

アンケート調査における文章回答の統計処理に関する研究

Statistical Analysis of the Document Data in Questionnaire Survey

中岡良司*・今尚之**・佐藤馨一***・五十嵐日出夫***

By Ryoji NAKAOKA, Naoyuki KON, Keiichi SATOH, Hideo IGARASHI

1. はじめに

本研究はアンケート調査における文章回答（フリーアンサーあるいは自由回答）の実用的な統計処理方法を提案するものである。その最大の特徴は回答文章中のキーワードを抽出し構造化する点にある。

文章回答は数値や選択肢回答と同様、アンケート調査票の一般的な回答形式であるが、その多くは、回答者に自由に意見を述べさせ参加意識を高めたり、選択肢の不備を補う形で導入されているに過ぎない。しかしながら、文章回答は調査主体の思い込みを排除したり、選択肢の設定が困難な設問に有効であり、回答者の生の声を得る有効な回答形式である。本研究においては文章回答を主体とした意識調査を実施しその有効性を実証している。

自由な意見（発想）を段階的に集約していく方法としてはKJ法¹⁾が広く知られている。しかしながら、KJ法の適用の正否は分析者（リーダー）の資質に負う所が大きく、安定した結果を得ることは難しい。また、本質的に統計的処理を求めるものではない。大井らは、回答内容を単語レベルに分解し近傍法で解析する自由連想法による調査方法を提案している²⁾が、自由連想法は「公害」や「住み良さ」など意味のあいまいな言葉の概念を明確にするのに有効な手法と考えられる。なお、アンケート調査の自由回答の設計・解析に関しては、既に著者らは回答内容を主語述語に要約してクラスター分析を適用する方法を報告している³⁾が、その作業は煩雑であり実用的とはいえない。そこで、本研究では回答文章からキーワードを抽出し分類、集計する実用的な方法を体系化した。分析の基本的手順はKJ法に

依存しているが、最終的にはアンケート調査における多重回答と同様の統計処理が可能となっている。

2. 文章回答の統計処理方法

文章回答の統計処理手順を図-1に示す。

1) 回答文章入力

最初に必要作業は回答文章（原文）をパソコンへ入力することである。入力作業自体は日本語ワープロやエディタを使用して構わないが、最終的にはデータベース化しておくことで、後の手順との対比や統計処理が容易となる。

2) 回答文章の要約・整形

回答文章（原文）には冗長な言い回しや修飾語句など多様な表現が混在している。そこで、文章内容を要約し共通の形式に整え単純なデータ構造にする。図-1の例では「〇〇を△△する」という形式に統一している。この手順で重要なことは、1人の文章中に現れる複数の意見は個々に分割して要約・整形することである。意見を要約する以上に分割することが重要である。この作業は担当者の文章理解力に負うところが大きい、そのバラツキは予想以上に小さいことが確認されている。

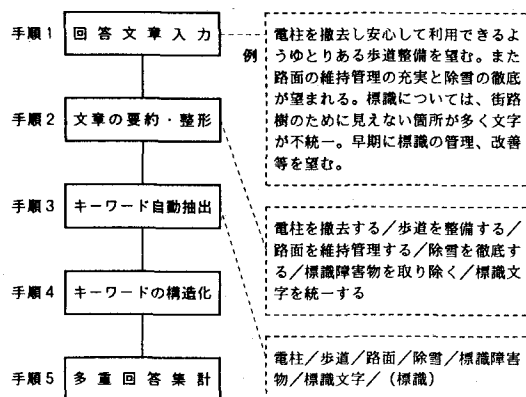


図-1 文章回答の統計処理フロー

キーワード：アンケート調査・文章回答・キーワード

* 正会員 北見工業大学助手 土木開発工学科
(〒090 北見市公園町165)

** 正会員 工博 小樽商科大学助手

*** 正会員 工博 北海道大学教授

3) キーワードの自動抽出

第3段階の第1ステップでは、整形されたデータから、単純なプログラムによってキーワード（上記の〇〇部分）を自動抽出する。ただし、これらのキーワード群には多くの複合的なキーワードが含まれているため集計には適さない。例えば、図-1の例では、「標識障害物」と「標識文字」が独立したキーワードとなっているが、これらは「標識」という語句で初めて関連づけられるキーワードである。意味を持つ最小単位のキーワードを本研究では単位キーワードと呼んでいる。

第2ステップでは、得られたキーワード群から単位キーワードを自動抽出する。そのためには1対のキーワードを1文字毎に比較して重複文字を抽出するプログラムが必要である。効率的なプログラムの作成には、情報処理技術の分野で開発されてきた文字列照合アルゴリズム（Brute-Force法）を利用すればよい。

4) キーワードの構造化

この段階では抽出された単位キーワード群をグループ化し分類名を付ける。この作業は単位キーワードとその複合キーワードの関係を利用して大部分が分類可能であるが、意味は同じで表現が異なるキーワードに対しては分析者が独自に判断し分類をする必要がある。たとえば、後述する事例における「休憩場に必要設備」を尋ねた設問においては、トイレを意味する表現（トイレ、便所、手洗場、洗面所、御不浄、お手洗いなど）が8種類使用されていた。

次に、単位キーワードを中心とする分類レベルを小分類とするならば、その内容を中分類、大分類へと集約しておくことによって、統計結果を総合的に判断することが容易となる。経験的には小分類数は20から30個程度、中分類数は10から20個程度、大分類数は10個以内が実用的な分類数である。

5) 多重回答集計

アンケート調査における多重回答形式（あるいは複数回答形式。「該当するものをいくつでも選んでください」という形式）は、選択肢は示すものの選択数は自由という質問形式である。前段階までの作業によって、自由回答形式の回答は最終的には多重回答形式によって得られたデータと同等となり、その後の統計処理が可能となる。

事前に選択肢を示す形式に対し本研究の方法においては、事後に任意の分類レベルで選択肢を設定できるという利点がある。

3. 文章回答を中心とした意識調査の事例

ここでは、文章回答を中心としたアンケート調査を通じて本研究の分析方法の有効性を立証する。

実施した「北海道 21世紀の道路づくり」アンケートは、低人口密度と廉価な地価の北海道において今後どのような道路づくりが期待されるかを調査したものである。今後の道路づくりにおいては、高速道路の早期延伸や自転車道の整備など対応が明確な課題ばかりでなく、交通情報の内容や高齢者に配慮した道路あるいは地域性ある道路など対応策自体が不明確な課題も多い。そこで、図-2に示す文章回答形式による調査が有効と考えられた。

調査対象は、①網走管内の一般ドライバー、②北海道内の道路専門家、③北海道外の道路専門家の3群とし、1,374票の郵送配布に対して511票を回収した（回収率37.2%）。

図-3は、「4車線化の利点や利用法」への回答のキーワードの構造図である。中分類の「渋滞防止」におけるキーワード群は単位キーワードとしての「交通」、「混雑」を中心に構成している。なお、本設問に関する回答者は335人（手順1）、整形した回答件数は597件（手順2）、キーワード数は360語（手順3）であった。

表-1は、各設問に対する文章回答の内容を大分

III. ここでは、北海道の幹線道路（マチとマチを結ぶ道路）の整備について具体的な提言をお聞きます。以下の質問には文章でお答えください。なお、集計の都合上、文章はできるだけ短く箇条書きで記入されるようお願いいたします。 (例: △△のために〇〇が必要。××を〇〇する。...)
a. 幹線道路の休憩場（駐車場）にはどのような設備を設けるべきですか。
b. どんな種類の交通情報（案内を含む）があれば便利ですか。
c. すべて片側2車線以上の広い道路としたらどんな利点や利用法がありますか。
d. 高齢者のドライバーが安心して運転できる道路とはどのような道路ですか。
e. 冬期間の自動車の運転が安全で確実であるためにはどんな対策が必要ですか。
注) 回答欄（3行の空白行）はいずれも省略

図-2 文章回答形式による調査票

高速度性向上	渋滞防止	交通混雑、交通混雑緩和、交通止め防止、交通渋滞、交通障害対策、交通容量、交通流、混雑、混雑しない、混雑解消、混雑緩和、混雑防止、渋滞、渋滞解消、渋滞防止、通行止め、片側交互交通、輸送量、輸送力、流れ
	追越し容易	安全な追越し、追い越し可能、追越、追越し、追越し禁止、追越し禁止、追越可能、追越緩和、追越規制、追越自由、追越車線
安全性向上	時間短縮	スピード向上、スムーズ、規制速度、経済、高速度性、高速走行、高速道路、時間、時間短縮、所要時間、走行時間、速度アップ、定時性、物流、旅行時間短縮
	事故防止	安全、安全性、安全性向上、巻き添え事故、交通安全、交通安全確保、交通死亡事故、交通事故、交通事故減少、交通事故処理、交通事故防止、死亡事故、事故、事故件数、事故減少、事故処理、自動車事故、車両事故減少、重大事故、衝突、衝突事故、衝突事故減少、正面事故、正面衝突、正面事故、走行安全性、大型事故、追越事故
快速性向上	追越し容易	安全な追越し、追い越し、追い越し可能、追越、追越し、追越し禁止、追越し禁止、追越可能、追越緩和、追越規制、追越自由、追越車線
	自由速度	イライラ、いらいら運転、イライラ運転、いらいら防止、イライラ防止、ブレーキ回数、マイペース、マイペースで運転できる、マイペースで走れる、マイペースな速度で走れる、マイペース運転、ゆっくり運転、ゆっくり走れる、ゆっくり走り、ゆったり走り、ゆったり運転、リラックス運転、安心運転、安全運転、安全運転可能、安全走行、運転、運転恐怖感、運転疲労、快適、快適運転、快適性、快適走行、緩急速度で運転できる、居眠り運転、緊張緩和、経済速度で走行できる、視界、視界確保、視距、視野、前方、定時運転、疲労解消、疲労度、余裕、余裕ある運転、余裕運転、余裕走行
専用車線	追越し容易	安全な追越し、追い越し、追い越し可能、追越、追越し、追越し禁止、追越し禁止、追越可能、追越緩和、追越規制、追越自由、追越車線
	低速車線	緩衝帯、緩速車、緩速車と高速車、緩速車線、緩速車線と低速車線、高低速車、高低速車分離、低速車線、低速車レーン、低速車線、低速車分離、低速車面と高速車面、低速専用車線、低速走行可能、低速分離
多目的利用	専用レーン	トラック専用車線、バスレーン、バス専用レーン、バス専用通行帯、バス優先レーン、右左折、右折、右折車線、右折車線確保、貨物車線、観光、景色、交通事故処理車線、公共専用車線、高齢者、高齢者ドライブ、高齢者専用、高齢者対策、作業車線、自転車、自転車道、車種分離、車種別走行、車線分離、専用レーン、専用車線、大型車、大型車専用車線、大型車線、大型専用車線、追い越し車線、通過専用車線、登坂車線、片側2車線、流入車面
	駐車帯	休憩、休憩場、緊急避難場所、待避所、駐車、駐車スペース、駐車場、駐車帯、駐停車可能、駐停車帯、停車、停滞、非常待停車帯、臨時駐停車場
冬期利用	イベント	イベント、イベント可能、イベント開催、ソーラーカーレース、ツールド、マラソン、マラソン大会、ロードレース、祭り、地域行事
	その他	ヘリポート、メインテナンス、滑走路、簡易滑走路、簡易飛行場、救急用道路、緊急車面、緊急用滑走路、軽飛行機、航空機、事故車処理スペース、騒音防止、地下埋設物用地、道路清掃、道路補修、臨時滑走路
堆雪帯	堆雪帯	1時的堆雪スペース、一時的滑雪スペース、除雪、除雪スペース、除雪作業、除排雪、吹雪、積雪対策用道路、雪捨場、堆雪スペース、滑雪スペース、滑雪処理、冬期

図-3 「幹線道路の4車線化の利点・利用法」への文章回答のキーワード構造図

表-1 大分類キーワードによる文章回答のクロス集計結果

質問内容	大分類	中分類	全体	地域別			性別		年齢別		
				管内	道内	道外	男性	女性	若年	中年	老年
休憩場に 必要な設備	衛生設備	トイレ、ゴミ、水道	89.6	87.2	93.9	69.2	88.3	92.5	83.7	89.5	92.7
	情報設備	電話、案内	44.5	38.3	48.9	33.3	46.7	39.6	32.6	48.5	39.0
	飲食施設	売店、飲食	25.8	13.8	26.4	51.3	27.2	22.6	25.6	23.4	32.9
	休憩施設	ベンチ、休憩	23.4	13.8	26.0	30.8	24.9	19.8	25.6	21.8	26.8
	環境整備	照明、樹木、園地	22.8	9.6	29.4	15.4	26.8	13.2	14.0	27.6	13.4
	その他	遊具、その他	9.9	4.3	9.1	28.2	12.5	3.8	4.7	10.5	11.0
欲しい 交通情報	気象情報	路面情報、気象情報	44.7	38.9	52.3	18.9	43.6	47.4	36.1	43.9	50.6
	渋滞情報	渋滞状況、通行規制	43.5	26.7	46.7	67.6	43.6	43.6	44.4	41.0	49.4
	地域案内	地域案内	41.0	36.7	39.0	62.2	45.7	26.9	36.1	40.5	44.4
	標識内容	距離情報、標識改善	31.1	24.4	30.8	48.6	35.4	17.9	27.8	33.7	25.9
	案内方法	案内方法	7.8	10.0	5.1	16.2	7.8	7.7	8.3	8.3	6.2
	その他	その他	1.9	0.0	2.6	2.7	1.6	2.6	0.0	1.5	3.7
4車線化 の利点	高速性向上	渋滞防止、時間短縮、追越し容易	51.6	47.3	55.0	42.9	52.2	50.6	47.2	52.1	52.3
	安全性向上	事故防止、追越し容易	49.0	37.4	57.4	28.6	51.4	42.0	33.3	46.0	62.8
	快適性向上	自由速度、追越し容易	37.3	35.2	42.1	14.3	39.1	32.1	47.2	32.4	45.3
	専用車線	低速車線、専用レーン	26.6	29.7	21.5	48.6	29.2	17.3	25.0	28.2	23.3
	多目的利用	駐車帯、イベント、その他	12.8	15.4	11.0	17.1	15.4	4.9	16.7	13.6	9.3
	冬期利用	堆雪帯	8.1	1.1	9.1	20.0	9.5	3.7	2.8	7.5	11.6
高齢運転者 への配慮	低速車線	4車線、専用車線	44.0	35.2	51.4	23.7	46.7	36.5	30.6	42.7	52.9
	線形改良	急カーブ、急勾配、視界	23.9	20.9	25.2	23.7	23.0	27.1	16.7	25.5	23.0
	安全管理	照明、交差点、路面管理、安全施設	22.2	20.9	22.9	21.1	24.1	16.5	19.4	21.8	24.1
	交通情報	交通情報	21.9	16.5	19.6	47.4	22.2	21.2	33.3	20.9	19.5
	幅員拡張	広幅員、歩道など	17.8	9.9	23.4	5.3	16.7	21.2	13.9	20.0	13.8
	休憩施設	休憩施設	8.2	1.1	12.6	0.0	8.2	8.2	0.0	10.5	5.7
	その他	その他	12.8	8.8	15.9	5.3	13.6	10.6	11.1	12.7	13.8
冬期交通の 安全確保	雪処理	融雪、除雪、防雪、わだち、排雪	86.7	86.3	87.3	84.6	86.0	89.5	88.4	85.3	89.7
	視覚情報	視界確保、情報、標識	22.7	13.7	24.4	35.9	27.2	10.5	14.0	23.7	24.1
	道路構造	線形改良、道路幅拡張	16.3	9.8	21.3	5.1	19.6	7.4	11.6	17.7	14.9
	その他	自動車改良、運転意識、その他	21.3	25.5	20.4	15.4	21.9	20.0	18.6	20.7	24.1

注) 地域：管内（網走管内）、道内（管内を除く北海道）、道外（北海道を除く全国）
 年齢：若年（20～39才）、中年（40～59才）、老年（60才以上）
 反応%：回答者総数に対する反応者（キーワード使用者）の割合（%）

類のレベルで分類し、地域別、性別、年齢別の反応を示したものである（手順5）。この段階の結果は多重選択肢回答の結果をクロス集計したものと何ら変わるものではない。例えば、「4車線化の利点」の設問に対しては、全体の51.6%の人が「高速性の向上」を指摘している。「安全性の向上」に関しては道内の57.4%に対して道外は28.6%とその認識の差が大きいことなどが分かる。

5. まとめ

本研究の成果と展望を以下にまとめる。

- ① 本研究においては、従来、統計処理が困難と言われてきたアンケート調査の文章回答データの実用的な集計プロセスを提示した。
- ② 本研究の方法の最大の特徴は回答文章のキーワードに着目している点にある。その客観的抽出に向けて、回答文章の整形化、単位キーワードの抽出、キーワードの構造化の方法を示した。

- ③ 最終的に文章回答データは多重回答データと同様に扱えることを示した。多重回答が事前に選択肢を与えるのに対し、文章回答は事後に任意の分類レベルで選択肢を与えることが出来る。
- ④ 文章回答を主体としたアンケート調査を実施し、文章回答形式による調査の有用性を示すとともに、分析を通じて本研究の方法の有効性を立証した。
- ⑤ 今後はキーワードの構造化の客観性の向上に努めるとともに、文献情報の分析に適用範囲を広げる予定である。

<参考文献>

- 1) 日本能率協会編：KJ法入門、日本能率協会、1981.7
- 2) 大井 紘・宮本定明・阿部 治・勝矢淳雄：生活環境に関する住民の認知の広がり構造、土木学会論文集、第389号/Ⅳ-8、1988
- 3) 中岡良司・森 弘：アンケート調査におけるアリーアンスーの設計・解析に関する研究、第5回土木計画学研究発表会講演集、1983.1
- 4) 中岡良司・森 弘：アンケート調査における自由回答の集計技法に関する研究、土木学会北海道支部論文報告集、第49号、1993.2