

言語学的考察および機械学習手法に基づく意見分析支援 - 要求意図を取り出す一手法 -

A Proposal of Supporting Processing for Opinion Analysis Based on Probabilistic Model and Linguistic Knowledge

大塚 裕子* 内山 将夫** 井佐原 均**

By Hiroko OTSUKA, Masao UTIYAMA and Hitoshi ISAHARA

1. はじめに

われわれは、アンケートの自由回答やコールセンターへの質問・苦情、web 掲示板への書き込みなどを「意見データ」と呼び、意見が集約されたテキストから、その発話者である話し手・書き手の意図を多様な観点で取り出すことを目標としている¹²⁾¹³⁾。観点の中には、アンケートのフェイス項目などから得られる年齢、性別、職業、居住地といった発言者属性ならびに予め予測された調査項目に基づいた多数意見、少数意見などが考えられる。また、発言者の意図として好き嫌いや賛成・反対、要求・提案、不平・不満などの表明があり、その対象内容や理由は調査者など意見傾向を知りたい立場にとって貴重な情報源となる。本研究では、意図の一つとして「要求」に注目している。要求意図の特定により、要求内容および要求動機の情報を取り出すことができる。しかし、意図を取り出すことは、単にテキストに文字通り表現されている要求表現を抽出することではない。この目的に対して現状の工学的な技術としての意図認識および意図理解は十分でない。

一方、意見データにおいてアンケート調査の自由回答は、行政や企業が行う数多くの調査によって蓄積されても活用される機会がほとんどないのが現状である。公開されないまま無用な情報となるか、あるいはすべての意見を一覧表示する利用に留まっている。われわれがとくに自由回答に着目するのは、適切な観点で情報を抽出し意見を分類し公開することによって、その結果を調査者と回答者で共有する仕組みを構築したいと考えているからである。これらのアプローチは、パブリック・インボルブメントのように公共的な事業計画における行政側と住民を

はじめとする計画関係者との知識情報共有ならびに意思決定過程に重要な役割を果たす。

これらの背景から、本稿では道路計画に市民の声を反映するためのアンケート調査の自由回答⁶⁾をコーパスとして最大エントロピー法を確率モデルとする機械学習手法により、要求意図の抽出およびそれに基づく回答文の分類を行う。とくに、学習データの作成方法として言語的な言い換えを用いた判定を提案する。これにより学習データの質が向上し機械学習の精度を高めることができる。このため、着目する回答文が典型的な要求表現「してほしい」を含む形式に言い換え可能であることの判定可能性を仮説とする。この仮説の妥当性検証のため、内容分析 (content analysis) の知見を踏まえて予測的妥当性、再現可能性、相関的妥当性の三観点から検証を行う。さらに判定基準と自動分類を統合し、基準を用いて作成した学習データによる機械学習手法の実験を行う。

2. 仮説としての要求意図判定基準

ここでは、本稿で提案する要求意図判定基準について述べる。まず「要求」の表現について演繹的に考察する。日本語話者の多くが要求と判断できる手がかり「～してほしい」「～してくれ」「～もらいたい」等は言語学的には直接依頼表現と言われる⁹⁾。一方、書き手の意図を示す「モダリティ」という構文カテゴリーの要素でもあるこれらは従来の構文論的言語学で要求・要望・依頼・願望・命令などに細分化されてきた。すなわち、言語学においても統一的な扱い方が確立されていない状況である。研究目的を鑑みると本研究で必要な情報は言語行為論的解釈に基づ

* 言語情報研究室 ** 独立行政法人情報通信研究機構

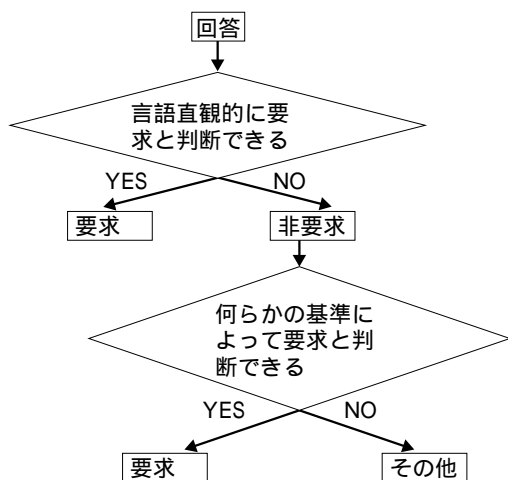


図 - 1 「要求」を判断する階層

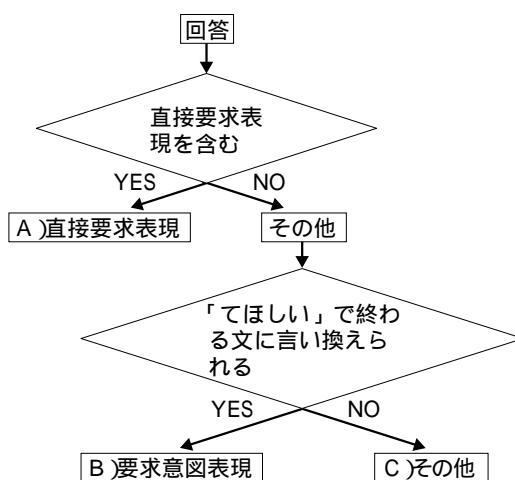


図 - 2 要求意図判定基準

いた視点での要求意図であると判断できるため、これらをまとめて直接要求表現と呼ぶ。これらが回答に現れた場合、調査者は要求と判断できる。この認知プロセスは図 - 1 の第一判定「言語直感的に要求と判断できるか否か」に相当する。本稿では、これらを図 - 2 の基準の第一判定「回答が直接要求表現を含むか否か」に相当するとみなす。よって、この判定で得られる要求は図 - 2 の「(A) 直接要求表現」を含む。

次に直接要求表現がない場合、すなわち図 - 1 の非要求の回答の中で要求と判断できるものについて述べる。「交通量の激しい道路の歩道にはガードレールをつけるべきだ」環境に配慮した道路づくりが重要である」といった回答は、調査者・回答者・アンケート調査という発話状況から「歩道にガードレールをつけてほしい」環境に配慮してほしい」といった要求と解釈することができる。しかし、図 - 1 の第二判定のとおり、この解釈の基準になっているのは暗黙の「何らか」の基準であり、その内容に個人差を含む可能性がある。従って、図 - 1 の要求②は判定者に依存して判定結果が変わる可能性があるため再現性のある結果が得にくい。そこで、本研究では人が解釈の際に行っている暗黙のプロセスを要求意図の判定基準として顕現化することを試みた。この顕現化が図 - 2 に示した第二判定「回答を『てほしい』を含む文に言い換えられるか否か」である。回答文を典型的な要求表現「てほしい」を含む文に言い換えることができれば、それは要求文であるといえる。この判定基準によって図 - 2 の(B) 要求意図表現を含む要求が得られる。要求を含む回答は基

準に従って取り出されるため再現性が保証される^(注1)。

(1) 予測的妥当性の検証

- 基準の適用と結果の分析 -

ここでは、要求意図判定基準を用いて実際の回答文から要求を取り出すことを一人の作業者が行い、この結果について分析・考察する。予測的妥当性は「ある方法で得られた予測が、その後直接観察された事実とどの程度一致するかを査定する」¹⁰⁾ことによつて測られる。ここでの予測とは要求意図判定基準に該当し、査定とは一定量の自由回答データに対してこの基準を適用した要求意図判定に該当する。

a) 分析方法および結果

分析手順は、判定基準に従って 1) 回答文に直接要求表現を含むか否か、2) 回答を「てほしい」で終わる文に言い換え可能か、を目視により判断、分類する。1) の判断には¹¹⁾らの分類において要求・依頼に列挙された表現を対象とした。2) の判断は筆者による。分析データは、複数文から成っている回答についても一文ごとに分け、無作為に取り出した 3,000 文を対象にした。分析結果を表 - 1 に示す。表 - 1 の①は直接要求表現を含む文である。これらは、すべて「～てほしい」への言い換えが可能であり、3,000 文中の約 2 割を占める。③の表現が図 - 2 における要求意図表現に相当する。

要求意図表現には動詞や名詞など、従来モダリティの表現とは認められていない品詞の語や、助動詞および助動詞相当語句によるモダリティ表現の中

表 - 1 要求意図判定基準の適用結果^(注2)

	直接要求表現	言い換え	3,000 文中
①	含む	可	547
②	含む	不可	3
③	含まない	可	1,190
④	含まない	不可	1,252

で構文論的な意味分類においては「要求」とみなされていなかった様々な表現が含まれている⁷⁾。従来の言語学的知見から得られたパターン分類的手法では、表 - 1 の①からのみ「要求」を取り出そうとしていた。しかし、提案手法を用いることにより、表 - 1 の③の要求意図表現を手がかりに「要求」を取り出せることがわかった。

(2) 再現可能性の検証

- 複数の作業員による言い換え判定 -

再現可能性は信頼性の確定条件の一つとして「ある過程を様々な状況の下で、様々な場所で、様々なコーダーを使って、再生することができる程度のことである。(略)その一例は複数の人がお互いに独立に同一の記録作業説明を一組の同一のデータに適用する場合である。これらの人々がデータを記録する仕方における不一致は、記録作業説明を解釈し、使用する仕方における観察者間の相違および観察者間の不整合の両方を反映している。」と定義されている¹⁰⁾。本節では、上記に示された方法に基づき複数の作業員による言い換え判定実験を行う。その作業結果の一致率を調べることで、判定基準の再現可能性を検証する。

a) 検証方法および結果

実験の被験者、すなわち言い換え作業員は日本語を母語とする大学生、20代男性3人である。3人の作業員ABCがそれぞれAB、BC、ACと二人ずつ共通する4,000文に対して作業したため、文の異なりは12,000である。一回答が複数文から成る場合は、一文ごとに分けて対象文とした。

尚、直接要求表現は「てほしい」に言い換え可能であるため、本節では図 - 2 の第二判定すなわち「てほしい」への言い換え判定だけを行う。各作業員が仕様に従って作業し、表 - 2 のとおり AB、AC、BC

表 - 2 判定による言い換え可否の事例数

		B		
A		可	不可	計
可		2,372	970	3,342
不可		36	622	658
計		2,408	1,592	4,000
		C		
A		可	不可	計
可		3,123	264	3,387
不可		171	442	613
計		3,294	706	4,000
		C		
B		可	不可	計
可		2,119	50	2,169
不可		934	897	1,831
計		3,053	947	4,000

のそれぞれの言い換え可否の結果を得た。例えば A の「可」と B の「可」の 2,372 は、両者が共に「てほしい」に言い換え可能な文であると判断し言い換え作業を行った事例数を示している。

一致率とは、同一の質問(テスト)による繰り返し調査あるいは同一対象に対する複数の判断の一致の指標であり、データの信頼性を保証する指標として用いられる。この指標の一つに、2人の評定結果の一致度を測定する指標としてコーエンの一致係数(Cohen's measure of agreement)がある。一致係数は、以下の式で表される⁴⁾。

$$\kappa = (p_0 - p_c) / (1 - p_c) \dots\dots\dots(1)$$

p_0 : 対応するカテゴリの観察頻度が一致する確率
 p_c : 対応するカテゴリが偶然により一致する確率

値は医療診断の一致率などに使われることが多いが、近年、言語処理の分野でも人が行うタグ付け作業や要約の一致率などに使われている²⁾³⁾⁸⁾。

表 - 2 に従って式(1)を解くと、例えば、AB の観察された一致率は、事例の合計値における、AB それぞれの「可」の和 2,372 と、AB それぞれの「不可」の和 622 を合計した値 2,994 の割合である。この値が表 - 3 の観察された一致率 0.749 にあたる。次に、偶然の一致率は AB それぞれの「可」の判定が生じる確率 0.503 と「不可」の判定が生じる確率 0.065 の合計値である。この値をもとに(1)式を解くと、表

表 - 3 作業の一致率

作業者	観察された一致率 p_0	偶然の一致率 p_e	κ 値
AB	0.749	0.568	0.417
AC	0.891	0.724	0.605
BC	0.754	0.522	0.485

- 3のとおり偶然の一致率0.568が得られる。

最大の一致の場合の確率は1なので、 κ 値は1に近いほど一致を示す。この値の解釈については一般に、 $0.80 < \kappa < 1.00$ ならばほぼ完全な一致、 $0.60 < \kappa < 0.80$ ならば実質的に一致、 $0.40 < \kappa < 0.60$ ならば中等度の一致、 0.40 以下ならば一致は見られないとされている。従って、言い換え作業の結果、作業者ACでは実質的な一致が、AB、BCでは中程度の一致が見られたと考えられる。

しかし、自然言語処理では自動要約の評価指標として判定者4人の一致率が0.24しかないこと⁸⁾、また、医学文献では $0.21 < \kappa < 0.40$ の値が「かなりよい」とみなされること³⁾、内容分析では0.70以下の一致しかない変数間の相関は統計的に有意でないこと¹⁰⁾などの見解から、一致係数の適用状況とその結果の解釈については議論の余地がある。一致率の解釈については、クリッペンドルフ自身が一致のレベルの高さについて決まった答えがないことを指摘している¹⁰⁾。以上の結果は、 κ 値の解釈を実質的に一致、中程度の一致と見るならば、異なる判定者による実験からも本手法を意図判定基準として要求か否かを判定できること、すなわち、本手法を用いることで、訓練された判定者以外にも要求の取り出しが再現できることを示している。

(3) 相関的妥当性の検証

- 意図判定基準を用いない意図判定 -

相関的妥当性は「ある方法で得られた結果が、他の方法で得られた結果とどの程度相関するかを査定する。すなわち方法に関する代替可能性を検証する」と定義される¹⁰⁾。この考え方に従って判定基準の有効性を検証するため、基準を用いずに直観的な判断で要求判定の実験を行い、その結果が複数の作業者間で一致するかを調べる。すなわち、判断基準を用いない方法との比較により方法の代替可能性が否定できれば本手法の有効性を示すことができる。

a) 検証方法および結果

実験は、日本語母語話者の大学生20代男性および20代女性の2人DとEを対象とした。2(2)節の実験で作業者BCが対象としたテキストと同じ4,000文を使用した。二者間で相談することなく個別に作業を行い、作業にあたって次の作業指示を提示した。

- ・例示する各回答文に文脈的つながりがないこと
- ・回答文が要求の場合は1、要求でない場合は0を直観的に判断すること
- ・無回答やその他の回答は無効になるので、必ずいずれかを選択すること

これらの作業によって、表-4の結果が得られた。被験者Eに無回答が2件あったため、得られた回答数は3,998件である。2(2)節の作業者BCと共通のデータであるため、表-4では結果を併記する。DEの作業の一致率は、BCより低く、かつ表-2に示した一致率(κ 値)のいずれよりも低い0.17となっている。また、この値は κ 値の観点からは一致がないことを示す。このことから、主観判断による要求意図の判定は結果が安定しないことがわかつ

表 - 4 有効性の実験結果

D	E			B	C		
	要求	要求でない	計		言い換え可	言い換え不可	計
要求	562	1,880	2,442	言い換え可	2,119	50	2,169
要求でない	39	1,517	1,556	言い換え不可	934	897	1,831
計	601	3,397	3,998	計	3,053	947	4,000
DEの κ 値	0.17			BCの κ 値	0.49		

た。これは基準がないためであるといえる。よって、言い換えを用いた判定基準の有効性が示された。

3. 機械学習による基準の検証実験

本節では、要求意図判定基準が機械学習を利用することにより自動化できることを示す。

(1) 機械学習による意図判定の概要

2節で述べた意図判定法は、機械学習の観点からは与えられた自由回答を「てほしい」に書換え可能なクラス(「可」)と不可能なクラス(「不可」)に分類する2値分類問題として定式化できる。

a) タグ付きデータ

機械学習のためのデータとしては、3,001文のタグ付きデータを2(1)節の分析対象とは別に用意した。このタグ付きデータにおいて「可」の回答数は1,944であり、「不可」の回答数は1,057である。ここでは、全ての回答を「可」と選んだ場合の正解率 $0.648(1,944 / (1,944 + 1,057) = 0.648)$ の正解率をベースライン(BL)とする。

b) 利用した機械学習手法

高精度な分類のために、本節で用い、かつ比較する機械学習法は最大エントロピー法(ME)³⁾とSupport Vector Machine(SVM)⁴⁾である。手法の詳細は省略するが、概要は以下の通りである。まず、回答 X は数値ベクトル $x = [x_1, x_2, \dots, x_n]$ により表現する。ここで、

$$x_i = \begin{cases} 1 & \text{素性 } i \text{ が回答 } x \text{ に含まれるとき} \\ 0 & \text{素性 } i \text{ が回答 } x \text{ に含まれないとき} \end{cases}$$

である。

MEやSVMでは、あらかじめ「可/不可」に分類された訓練データを利用して、回答における素性値の0/1のパターンから「可/不可」への写像を学習し、それを利用してテストデータにおける回答を分類する。なお、本節では、MEの学習にはmaxent^(注3)を用い、SVMの学習にはTinySVM^(注4)を用いた。SVMの学習においては、1次の多項式カーネルを用いた^(注5)。

学習に利用した素性の集合について、本節では3通りの素性集合を比較した。これらを、 $n=1$ 、 $n=2$ 、 $n=3$ と示す。 $n=1$ は任意の形態素の集合、 $n=2$

は $n=1$ に任意の形態素の2連鎖を追加した集合、 $n=3$ は $n=2$ に任意の形態素の3連鎖を追加した集合である。また、ある回答 x に含まれる素性とは、 x が形態素列により w_1, w_2, \dots, w_m で表現されているとすると $w_1 = b$ 、 $w_m = e$ という文頭と文末を示す特殊形態素として、 $n=1$ については $\{w_i \mid 2 \leq i \leq m-1\}$ に含まれる形態素であり、 $n=2$ については $n=1$ の場合に加えて $\{w_i w_{i+1} w_{i+2} \mid 1 \leq i \leq m-2\}$ を追加した集合に含まれる形態素連鎖であり、 $n=3$ については $n=2$ の場合に加えて $\{w_i w_{i+1} w_{i+2} \mid 1 \leq i \leq m-2\}$ を追加した集合に含まれる形態素連鎖である。なお、回答は茶釜^(注6)を利用して形態素に分割した。

c) 実験手順と実験結果

実験の手順としては、10分割交差検定(10 fold cross validation)を利用した。MEとSVMの全体の正解率 P_{ME} 、 P_{SVM} は各分割での正解率の平均により定義する。また、MEとSVMの精度を統計的に比較するためウェルチ(Welch)検定を利用した。なお、MEとBL、およびSVMとBLとの比較においては、MEとSVMとがBLの正解率0.648よりも大きな値をとるかをt検定で比較した。これらの検定はいずれも片側検定とした。このようにして求めた正解率を表-5に示す。表の「 $n=1$ 」「 $n=2$ 」「 $n=3$ 」の行はそれぞれの素性集合を用いた正解率であり、「ME」「SVM」「BL」の列はそれぞれの学習法による正解率である。表-5は、1)どの n についても、MEとSVMの両手法ともBLよりも正解率が高いこと、2)MEとSVMとで正解率に有意差はないため、学習法としてはどちらを用いても良いこと、また、3) $n=1$ と $n=2$ 、 $n=1$ と $n=3$ 、 $n=2$ と $n=3$ について、 $n=2$ と $n=3$ では有意差はないが、 $n=1$ と $n=2$ 、 $n=1$ と $n=3$ には有意差(有意水準 1×10^{-3})があり、素性としては $n=2$ か $n=3$ が良いこと、を示している。以上は相対的な

表-5 正解率の比較

	ME	SVM	BL
$n=1$	0.892	0.887	0.648
$n=2$	0.912	0.909	0.648
$n=3$	0.913	0.915	0.648

正解率の比較である。絶対的な正解率では、 $n = 2, 3$ がMEもSVMも91%前後の正解率である。この正解率はBLよりも十分に高いので、機械学習が有効であったことを示している。また、これは自動的に精度良く書き換え可能性を判定できることも示している。

(2) 異なる種類のデータへの判定基準の適用

ここでは、本研究での対象データ以外の自由回答をデータとした場合にも判定実験で十分な精度を得られるかどうか調べる。異なるデータとして対象としたのは、ある地方都市の国道の満足度に関する意識調査の自由回答である。該当の国道について、信号の長さの適切さ、バス専用レーン運用の適切さ、街路樹の有無など20項目に対して「満足/やや満足/どちらでもない/やや不満/不満/無回答」といった5段階の顧客満足度を測定するとともに、各項目に対する自由回答欄を設けている。自由回答は回答数577件全文数1,290文で、一回答あたりの平均文数は2.24文である。このデータに対して前節同様の分類実験を行い表-6の結果を得た。これらの結果の全文に対して目視による要求判定を行った。この結果、異なるデータへの適用においても①の正解率94.46%、②の正解率83.06%、全体の正解率87.21%という高精度の結果が得られた。

本節の実験で興味深いのはデータの性質の違いが統計的に示されたことである。これまで対象とした将来の交通に対する要望・提案型のアンケートでは、2(1)節の分析対象データおよび3(1)節の学習コーパスともに言い換え不可の割合が35.2~41.7%、2(2)節のABCの作業員においても15.3~45.8%だったのに対し、本節のデータでは63.6%を占めている。ここで、言い換え不可すなわち非要求文と判定された回答文は「困る」「～づらい」「～にくい」など明らかに不満を表明した文であった。これは、本研究で扱った将来の交通に対する要望・要求では回答者が意見として明示的な要求を示す傾向があるのに対し、顧客満足度という観点で測られる現状の

表-6 機械学習による分類の正誤判定

	正	誤	計
①要求	443	26	469
②非要求	682	139	821

意識調査では意見に不満が表明される傾向が強いことを示す。目的など調査の状況によって得るべき情報を判断するという考察にも即したものとなった。

4. おわりに

要求意図判定基準を仮説として2節では内容分析の知見を踏まえて予測的妥当性、再現可能性、相関的妥当性の三観点から検証を行い、判定基準の有効性を示した。3節では判定基準と自動分類を統合し、判定基準を用いて作成した正解データによる機械学習手法の実験を再び行うことにより91%の精度で要求か否かを判定できた。すなわち、機械学習により要求を含むテキスト・含まないテキストへの自動分類を91%の精度で実現した。さらに、異なる種類のデータで実験した場合も87%の高精度で要求を判定できた。今後は、要求内容や動機の抽出を行い、さらに他の観点による分析結果ならびに統合結果の提示に関する検討も行う。

注

注1) 再現性については、2(2)節で検証する。

注2) 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である。②は直接要求表現を含むが、引用文中に現れているため言い換えは不可であった。これは例外事例といえる。

注3) <http://www.nict.go.jp/jt/a132/members/muti-yama/software.html>

注4) <http://cl.aist-nara.ac.jp/~takuku/software/TinySVM/>

注5) 2次の多項式カーネルも用いたが、その精度は、1次の場合よりも統計的に有意差がないかもしくは低かったため、ここでは割愛する。

注6) <http://chasen.aist-nara.ac.jp/>

参考文献

- 1) Berger, A. L., Della Pietra, S. A. and Della Pietra, V. J. (1996) A Maximum Entropy Approach to Natural Language Processing. Computational Linguistics, Vol. 22 No. 1, pp. 39 - 71.
- 2) Carletta, J. (1996) Assessing Agreement on Classification Tasks: The Kappa Statistic. Computational Linguistics, Vol. 22 No. 2, pp. 249 - 254.
- 3) Carletta, J., Isard, A., Isard, S., Jowtko, J. C., Doherty - Sneddon, G. and Anderson, A.H. (1997) The Reliability of a Dialogue Structure Coding Scheme.

-
- Computational Linguistics 23 (1) pp .13 - 32.
- 4) Cohen, J . (1960) A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20, pp. 37 - 46.
- 5) Cristianini, N. and Shawe - Taylor, J. (2000) An Introduction to Support Vector Machines. Cambridge University Press.
- 6) 道路審議会基本政策部会 (1996) ボイス・レポート . 21 世紀の道を考える委員会 建設省道路局建設省都市局 .
- 7) 乾裕子・井佐原均 (2002) 拡張モダリティの提案 - 自由回答から回答者の意図を判定するために - , 電子情報通信学会信学技報 NLC 2002 - 43 (2002 - 10) , pp 31 - 36 .
- 8) Mani, I (2001) Automatic Summarization, John Benjamins Publishing Company , 奥村学・難波英嗣・植田禎子 (2003) 『自動要約』 共立出版 .
- 9) 国立国語研究所 (1960) 『話しことばの文型』 秀英出版 .
- 10) Krippendorff, K. (1980) Content Analysis: An Introduction to Its Methodology, Sage Publication, K・クリッペンドルフ著 三上俊治・椎野信雄・橋元良明訳 (1989) 『メッセージ分析の技法 ; 「内容分析」への招待』 勁草書房 .
- 11) 森田良行・松木正恵 (1989) 『日本語表現文型』, アルク
- 12) 大塚 (乾) 裕子 (2004) 自由記述アンケート回答の意図抽出および自動分類に関する研究 - 要求意図を中心に - , 神戸大学大学院自然科学研究科博士論文 .
- 13) 内山将夫・大塚裕子・井佐原均 (2004) フェイスシートとの関係を利用した自由回答アンケートの分析, 情報処理学会研究報告 2004 - NL - 164 , pp 47 - 52 .