

表層情報からの意図タグ判定の試み - 自由記述アンケートを対象に -

乾 裕子

井佐原 均

九州工業大学大学院情報工学研究科, 計量計画研究所

総務省通信総合研究所

h\_inui@pluto.ai.kyutech.ac.jp, hinui@ibs.or.jp

isahara@crl.go.jp

1. はじめに

1.1 背景

自由記述アンケートは、選択型回答と異なり、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[3,4,9]。従来の高頻出語などをキーワードにした解析[2,3]に加え、回答者の印象[8]・不満[8]を示す表現に着目した研究なども進められている。しかし、自由回答の解析方法は現時点で十分に確立されておらず、回答の内容にしたがった分類は人手による作業が中心である。作業コストの大きさに加え分類時の判断主観性についても懸念されているため、自由記述アンケートは十分に活用されていないのが現状である。

テキストからの情報抽出や、要約・自動分類などの要素技術が蓄積されてきた言語処理技術を用いれば上記の問題を解決できる可能性がある。一方、言語処理で主に対象としているテキストは新聞データが多く自由記述アンケートのテキストに関する特性は明らかになっていない。

以上の状況を踏まえ、本研究では自由記述アンケートの自動分類に関する研究を進めている[5]。大量の自由記述アンケートの回答文から回答者の意図を機械的に推定し、意図にしたがって分類することを目指す。そこで本研究では、文末表現、副詞、感情語などのモダリティ表現に回答者の意図が表出されていると仮定し、その表現の意味が回答者の意図を適切に示しているかどうか分析する。この分析結果をもとに、表層情報による意図タグ判定モデルの作成とその実装を目指す。本稿では、文末表現を中心に、回答中のどのような情報をもとに、どのようなタグを決めればよいかを検討する。文末表現に着目した研究には、Web ページを対象に現れる表現を網羅的に調査した

ものもある[6]。

1.2 自由記述アンケート

ここでは、本研究で対象とした自由記述アンケート(ボイスレポート)について説明する。われわれが使用したボイスレポートとは、道路審議会基本政策部会の「21世紀の道を考える委員会」が平成8年5月から7月末に実施した全国規模のアンケート調査で得た回答であり、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。回答者数は 35,674 人、回答数(意見数)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ものである。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記述し書ききれない場合には別紙に記入する。

2. 研究方法

自由記述アンケートを分析するにあたり、まずは予備調査として既存の助詞・助動詞相当表現事例と意味がどのように分布するかを調べた。この結果から、個々の回答に付与された意味タグが回答の意図を示すものとして妥当かどうかを検討する。次に、意味タグを付与した回答文の予備的分析をもとに本研究での分析手順を示す。

2.1 『日本語表現文型』の助詞・助動詞相当表現

実際の用例から複合辞(助詞・助動詞相当表現)を集めた『日本語表現文型』[1]では、720 例の基本的表現(助詞相当 466、助動詞相当 254)を中心に、助詞相当を格助詞、副助詞、係助詞など、また助動詞相当を禁止・義務・可能・許可・推量・提案・要求などそれぞれの意味機能にしたがって分類し、延べ 1659 語の表現事例を挙げている(表 1)。

われわれは、これら 1659 語の表現事例を見出し語として、それぞれに割り当てられた意味機能を意味タグとして電子データ化した(以下、表現データと呼ぶ)。表 1 に挙げられた意味機能とそれぞれの意味を担う表現の数から、これらが網羅的なデータであることがわかるが、意味の違いが生じる条件や多義の

表1 『日本語表現文型』の内容

複合辞		1659	
助詞相当表現			
格助詞		副助詞	
係助詞		終助詞	
接続助詞		並立助詞	
助動詞相当表現			
アスペクト	105	意志・超意志	82
可能・不可能	27	義務・当然・当為・必然・必要・勧告・主張 およその否定	188
許容・許可	53	禁止	35
経験・回想・習慣	18	限定	40
行為の授受	21	推量・推測・推定	46
程度	36	適当・願望・提案・勧誘・勧告	49
伝聞	8	要求・依頼	74
他	75		

表2 表現の意味の差異

な け れ ば	ならない	義務・責
	ならなかった	任・当然・
	ならぬ	決意・強制
	ならない	義務・当然
	ならなかった	
		強制・義
	いけない	務・責任・
	いけません	当然・決意
	だめ	
	だめだ	
だめです		
ならなかった		
ならぬ		
なりません		

場合(例えば「なければならない」の「義務・責任・当然・決意・強制」「義務・当然」)の判断条件については明記されていない(表2)。

## 2.2 予備的分析

予備調査では、表現データの見出し語が回答にどのように分布するかを観察するために、見出し語と回答文の文末表現のボタンマッチングを行った。対象とした回答文は、上述した自由記述アンケートのうち、手紙で郵送された回答ファイルの回答数3116件、8328文である。これらに対し、文末から最長一致法で見出し語に相当する表現部分を取り出し、見出し語の意味タグを自動的に付与した。マッチしたデータ 3472 文の中で、高頻度の見出し語と意味タグの関係を表3に示す。表中、二重線以下の見出し語は文献[1]の見出し語としては記述されていないが、参照事例として比較対照されている表現であるため、表現データに登録されているものである。表3から、アンケートの回答文に関する以下の特徴が客観的に明らかになった。

- ・願望、提案、要求を示す表現が多く、とくに「てほしい」という直接的な要求表現で示されることが多い
- ・推量、推測、推定などの表現は比較的少ない

表3 自由回答文における表現データの出現頻度

意味タグ	見出し語	出現頻度
(アスペクト)	する	331
	ている	152
	ています	142
	ております	33
	れている	20
(適当・願望・提案・勧誘・勧告) (要求・依頼)	てほしい	519
(義務・当然・当為・必然・必要・勧告・主張、およびその否定)	べきである	90
	こと	44
	のです	26
	べきだ	26
	べきです	23
(反語・反駁・否定・非難・後悔) (願望・勧誘)	うか	177
(要求・依頼)	てもらいたい	76
	ていただきたい	35
(適当・願望・提案・勧誘・勧告)	ばいい	29
	がいい	19
(感動・詠嘆・驚異) (疑問・問いかけ・確認)	ではないか	62
	かな	7
(願望・勧誘)	ないか	24
(参照事例: 陳述性の高い表現)	う	739
(参照事例: アスペクト)	た	147
(参照事例: 格助詞)	を	57
(参照事例: 「なくてはならない」の 等価表現)	必要である	25

・意志表現も表現データに示す割合からすると比較的少ないと考えられる

しかし、同時に下記の問題点も明らかになったといえる。

- ・アスペクトと示される表現を持つ文では回答は何を意図しているのか。また、これらの表現は他の意味タグと同様に扱ってよいのか
- ・「適当、願望、提案、勧誘、勧告」「要求、依頼」「義務、当然、当為、必然、必要、勧告、主張」「願望、勧誘」などの意味の違いは回答の意図の違いとしても現れるのか
- ・意味タグはついていないにもかかわらず頻出する表現をどう扱うか
- ・助動詞相当表現「う」として取られているものの多くはボタンマッチングによる誤りで「と思う」の活用語尾。しかし、表現データには明示されていない「と思う」は頻出する表現であり、どのような意図を示すか疑問である

上記の問題を明らかにし、一貫性のあるタグ判定を行うために、本研究で行った分析方法を示す。

## 2.2 分析手順

対話データにおける談話タグ[7]や、回答者の意図を反映したタグ[5]を決定する作業が困難であることは従来から指摘されている。そこで、本研究ではゆれのない判断をするために表3にも示されていた直接的な要求表現「てほしい」への言い換え可否をチェックすることにより、等価な意図であるかどうかを判定する。また、なぜ言い換え可否の差異が生じるのか、言い換えられない回答文は何も要求していないのかを検討することで、2.1 で述べた問題点を明らかにする。

- 1) 回答文を「てほしい」という表現に意味的に言い換えることが可能かどうか
  - 2) 回答文に含まれる見出し語を下記いずれかに分ける
    - ① 要求表現(ある見出し語を含む回答文のうち9割以上に対して「てほしい」への言い換えが可能だったもの)
    - ② 要求以外の表現(ある見出し語を含む回答文のうち9割以上に対して「てほしい」への言い換えが不可能だったもの)
    - ③ 上記①②いずれにもあてはまらず言い換え可と不可能な場合が文脈に応じて生じる表現
  - 3) 上記①～③それぞれについて下記を検討する
    - ① 「てほしい」への言い換えを考える際、等価な働きをする文末表現は表現データが見出し語とする単位でよいかどうか
    - ② 回答の意図が要求以外の何であるかを調べる。あるいは「てほしい」に言い換えられなくても要求の可能性もあるか。他の表現(例: すべきである/なければならない/等)との言い換えが可能かどうか検討する
    - ③ 回答の意図の違いと表層情報の関連性を調べる
- 尚、1)の言い換え可否の判断条件は、着目する意図表現から「てほしい」への機械的な置き換えではないため例#1のような言い換えも含む。

< 言い換え可と判断する例 >

- #1 (思)う:信号を少なくして歩行者は地下道を通るようにしたら渋滞が解消されると思う。→信号を少なくして歩行者は地下道を通るようにし、渋滞を解消してほしい。
- #2 ないか:融雪により除雪できないか。→融雪して除雪してほしい
- #3 てもらいたい:高速道路の料金所ではもつとすばやくスムーズに通れるように工夫してもらいたい。

< 言い換え不可と判断する例 >

- #4 (思)う:交差点と踏切がいっしょになっている所は、渋滞しやすいと思う
- #5 ないか:自動車の増産によって道路の狭隘が増加、渋滞の発生が多くなっていないか。
- #6 ている:高速道路については、道路公団での建設で止むを得ないと思っている / たったひとりの運転者の為に渋滞する事をよく経験している。

本研究では、1)2)の作業結果を示し、3)についてはいくつかの事例を挙げ今後の分析見通しを述べる。

3. 結果と考察

表4は、分析手順1)の作業結果である。回答文に頻出する見出し語を高頻度順に並べている。「?」はもとの例文の意味が取れないもの、「?可」「?不可」はそれぞれどちらかといえば言い換え可能・不可能と考えられるが、いずれも可否を判断しにくい例である。表中網がけの列の可否は比較的迷わず判断している。これらの数字は、総出現数にあたる総計個数に対する割合で示されている。累計比率は回答に現れた表現のうち高頻度の上位 22 表現が 3472 文全体に対して占める割合である。表4では約 8割までを示した(表現「う」~「ないか」)。次に、分析手順2)にしたがって、表現を「要求表現」「要求以外の表現」「文脈に依存する表現」に分けた結果を示す。

①「要求表現」の文末表現(出現5回以上:9割可)

- べきである/てもらいたい/を/ていただきたい/ようにする/ばいい/必要である/べきです/がいい/すべきだ/てください/たらい/ばよい/よう/がよい/てくれ/させてほしい/といい/てもらう/ほうがいい/ものか/ように/たいものです/ほうがよい

#7 べきである:人・自転車・車の通行区分を明確にするべきである。

#8 を:中山間地域では、都市へのより良いアクセスとともに高齢者に歩きやすい道を。

②「要求以外の表現」の文末表現(出現5回以上:9割不可)

- た/ております/から/れている/について/てくる/ことがある/ため/まい/ので

#9 た:先日、尾瀬に行ってきた。/自分で、車に乗ってみて、危険であると、分かるようになりました。

#10 から:体が悪い人、重い荷物を運ぶ際など、どうしても車が必要な場合もあるのだから。

③文脈に応じて可否の分かれるもの(出現5回以上)

- (思)う/する/うか/ている/ています/ではないか/こと/のでは/のです/べきだ/ないか/ものです/が/なければならぬ/だろう/ていない/など/しか

表4「てほしい」への言い換えの可否

aux	?	?可	?不可	可	不可	総計(個)	累計比率
(う)	0.8	8.8	1.4	66.2	22.9	739	21.3
てほしい	0	0	0	100	0	519	36.2
する	0.6	5.1	0	83.4	10.9	331	45.8
うか	0	4.5	1.7	68.4	25.4	177	50.9
ている	0.7	5.9	2.6	1.3	89.5	152	55.2
たた	0	0	0	4.8	95.2	147	59.5
ています	0	4.2	1.4	12.0	82.4	142	63.6
べきである	1.1	3.3	5.6	90	0	90	66.2
てもらいたい	0	0	0	100	0	76	68.3
ではないか	1.6	8.1	0	67.7	22.6	62	70.1
を	0	0	0	100	0	57	71.8
こと	2.3	25.0	0	68.2	4.5	44	73.0
ていただきたい	0	0	0	100	0	35	74.0
ております	0	3.0	0	3.0	93.9	33	75.0
ようにする	0	6.1	0	93.9	0	33	76.0
のでは	3.2	9.7	3.2	71.0	12.9	31	76.8
ばいい	0	0	0	100	0	29	77.7
のです	0	0	0	30.8	69.2	26	78.4
べきだ	0	0	7.7	88.5	3.8	26	79.2
必要である	0	0	0	100	0	25	79.9
から	0	4.2	0	0	95.8	24	80.6
ないか	0	12.5	0	41.7	45.8	24	81.3
ものです	0	9.1	0	45.5	45.5	22	
すべきだ	0	0	0	100	0	18	
なければならぬ	0	0	11.1	66.7	22.2	18	
てください	0	0	0	100	0	15	
だろう	0	7.1	0	35.7	57.1	14	
たらい	0	0	0	100	0	12	
てくる	0	0	0	0	100	12	
ばよい	0	0	0	100	0	12	
ことがある	0	0	0	0	100	9	
しかない	0	11.1	0	22.2	66.7	9	
ないかな	0	22.2	0	44.4	33.3	9	
ものである	0	0	22.2	33.3	44.4	9	
s	0	0	0	100	0	9	
がよい	0	0	0	100	0	8	
てくれ	0	0	0	100	0	8	
てしまう	0	100	0	0	0	8	
まい	0	0	0	0	100	8	
総計	1.0	5.6	1.2	63.1	29.2	3472	

単位: (%) (個) (%)

- ない/ないかな/ものである/で/てしまう/かな/たらよい/ていく/はず

#11 する(可):車の流れをスムーズにするため、道巾の狭い所では、右折禁止にする。

#12 する(否):道は人々のものであるという基本理念を再認識させねばならないことを痛感する。

#13 なければならぬ(可):そのためには公共交通機関をより充実させなければならぬ。

#14 なければならぬ(否):その為、車線変更した後さらにスピードを出して追い越さなければならぬ

上記の例にもとづき、分析手順3)について検討する。

①先に挙げた言い換え可能な25表現のうち、表現ダ

一タの単位で不十分だと思われるのは「ものか」(6例)である。4例が「できないものか」、あとの2例が「ならないものか」「のはいかなものか」など否定・疑問表現をともなって現れることによりこれらの表現自体は反語的表現となり要求として成り立っている。否定や疑問がつかない場合は、「～するものか」のように相手の言葉・考えに対する強い反対・否定、ある動作・行為を行わないことへの固い決意を示すことになる[1]。また、「がいい」(18例)は「方がいい」の形で取れたものが15例あり、「(名詞)がいい」は3例だけであるため、この表現が要求に結びつく判断するためにはさらに事例を集める必要がある。「べきである」「べきです」は「てほしい」に文脈依存せず言い換え可能な表現となっているが、類似表現である「べきだ」は文脈依存の結果が出ているため、これらに対する他事例の更なる検討も必要である。

②事例#9,10 からわかるように、「てほしい」への言い換え不可の表現を含む文は、自分の経験や現状認識など要求の導入になりうるもの、また要求の根拠が述べられているものが多い。したがって、要求と根拠、現状認識と要求のように回答が二文以上の文から成るとき、「てほしい」への言い換え不可の文が頻出すると考えられる。実際に、一文の回答と二文以上から成る回答との間でその差を調べて見ると次のような傾向が見られた(表5)。すなわち、一文の回答は9割近くが「てほしい」に言い換え可能な要求表現であるのに対し、二文以上の場合は要求とは異なる意味役割の文を含むことを示す。この差は、副詞節などで因果関係を一文で表現している複文の扱いを二文以上と考えることにより、さらに明確になると考える。

表5 一回答における文数の相違(カッコ内は割合)

	一文	二文以上
全体(8328文)の比率	1005文(0.12)	7323(0.98)
表現データの見出し語を自動的に付与できた回答(3472文)	423(0.12)	3049(0.98)
回答文末を「てほしい」に言い換えることが意味的に可能であった文	380(0.898) /423	1810(0.59) /3049

一文でも、「高速道路を使わないからどうでもいい」「運転する人や歩く人、自ら問題が多いと思う」「歩道がせまい」など明確な要求でなくても、不満を表すことによって現状に対する満足度の低さを示す場合もある[9]。下記についてはさらに検討が必要である。

- ・二文以上から成る回答文の談話構造
- ・明確な要求でない場合の回答の意図

③「てほしい」への言い換え可否が文脈依存の場合、可否を決める条件は共起する表現であることも多い。「思う」を例にとると、先に挙げた「たらない／べきだ／てもいい」などと結びつく場合は殆ど言い換え可能になるのに対し、「(形容詞)と思う」は27例中24例、「(動詞)と思う」は36例中30例、言い換え不可である。一方、直前が動詞や形容詞であっても「～(ば／と／たら)～と思う」のように条件表現を作る副詞節と共起する場合には言い換え可能になることが多い。文脈依存によって言い換え可否が分かれる表現の場合、さらに詳細な共起表現などの制約によって回

答者の意図を示す表現を取り出すことができると考える。

#### 4. おわりに

大量の自由記述アンケートから機械的に回答者の意図を推定し、意図にしたがって分類するために、どのような表層情報に着目すると回答者の意図を取り出すことができるのか検討した。今回はとくに文末表現に着目し、文末を「てほしい」という要求表現に言い換えることができるかどうかを分析のポイントにした。この結果、表現自体が固有に「要求」を示す場合、示さない場合、文脈に依存する場合などが明らかになってきた。今後は、言い換え不可能な表現に対する詳細な分析により要求でない回答が何を示しているのか、また文脈依存である場合、可否を決める条件は文脈にどのように現れているかなどの分析を進める予定である。また、副詞、感情語などにも注目することによって、表層表現から機械的に回答の意図を示す意図タグを判定できるモデルを作成する。

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。また、本研究を進めるにあたり、議論に参加し研究に協力してくださった九州工業大学大学院情報工学研究科情報科学専攻博士前期課程の乾孝司さんに感謝いたします。

#### 参考文献：

- [1] 森田良行・松木正恵：『日本語表現文型』アルク、1989。
- [2] 野崎進・増田尚・福田暢行・白山麗：アンケートにおける日本語自由文の情報分析、情報処理学会第47回全国大会論文集、3pp165-166、1993。
- [3] 須賀伸介・大井絃：自由回答記述データを用いた瀬戸大橋に対する住民意識の解析、土木計画学研究・講演集 No.20(2)、pp31-34、1997。
- [4] 松井和香・石田東生：道路審議会建議策定過程におけるパブリックインボルブメント方式の効果、土木計画学研究・講演集 No.21(2)、pp361-364、1998。
- [5] 乾裕子・内元清貴・村田真樹・井佐原均：文末表現に着目した自由回答アンケートの分類、情報処理学会自然言語処理、No.128、pp181-188、1998。
- [6] 土井晃一：文末態度表現に注目した Web Page の調査、情報処理学会自然言語処理、No.130、pp49-56、1999。
- [7] 荒木雅弘・伊藤敏彦・熊谷智子・石崎雅人：発話単位タグ標準化案の作成、人工知能学会誌、vol.14 No.2、1999。
- [8] 月出奈都子・石崎俊：TV番組に対する自由回答文の印象抽出システム—インターネットアンケート調査による自由回答文の解析—、第6回言語処理学会年次大会発表論文集、pp249-251、2000。
- [9] 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、第35回日本都市計画学会学術研究論文集、p571-p576、2000。