

# 形態素解析を用いたアンケート調査自由記述欄 の分析手法に関する研究 ～路面電車利用意識調査データを用いたケース スタディ～

永野 峻祐<sup>1</sup>・小根山裕之<sup>2</sup>・大口敬<sup>2</sup>・鹿田成則<sup>2</sup>

<sup>1</sup>正会員 東京都交通局 建設工務部（前・首都大学東京大学院）（〒163-8001 東京都新宿区西新宿 2-8-1）

E-mail: Shunsuke\_Eino@member.metro.tokyo.jp

<sup>2</sup>正会員 首都大学東京大学院 准教授 都市環境学部都市基盤環境学域  
（〒192-0397 東京都八王子市南大沢 1-1）

E-mail: oneyama@tmu.ac.jp

<sup>3</sup>正会員 東京大学生産技術研究所 教授 先進モビリティ研究センター  
（〒153-8505 東京都目黒区駒場 4-6-1）

E-mail: takog@iis.u-tokyo.ac.jp

<sup>4</sup>正会員 首都大学東京大学院 准教授 都市環境学部都市基盤環境学域  
（〒192-0397 東京都八王子市南大沢 1-1）

E-mail: shikata@tmu.ac.jp

本研究は、自由記述欄から得られるテキストデータについて、言語処理手法の一つである形態素解析手法を用いた解析を行うとともに、プリコードデータと自由記述欄の分析結果との比較検証を行い、アンケート自由記述欄の特性及び分析手法について考察したものである。路面電車利用意識調査のアンケート結果を事例として用い、テキストデータから単語の抽出を行い、抽出単語の出現頻度分析や、Jaccard 係数を適用した共起ネットワーク分析により、自由記述欄の内容把握において形態素解析を適用することの有用性と課題を示した。また、プリコードデータの因子分析等によるグループ形成を行い、グループ毎に自由記述欄の定量化分析を行うことにより、プリコードデータと自由記述欄の記載事項の関係を分析し、その特性を明らかにした。

***Key Words:*** Free Answer, Questionnaire Survey, Morphological Analysis, Tram

## 1. はじめに

人々の価値観や行動様式の多様化している現代社会において、企業・行政の両方で住民や利用者の意見や意識等の把握に対するニーズの高まりが顕著に見られるようになった。近年のインターネットの普及などの技術革新もあって、人々の意見・意識もより収集しやすくなっている。社会基盤分野においても、現状のインフラへの評価を把握したり、社会資本整備を適切に進めたり、改善を図ったりなどしていく上で、国民の意見や意識を収集し、把握することは非常に重要であると考えられる。意見や意識等

の把握・抽出には様々な方法を考えることができるが、やはりアンケート調査が一般的である。しかし、通常のアンケート調査票からでは多様な意見や意識等を把握・抽出するには限界が多い。また、調査主体による調査票の設計方針自体が把握すべき意見や意識等に影響してしまう場合も考えられる。このような懸念事項を補完する役割を期待して自由記述欄を設け、ありのままの意見を書いてもらうという方法が採られることが多い。

しかしながら、自由記述欄のテキストデータには統計的処理を適用しにくいという問題があり、確立された分析方法や分析基準がないというのが現状で

表-1 アンケート調査概要

配布日時	平成20年10月～12月
配布方法	郵便受けに調査票を投函、後日郵送による回収
配布対象	荒川線沿線居住者、1世帯3名まで 世田谷線沿線居住者、1世帯2名まで
調査項目	①対象路線の利用頻度に関する項目
	②対象路線に対する意識・満足度の5段階評価
	③個人属性
	④自由記入欄

ある。数々の意見を人手により分類し整理するという方法がとられることが多く、莫大な時間とコストがかかること、分類する個人の主観に偏る可能性があること、などという課題がある。以上のようなこともあり系統的に分析されず、これまでは必ずしも十分に活用されてこなかった。

そこで、本研究では、有益な情報が豊富に含まれている可能性のある自由記述欄から得られるテキストデータに着目し、形態素解析の手法により路面電車利用意識調査データを用いたアンケート調査自由記述欄の分析に関するケーススタディを行う。また、プリコードデータと自由記述欄の分析結果との比較検証を行い、両者の関係性を把握することにより、アンケート自由記述欄の分析手法を検討するための知見を提供することを目的とする。

## 2. 既存の研究と本研究の位置づけ

計算言語学・社会学・心理学・教育学等の分野においては、自由記述型のテキストデータからの意見の分析や分析方法について多くの研究がなされてきた。また、近年、土木工学分野・社会基盤分野においても、社会資本整備を適切に進める上で、国民の意識や意見を自由記述形式のアンケートによって収集し、これを言語処理的に分析して把握しようとする試みがある。

本研究で取り扱うアンケート自由記述欄の分析に関して、高田ら<sup>1)</sup>はドライバーを対象として利用者のニーズなどを反映した事業執行や施策展開のため、道路が提供するサービスに対しての利用者の実感に基づく意見を直接、評価指標とする調査分析方法の検討を行っている。結果として、自由記述データから道路の問題視されている内容を抽出できたこと、自由記述回答データを用いた表現強度の分析方法の提案と、抽出した問題点の不満レベルの計測ができたことなどが挙げられ、問題点の抽出には自由記述データを用いることが良いことを示している。しかし、属性の大きな偏りがあり属性間の比較はなされていない。

表-2 アンケート配布・回収結果

	荒川線沿線配布	世田谷線沿線配布
配布数	15000	6000
回収数	2491	504
回収率	16.6%	8.4%

分析における形態素解析の適用に関しては、大塚ら<sup>2)</sup>の研究がある。水・緑環境が都市のイメージに与える影響を市民の視点からのイメージを把握することで定量的に明らかにするために、住民に対しアンケート調査を行い、自由記述欄から得られたテキストに対し形態素解析を適用し、その有用性を示している。しかし、実施されたアンケートには評価項目のような設問は存在せず、そういった設問と自由記述欄の関係は明らかになっていない。

他にも大村ら<sup>3)</sup>は、現場の技術者の経験を抽出・分析するために、インタビューから得られたテキストデータに対して形態素解析の手法を適用している。

本研究では、アンケート自由記述欄の分析に、既に土木計画学も含む様々な分野で適用の適用が進んでいる形態素解析の手法を適用しつつ、これらを用いたアンケート自由記述欄の分析の有用性を明らかにするとともに、プリコードデータとの比較分析により、両者の関連と相違を明らかにしようとする点が新しい点であると言える。

## 3. アンケート調査の概要と基礎集計結果

### (1) アンケート調査の概要

本研究で分析対象となっているテキストデータは、2008年10月～12月に都電荒川線・東急世田谷線の沿線住民を対象として、交通行動と公共交通機関の利用意識を調査する目的で行われたアンケート結果から得られた自由記述欄テキストデータである。主な調査項目は表-1の通りとなっている。配布実施箇所は、都電荒川線沿線においては早稲田停留所～三ノ輪橋停留所間を中心とするエリア、世田谷線沿線においては世田谷停留所～松原停留所を中心とするエリアであり、設定した各配布エリアを図-1および図-2に示す。アンケート配布・回収結果は表-2に示すとおりである。調査項目のうち、『対象路線に対する意識・満足度の5段階評価』については、5：思う～1：思わないの5段階で回答させた。

### (2) 個人属性について

得られたサンプルの男女比は男性：女性で約4：6、年齢構成は50代～70代で全体の5割を占め、中年

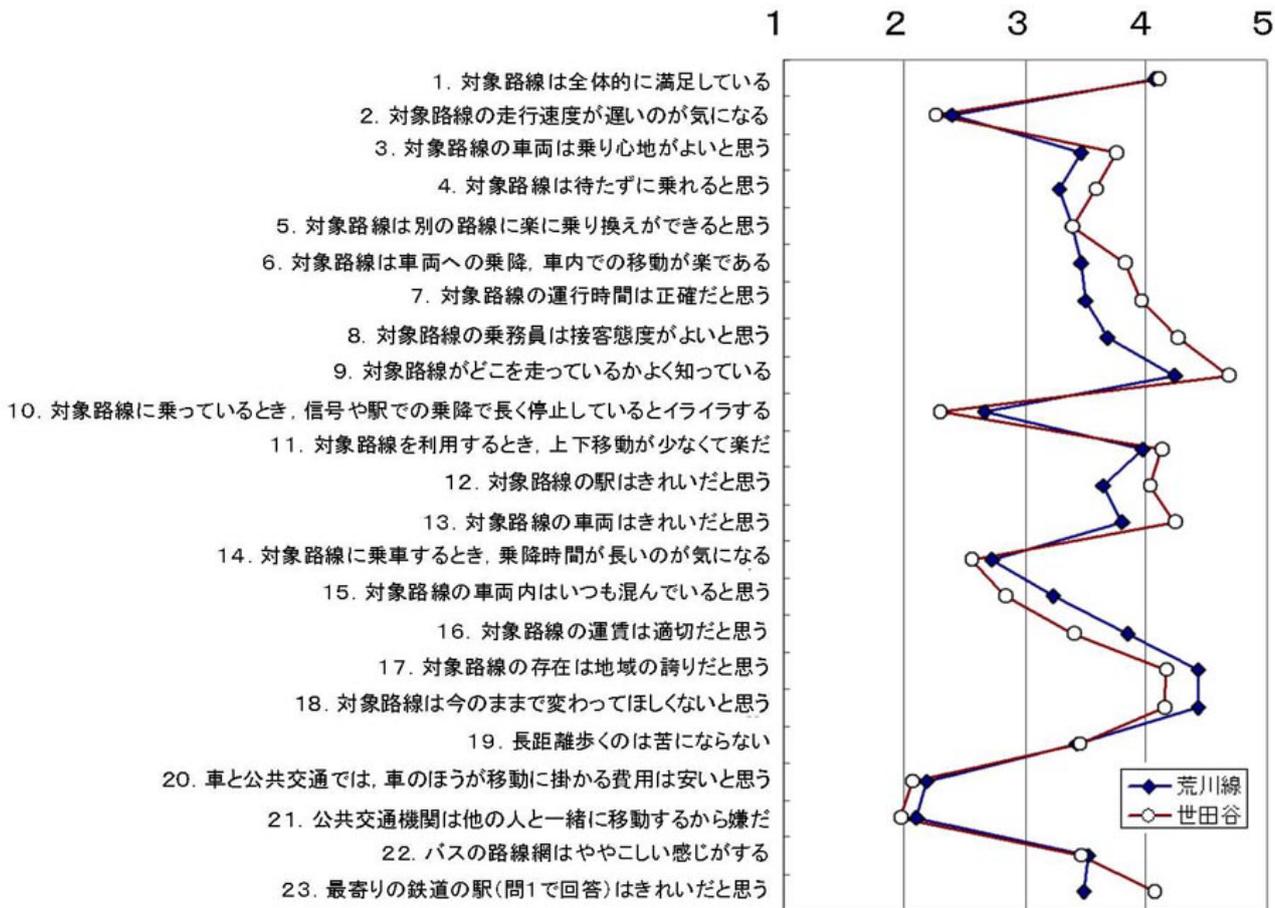


図-3 路線別 意識・満足度の5段階評価値

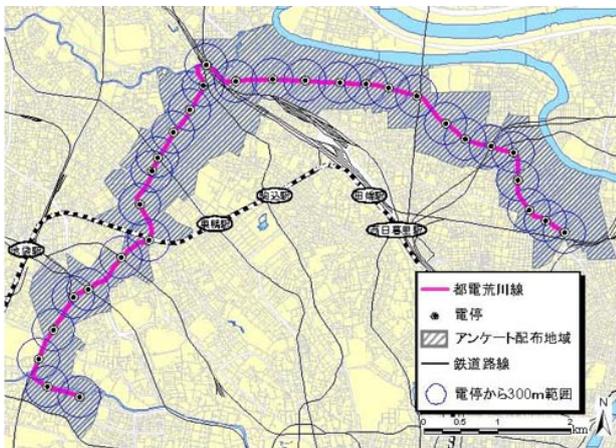


図-1 荒川線 配布地域



図-2 世田谷線 配布地域

層からの回答が多い結果となった。利用頻度については、いずれの路線においても高いが、世田谷線の方は代替経路が周囲に少ないことから、高頻度利用者の割合が多くなっている。

### (3) 対象路線に対する意識・満足度の5段階評価について

対象路線に対する意識・満足度の5段階評価23項目の内容及び荒川線、世田谷線それぞれの項目別平均値を図-3に示す。両路線ともに総合満足度に関す

る評価(質問項目:1番)が4以上となり、路面電車に対して全体的に満足感が大きい傾向が見られる。ただし、それ以外の項目については、それぞれの路線の特性を反映して、多少異なる傾向を示している。

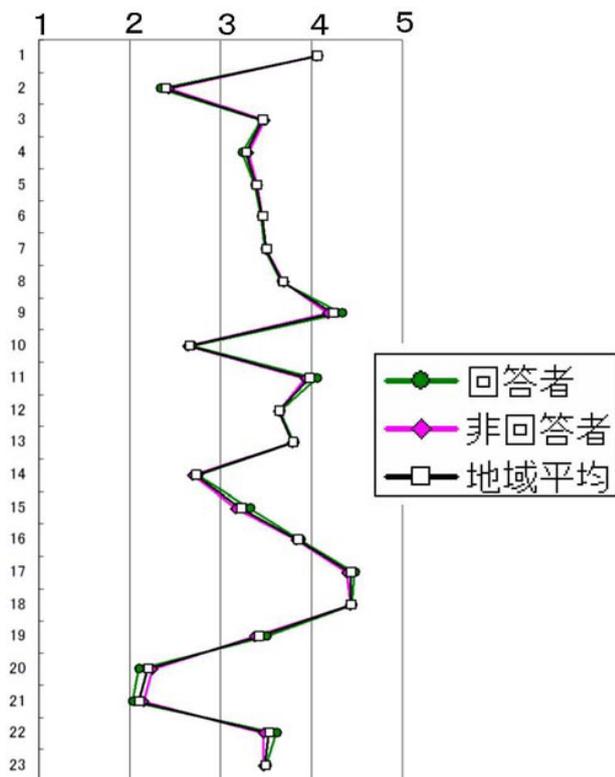


図-4 意識・満足度の5段階評価23項目 回答者・非回答者の比較【荒川線】

#### 4. 自由記述欄回答の有無による属性の比較分析

本研究では、アンケート自由記述欄から得られたテキストデータに着目し分析を行うが、そもそも自由記述欄の回答者と未回答者で特性が異なる可能性がある。もし、特性が異なる場合、アンケート自由記述欄の分析結果の一般性が担保できなくなる。そこで、アンケート自由記述欄への回答者と未回答者が特殊な回答者であるか否かを属性やプリコードデータの比較分析により検証した。

その結果、性別（男女）の違いによる記入率の変化は見られず、年齢別には、若い年代でやや記入率が高く、高齢になるにつれて、記入率が少し低下する傾向が見られた。さらに、利用頻度別には、普段の利用頻度が多いほうが記入率が高くなる傾向がみられた。これは、普段から利用していると路線に対して関心が強くなり、記入率が高くなる傾向があるものと考えられる。

次に、意識・満足度の5段階評価23項目を用いて、回答者と未回答者とで2グループに分け、それぞれの考え方・評価の違いなどに特性があるかどうかについて分析を行った。その結果、一部の項目についてわずかな差がみられるものの、5段階評価平均値の違いはほぼみられなかった。荒川線の例を図-4に

表-3 テキストデータの概要

配布地	記入数	総語数	総異なり語数	平均使用数(語/人)
荒川線沿線	957	39534	2849	41.3
世田谷線沿線	266	14651	1686	55.1

示す。

以上の結果より、自由記述欄の回答者はアンケート回答者全体の集団と大きな特性の違いはないものとして、以降の分析を行う。

#### 5. 形態素解析による単語の抽出

これより、アンケート自由記述欄から得られたテキストデータを用いて分析を行っていく。形態素解析器として『KH-coder』<sup>4)</sup>を用いて分析を行った。表-3は、地域ごとに回収したアンケート自由記述欄のテキストデータに形態素解析を適用した結果の基本情報である。この結果をみると、記入数（自由記述欄へ回答があったサンプル数）が多いほうが総語数は多くなっている。なお、総異なり語数の数値は助詞などの細かいものは含まれていない。また、回答者1人当たりの使用単語数は世田谷線沿線の方が高い結果となった。

##### (1) 荒川線沿線

荒川線沿線配布から回収したテキストデータから抽出した頻出単語を見ると、『する』『思う』『ない』『ある』など、どんな文章でも頻出する一般的な単語が最上位に位置し、要望に関する『ほしい』『欲しい』という単語も上位に位置している。また、対象路線が路面電車である都電荒川線ということもあり、それに関連した単語も多く抽出されている。具体的には、荒川線にある停留所名が多く抽出されていることや、荒川線が『チンチン電車』と呼ばれていたこともわかる。また、アンケート配布時期と前後して、荒川線では新型車両を導入しているからか『新型』という単語や、路面電車ならではの表現として『レトロ口』という単語も上位に現れている。他には、走行速度が他の高速輸送電車と比べても遅いことから『のんびり』や『ゆっくり』という単語も上位に現れている。路面電車は自動車社会においては、存廃が議論されやすく『存続』や『廃止』という単語も上位に現れている。存廃の議論とは全く関係のない目的でアンケートをとったにも関わらず、こういった単語が抽出されるということは、アンケート調査そのものが敏感なものを刺激してしまう恐れがあると推測される。対象路線に対する評価として、『便利』、『好き』、『親切』、『満足』、『大好き』、『ありがとう』

表-4 Jaccard 係数

$$Jaccard(x, y) = \frac{f_{11}}{f_{01} + f_{01} + f_{10}} = \frac{\text{積集合}}{\text{和集合}}$$

	語句yが現れる	語句yが現れない
語句xが現れる	$f_{11}$	$f_{10}$
語句xが現れない	$f_{01}$	$f_{00}$

などの肯定を表す単語が上位に位置している。一方で、『不便』、『残念』、『迷惑』、『困る』という不満を表す単語も比較的多く抽出されている。他にも、『混雑』や『混む』といった単語がかなり上位に位置することから、利用時の混雑に対する不満が大きい可能性がある。

## (2) 世田谷線沿線

世田谷線沿線配布から回収したテキストデータから抽出した頻出単語を見ると、荒川線と同様に『する』・『思う』・『ない』・『ある』など、どんな文章でも頻出する一般的な単語が最上位に位置し、要望に関する『ほしい』、『欲しい』という単語も荒川線ほどではないが上位に位置している。対象路線が同じく路面電車である東急世田谷線ということもあり、それに関連した単語も多く抽出されている。世田谷線に関連する地名も多く抽出されているし、路面電車ならではの表現として『レトロ』という単語もやはり上位に現れている。他に共通する点として、『ゆっくり』という単語も上位に現れている。また同様に対象路線に対する評価として、『便利』、『好き』、『親切』などの肯定的単語が上位に位置している。一方で、『不便』、『残念』、『悪い』、という不満な単語も比較的多く抽出されている。他にも、『混雑』や『混む』といった単語がやはり上位に位置することから、利用時の混雑に対する不満が大きい可能性がある。

世田谷線沿線からの抽出結果は荒川線沿線と比べてみると、アンケート項目に影響されたであろう単語の抽出が目立つ。『旧型』『新型』『140円』『寄付』『値上げ』がそれに該当すると思われる。世田谷線沿線へのアンケートは、交通行動と公共交通機関の利用意識の調査以外にも、過去に行われた車両更新への意識の調査も兼ねている<sup>5)</sup>。その車両更新の意識調査では、昔の『旧型』車両から現在の『新型』車両への更新に対する評価、CVM（仮想評価手法）による設問項目で、現在の運賃である『140円』から運賃の仮『値上げ』、沿線住民からの仮の『寄付』を募る、以上のような項目が荒川線アンケートとは大きく違っている点である。荒川線沿線においては、アンケートのほかの設問から影響された単語は『舍人』

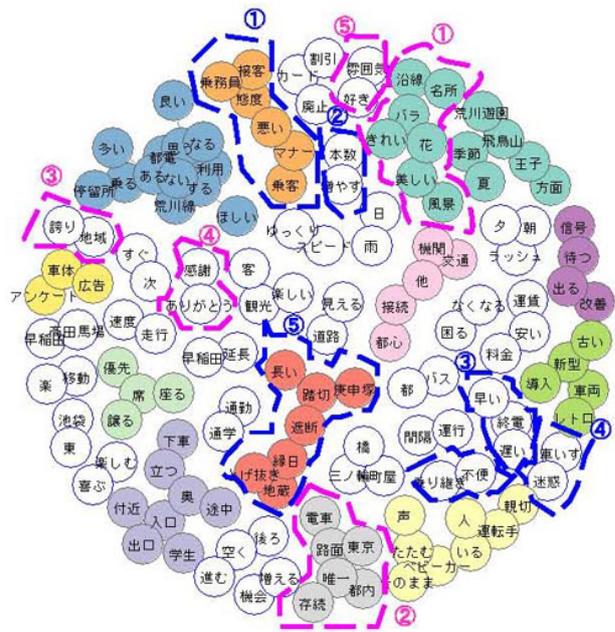


図-5 共起ネットワーク図【条件:出現数10回以上, Jaccard ≥ 0.15】【荒川線沿線】

『ライナー』以外にはほとんどなかったが、世田谷線沿線においてはかなり影響されているようである。自由記述欄以外のアンケート項目に強烈なインパクトを与えるものがあると、自由記述欄の内容はそれに影響された内容に偏ってしまう可能性がある。自由記述欄から本当に“自由”な意見を把握するためには、他のアンケート項目との兼ね合いも考慮してアンケート設計を行う必要がある。

## 6. 共起ネットワーク分析

単語の出現回数結果からも得られる情報は多いが、解釈の容易性という点で難が残る。また、単語と単語のつながりは全く把握できない。そこで、共起関係の強さの尺度として Jaccard 係数を適用した共起ネットワーク分析により、視覚的に単語同士のつながりを把握し、より詳細な情報の抽出をする。Jaccard 係数とは、ある言葉に対する積集合を和集合で除したものとなり<sup>6)</sup>、解釈整理すると表-4のようになる。また、この共起ネットワークは、布置された位置よりも、線で結ばれているかどうかということと意味があり、重要である。つまり、単に近くに布置されているだけで、線が結ばれていなければ共起の程度が強いことは意味しない。

ここでは、荒川線沿線における分析事例を示す。自由記述欄テキストデータより作成した共起ネットワークを図-5に示す。全ての単語をネットワークに組み込むことは不可能なので、単語の出現数10回以

表-5 『週4回以上利用』によく出現した単語

単語	週4回数	週4順位	年1回数	年1順位	順位差	回数差
欲しい	27	9	2	262	-253	25
運転手	23	11	2	262	-251	21
鉄砲	17	19	3	183	-164	14
混雑	15	24	3	183	-159	12
車いす	8	46	1	460	-406	7
垂れる	8	46	1	460	-406	7
乗客	10	35	3	183	-148	7
間隔	9	40	3	183	-143	6
朝	7	55	1	460	-395	6
やましい	7	55	2	262	-207	5
マナー	9	40	4	140	-100	5
現在	7	55	2	262	-207	5
後ろ	7	55	2	262	-207	5
信号	8	46	3	183	-138	5
線路	6	70	1	460	-380	5
遅い	9	40	4	140	-100	5
不便	6	70	1	460	-380	5
アナウンス	5	84	1	460	-366	4
出す	5	84	1	460	-366	4
乗務員	6	70	2	262	-192	4
態度	6	70	2	262	-192	4
定期	5	84	1	460	-366	4
付近	7	55	3	183	-128	4
きれい	4	106	1	460	-345	3
ホーム	4	106	1	460	-345	3
荷物	4	106	1	460	-345	3
今後	5	84	2	262	-178	3
困る	5	84	2	262	-178	3
止まる	6	70	3	183	-113	3
待つ	5	84	2	262	-178	3
遅れ	4	106	1	460	-345	3
入口	4	106	1	460	-345	3
方面	4	106	1	460	-345	3

上かつ、Jaccard  $\geq 0.15$  という条件を満たしたもののみのプロットとなっている。

まず、肯定的表現の解釈として、「赤① 沿線に咲くバラなどの季節の花風景が美しい」「赤② 都内（東京）で唯一の路面電車であるので存続を」「赤③ 地域の誇りである」「赤④ 感謝，ありがとう」「赤⑤ 雰囲気が好きだ」という項目が挙げられる。

また、要望・否定的表現として、「青① 乗務員の接客態度が悪い，マナーの悪い乗客がいる」「青② 本数を増やしてほしい」「青③ 終電が早いので遅くまで」「青④ 車いすが迷惑だ」「青⑤ とげ抜き地蔵の緑日では庚申塚の踏切による遮断が長い」「青⑥ 乗り継ぎが不便だ」という項目が挙げられる。

共起ネットワークによる分析の結果を見ると、荒川線沿線からは、肯定的表現・否定的表現は同量程度の出現となったが、世田谷線沿線では要望や不満表現の抽出のほうが多くなった。荒川線関連と世田谷線の共通項目として、乗務員の態度，踏切による長時間の遮断，乗客の乗車マナー，乗り継ぎシステムの改善などが挙げられる。また、具体的な解釈とまではいかなかったが、『ベビーカー』や『車いす』への対応も問題点として存在していると思われる。

以上のことから、自由記述欄には肯定表現よりも不満表現が多く存在する可能性が高く、その視覚的把握には共起ネットワーク分析の適用が有用であることを示した。しかし、文章量が多い場合には、ある程度容易に回答者の意見も抽出することができる

表-6 『年1回以上利用』によく出現した単語

単語	週4回数	週4順位	年1回数	年1順位	順位差	回数差
混雑	3	130	22	16	114	19
今	2	193	21	17	176	19
地域	3	130	16	26	104	13
感じる	3	130	14	32	98	11
踏切	1	399	11	41	318	10
機会	1	399	10	48	311	9
言う	3	130	12	34	96	9
交通	2	193	11	41	152	9
残す	3	130	12	34	96	9
ずっと	1	399	9	54	306	8
ベビーカー	3	130	11	41	89	8
観光	2	193	10	48	146	8
広告	1	399	9	54	306	8
考える	1	399	9	54	306	8
魅力	1	399	9	54	306	8
あまり	3	130	10	48	82	7
場所	1	399	8	64	295	7
うれしい	1	399	7	71	288	6
残る	1	399	7	71	288	6
自転車	2	193	8	64	129	6
美しい	2	193	8	64	129	6
イメージ	2	193	7	71	122	5
見る	2	193	7	71	122	5
必要	2	193	7	71	122	5
無い	1	399	6	87	272	5
よい	2	193	6	87	106	4
イベント	1	399	5	106	253	4
演習	1	399	5	106	253	4
楽しい	1	399	5	106	253	4
好き	1	399	5	106	253	4
行ける	1	399	5	106	253	4
手段	1	399	5	106	253	4
場合	1	399	5	106	253	4
窓	1	399	5	106	253	4
存在	1	399	5	106	253	4
飛鳥山	2	193	6	87	106	4
木	1	399	5	106	253	4
郵便	1	399	5	106	253	4

が、文章量が確保できていないと、単語同士のつながりからだけではよく分からないことになってしまい混乱を引き起こす場合がある。分析対象となっているデータ量と分析目的を確認しつつ作業をすすめる必要がある。

## 7. プリコードデータと自由記述欄の比較分析

プリコードデータで得られる情報と自由記述欄から得られる情報の関連性を分析するため、双方のデータを用いて比較検証を行った。分析に使用したプリコードデータは属性項目（性別・年齢・利用頻度）と5段階評価23項目である。なお、以下では、サンプル数を十分に確保できる荒川線沿線を中心に分析を行った。

### (1) プリコードデータによる分類

プリコードデータにより特性の異なるグループに分類し、グループ間の単語出現頻度等の違いにより、グループの特性抽出を試みるとともに、プリコードデータと自由記述欄の関係を分析した。ここでは、出現回数の比較だけでなく、順位差も考慮した。順位差とは、各々のグループのテキストデータにおける出現単語を頻度順に並べ、双方に出現する単語の

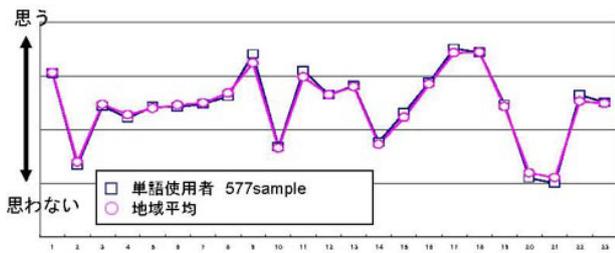


図-6 単語使用の有無による5段階評価値の比較  
【抽出単語：する】

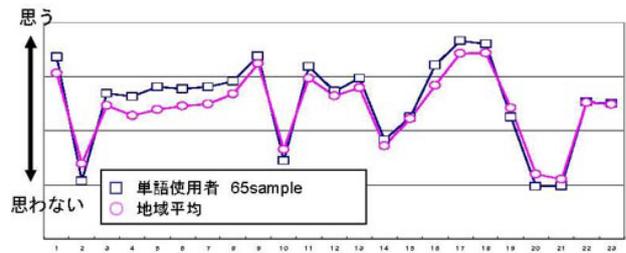


図-7 単語使用の有無による5段階評価値の比較  
【抽出単語：便利】

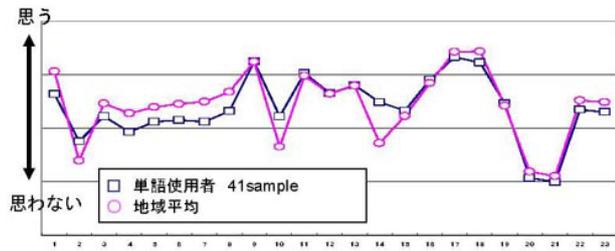


図-8 単語使用の有無による5段階評価値の比較  
【抽出単語：踏切】

中で順位の違いを表す。このことにより、グループの特徴を示す単語の抽出を試みた。

分類軸の設定方法としては、性別、年齢別、利用頻度別、意識・満足度の5段階評価別、意識・満足度の統合指標としての因子得点ランク別（5段階評価23項目による因子分析を行い、因子得点のランク別）などを行った。

以下、利用頻度別分析として、『週4回以上利用』と『年1回以上利用』の2グループでの比較分析例を示す。表-5は『週4回以上利用』によく出現した単語、表-6は『年1回以上利用』によく出現した単語を表す。これらの比較から、『週4回以上利用』のグループでは、速度に関する単語や混雑といった不満表現が上位に抽出されている。また、普段接する機会が多いためか、『運転手』『乗務員』という単語も目立つ。年1回と比べて『欲しい』の使用数が圧倒的に違うことから、要望に関する文章が多くを占めていると思われる。一方、『年1回以上利用』のグループでは、『のんびり』といった表現や、『楽しい』『好き』といった明るい表現が目立つ。その他の単語を見ても、『都内唯一の存在』としての評価が高いように思われる。

これらの分析結果には、利用頻度によって荒川線に対する見方や要望等が異なることが明確に表れており、大変興味深い結果であると言える。特に、単語の出現頻度や順位差などの定量的な指標によって評価できる点が、本手法の大きな特徴である。

## (2) 自由記述の抽出単語による分類

続いて、自由記述の抽出単語によりグループ形成し、プリコードデータ、主に意識・満足度の5段階評価23項目による比較を行った。抽出単語によるグループを形成し、5段階評価23項目のグループ平均値と地域平均値を比較した。分析結果の一例を図-6（する）、図-7（便利）、図-8（踏切）に示す。『する』などの一般的な頻出単語を除けば、単語の意味に応じてプリコードデータの評価に影響している可能性がみられた。『便利』という単語は、ある程度肯定的な表現である。そういった単語を自由記述欄で使用している人は、対象路線を肯定的に捉えており、5段階評価に対しても評価が高い傾向にあると考えることもできるし、逆に、5段階評価で表されるような個別の要因に対する評価があり、それが『便利』という用語に総合評価として表れているという見方もあり得る。

一方、『踏切』という単語の場合、踏切による長時間の通行遮断といった不満表現の1つであると考えられる。このような不満要因がきっかけで、必ずしも踏切とは直接的な関係がないと思われる質問項目についても、5段階評価が低くなっているとも考えられるし、逆に、荒川線に全体的に否定的な見方を持っているが、その中心的又は具体的な要因が『踏切』である、という可能性もあり得る。

このような事例から、自由記述欄で用いられる単語と、意識・満足度の5段階評価などのプリコードデータでは、意味づけとして概ね対応が取れているものの、それらの関係をより詳細に分析することにより、それぞれの内容を相補的に分析することが可能になることが示唆される。

## 8. おわりに

本研究では、自由記述欄から得られたテキストデータに着目し、言語処理分析を中心としたアンケート解析を行った。分析結果を以下にまとめる。

自由記述欄から得られたテキストデータに対し、形態素解析の適用や、共起ネットワーク分析により、単語同士のつながりを定量的、視覚的に分析・把握できることを示し、テキストデータにおける形態素解析の適用の有用性を実証した。また、プリコードデータと自由記述の組み合わせによる分析により、単語の出現頻度や順位差などの定量的な指標によってグループ間の相違を分析したり、単語毎のプリコードデータとの関係をより詳細に分析することにより、それぞれの内容を相補的に分析できる可能性があることなど、プリコードデータだけでは把握しきれない様々な意識の抽出に役立つ可能性を示すことができた。

従来、人手により定性的かつ曖昧に分類されてきた数々の自由記述欄の意見を、形態素解析等の手法を用いることにより、単語レベルで定量的に分析でき、単語同士のつながりも本研究で使用した共起ネットワークにより把握することができる。また、数々の分析により、どちらかといえば自由記述欄から不満表現の抽出が多くなったが、プリコードデータからだけでは把握できない不満要素を抽出することにより、様々な課題解決を図ることのできる可能性がある。

一方、テキストデータの電子化や形態素解析に適用するための前処理に多くの労力と時間がかかること、異なる表現で同じ意味の単語の扱いなど、実際に適用する際の課題は多い。また、ある程度定量的に意味のある内容を扱うためにはある程度のサン

ル数が必要になる点や、定量的に扱うことにより、少数だが貴重な自由意見が埋没してしまうなどの課題もあり、実際の適用には調査の設計手法、分析手法の更なる工夫と事例の蓄積が必要と考えられる。

#### 参考文献

- 1) 高田伸二, 屋井鉄雄: アンケート自由記述による道路ニーズ・不満の把握手法の研究, 日本都市計画学会学術研究論文集, No. 35, pp. 571-576, 2000.
- 2) 大塚裕子・森田哲夫・吉田朗・小島浩・塚田伸也: テキストマイニングによる都市・景観イメージ分析－水緑環境に着目して－, 土木計画学研究・講演集, Vol. 41, CD-ROM, 2010.
- 3) 大村陽・大塚裕子・伊藤裕美・川野佐江子・室町泰徳: 自由回答インタビューの分析に基づく都市交通メガプロジェクトの評価に関する研究, 土木計画学研究・講演集, Vol. 40, CD-ROM, 2009.
- 4) KH-coder HP, <http://khc.sourceforge.net/> (2011/01/31 付 閲覧確認).
- 5) 永野峻祐・小根山裕之・大口敬・鹿田成則: 東急世田谷線における新型車両導入効果に関する研究, 土木計画学研究・講演集, Vol. 40, CD-ROM, 2009, Nov. .
- 6) Pang-Ning Tan (著), Michael Steinbach (著), Vipin Kumar (著): 出版社: Addison Wesley; US ed 版 (2005/5/12), 『Introduction to Data Mining』, 2005, ISBN -10: 0-321-32136-7.

(2011. 5. 6 受付)

## STUDY ON ANALYSIS METHOD OF FREE ANSWERS IN THE QUESTIONNAIRE SURVEY USING MORPHOLOGICAL ANALYSIS : A CASE STUDY OF AN ATTITUDE SURVEY FOR UTILIZATION OF TRAM

Shunsuke EINO, Hiroyuki ONEYAMA, Takashi OGUCHI and Shigenori SHIKATA