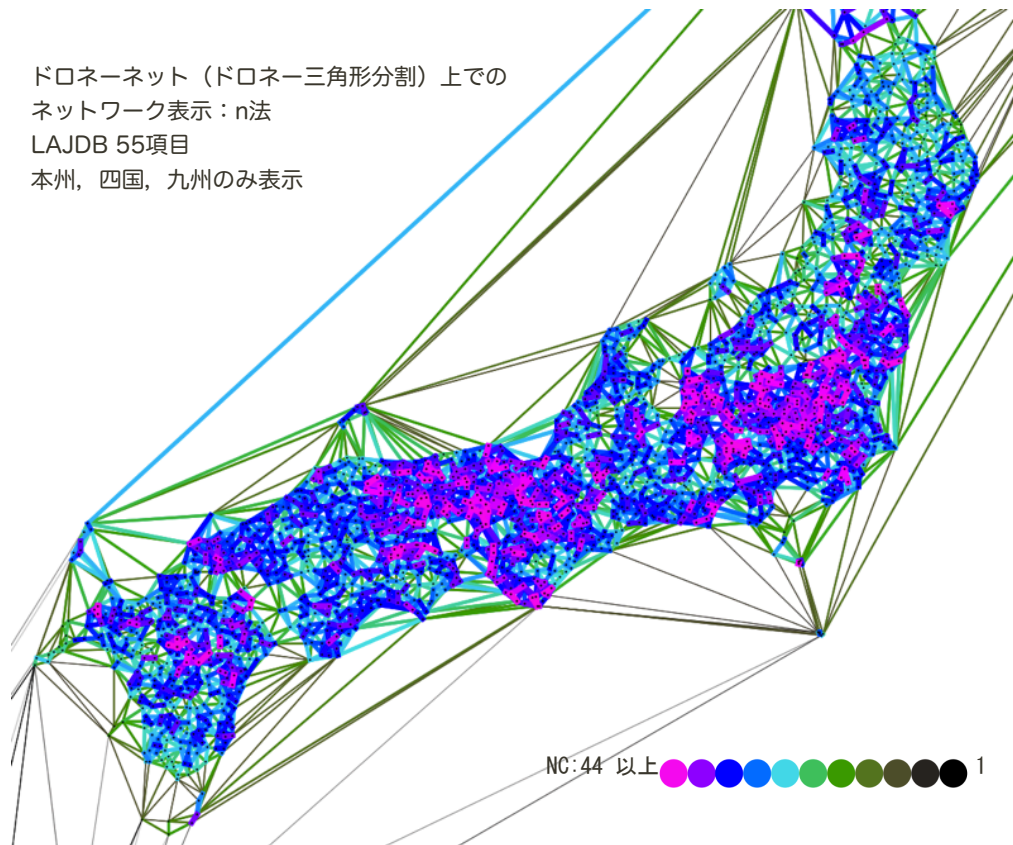


図 1 2 ドロネーネット上でのネットワーク表示：n法，55項目

7. 併用現象, 併用処理, 標準語形の分布と交通網, ネットワーク法の観察

図 1 1 は LAJ 参考図 V の「近代道路図」をシェープファイル化したものである。シェープファイルは GIS など標準的に利用されているファイルの形式である。LAJ の分布の分析に役立つために、LAJ の参考図 6 枚の中から、調査地点番号図、地勢図、藩領図、近代道路図を電子地図化した中の一つである。この近代道路図は「明治中期（1890 年・明治 23 年ごろ）の日本における道路網と、調査地点との関係を概観することを目的として作成した。この地図は、日本言語地図作成のための調査に被調査者として方言資料を提供した人たち（その生年は 1889 年・明治 22 年から 1903 年・明治 36 年までの 15 年間に集中している。—中略—）の育った時代、ないしは、かれらの養育者たちの活躍した時代の陸上交通路の概要を示したものと見えよう」（『日本言語地図』参考地図 V の凡例欄の解説より）というものである。この「近代道路図」のシェープファイルを用いて、LAJDB のデータを用いた集計と重ね合わせて、道路網との関係を観察した。

標準語形数の分布（図 8）と近代道路図（図 1 1）を重ねて見ると、標準語形の分布と街道の関係をはっきり観察できる。これは、言わば、多数の項目についての集計を LAJ の「2400 地点の解像度」による分布として見ることができることによる。都道府県別の数が 47 であるから、約 51 倍の「解像度」で集計値の地理的な分布を見ることができるとことになる。これは、これまでのデータの状態に対して、LAJDB がもたらす利点である。

関東と近畿の関係を見てみると、東海道やそれ以上に中山道が強く浮かんでくる様子が見える、脇街道など、より仔細に追うこともできる。街道筋のネットワークに沿った分析

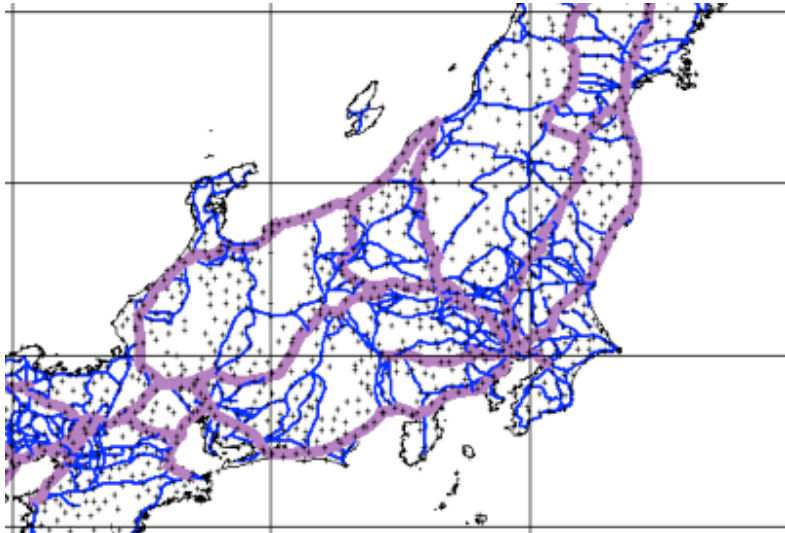
も可能性がある。脇街道の事例としては、一つあげれば、中仙道の脇往還である三州街道（別名伊那街道。中馬の通商の路として盛んに利用された。塩尻と岡崎を結ぶ。）は中山道と同等あるいはそれ以上に強く浮かんでいる様子が観察できる（他にも興味深い対応がある）。図13に関東から近畿にかけての「標準語形数の分布」と「近代道路図」の比較のための図を示す。図13(3)は、「標準語形数の分布」上へ「近代道路図」の重ね合わせた図です。見やすくするため、図13(1)上の道路の線幅を調整して表示した。このような観察は今後の伝播のルートを追跡する分析の可能性を示す。また、交通網は人口の分布とも関係する。これらも含めたより総合的な検討は語の伝播の分析につながる。

併用現象の項目数の図について見てみると、大括りだが、関東が山なら近畿は谷（あるいはカルデラ）のような様子を示している。その様子は図5（俚言形も含む併用）、図6（削除された標準語と特殊語）、図7（削除された標準語）とよりはっきりする。図14に関東から近畿にかけての部分での比較の図を示す。図7や図14(3)（標準語形で併用処理された項目数）を見ると、近畿地方では、近畿地方の周辺に、関東地方は関東から新潟方面に向かってあるいは、中山道にそった方面の分布が目立つ。井上史雄をはじめ、すでになされている関東と近畿との比較、対比と突き合わせて見ることもできる。語形の分布の中心別の集計なども未着手だが、語の伝播を探る分析全体の中で位置付けていきたい。

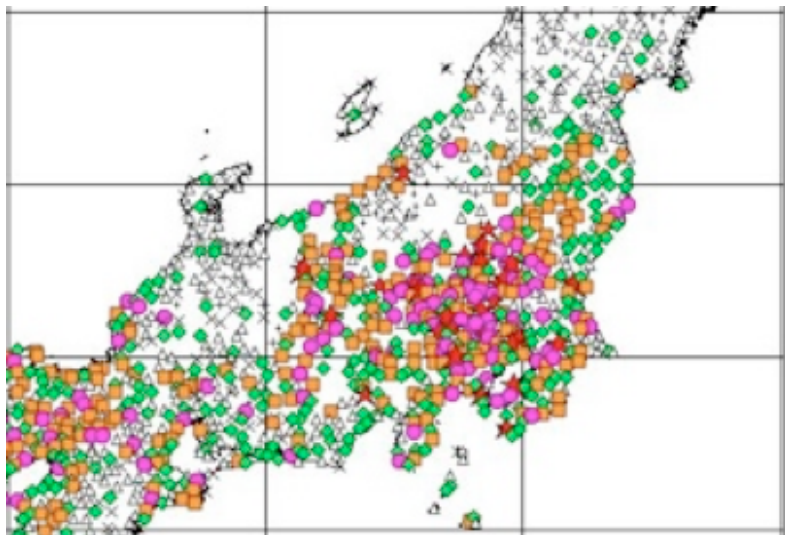
標準語と併用処理の関係は直接的である。図8（標準語形数分布）と図7（併用処理の項目数の分布：標準語形の併用処理のみ）を比較してみる。標準語形の分布の山と谷の部分と併用処理の分布の山と谷の対応を見ると、同じ方向の部分と逆方向の部分とが見られる。図7、図8を交互に表示するアニメーションを作成して観察したが、紙の上での表示の仕方を工夫する必要がある。図15の(1)、(2)を比較すると、例えば、いくつか図15(3)の○印をつけたところは、図15(1)の標準語形数の分布の山の空きにはまり込むように、あるいは接するように、標準語形の併用処理の項目数の山が分布している。一方、図15(3)の◇印のところは、標準語形の山と併用処理の山がほぼ重なっている。語の地理的伝播の先端部分で併用（併用処理）の現象が起きているか、そうならその様子を明瞭に視覚化してみようと考えているが、○印や◇印のより詳しい位置付け、分析を続ける。

図12はLAJDBの55項目を使ったネットワーク法によるドロネーネット上でのネットワーク表示(n法)である(熊谷2002)。言語的に類似した地点が連続する様子を地点間の近似的な隣接関係のネットワーク上で捉えたものである。より類似している(n法では地点間で語形が一致した数が多い)地点同士がより太い線で結ばれている。ここでも上で見た街道との関係が再度見えている。図8をはじめ上の各図との突き合わせもできる。

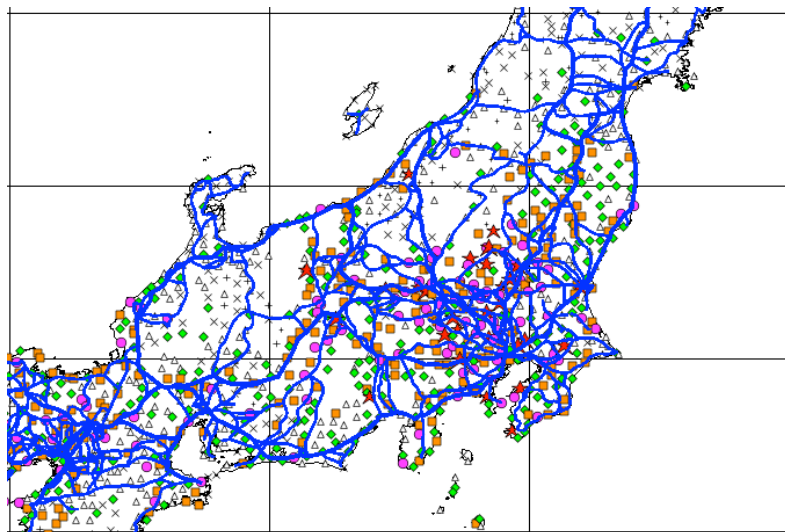
図16に関東から近畿にかけての部分で、ドロネーネット上でのネットワーク表示と近代道路図、標準語形使用数を比較できるようにした図を示す。俚言形（標準語形も含む）の語形の一致の度合いによって描かれた地点間の言語的な類似関係のネットワークの表示(図16(1))の上に観察できる、言語的に類似した領域のつながっていく様子と標準語使用数(図16(3))とを比べることができる。例えば、俚言形（標準語形を含む）で相互に類似した地点の連続が観察される領域の広がりや標準語形数の分布が示す広がりや、関東方面ではかなり重なるが、近畿方面では異なる。近代道路図(図16(2))との関係も相互に比較できる。これら3つの図の異同関係の分析、位置付けも進めていく。



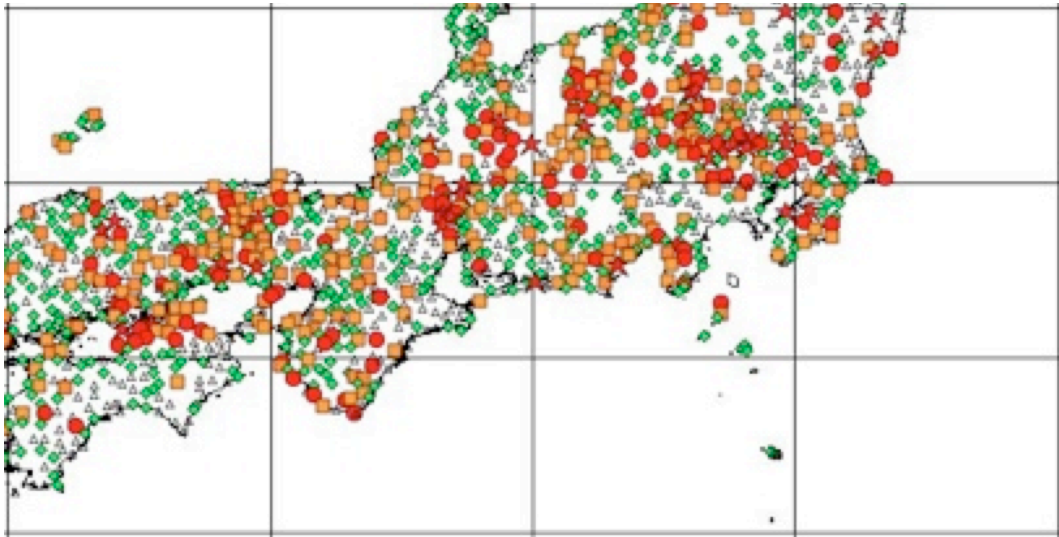
(1) 近代道路図 (図11部分表示)



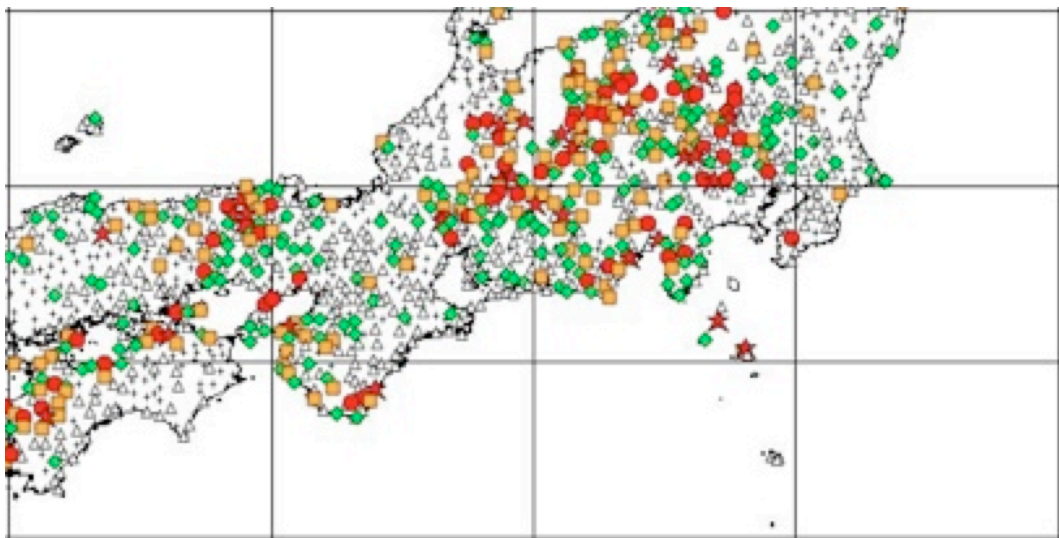
(2) 標準語形数の分布 (図8部分表示)



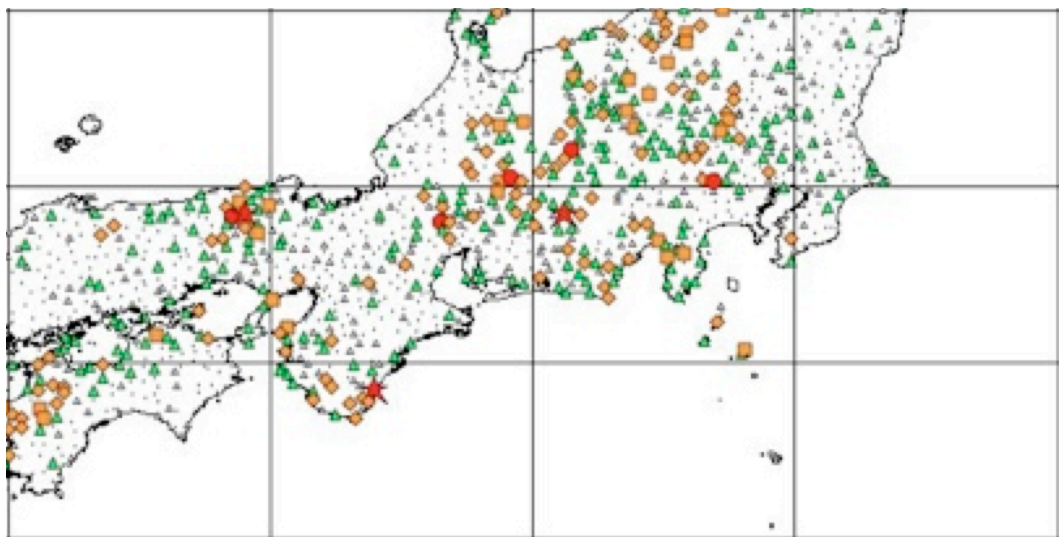
(3) 「標準語形数の分布」上への「近代道路図」の重ね合わせ (道路の線幅を調整)
 図13 標準語形数の分布 (図8部分表示) と近代道路図 (図11部分表示) の比較



(1) 図5 (俚言形も含む併用) の部分表示

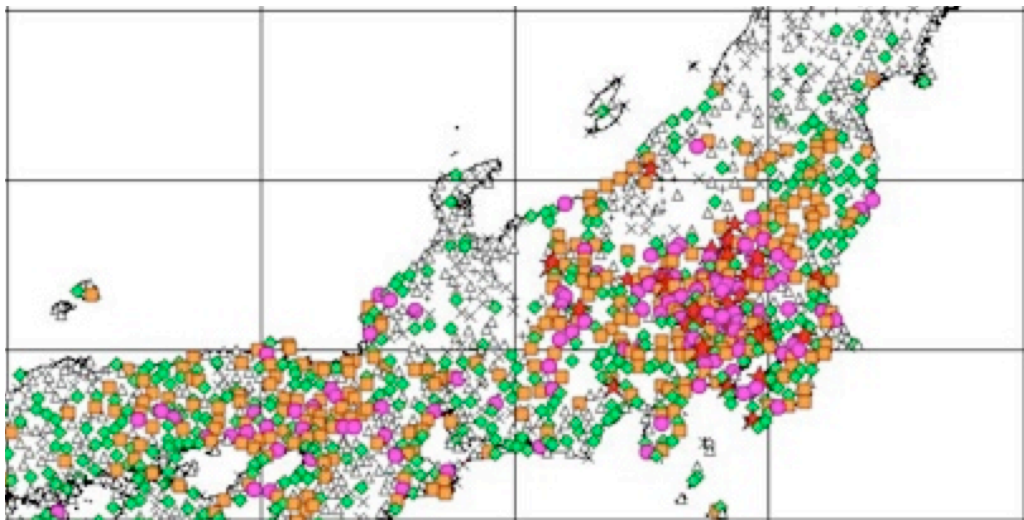


(2) 図6 (削除された標準語あるいは特殊語の併用) の部分表示

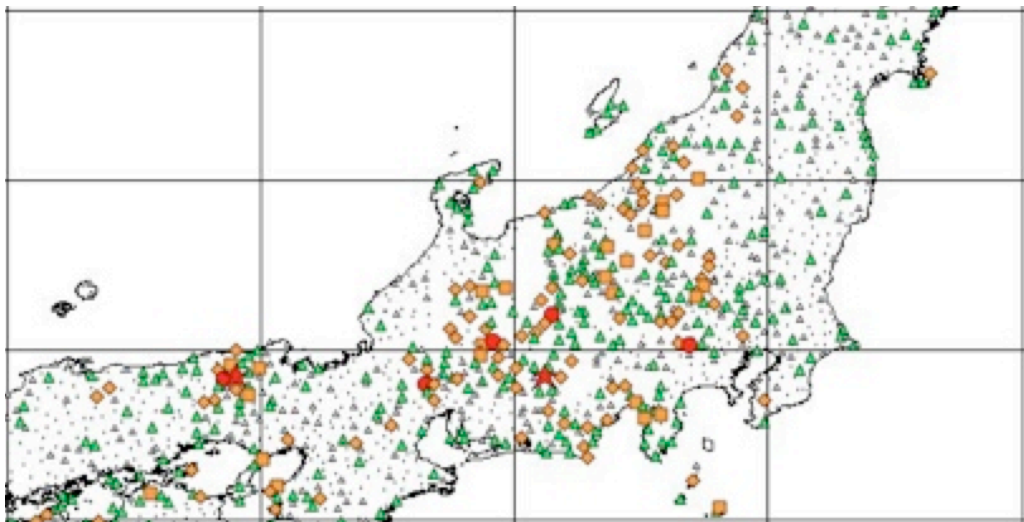


(3) 図7 (削除された標準語の併用) の部分表示

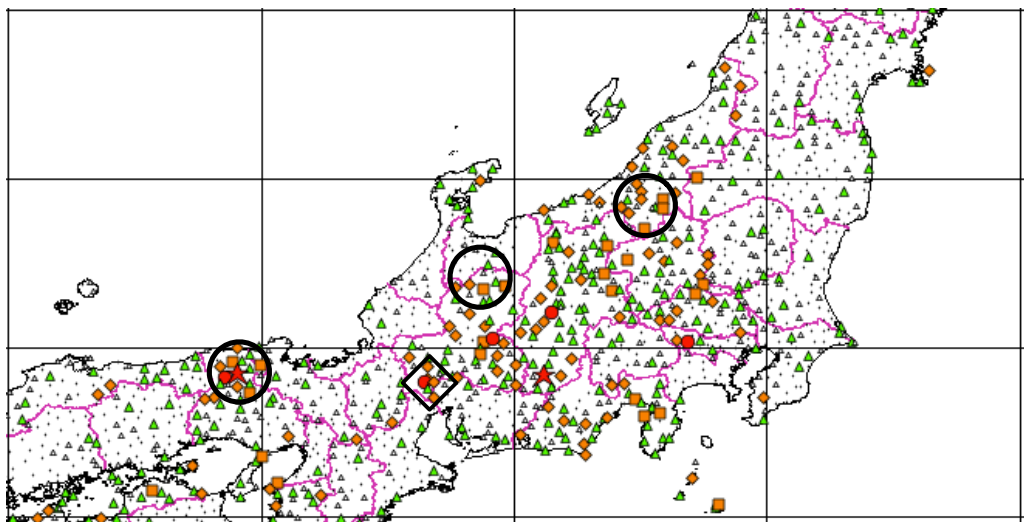
図14 関東と近畿の併用現象の比較 (図5, 6, 7)



(1) 標準語形数の分布 (図8の部分表示)

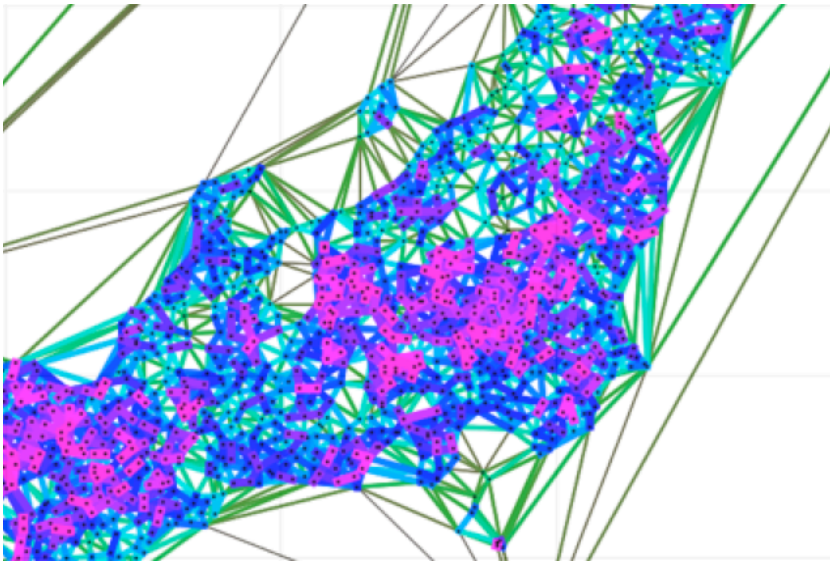


(2) 標準語形の併用処理の項目数の分布 (図7部分表示)

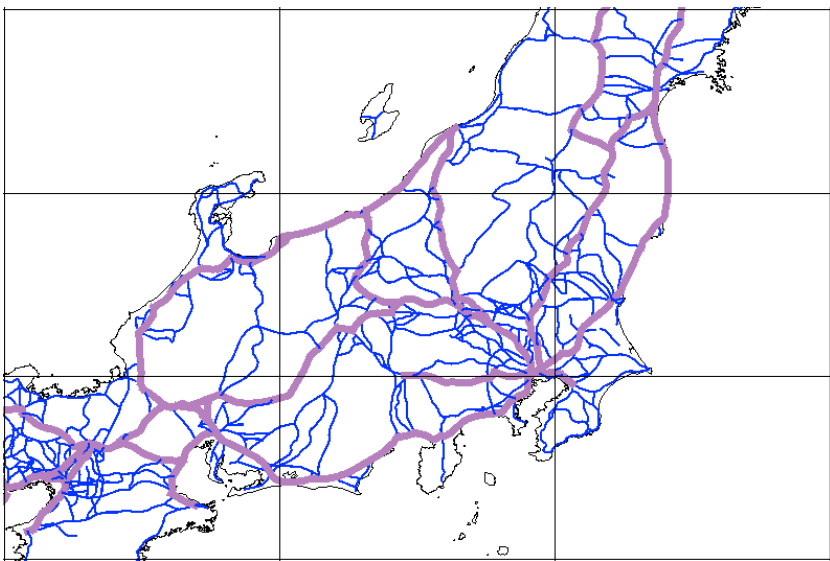


(3) 標準語形の併用処理の項目数の分布 (図7部分表示) に県境を表示

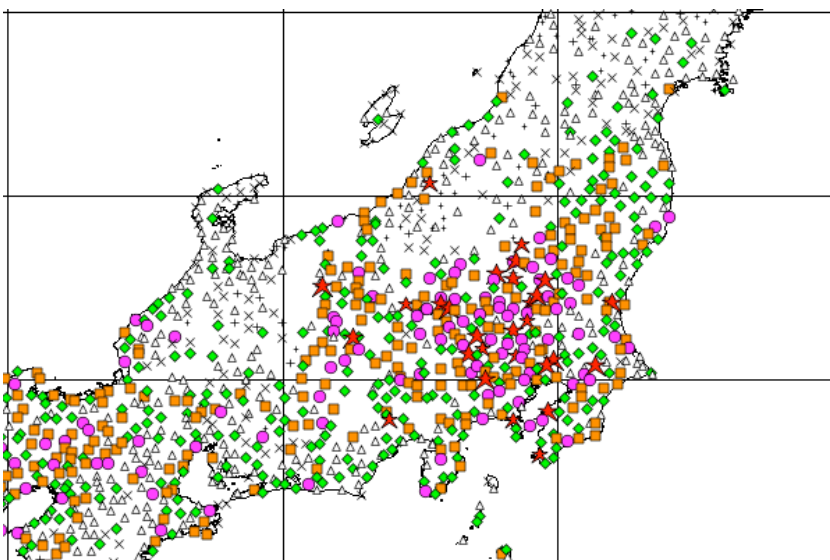
図15 標準語計数の分布と併称処理の項目数の分布



(1) 図12部分表示



(2) 図11部分表示



(3) 図8部分表示

図16 ドローネーネット上でのネットワーク表示と近代道路図，標準語形使用数

県別のデータでは得られなかったLAJ2400地点のデータの描く詳細な模様は、併用現象、併用処理、標準語形、交通網、地点間の方言的な類似度という関係の中でも、より詳しい観察が可能である。伝搬経路との関係を追跡し、語の伝播の分析につながる。言語地図では調査地点の回答のから面としての広がりを持つ分布パターンが描かれ、観察されるが、これは地点間の接触（交通網、人的交流等）を背景に形成されたものである。LAJの2400地点の細かさで、多数の項目を扱うことで、逆に、その面的な分布の背後にある交通網などネットワーク的な関係、その関係の上での語の伝播の様相が全国的な広がりの中で浮かび上がってくる。可能性を発掘していきたい。

8. おわりに

本稿ではLAJDBによる新たな可能性を探った。LAJDBにより、『日本言語地図』2400地点の項目別のデータが使える、また、注記や原表記が記載されている原資料であるカードに簡単にアクセスできる。このことにより、これまでには見えなかったより詳細な現象の観察、新たな観点による分析、これまでになされて来た研究の再評価や位置付けなど、さまざまな可能性があるが、全体像を明確に描き出し、現象の観察、分析を進めるには、LAJDBの完成を目指してデータの項目数を増やす必要がある。今回は、現状の暫定的な段階であり、データを詳細に分析、追跡し、結論を導くことには控え目であるべき段階だが、LAJDBの整備を継続して、完成を急ぎ、詳細な分析につなげたい。

本発表は国立国語研究所の共同研究プロジェクト「大規模方言データの多角的分析」(プロジェクトリーダー:熊谷康雄)の研究成果の一部である。なお、LAJDBには科学研究費研究成果公開促進費(データベース)[平成13,14,15,16,17,20年、『日本言語地図』データベース(研究代表者:熊谷康雄)]の補助を得た。地図の描画にはQGIS 1.6.0を利用した。また、本稿は、JLVC2013で報告した原稿(熊谷2013)にさらに加筆したものである。

参考文献

- 稲垣滋子(1980)「方言接触と語形併用―「日本言語地図」の分布から―」『都立大学方言学会会報』92, 1-10.
- 井上史雄(2001)『計量的方言区画論』明治書院.
- 井上史雄(2004)「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」『語学研究所論集(東京外国語大学語学研究所)』9, 1-18.
- 河西秀早子(1981)「標準語形の全国分布」『言語生活』354, 52-55.
- 熊谷康雄(2002)「方言区画論と方言地理学―計量的方言区画のためのネットワーク法の開発を通して―」馬瀬良雄監修『方言地理学の課題』明治書院, 150-164.
- 熊谷康雄(2007)「『日本言語地図』のデータベース化」『日本方言研究会第85回研究発表会発表原稿集』, 27-34.
- 熊谷康雄(2012)「言語地図の計量的分析:『日本言語地図』データベースによる併用現象の地理的分布の探索」『日本行動計量学会 第四十回大会抄録集』, 411-414.
- 熊谷康雄(2013)「『日本言語地図』のデータベース化と計量的分析」『国立国語研究所時空間変異研究系合同研究発表会 JLVC2013 予稿集』, 73-82.

- 国立国語研究所 (1966-1974) 『日本言語地図』(全6巻) 大蔵省印刷局 (縮刷版 1981-1985).
- 小西いずみ (2007) 「『方言文法全国地図』における回答語形数」『日本語学』26(11), 35-43.
- 佐藤亮一 (1986) 「地域社会の共通語化」飯豊毅一, 日野資純, 佐藤亮一編『講座方言学3 方言研究の問題』, 145-178.
- 澤木幹栄 (1988) 「『日本言語地図』の語形の数量的性質」, 『方言研究法の探索』秀英出版, 15-40.
- 柴田武 (1969) 『言語地理学の方法』筑摩書房.
- 徳川宗賢 (1973) 「言語地図における弧例」『ことばの研究』4, 秀英出版, 133-150.