

「意見の強さ」を分析するための特徴抽出

柳瀬 隆史¹ 丸元 聡子² 乾 裕子^{2,3} 落谷 亮¹

¹ 富士通研究所 ² 計量計画研究所 ³ 通信総合研究所

tyana@jp.fujitsu.com

1 はじめに

企業が顧客に対して製品に関する意見を聞いたり、行政機関が住民に対して行政サービスに関する意見を聞く場合にアンケート調査がしばしば用いられる。従来のアンケート調査では、回答の集計や分析の正確さや簡便さを求める観点から、想定される回答を調査側があらかじめ選択肢として用意する選択回答方式が主であった。これに対し近年では回答者個々の意見の収集が重視されるため、自由記述方式による調査も増えている。製品やサービスに関する意見を自由記述方式で集めると、それらに対する不満や改善要求が回答として書かれることが多い。しかし、調査側は各回答文から回答者の不満や要求の内容およびその程度を判別する必要があるが、従来はほとんど人手で判別していたために多大なコストを必要としていた。

そこで本研究では、アンケート調査によって収集された個々の自由記述回答文（以降では「意見文」と呼ぶ）から得られる不満や要求の程度について、その分析を試みた。分析には意見文中の単語や表層的な構文パターンなどから得られる特徴を用いる。また、このような不満や要求の程度の分析にしばしば用いられる文末表現がどのように関連するかについての考察も行う。

2 「意見の強さ」と特徴抽出

2.1 「意見の強さ」の評価基準

多数の自由記述式のアンケート回答から代表意見を選別する際に、どの意見を代表とするかの尺度にはさまざまなものが考えられる。その意見が多数派かどうかというの也被考えられるが、不満や要求の程度も一つの尺度として考えられる。この不満や要求の程度を「意見の強さ」と呼ぶことにする。

また意見の強さの評価基準にもさまざまな観点があり明確な定義は難しいが、本研究ではその観点として次の2つを提案する。

- 説得力

意見文にはその意見を持つに至る理由や目的が

書かれているものもある。それらの記述を含む意見文は、含まないものと比べて読み手にとってはより高い説得力を持つ。説得力が高ければ不満や要求の程度も高く感じられると考えられる。

- 具体性

ある意見文について、同じ意見であっても内容に具体的な記述を含むほうが、含まないものに比べて回答者にとっての重大性が高く、その程度も高く感じられると考えられる。

2.2 意見文から抽出する特徴

前節で説明した評価基準に基づき、意見文からは以下のような特徴を抽出する。

- 根拠記述

意見文中において、ある意見に至る理由や目的に関する記述を「根拠記述」と呼び、根拠記述の有無を説得力に基づく特徴とする。

- 具体性を示す語

意見文の具体性は一般には評価が難しいが、今回は具体性を示す語の有無によって判断するものとする。意見文中にその語が含まれているかどうかを具体性に基づく特徴とする。

2.3 アンケート回答からの特徴抽出方法

本研究で分析の対象としたデータは、将来の道路計画の策定に市民の声を採り入れるために「21世紀のみちを考える委員会」が平成8年5月から7月に実施したアンケートである[1]。回答者はあらかじめ設定された12のテーマ（渋滞解消、高速道路、交通安全など）から選択したいいくつかのテーマについて、提示されている資料や参考意見（各テーマに対して、意見の例がAさんの意見、Bさんの意見、という形で自由記述回答と同様の形式で提示されている）を基にして各自の意見を自由記述形式で回答する。回答者数は35,674人で、回答件数はのべ113,316件である。このうち、回答が複数文から構成されている

表 1: 各特徴を含む意見文の件数と全体 (36,265 件) に対する割合

根拠記述	件数	割合
理由	3,876	10.7%
目的	2,459	6.8%
具体性を示す語	件数	割合
固有名詞	2,963	8.1%
数値	2,471	6.8%

ものは複数の意見が書かれているものもありその切り分けが困難であること、1文のみの回答に絞り込んで十分な件数のデータが得られることなどの理由から、本研究ではこのうち回答が1文だけで記述されているもののみを対象にした。さらにA,Bなどの記号のみしか書かれていないものを除外し¹、残った36,265件を分析対象とした。

以下では、説得性・具体性それぞれに基づく特徴の抽出方法について説明する。また、表1に各特徴を含む意見文の件数と分析対象の意見文全体に対する割合を示す。

根拠記述の抽出

根拠記述の切り出しは、理由や目的を表す接続詞、接続助詞、複合語を用いて行う。これらの語をパターンとしてマッチング規則を作成し、各意見文に対して文頭から機械的にパターンマッチングを行う。そしていずれかのパターンにマッチすれば文頭からマッチした箇所までを根拠記述とみなす²。

根拠記述は理由と目的に分けて切り出す。理由を表すパターンは「ので」「だから」など5種類、目的を表す語は「ために」「為に」「ためには」など8種類である。各意見文に対しては、それぞれ1つ以上切り出されたときに、理由あるいは目的を表す根拠記述を含むという特徴を与える。例えば、「渋滞がひどいのでバイパスを設置すべきだ。」という文に対しては「ので」がマッチし、「渋滞がひどいので」が理由を表す根拠記述となる。

具体性を示す語の抽出

今回は、分析対象のアンケートが道路行政に関するものであるという点を考慮して、固有名詞および数値を具体性を示す語とした。各意見文を形態素解析して固有名詞および数値の有無を調べ、それぞれ1つ以上ある場合に固有名詞あるいは数値を含む文、という特徴を与える。

¹記号のみが回答欄に書かれていた場合は、その記号に対応する参考意見に賛成であるとみなすことができるので、本研究では分析の対象から除外した。

²例えば「Aの意見に賛成なのです。」という文に「ので」がマッチして意図しない箇所が切り出されることを避けるために、切り出し用のパターンとは別に例外パターンを用意して、例外パターンがマッチしたときには切り出しをしないようにしている。

表 2: 各特徴を持つ意見文の不満・要求程度ランク別分布

	程度ランク			合計
	0	1	2	
全体	199	656	145	1000
根拠記述				
理由	8	51	37	96
目的	5	27	30	62
具体性を示す語				
固有名詞	10	129	53	192
数値	11	110	43	164

2.4 特徴抽出の有効性の評価

前節で説明した根拠記述および具体性を示す語についての特徴が意見の強さの分析に有効かどうかを評価する。まず、分析対象としている36,265件の意見文のうち、任意に選択した1,000件に対して何らかの不満や要求が述べられているか、またその程度は大きいかについて調べ、以下に示す基準で3段階に人手でランク付けした。

- ランク 0
不満や要求は述べられていない
- ランク 1
不満や要求は述べられているが、その程度は大きくないと思われる
- ランク 2
不満や要求が述べられており、かつその程度は大きいと思われる

次に、各特徴を持つ意見文について程度ランク別分布を調べ、その結果を表2にまとめた。根拠記述に関する特徴に着目すると、ランク2の意見文の占める割合が理由で38.5%(=37/96)、目的で48.4%(=30/62)であった。具体性を示す語についても固有名詞が27.6%(=53/192)、数値が26.2%(=43/164)となり、いずれの特徴についても全体におけるランク2の割合(14.5%=145/1000)を大きく上回っている。従って、根拠記述および具体性を示す語に関する特徴は意見の強さを分析するための特徴として有効であると考えられる。

3 文末表現と「意見の強さ」

「～は高すぎる」や「～するべきだ」などの文末表現には、賛成・反対・不満・要求といった意見の性格を端的に表すものが多い。また、ほぼ同じ意味であっても微妙なニュアンスの違いを含むものもあり、それらの違いを詳細に検討すれば、文末表現が意見の強さを評価するための特徴になり得ると考えられる。

ここでは、意見文から収集した文末表現を表現類型によって分類し、各分類に属する文末表現を含む意見文の「意見の強さ」や、説得性および具体性に基づく特徴を調べることにより、文末表現が意見の強さにどのように関連するかについて考察する。

3.1 文末表現の収集と整理

文末表現を分析に用いるには、まずアンケート回答文で用いられている文末表現を収集し、それらを整理する必要がある。本研究では、以下の手順で収集・整理した。

1. 分析対象の36,265件の意見文を対象に文末からの最長一致法により、10回以上出現した文末表現を候補とする。
2. 1.のうち、以下に示すいずれかの意味合いを持つと判断される文末表現を手作業で選別する。
 - 賛成表現
「に賛成です」「に同感だ」など
 - 反対表現
「に反対です」「はおかしいと思う」など
 - 不満表現
「困っている」「高すぎる」など
 - 要求表現
「が必要である」「するべきだ」など
 - その他（不必要、禁止表現など）
「は必要ない」「してはいけない」など

以上の手順で収集された文末表現は1,009種類で、分類ごとの内訳は賛成表現63種類、反対表現13種類、不満表現118種類、要求表現779種類、その他36種類であった。文末表現が賛成表現あるいは反対表現であった意見文のほとんどは、参考意見に対する賛成・反対のみを述べたものであり、回答者独自の意見を述べたものではなかった。従って、意見の強さを分析するのに手掛かりとなり得る文末表現は不満表現、要求表現などである。以下では、これらのうち要求表現に焦点を当てて、意見の強さとの相関があるかどうかについて検討する。

3.2 要求表現と「意見の強さ」との相関

まず要求表現に分類された779種類の文末表現について、表現形式によって表3に示す12種類に分類した。分類に当たっては、森田ら[2]の表現類型分類や乾ら[4]の分類を参考にした。

次に、2.4節での評価実験に用いたランク付きの1,000件の意見文のうち要求表現を含むものについて、表3の分類に従って分類した。その結果を表4に示す。この表には、ランク付き意見文全体および「意見の強さ」の大きいランク2の意見文について、

表 3: 要求表現の分類

分類	文末表現数	文末表現例
よい型	159	してもよい、したらよい、ほうがよい
ない型	38	しななければならない、しかない
疑問型	74	できないものか、してはどうか
べき型	60	べきである、べきだと思う
依頼型	30	して下さい、していただきたい
望む型	60	望む、期待する、お願いします
急務型	37	急務だ、第一である、先決である
重要型	78	重要だ、大切である、大事である
必要型	120	必要である、不可欠である
欲しい型	28	欲しい、欲しいと思う
終止形	79	設置する、広くする、など動詞の終止形
述語省略型	16	「(～の設置)を。」「(～を設置)すること。」など後ろの述語が省略されている形

文末表現が各分類に該当する件数および要求表現全体に対する割合が示されている。従って、各分類における割合が全体よりもランク2のほうが高い場合に、その分類に属する文末表現を含む意見文が「強い」傾向があると言えることになる。

この表によれば、「必要型」「依頼型」「望む型」などでランク2のほうが割合が高くなっており、比較的「強い」表現であると言える。逆に「よい型」「欲しい型」「終止形」などでは割合が低くなっており、比較的「弱い」表現であると言える。このように、要求表現は表現形式によって分類することによって、意見の強さと相関を持つと考えられる。

3.3 要求表現を含む意見文の特徴分布

次に、要求表現と意見の強さとの間のさらに詳細な相関関係を調べるために、表現形式による分類ごとに説得性および具体性に基づく特徴の分布を調べる。

今回対象とした意見文36,265件のうち文末に要求表現を含むもの18,407件について表3に従って分類し、各分類ごとに根拠記述（理由、目的）および具体性を示す語（固有名詞、数値）の特徴を持つ意見文の割合を求めた。さらに、比較対照として要求表現を含む意見文全体に対しても同様に特徴を持つ意見文の割合を調べた。

結果を表5に示す。ある特徴について、要求表現全体での割合よりも高い割合を持つ要求表現分類は、その特徴との相関が高いと判断することができる。まず全体の傾向を見ると、「依頼型」「望む型」などでどの特徴についても要求表現全体での割合よりも高い割合となっており、前節での結果と合わせても比較的「意見の強さ」の高い表現類型とすることができる。逆に、「終止形」などではいずれの特徴でも比較的低い割合となっており、「意見の強さ」は低い

表 4: 「意見の強さ」ランク付き意見文の要求表現分類別分布

要求表現分類	全体	うちランク 2
よい型	77(15.5%)	6(6.2%)
ない型	3(0.6%)	1(1.0%)
疑問型	15(3.0%)	2(2.1%)
べき型	52(10.4%)	13(13.4%)
依頼型	44(8.8%)	14(14.4%)
望む型	28(5.6%)	13(13.4%)
急務型	0(0.0%)	0(0.0%)
重要型	9(1.8%)	4(4.1%)
必要型	40(8.0%)	20(20.6%)
欲しい型	167(33.5%)	18(18.6%)
終止形	43(8.6%)	3(3.1%)
述語省略型	20(4.0%)	3(3.1%)
要求表現合計	498(100%)	97(100%)

と考えられる。「べき型」についても、語自体の持つニュアンスとしては強いイメージを持たれがちであるが、例えば単に「バイパスを設置するべきだ。」のように根拠記述や具体性を示す語は含まない場合が比較的多いということが読みとれる。

次に個々の特徴について考察する。理由の根拠記述については「ない型」や「欲しい型」で割合が高くなってきている。理由の根拠記述には「渋滞がひどいので」など回答者の実体験に基づく不満が書かれており、これらの表現類型はその不満を解消するための方策を行政に要求するときに用いられる傾向が高いと考えられる。また、目的の根拠記述に対しては「必要型」や「急務型」で割合が高い。目的の根拠記述は「モラル向上のためには」のように将来に向けての願望が書かれることが多く、この願望を実現するための方策を要求するときにこれらの表現類型が用いられることが多いと考えられる。

具体性を示す語については、「依頼型」や「望む型」で固有名詞を含む意見文の割合が高いが目立っている。詳細に調べてみると、「〇〇市にバイパスの設置をお願いします。」のように具体的な地名を挙げて行政側に対して要求を行う意見が多かった。

このように、今回提案した説得力や具体性といった意見の強さの評価基準や、要求の動機などによって高い相関を持つ要求表現が異なることが分かる。

4 おわりに

本稿では、アンケート自由記述回答文から不満や要求の程度を示す「意見の強さ」を分析するための特徴として読み手への説得力という観点からの根拠記述の有無や、内容の具体性という観点からの具体性を示す語の有無による特徴を提案し、その有効性を検証した。また、要求表現を中心とした回答文の文末表現と提案した特徴との共起関係を調べることによって、これらの観点における「意見の強さ」と文末表現との関係について考察した。しかし、回答文

表 5: 要求表現を含む意見文の件数と特徴を含む割合

要求表現分類	件数	根拠記述		具体性を示す語	
		理由	目的	固有名詞	数値
よい型	2803	17.6%	6.8%	5.6%	4.6%
ない型	292	22.9%	6.8%	10.6%	5.8%
疑問型	463	12.3%	9.5%	12.1%	13.2%
べき型	3287	12.5%	8.3%	8.8%	6.8%
依頼型	1018	21.6%	9.1%	14.2%	8.6%
望む型	998	14.2%	11.0%	20.1%	11.1%
急務型	103	11.7%	13.6%	11.7%	13.6%
重要型	580	11.7%	11.9%	7.1%	5.5%
必要型	2906	12.1%	15.3%	7.4%	7.4%
欲しい型	4202	20.9%	6.7%	10.5%	6.9%
終止形	1294	3.9%	7.3%	4.9%	5.6%
述語省略型	461	4.8%	8.0%	11.7%	7.4%
要求表現全体	18407	15.1%	9.1%	9.3%	7.0%
全体	36265	10.7%	6.8%	8.2%	6.8%

から抽出できる特徴は他にも考えられる上、「意見の強さ」をどのような観点から分析するかによって有効な特徴も変わってくるであろうと考えられる。今後は他の観点も含めた特徴抽出についてもさらに検討していきたいと考えている。

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下さった(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] 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。
- [2] 森田良行、松木正恵、日本語表現文型、アルク、1989。
- [3] 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、pp.571-576、2000。
- [4] 乾裕子、井佐原均、拡張モダリティの提案、信学技報、NLC2002-11、2002。
- [5] 乾裕子、高梨克也、井佐原均、自由記述型アンケート回答を対象にした意図特定スキーマの提案、言語処理学会第8回年次大会、pp.104-107、2002。