

自由回答アンケートにおける要求内容の分析

内山 将夫 大塚 (乾) 裕子 井佐原 均
通信総合研究所

1 はじめに

本稿では、自由回答アンケートの回答として記述されている要求文から要求内容を同定するための分析方法について検討する。

まず、本稿で想定する自由回答アンケートとは、アンケートを主催する調査者側が与えたテーマについて、回答者側が、意見等を自由に記述する形式のアンケートを言う。特に、本稿では「21世紀の道路」に関するアンケートを題材とするが、このアンケートにおいては、たとえば「渋滞の解消」とか「高速道路の料金と道路整備」のようなテーマがあり、そのテーマにそって、回答者側が、たとえば「車道をもっと広くすれば良いと思う」とか「高速道路の料金償還の済んだ道路は、無料でも良いのではないか」とかの回答をしている。

このときの回答には、テーマに関して、要求、不満、提案、希望など、回答者の種々の態度が記述される。これらの態度の中で、我々が、特に注目しているのが要求を表現する回答である。より詳細には、我々は、回答を構成する各文に着目しており、その文でも、特に、要求を表現する文(要求文)に注目している。

ここでの要求文とは、典型的には「クルマは横断歩道で止まってほしい」のように「～てほしい」で表されるものである。しかし、このように明示的に「～てほしい」となっていない回答文であっても、たとえば、「路上駐車をなくすべきだ」とか「信号を減らして歩道橋を増やしたらどうか」のように、言語形式上は、義務・当然や提案といえる[10]のものであっても、回答者側や調査者側から解釈すれば、要求と考えて良い回答がある[9]。そのような要求文の認定方法については後述する。

このような要求文が認定されたあとで、次に分析すべき事柄の1つは、その要求文における要求内容を同定することである。そのための分析方法について検討することが本稿の主要な目的である。そこで、まず「要求内容を同定する」とはどういうことかを定義する。本稿では、要求文が与えられたとき、それを、ある種のパターン(あるいは、意味表現、テンプレート)に言い換えることを「要求内容を同定する」と言う。そして、

その言い換え結果をもって「要求内容」と考える¹。

具体的には、我々は、要求文は、名詞句Aと動詞句Vとに関して「Aについて、Vしてほしい」²に言い換え可能であると考えているため、要求文を、このパターンに言い換えることをもって、要求内容の同定と呼ぶ。たとえば、要求文として「クルマは横断歩道で止まること!」が与えられたとき、それを「クルマについては、横断歩道で止まってほしい」へと言い換える作業を要求内容の同定と呼ぶ。この結果、たとえば「A=クルマ」「V=横断歩道で止まる」などが分かる。本稿では、この言い換えパターンを用いた要求内容の同定について、その性質を検討する。

以下では、まず、我々が分析方法に必要なであるとする性質について述べ、つぎに、それらの性質を満すと考えている分析方法を述べる。それらは、言い換えに基づくものである。そのうち、まず、要求文を認定するときに利用する言い換えパターンを述べ、次に、要求内容を同定するときに使う言い換えパターンについて述べる。後者については、更に、実験を通して、その性質を調べる。最後に関連研究と結論を述べる。

2 分析方法に必要な性質

我々は、自由回答アンケートの分析方法に対して、3つの性質「再現性」「網羅性」「妥当性」が必要である³と考える。まず、再現性とは、ある与えられた回答文に対して、その分析方法を適用したとき、その適用者(作業員)が誰であっても、同一の分析結果が得られることを言う。次に、網羅性とは、全ての回答文に対して、適切な分析結果が期待できることを言う。最後に、妥当性とは、分析の結果として得られる内容が、そもそも、望むものとなっていることを言う。

再現性があるとは、たとえば、要求文を認定するときには、ある回答文において、任意の作業員間で、そ

¹したがって、分析が精密であれば、「要求内容」も精密になる。なるべく精密な分析をすることにより、適切な要求内容を同定することが、我々の目標である。

²「Aについて」により「Aについては」「Aについてから」「Aについてまで」等を代表させる。

³ここでの再現性と妥当性は心理測定における信頼性と妥当性[2]に相当する。

の文が要求文か否かの判定が一致することをいう。また、網羅性があるとは、任意の回答文を適切に処理できることを言う。たとえば、「道路を整備してほしい」を要求と判定するときには、「道路の整備!」のような、典型的とは言えないような文についても、それを要求文と判定してほしい。最後に、妥当性とは、分析の結果として、要求文と認定されたとき、それが本当に要求を表現しているかどうかである。これと同様なことは要求内容の同定についても言える。

本稿での主要な関心は、要求内容の同定についての、再現性と網羅性とである。妥当性の検証は今後の課題とする。以下では、言い換えパターンを用いた、要求文の認定、および、要求内容の同定について述べ、その後、要求内容の同定について、その再現性と網羅性について検討する。

3 要求文の認定

要求文の認定には、「～してほしい」という言い換えパターンを用いた。つまり、与えられた回答文が「～してほしい」に言い換えられるときには、それは要求文であり、そうでないときには、要求文でないとして判定した。たとえば「歩道を広くしてください」「歩道を広くできないのですか」「歩道を広くすべきだ」「歩道を広くしたらどうですか」などは、全て「歩道を広くしてほしい」に言い換え可能であるので、これらは要求文と判定する。一方「歩道が掃除されていて気持ちが良い」のような文は「～してほしい」には言い換え可能ではないので、非要求文と判定する。

ここで「～ですか」や「～べきだ」等のように、言語形式としては疑問や義務・当然等である文が、要求文と解釈できる理由は、その文が書かれた語用論的状况による。すなわち、このアンケートに対する回答という状況においては、「歩道を広くできないのですか」というような疑問形式の回答は、調査者側や回答者側からみれば、実際には「歩道を広くしてほしい」という要求と解釈することが適切である。その語用論的な根拠については、文献[9]を参照せよ。

文献[9]では、更に、このような言い換えによる要求文の判定が、言い換えを用いずに直感的に要求文を判定したときと比べて、作業員間の一致率、すなわち、再現性が高く要求文を判定できることを示している。また、判定基準をコーパスから学習することにより、約91%の精度で、要求文かどうかを自動判定できることを示している。

これらの結果は、言い換えを自由回答アンケートの分析に用いることの有効性を示している。以下では、この「～してほしい」というパターンを更に拡張することにより、要求内容の同定ができることを示す。

4 要求内容の同定

要求内容の同定に利用するパターンは、1節でも述べたように「Aについて、Vしてほしい」である。このパターンは、より詳細には、以下の2つの下位パターンに分けることができる。

1. Sが、Pに、Aについて、Vしてほしい
2. Sが、Aについて、Vしてほしい

ただし、Sは「Aについて、Vしてほしい」ということ（これを要求命題と呼ぶ）が実現することを要求している者（要求者）であり、Pは、Sから、要求命題を実現することを要求されている者（被要求者）である。また、Aを要求テーマと呼び、Vを要求行為と呼ぶ。

一般に、自由回答アンケートに回答する状況においては、Sは必ず存在すると考えられる。これは、通常、回答者（私）である。しかし、回答においては、Sが記述されることは、ほとんどない。

一方、Pもほとんど記述されることはないのだが、「ドライバーにマナーを守ってもらいたい」の例における「ドライバー」のように、記述される場合もある。

次に、Pが記述されていないときに、そもそもPが存在するかどうかを考える。これは、大体的場合には、存在していて、それは調査者である。しかし、回答者側の意識としては、Pが存在していないと解釈した方がよい回答もある。たとえば「道路は皆に公平であるべきだ」という文を、Pが調査者（この場合は行政）であると考えたと「道路については、皆に公平であるようにしてほしい」のように言い換えることができ、「A=道路」「V=皆に公平であるようにする」となる。一方、Pが存在しないとすると、これは、Vの部分が発言的な表現となる。つまり「道路については、皆に公平であってほしい」となり、「V=皆に公平である」となる。

いずれにしても、回答においては、SとPとは、ほとんど省略されるため、言い換えにおいては、Pの存在非存在に関らず、「Aについて、Vしてほしい」への言い換えを考慮すれば良い。

この言い換えにより同定されるAとVの性質について理解するために、複合格助詞である「について」の

用法を述べる。まず、複合格助詞の特徴は、それが格を表示することである [7]。さらに、「あいまいな関係、あるいは格助詞によって結び付けられないような結び付きの弱い関係にある二者をことさらに関係付ける」働き [8]がある (文献 [5]より孫引き)。また「について」は、それが関係する動詞にとっての「対象・関連」を表すものである [4]。加えて、「について」のあとには、「は」や「から」などが後接できるので、対象・関連に加えて、動作主や、起点などを表現することもできる。更に、Vとして「公平である」のように状態を表現するものが来ても良い。このときには「道路については」のように「は」が繫辞として後接する。

以上の性質より「Aについて、Vてほしい」に言い換えられたときのAは、Vと関連する名詞句であり、かつ、Vと様々な関係を取りうる⁴。そのため、網羅性高く、関連する名詞句を同定できる。たとえば「よりよい道が出来るよう、住民の代表と話し合う」における「話し合う」における、その話題のように、単独の格助詞ではVとの関連を付けるのが困難な例についても「よりよい道を作ることについて、住民の代表と話し合う」のように「について」によれば、関連付けをして、同定することが可能となる。

5 実験

「Aについて、Vてほしい」への言い換えが、網羅性や再現性高く実行できるかを調査した。そのときに利用した自由回答アンケートのデータは、道路審議会基本政策部会「21世紀の道を考える委員会」が平成8年5月から7月にかけて実施したアンケート調査の自由回答であり、回答人数35674人、回答数113316件の大規模調査である。このデータから、無作為に回答を抽出し、複数文からなる回答については1文ずつに分割した結果3001文の回答文が得られた。それから、3節で述べた方法により、要求文を認定した結果、1944文が要求文であった [9]。この1944文より、100文を無作為に抽出し、要求内容を同定することを試みた。

このときの作業者は2名である。1名(作業者1)は、本稿の第1著者であるので、言い換えの方法等につい

⁴後述の実験では、2名の作業者が、AとVを同定するとともに、その間にある関係も同定した。その関係としては、格助詞、および格助詞で適当なものがない場合には「は」を選択した。その結果、作業者間でAとVの双方が一致する89組について、その関係の数を調べたところ「を」が73「は」が7「に」が4「で」「が」「から」がそれぞれ1であった。なお、2組についてはAが推定不可能であったため、関係も推定不可能であった。また、このときの関係については、作業者2が選択したものについて数えた。

て良く把握している。もう1名(作業者2)は、大学の文学部の4年生である。この作業者には、作業仕様において、言い換えの事例を提示した。その主な作業内容は、言い換えの結果として、AやVに何がくるかを同定することである⁵。また、もし、1つの回答に複数のAとVの組がある場合には、その全てを同定した。

まず、網羅性について述べると、100文について、作業者1は、全てが「Aについて、Vてほしい」に言い換え可能であると判定し、作業者2は、1文を除いて言い換え可能であると判定した。これより、網羅性は高いといえる⁶。

次に再現性について述べる。それには、作業者間での一致と不一致の度合を調べる。一致の程度には、AとVの双方一致(BOTH)、Aのみ一致(MA)、Vのみ一致(MV)、AとVとの双方が不一致の4種類がある。また、双方が不一致のときには、作業者1のみ(O1)、あるいは、作業者2のみ(O2)に同定されたAとVの組がある。次に、BOTHとMAとMVとを合わせてMATCHとし⁷、O1とO2とMATCHとを合わせてTとすると、その統計量は表1になる。

	BOTH	MA	MV	O1	O2	MATCH	T
平均	.89	.04	.18	.28	.25	1.11	1.64
割合	.54	.02	.11	.17	.15	.68	1.00

表 1: 100 文中の諸元

ここで平均とは、100文について、各文における各項目に該当するAとVの組の平均であり、割合は、そのTにおける割合である。これより、1文あたり、平均1.64個の要求内容が同定されており、そのうち、1.11個が作業者間に共通して同定されていることがわかる。

これをもう少し詳しくみるために、回答文を、Tの数が1のもの58文とそれ以外のもの42文とに分けて、同様な統計量をとると表2と3のようになる。なお、T

⁵言い換えのパターンとしては「Sが、Pに、Aについて、Vてほしい」のみを指定したが、作業者2の同定したVの中には、その態(voice)が作業者1と異なるものがいくつかあった。そして態が異なる場合には、作業者1は使役(「色を付けさせる」等)であり、作業者2は自発(「色が付いている」等)等であった。この食い違いは、4節で述べたように、作業者1が「Pに」を意識しているのに対して、作業者2は「Pに」を意識していないためだと解釈した。

⁶作業者2が言い換え不可能と判定した回答文は「人専用、普通車専用、大型車専用など」というものである。これについて、作業者1は、この回答文に対するアンケートのテーマが「渋滞の解消」であることから、この回答は専用道路の建設に関することであると考え、「人専用、普通車専用、大型車専用などについて、作ってほしい」と言い換えた。

⁷A同士やV同士が作業者間で一致するとは、それぞれの主辞が(表記のゆれを許容して)一致することと定義する。こうした場合には、AかVかのどちらかが一致すれば、意味的には、AとVの組同士はほぼ同一の要求内容であった。そのため、これらをまとめてMATCHとする。

は、MATCHとO1とO2の和なので、Tは、1文中に同定された要求内容の総和となる。

	BOTH	MA	MV	O1	O2	MATCH	T
平均	.81	.02	.16	.02	0.0	.98	1.0
割合	.81	.02	.16	.02	0.0	.98	1.0

表 2: $T = 1$ である58文中の諸元

	BOTH	MA	MV	O1	O2	MATCH	T
平均	1.0	.07	.21	.64	.60	1.29	2.52
割合	.40	.03	.08	.25	.24	.51	1.0

表 3: $T > 1$ である42文中の諸元

これより、要求内容が1つしかない場合には、その同定結果は、ほぼ一致するといえ、複数ある場合には、半数程度が一致すると言える。なお、O1やO2とされたなかには、意味的には同等なものもあるが、その同等性の判断は難しいため、主辞の不一致により不一致と判定したものもある。そのため、もし意味的な同等性を考慮すれば、もう少し一致の割合は高くなる。

表1,2,3から、1つの要求文について、作業員間では、1つの共通する要求内容を同定するが、その他の要求内容については、必ずしも一致しないと言える。作業員間で一致して同定されやすい要求内容は、主節で記述されているものであり、不一致の割合が高いものは、従属節で記述されているものである。不一致なものの詳細な分析と、その不一致の解消については、今後の課題とする。

6 関連研究

自由回答アンケートにおける要求内容を同定する標準的な方法は、確立していない。関連するものの1つは、テキストマイニングのように、アンケート全体において頻出する言語表現に注目し、それを抽出する方法である[1]。しかし、頻出する言語表現に注目した場合には、ただ1回しか出現しないようなものを拾うことはできない。

また、情報抽出[6]のように、ある種のテンプレートを設けて、そのスロットを埋める方法もあるが、この方法では、そのスロットにあらかじめ定められた言語表現の型に合致しないようなものは取れない。

そのため、これらの方法では、我々の目標である「全ての回答文を適切に処理すること」はできない。

ここで、要求内容を同定するために、格解析をすることも考えられる⁸。しかし、格解析は、人間にとって

⁸アンケート分析への格解析の利用例には文献[11, 3]がある。

も困難な場合がある⁹ので網羅性は低くなると考える。また、たとえ格解析をしたとしても、その後で、複数ある動詞句のなかから、どの動詞句が要求行為かを判定するとともに、どの格の名詞句が要求にとって重要であるかを定める必要があるため、格解析だけでは不十分である。ここで、名詞句の重要性の判定の必要性は、たとえば、アンケート全体から要求内容を抽出し、それを要約して提示しようとするときなどに生じる。これについては、我々は「Aについて」で取り出されるAが要求にとって重要な名詞句であると考えているが、それについての検証は今後の課題である。また、AとVにおける、単なる「関連」よりも詳細な関係の分析も今後の課題とする。

7 おわりに

本稿では、自由回答アンケートにおける要求文から、要求内容を、言い換えに基づき同定する方法を述べた。その方法は、要求文を「～について～してほしい」に言い換えるというものである。これにより、網羅性高く要求文を取り扱うことができ、また、作業員間で、1文につき、1つは共通して要求内容が同定できることを述べた。今後の課題は、提案手法の自動化、および、提案手法を実際のアンケート分析に応用することである。

参考文献

- [1] 特集「テキストマイニング」. 人工知能学会誌, 16(2), 2001.
- [2] 繁樹算男. 心理測定法. 放送大学教育振興会, 1998.
- [3] 松村真宏, 川原大輔, 岡本雅史, 黒橋禎夫, 西田豊明. 文の背後に潜む「問い」の抽出. ことば工学研究会資料 SIG-LSE-A301-1, 2003.
- [4] 松木正恵. 複合辞の認定基準・尺度設定の試み. 早稲田大学日本語研究教育センター紀要, Vol. 2, pp. 27-52, 1990.
- [5] 松木正恵. 複合辞性をどうとらえるか. 辻村敏樹教授古稀記念論文集刊行会(編)『辻村敏樹教授古稀記念日本語史の諸問題』, 明治書院, pp. 590-606, 1992.
- [6] 関根聡. テキストからの情報抽出. 情報処理, 40(4), 1999.
- [7] 佐伯哲夫. 複合格助詞について. 言語生活, 178:80-88, 1966.
- [8] 砂川有里子. 複合格助詞について. 日本語教育, 62:42-55, 1987.
- [9] 大塚裕子, 内山将夫, 井佐原均. 自由回答アンケートにおける要求意図判定基準. 自然言語処理(採録), 2004.
- [10] 森田良行, 松木正恵. 日本語表現文型. アルク, 1989.
- [11] 高橋和子. 自由回答のコーディング支援. 理論と方法, Vol. 15, No. 1, pp. 149-164, 2000.

⁹たとえば以下の回答文は格解析が困難だろう「途中迄完成使用され、残存未着工部分の該当地域住民には、高層住宅を建築入居して働き速やかに完成しなければ当初の建設目的も達成されず協力立退きに応じて載いた人達の事も考慮すべきでなかろうか」。ちなみに、これは「道づくりと合意形成」というテーマについての回答文であり、これから同定された要求内容は「残存未着工部分について、速やかに完成してほしい」と「協力立退きに応じて載いた人達の事について、考慮してほしい」とである。