

自由記述型アンケート回答を対象にした要求意図特定スキーマの提案

乾 裕子^{1&2} 高梨 克也¹ 井佐原 均^{1&2}通信総合研究所¹ 神戸大学大学院²

{hinui,takanasi,isahara}@crl.go.jp

1. はじめに

筆者らは、柔軟かつ確実な方法で、より多種多様な意見を収集蓄積し、これらの意見から発言者の意図を適切に取り出すことに関心がある。本研究では、意見収集の方法としてのアンケート調査に着目し、自由記述アンケートの回答テキストから回答者の意図を的確に取り出す方法について、また、それらの意図をどのような体系に整理すればよいかについて検討している[6][8][9]。

本稿の課題は、アンケートの回答から読み手(調査者)が回答者の要求意図を認識・理解するために、1) どのような知識構造を用意すべきか、2) 知識構造を構成する要素は何か、3) 各要素は回答からどのように発見できるのか、を明らかにすることである。そこで、回答には回答者の要求意図が必ず含まれるという仮説を前提とすることにより、回答の要求内容および動機・根拠と、焦点である要求意図との間の有機的構造と推論関係を明らかにする。このため、回答の要求意図、内容、根拠について、およびこれらの要素間の関連について分析を行う。分析で得られた知見をもとに、この構造と関係を記述する方法として要求意図特定スキーマを提案する。

本研究全体の目標として、要求意図に関連する上記要素のタグ化や、自動学習が可能なタグの弁別、これらのタグに基づく意図の自動抽出やテキスト分類を目指しており、自動化の可能性についても検討している[7]。

2. コミュニケーションとしてのアンケート調査

2.1 アンケート調査と自由記述回答の特徴

アンケート調査は、従来から心理学・社会学の分野においてデータの収集方法として発達し、社会的にも市場調査や意識調査などで多用されてきた。一般に「質問を行い」「回答としてデータを収集し」「解析することで役立つ情報を引き出す」といった表現で定義されるように[3]、アンケート調査はしばしば、調査者の一方的な行為とみなされている。しかし、本来「質問を行う」という行為の反応として「回答する」という回答者の行為があり、その行為に対して調査者は「データを収集し解析する」という反応を返している。さらに、アンケートの目的や種類によっては、解析した結果を調査側が回答者に対して公開し、それにより回答者は自分の意見や考えがどのように位置付けられたかを知るといった段階を想定することもできる。すなわち、アンケート調査とは調査者と回答者との間のコミュニケーションの一種と捉えることができる。

アンケート調査の方法論は、質問紙調査法としてさまざまな種別・手法が研究考案されてきた。自由回答は一般に自由記述形式とも呼ばれる回答形式のひとつである。自由回答には、回答しにくいために無回答が増える、的外れの回答がある、分類(コーディング)に時間がかかること[1]から、心理学や社会学におけるアンケート調査では対象外にされる傾向にあった。

2.2 他者想定型テキストとしてアンケート回答

近年、自由記述形式のアンケート調査は、選択型回答形式

と異なり、回答者の自由な意見を集約できる効果があるため社会的にも注目されている。そこで、本研究では、従来の自由記述型アンケート回答の分析・分類における、1) 特定の要素のみに着目して情報を抽出することに重点を置いたため自由記述である長所を活用できない、2) 回答テキストからどのような情報を取り出すかの焦点化が十分でない、といった問題点を解消するため、次のような研究方針を掲げる。それは、先に述べたようにアンケートを調査者と回答者のインタラクションとみなすことにより、アンケート回答を他者想定型テキストと捉えることである。

他者想定型テキストとは、新聞などの方向型テキストと異なり、テキスト中に「読み手への働きかけ」を含む。とくに、新製品の開発、あるいは地域開発・将来の事業計画などの意見集約におけるアンケートの場合には、この発語行為の多くが要求であると推測できる。本研究で扱う将来の道路計画に関するアンケートの自由記述においても同様の傾向を観察することができる。同時に「読み手への働きかけ」である以上、読み手である調査者は、回答者の要求意図を適切に汲み取り、要求意図は何か、なぜそのような要求がなされたかを判定できなければならない。

2.3 分析対象データ

本研究で分析対象としたデータは、「21世紀の道を考える委員会」が平成8年5月から7月末に実施したもので、将来的な道路計画に市民の声を活かす目的で行われている[5]。回答者数は35,674人、回答件数は延べ113,316件である。意見は、ハガキ、封書、FAX、電子メールによる回答の他、ホームページへの書き込みによって集められている。回答方法は、あらかじめ設定された道づくりに関する12個のテーマ(渋滞の解消、交通安全の確保、情報通信技術と交通など)の中から、回答者が各自関心の高いテーマを選択する。各テーマに対し、4個程度の参考意見およびグラフや図などの参考資料が提示されている。参考資料をもとに、120字程度の文字を記入できる回答欄に自由に記述し、書ききれない場合には別紙に記入する。

3. 要求意図の分析

ここでは、まず本研究における分析の方針と概要について述べ、次に事例を挙げて具体的に分析過程を説明する。

3.1 分析方針

アンケート調査が調査者と回答者のインタラクティブな行為であることに着目し、プランニング研究の枠組みである信念欲求モデルを、要求意図を解釈するためのスキーマ(知識構造)として整形する。このスキーマを用いて、明示されていない意図内容や動機を要求意図からアブダクティブに推論する方法を提案する。本研究では、このスキーマを「要求意図特定スキーマ」と呼ぶ。要求意図特定スキーマは、意図が信念と欲求から生成される一般的な信念欲求モデルを利用して、下記のように記述することができる。

回答者の要求意図特定スキーマ

= 要求根拠 (信念1, 信念2, … 信念n)
+ 要求内容 (内容1, 内容2, … 内容n)
+ 要求意図 (社会的/個人的欲求)

構成要素である要求意図、要求内容、要求根拠を特定するにあたって、理論上、各要素がすべて表現されている場合から、いずれか二要素がある場合、一要素だけある場合、すべて表現されていない場合まで8通りの組み合わせが考えられる(表1)。また、表の要求意図の有無に着目すると、a~dを明示的な要求意図、e~hを暗黙的な要求意図と区別できると考えられる。

表1 作業仮説：回答に現れる要求要素のパターン

要求要素のパターン	要求意図	要求内容	要求根拠	要素の上位範疇
a	○	○	○	明示的 要求意図
b	○	○	×	
c	○	×	×	
d	○	×	○	
e	×	○	○	暗黙的 要求意図
f	×	×	○	
g	×	○	×	
h	×	×	×	

上記を仮説として、実データのアンケート回答を対象にこれらのパターンの違いが回答中にどのように現れるか、また、現れていない要素が他の要素から特定できるかどうか分析する。それぞれの違いについては後で詳しく述べる。

3.2 分析作業手順

ここでは、分析の手順を決め、これに基づいて「要求意図分析データベース」を作成する。以下の手続きにしたがって、分析を進めながら項目列を増やすという発見的方法によって作成している。最終的に構築されるデータベースの項目は、要求意図特定スキーマの記述項目でもある。

要求意図分析およびデータベース構築の作業手順

- 1) 要求意図を取り出す
 - 1A) 回答文に直接要求表現(てほしい、ください等)が含まれるか否か調べる
 - 1B) 回答文の文末を「てほしい」という直接要求表現、もしくは「(~する)ことを要望する」という表現に言い換えることが可能か否か調べる
- 2) 回答の1以外の部分を要求内容とみなす
- 3) 要求内容の中に、内容の根拠に相当する要素があれば区別する(原因・理由、動機を示す接続助詞などに着目。それ以外は目視により意味から判断する)
- 4) 要求内容を異なる内容ごと(節単位)に分割し、内容間、および内容と根拠との順序関係(因果関係含む)を記述する
- 5) 順序関係を記述する際、内容や根拠が回答に明確に記述されていなくても、前後関係からこれらの各要素が十分推論可能な場合には、分析者が適宜補充する
- 6) 内容や根拠の中に、回答者の感情や信念に対する判断・評価などが現れていれば記述する
- 7) 根拠が複数の信念から成っていれば分け、それらの間の関係を記述する
- 8) 7のうち、要求の根拠に直接結びつかない信念は区別する

分析にあたってまず、アンケート回答の各事例を対象に要求意図の有無を区別する(手順1)。1のA,Bはいずれも表層的な情報を手がかりとはしているものの、本研究が目指すのは表層表現からの意図抽出ではない。上記の組み合わせから、少なくとも1A,1Bは下記四つの型に区別できる。

しかし、表層情報だけを手がかりにするならば、約40%を占める④を対象とすることはできないからである。

表2 明示的要求意図と暗黙的要求意図の区別

要求要素のパターン	要素の上位範疇	直接要求表現	言い換え	3000文中
① a~d	明示的	含む	可	547
② —	—	含む	不可	3
③ a~d	明示的	含まない	可	1190
④ e~h	暗黙的	含まない	不可	1252

注：①~④に含まれない8文については、前後文脈を見ないと判断に迷う例。②の要求表現は引用文中に現れているため対象外とする

次に手順2、3では要求の内容と根拠を特定する。ここまでの作業を行った段階で、要求意図分析データベースの基本項目列が整う。

この基本の骨組みをもとに、例えば、要求の根拠となる信念が形成されるプロセス(手順5)や、要求の根拠が複数の信念から構成されている場合には各信念間の関係(手順7)、また、要求には直接結びつかない信念(手順8)などを、必要に応じて基本項目列と区別しながら項目列を追加している。

3.3 要求意図の分析

分析の結果、任意の約3000事例を対象にした分析では、表1の8パターンのうち、要求意図(C)のみ以外のパターンは実際にデータとして現れた。これら残りの7パターンの回答では、1)要求意図の明確な事例の場合、1a)要求根拠から内容を、1b)要求内容から根拠を、2)要求意図の不明な事例の場合、2a)根拠から内容と要求意図を、2b)内容から根拠と要求意図を、それぞれ特定できることが判明した。以下、事例を挙げながら具体的に説明する。

3.3.1 明示的要求意図

ここで扱うのは、「てほしい」「てください」あるいは動詞の命令形といった表現で表されるような直接要求表現を含む文である。直接要求表現は、[4]に挙げられた80表現を対象としている。また、直接要求表現でなくても、ほぼ同様の意味機能を持つ表現を含む場合には要求意図を持つと判断できる[6]。

a) 要求意図(I)+内容(C)+根拠(R)

- #1 国道10号線の渋滞がひどいから、早くバイパスを整備し、(c)てもらいたい。(r)
- #2 歩道がせまいので、(r)広くし、(c)てくれ。(r)
- #3 車の駐車違反がとんでも多くて、自転車をのっているとその車がじゃまでその車をよけると車にぶつかりそうになってとても危ない。(r)。もっと駐車違反をきびしくとりしめる。(c)必要があると思う。(r)

このパターンを最初に観察する動機は、aの例に現れるすべての要素とその関係を、他のパターンの場合にも推論によって発見できるかどうかを検討することにある。

#1で「国道10号線の渋滞がひどい」という根拠(R)は、回答者の事実認識から形成された信念に基づいている。また「早くバイパスを整備する」(C)「てもらいたい」(I)のように、すべての要求要素が明示的で、要求要素と表層の表現が一致している例である。#2では「歩道がせまいので(歩道)を広くしてくれ」のように、既出の要求内容の対象を省略補充することで、要求要素のスロットをすべて埋めることができる。

なお、本研究での分析は、#3に見られるように、根拠と内容・要求意図が二文に渡っている事例も、また、#1、#2には示されていた「から」「ので」のような原因・理由の接続表現が明示されない事例も対象とする¹。表層的な処理だけでは十分に明らかにできない意図の取り扱いについて検討している。

b) 要求意図(I)+内容(C)→根拠(R)

#3 歩道を広くし、(c)てほしい_(i)

この事例を見て「この回答者はなぜ歩道を広くして欲しいのか」と問われたとき、回答者本人でなくても「現在ある歩道が狭いからだろう」と推測することができる。実際には「歩道が狭くて危険である」「歩道が狭いので歩行者が車道を歩いて困る」など、根拠の中にも因果的に関係づけられる複数の信念を含む場合もある。これらを図式化したものが要求形成および解釈プロセスである(図1)。

このボタンでもっとも重要なことは、意図(I)と内容(C)によって示された要求(D)が提示される際、要求の前提には根拠(R)を形成する信念(r)がある、信念(r)は内容(C)を否定するものである、という推論によって、明示されていない根拠をアブダクティブに発見できることである。

この推論は(C)の否定が回答者の信念(r)であること(図1③)、信念(r)がネガティブな評価(狭い/狭いのは嫌だ)に裏付けられていること(図1④)によって保障されている。また、前述したaタイプのように、根拠が明示されている場合、①から④までのように信念間の推論関係を想定することもできる。

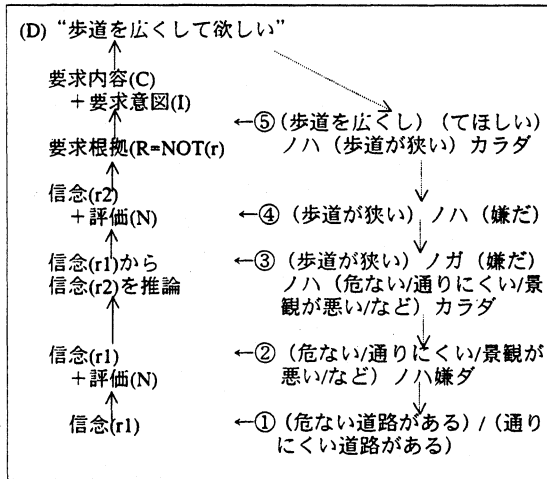


図1 要求形成および解釈のプロセス

c) 要求意図(I)→?

このタイプの事例は、対象とした3000文のデータには現れなかった。しかし、要求意図しか表れない文として例えば下記のような例が想定できる。

#4 ちゃんとしてほしい_(i)/なんとかしてほしい_(i) (作例)

本研究で使用した回答テキストのアンケート調査は、先に述べたようにあらかじめ設定された道づくりに関する12

¹ 表層における因果関係は実際には要求と根拠の関係であることが多い。両関係の区別については[7]を参照のこと。

個のテーマがあり、回答者はそれにしたがって回答することができる。同時にこれは分析者(あるいは調査者)にとって、要求内容が明示されていなくても回答者がテーマを選択した段階で何について「ちゃんとしてほしい」と述べられたのかを推論できるということである。したがって、このタイプの場合、要求内容(C)などを補完するためには、質問紙の設計を考慮する必要がある。

d) 要求意図(I)+根拠(R)→内容(C)

#5 週末になると、小倉競馬場周辺の車の渋滞がはげしいので、(R)、なんとかしてほしい_(i)

#6 道路が凸凹になって水たまりができていますので、(R)、ちゃんとしてほしい_(i)

上記に見られるように、「何に対して」「何をするのか」といった要求内容が明示的でない例がある。このような場合、aの#2の例と同様に、要求根拠に現れている既出の対象(「小倉競馬場周辺の車の渋滞」「道路の凸凹」)を省略補完する。さらに「渋滞をなんとかする/道路の凸凹をちゃんとする」から「渋滞を解消する/道路の凸凹をなくす」を推論するという二つの段階を踏むことにより、要求内容を特定できる。

以上の分析を踏まえ、明示的要求意図における要求要素の特定を下記のようにまとめる。

記号の定義: Demand:調査者が理解すべき回答者の要求,
I:要求意図, C:要求内容, R:要求根拠(r:信念),
φ:空きスロット

要求および要素特定のパタン:

- a) $R(r1, r2...rn) + C(c1, c2...cn) + I(i) = Demand$
但し、C中の行為対象はRから省略補完する場合がある
- b) $R(\phi) + C(c1, c2...cn) + I(i) = Demand$
但し、NOT(C)に対するネガティブな評価からRをアブダクティブに特定する
- c) $R(\phi) + C(\phi) + I(i) \neq Demand$
但し、質問紙の設計に依存してCを特定できる場合がある
- d) $R(r1, r2...rn) + C(\phi) + I(i) = Demand$
但し、C中の行為対象をRから省略補完しCを形成する場合がある。さらにCからCをアブダクティブに特定する

3.3.2 暗黙的要求意図

ここでは、直接要求表現を含まず、かつ言い換え不可の文を対象としている。

e) 内容(C)+根拠(R)→要求意図(I)

#7 高速道での事故対面交通箇所通過時に「ぶつかる」という不安を感じます_(R)。「八代〜人吉」間、同じ料金を支払うのであれば早期解消を_(i)

これは、二文からなる一回答中、前文が要求の根拠を、後文が内容を示している例である。この事例で「早期解消(してほしい)」という要求意図が特定できるのは、格助詞で終わる表現的特徴に依存するところが大きい。現在のところ、このボタンにおいて根拠と内容からアブダクティブに要求意図が特定できる例は見つかっていない。

f) 根拠(R)→要求意図(I), 内容(C)

#8 歩道がせまい。

これは、#3の解釈プロセスを逆方向に辿る、すなわち「歩

道が狭い→(歩道が狭いのは危険である)→(危険は嫌だ)→(広くしてほしい)」という推論プロセスにより要求意図と内容が特定できる事例である。このように要求意図特定スキーマを用いた推論では双方向の推論が可能である。

g) 内容(C)→要求意図(I), 根拠(R)

#9 高速道路の料金は JH が決めているが、もっと経済原則が反映するような決め方、市場原理が働くような値決め方式があるのではないか。

この事例は、要求意図と根拠だけでなく正確に言えば要求内容自体も下記に示す推論プロセスを経ることによって明らかになるものである。これについて、詳しく説明する。

1. もっと経済原則が反映するような決め方、市場原理が働くような値決め方式があるのではないか
2. (いや、あるはずだ_(x))
3. (しかし、実際には存在しない_(r1-R))
4. (r1 が無いのは嫌だ)
5. (よって、r1 の状態を要求する)
6. (すなわち、もっと経済原則が反映するような決め方、市場原理が働くような値決め方式を作_(C)てほしい_(I))

上記推論プロセスでは、1~2の間でいわゆる反語表現の解釈を行っている。この補完によって、#9における信念 x を決めることができ、また、この信念 x により x の否定である信念 r1 を決めることができる。これによってボタン b の推論プロセスを経ることができるため、上記 3 や 6 に見られるような根拠や内容の役割を担うことが明らかになる。推論により、x の否定が根拠であり、x が結果的に内容となる。最終的に現れる内容が表現として明示されているという点から、この事例を g のボタンと位置付けている。

#10 国道 200 号線の渋滞

要求意図が明示されていないものの、#10では「渋滞を解消してほしい」のように、要求意図を表現として補完することにより b のボタンになる。したがって、根拠も特定することができる。#9ではネガティブな評価が「のはいないか」という反語的表現からの推論によって行われるのに対し、#10の場合には「渋滞」という事象を示す表現から推論される。

h) 要求要素がないもの

#11 別に今のままでいいと思う。

すべての要求要素がなければ回答者の要求は無しとする。以上の分析から、暗黙的要求意図における要求要素の特定を下記のようにまとめる。

記号の定義：(前述のとおり)

要求および要素特定のボタン：

$$e) R(r1, r2, \dots, rn) + C(c1, c2, \dots, cn) + I(\phi) = Demand$$

但し、I は表層的に特定できる場合がある

$$f) R(r1, r2, \dots, rn) + C(\phi) + I(\phi) = Demand$$

但し、R に対するネガティブな評価により NOT(R) から C をアブダクティブに特定する

$$g) R(\phi) + C(c1, c2, \dots, cn) + I(\phi) = Demand$$

但し、NOT(C)に対するネガティブな評価から R をアブダクティブに特定する。また、C から I を特定するとき、機能的側面に拠る場合と内容的側面に拠る場合がある。

$$h) R(\phi) + C(\phi) + I(\phi) \neq Demand$$

4. 要求意図特定スキーマの提案

以上の分析から明らかになってきたように、明示されていない要求要素を特定する際に必要になるいくつかの情報がある。これらの情報が要求意図特定スキーマの項目であり、要求要素をアブダクティブに特定するために利用できる情報であると共に制約や条件となる。

要求意図分析データベースの作成過程の説明で述べた項目、および分析の説明で述べた項目を下記項目に挙げる。

- ・要求根拠になる信念
 - 信念 p1~信念 n
 - 信念間の関係 1~n
- ・要求内容か根拠かを推論する過程で仮定する信念 x
- ・要求根拠にならない信念 (信念 0, 評価, 態度, 態度表明表現)
- ・信念に対する回答者の評価 (ポジティブ/ネガティブ)
- ・要求解釈プロセス
 - ・文間の関係 (因果的/並列的)
 - ・節間の関係 (因果的/並列的)
- ・要求内容
- ・要求意図 (要求程度, 要求意図表現, 直接要求表現を含むか否か, 直接要求表現への言い換えの可否)

なお、上記以外の項目として；

- ・回答者がどのような視点で要求しているか (個人的欲求/社会的欲求)
- ・誰に対する要求か
- ・信念は経験的か概念的か

といった、項目についても試行的に記述を進めている。

5. おわりに

本研究では、アンケートの回答から回答者の要求意図を特定するために、要求意図特定スキーマによる解釈のための知識構造を提案した。それにより、要求を構成する意図、内容、根拠が明示されていない場合、必要な要素を発見的に特定できることを実データの分析をもとに報告した。今後は、さらにスキーマの記述を進め、作成したスキーマによる記述の精度と簡潔さを高めると共に、スキーマ内の生起項から暗黙的な項目を自発的に発見できることのできる規則体系の構築を目指している。

謝辞：研究データとして道路審議会基本政策部会「21世紀の道を考える委員会」が実施されたボイス・レポートについて研究利用を快諾してくださった(財)国土技術研究センター調査第二部の前田様、川原様のご厚意に深謝いたします。

参考文献：

- [1] 続 有恒・村上英治：『心理学研究法 9 質問紙調査』, 東京大学出版会, 1975
- [2] 浅井 晃：『調査の技術』, 日科技連出版社, 1987
- [3] 辻新六・石馬昌宏：『アンケート調査の方法』朝倉書店, 1987
- [4] 森田良行・松木正恵：『日本語表現文型』アルク, 1989.
- [5] 道路審議会基本政策部会「21世紀の道を考える委員会」, ボイス・レポート, 建設省道路局 建設省都市局, 1996
- [6] 乾裕子・井佐原均：表層情報からの意図タグ判定の試み-白山記述アンケートを対象に-, 第7回言語処理学会年次大会発表論文集, pp437-440, 2001.
- [7] 高梨克也・井佐原均：発話理解の多重性と会話進行：「説明的発話連鎖」の分析から, 第7回言語処理学会年次大会発表論文集, pp417-420, 2001.
- [8] 乾裕子・兼重賢太郎・欠嶋宏光・井佐原均：アンケートというコミュニケーション-PI 手法を取り入れた意図の抽出方法-1-, 人知能学会研究会資料 SIG-SLUD-A102-9, pp51-pp56, 2001
- [9] Inui Hiroko, Murata Masaki, Uchimoto Kiyotaka and Isahara Hitoshi, Classification of Open-Ended Questionnaires based on Surface Information in Sentence Structure, NLPRS2001, pp-pp 2001.