社会情報解析への一寄与：形式概念によるデータ解析

A Contribution to Social Information Analysis: Data Analysis using Formal Concept Analysis

長田　博泰

Formal Concept Analysis (FCA), introduced as a formalization of the concept of 'concept', has grown to a powerful theory for data analysis, information retrieval, and knowledge discovery. In this paper, we present a method based on the use of FCA for the data analysis when dealing with real-world data sets. The usefulness and clarity of the method are illustrated by data analysis of the questionnaire on junior high school and high school students' use and awareness of *keigo*, or Japanese honorific/polite forms.

1．はじめに

社会の現象は複雑・多様であり，その要因を明らかにすることは容易ではない．それは社会現象と人間の意識・行為が互いに影響しあいその絡み合いを分離することが難しく，さらにその解明に関わっている観察者自身も間接的であれその現象に取り込まれているためである．このような特性を有する社会現象を研究対象とする社会科学が社会認識の客観性を保つには何らかの方法が必要となる．その有力な手段が社会調査であり，その中には観察，聞き取り調査あるいは質問紙調査など多様な方法が含まれる．とくに質問紙調査に基づく研究では調査対象者を多くすることによって大数の法則に基づいた分析が可能であり，さらに統計的計量的方法にもとづくデータ解析が可能になる．

しかし，この種のデータ解析は，調査項目の特性を構成比，平均等の単純な代表値で表し，そのわかりやすさもあってそれだけがひとり歩きしがちであり，調査項目間の関連を分析するにしても比較的少ない項目しか対象にしない傾向があるように思われる．さらに，多変量データの解析方法にはつぎの問題点がある．すなわち，一般に元のデータを数量化する際に元のデータの情報を一部失っており，また，データ間に設定された〝距離〟がどのような〝意味〟をもつかを解釈することが難しく，結果の解釈も多義的である（Wolff, 1996）．

社会情報過程は「価値と論理の織りなす情報過程」であり，これが社会情報に一つの特徴をもたらすこと，すなわち「現実の社会情報が多くの場合互いに矛盾を含む複数の価値システムから構成されていることである」（田中，1999，p.87）．そして社会情報の論理過程と対立する複数の価値システムの相互連関を明らかにするアプローチを社会情報解析と呼び，その分析を試みている（大國他，1999）．この立場に立つならば，質問紙調査等にもとづく情報は複数の価値システムからなる価値

NAGATA Hiroyasu　札幌学院大学社会情報学部

と論理にもとづく過程を含む社会情報である．

以上を踏まえて，質問紙法等による調査の解析方法を改めて考えるために，いま調査項目を Qj，各調査項目に対する回答を Aj（j＝1，…，m）とし，調査対象に対するひとりの回答者の意識・意見等を組合せ（A 1，A 2，…，Am）で表すことにしよう．調査によってこのような回答の組合せ（Ai 1，Ai 2，…，Aim）（i＝1，…，n）が多数得られる．回答の組合せの集合が対象の全体的記述である．この集合を特徴づけるには 2 つの視点が必要である．ひとつは，回答集合の全体的構造を表現し，その特徴を論理的に把握することである．もうひとつは，調査項目間の関係を引き出すことである．後者は全体の特徴を捉えることにも関連するが，これに尽きるものではない．項目間の関係には何らかの関係，例えば含意関係などを見出すことによって相互の意味的考察が可能になり，その場合さらに新たな考察を示唆することになるので，この視点を欠くことはできない．

上述の観点からデータ解析を行うことを可能にする有効な方法のひとつが形式概念分析である．この方法は，1980 年代前半に提案されたものであり（Wille, 1982），対象を属性記述表現し，属性間に成立する（集合的）包含関係を明らかにし，またその属性間に成立する（論理的）含意関係を見出そうとするものである．この方法は多くの分野に適用可能である．実際，データ解析，情報検索，知識発見など様々な分野で適用が試みられている（Ganter, et al (eds): 2005）．長田（2004）は，社会調査データに適用し，その有効性を示した．しかし，そこで分析対象としたデータセットは比較的小さく実用規模の大きさのデータセットではなかった．形式概念分析をある程度の大きさのデータセットに適用しようとすると，克服すべき固有の問題，たとえば，概念数が非常に多くなり，見通しのよい図示が不可能になるなどの問題がある．

本稿の目的は，実用規模のデータ，国立国語研究所が 1989 年度～1992 年度に行った東京・大阪・山形の中学生・高校生を対象に実施した敬語意識に関する調査データに形式概念分析を適用する方法を具体的に展開するとともに，調査対象の全体構造と特徴を把握することである．とくに，調査項目間に成立する含意などの関係に基づいて，論理と複数の価値システムの相互連関を明らかにしようとする社会情報解析の一つの方法を提示しようとするものである．

以下，2 節で形式概念分析の展開に必要な定義およびデータ解析方法を説明し，3 節では具体例として上で述べたアンケート調査データに適用し，全体的構造と特徴を捉え，含意関係の分析から導かれる特徴点を指摘する．4 節では国立国語研究所の報告書（国立国語研究所，2002）で用いられている数量的方法と形式概念分析による結果を比較し，その長短を論ずる．

2．形式概念とデータ解析

順序は日常生活のあらゆる面に浸透している．一番，二番，……はもちろん，大きい－小さい，よい－わるい，満足－不満など．これらにその程度を表す「まあ」あるいは「やや」などの形容詞をつけることも可能である．このような順序とその順序関係[1]にもとづく構造は数学の一分野である束論を用いて扱うことができる．とくに〝概念〟を束論の枠組みの中で形式的に扱う形式概念分析（formal concept analysis）が提案されてから（Wille, 1982: Ganter & Wille, 1999），これを用いてデータの集合の複雑な構造を解析することが可能になった．

2.1 形式概念

形式概念分析の議論に必要な用語等を太陽系惑星の属性記述を用いて説明する（Davey

表１　太陽系属性表

	小	中	大	近い	遠い	衛星有	衛星無
水　星	×			×			×
金　星	×			×			×
地　球	×			×		×	
火　星	×			×		×	
木　星			×		×	×	
土　星			×		×	×	
天王星		×			×	×	
海王星		×			×	×	
冥王星	×				×	×	

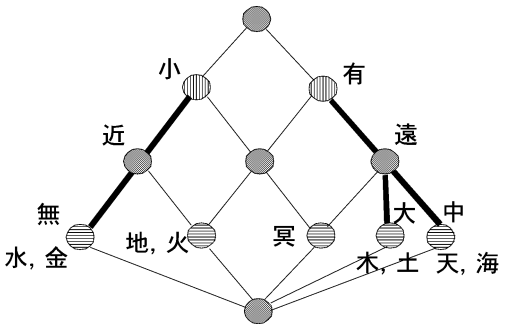

図１　太陽系 Hasse 図

& Priestley, 2002)．ここでは惑星の属性のうち，大きさ，太陽からの距離および衛星の有無だけに着目し，各属性はそれぞれ，小/中/大，近い/遠い，有/無の属性値をとるものとする．以下，属性値まで含めて属性と呼ぶことにする．９個の惑星の属性を表１のように整理することができる（網掛け部分についての説明は後述する)．

対象の集合を G，属性の集合を M，対象 g∈G が属性 m∈M をもつことを２項関係 I で表し，gIm と記す．K＝(G，M，I) を形式文脈といい，簡単なものは表１のようなクロス表で表すことができる．

A⊆G に対し，A のすべての対象に共通な属性の集合を A’ で表す．同様に，B⊆M に対し，B のすべての属性をもつ対象を B’ で表す．このとき，文脈（G，M，I）の形式概念を A’＝B かつ B’＝A である A，B の対(A，B)で定義し，A を外延，B を内包という．表１において，A 1＝{水星，金星，地球，火星}，B 1＝{小，近い}とすれば，A 1’＝{小，近い}，B 1’＝{水星，金星，地球，火星}であるから，A 1’＝B 1，B 1’＝A 1 であり，(A 1，B 1)は形式概念である．また，同様に，A 2＝{水星，金星，地球，火星，冥王星}，B 2＝{小}とすれば，（A 2，B 2）も形式概念である．

２つの概念（A 1，B 1)，（A 2，B 2）に対し，A 1⊆A 2(あるいは B 2⊆B 1)なら，(A 1，B 1)≦(A 2，B 2)と定義する．上で例示した２つの概念(A 1，B 1)，(A 2，B 2)で A 1⊆A 2 であるから，(A 1，B 1)≦(A 2，B 2)である．≦は概念の順序関係であり，(G，M，I)の概念の集合全体 B(G，M，I)は完備束[2] であり，とくに概念束という．概念束を図式(Hasse 図)で表すことができる．表１の概念束は図１である．

図１の各丸印（以下，ノードという）が形式概念を表す．各ノードの上側に属性，下側に対象を記入してある．最下位の概念のすぐ上の概念をアトム，最上位のすぐ下の概念をコアトムという．図１でいえば，横ハッチングの概念がアトム，縦ハッチングの概念がコアトムである．

2.2　含意規則

クロス表から形式概念間の関係だけでなく，属性間の含意規則を見つけ出すことができる．すなわち，ある属性（A 1，A 2，…，Am）をとるとき，必ず有する他の属性(B 1，B 2，…，Bn)を見出すことである．この関係を A 1，A 2，…，Am　⇒　B 1，B 2，…，Bn で表す．表１ではつぎの６個の属性含意規則が成立する．これは，図１中の太線を上向きにたどることを意味する．

１）衛星無　⇒　小，近い
２）遠い　　⇒　衛星有
３）近い　　⇒　小
４）大　　⇒　遠い，衛星有

5）中 ⇒ 遠い，衛星有
6）小，近い，衛星有，衛星無 ⇒ 小，中，大，近い，遠い，衛星有，衛星無

これらの含意規則から表1を再構成することができる．実際，1）から5）の含意規則によって表1中の網掛け部分が埋まる．6）の左辺には両立しない属性(「衛星有」,「衛星無」)を含んだ，いわゆる矛盾則である．矛盾則からは何でも導くことができる．6）の右辺はこれを示している．表1の網掛け以外の部分はこの矛盾則を用いてその属性値を決めればよい．

2.3 形式概念によるデータ解析手法

形式概念によってアンケート調査等のデータを解析するということは，回答の組合せ集合を概念束として構成するとともに，回答項目間の関係（含意規則）を導出することである．しかし，対象数が大きくなると属性の組合せも多様になり全体の見通しがわるくなる．また，含意規則も多数にのぼり，その上各規則も適用範囲が狭く，全体的傾向を特徴づけるものではない場合も少なくない．したがって，全体を把握するには実用上もう少しキメの粗い捉え方をする必要がある．

ここでは，データマイニングの分野で用いられているアイテム集合の考え方を採用する(Kantardzic, 2003)．アイテム集合とは，探索集合の中である条件を満たしている部分集合である．これを次のように形式概念に導入する．すなわち，B⊆M を属性集合とし，属性集合 B の支持度 (support count) を以下のように定義する．

$$\mathrm{supp}(B) = |B'|/|G|,$$

ここで｜｜は個数を表す

支持度のしきい値を表す最小支持度 (minimum support) 以上の頻度で現れる属性の集合 B，すなわち $\mathrm{supp}(B) \geqq \mathrm{minsupp} \in [0, 1]$ ならBを多頻度アイテム集合 (frequent item/attribute set) という．

形式文脈(G，M，I)と与えられた minsupp に対し，

$$\{(A, B) \in B(G, M, I) \mid \mathrm{supp}(B) \geqq \mathrm{minsupp}\}$$

を考える．最小支持度以上の支持度をとる属性集合とその対象集合との対の集合に最小要素を付加すると，束を構成することを示すことができる．これを粗い概念束（氷山概念束 ── iceberg concept lattice）とよぶ（Stumme, 2002)．要するに，概念束の中で，ある個数以上の対象を含む外延と内包から構成される束であり，一部だけが海面に現れ，大部分は海面下にある氷山のごときものである．

表1の概念束から粗い概念束を構成して適用してみよう．まず，minsupp＝0.5 とすると，つぎの概念が選ばれる：

$|\{有\}'| = 7$，$\mathrm{supp}(\{有\}) = 7/9 = 0.78$
$|\{遠い, 有\}'| = 5$，
$\mathrm{supp}(\{遠い, 有\}) = 5/9 = 0.56$
$|\{小\}'| = 5$，$\mathrm{supp}(\{小\}) = 5/9 = 0.56$

この三つの属性から概念束を描いた結果が図2 a) である．つぎに minsupp＝0.3 にすると，以下の属性集合が追加される．

$|\{小, 近い\}'| = 4$，$\mathrm{supp}(\{小, 近い\}) = 4/9 = 0.44$
$|\{小, 有\}'| = 3$, $\mathrm{supp}(\{小, 有\}) = 3/9 = 0.33$

この図を描くには，0.3 以上のノードをすべて含んだ概念束を直接描くのではなく，0.5 に対して追加された概念のみを描き（図2 b)，共通の概念を結合するのがよい．こうすると図の構成が容易になり，また全体の見通しが得やすい．

さらに minsupp＝0.2 とすると，以下の4つの属性集合が追加される（図2 c)．

$|\{小, 近い, 無\}'| = 2$，
$\mathrm{supp}(\{小, 近い, 無\}) = 2/9 = 0.22$
$|\{小, 近い, 有\}'| = 2$，
$\mathrm{supp}(\{小, 近い, 有\}) = 2/9 = 0.22$
$|\{大, 遠い, 有\}'| = 2$，
$\mathrm{supp}(\{大, 遠い, 有\}) = 2/9 = 0.22$

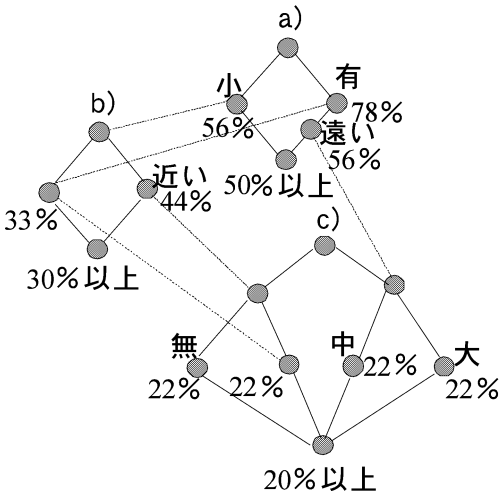

図2　表1の粗い概念束

|{中，遠い，有}'|＝2,

supp({中，遠い，有})＝2/9＝0.22

以上の属性集合からなる概念束を描くと図2になる．

上で述べた処理を minsupp＝0.1 まで続けると，残りの属性集合 {有，遠い，小} を図示することができるが，このプロセスは，結局のところ対象を属性によってクラスタリングしているのである．樹木図で表せば図3のようになる．この方法を概念的クラスとリングというが，通常のクラスタリングに比べつぎの利点を有する．

1）データの入力順序によってクラスタリングが変わることはない．

2）変換によってデータの有する情報が失われることはない．

3）通常のクラスタリングではクラスターの意味を属性によって推測する必要があるが，概念的クラスタリングではその必要はない．

通常の概念束では含意規則は恒に成立するが，粗い概念束では恒に成立するとは限らない．そのために含意規則を少し弱め，データマイニングで用いられている連関規則とみなすことにする．M を属性集合とし，X⊆M,

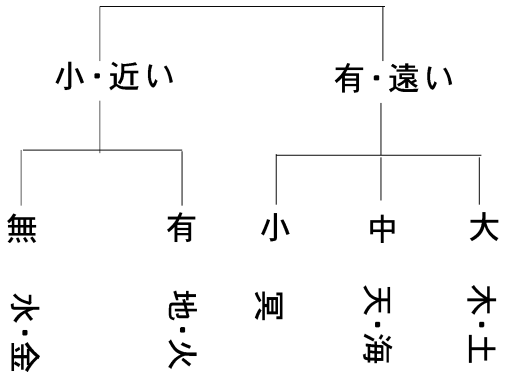

図3　太陽系惑星のクラスタリング

Y⊆M かつ X∩Y＝ϕ であるなら，X⇒Y を連関規則と定義する．さらに，この連関規則の支持度（supp）および確信度（conf）をつぎのように定義する．

supp(X⇒Y)＝def|(X∪Y)'|/|G|

conf(X⇒Y)＝def|(X∪Y)'|/|X'|

とくに確信度1の連関規則を含意規則あるいは正確な連関規則という．たとえば，

小，衛星有　⇒　近い

は連関規則であるが，その支持度および確信度は次のようになる．

supp(小，衛星有 ⇒ 近い)＝|{小，衛星有，近い}'|/|G|＝2/9＝0.22…

conf(小，衛星有 ⇒ 近い)＝|{小，衛星有，近い}'|/|{小，衛星有}'|＝2/3＝0.66…

3．適用例

上述の方法を実際の調査データに適用する．データは国立国語研究所が1989年度～1992年度に行った東京･大阪･山形の中学生･高校生を対象に実施した敬語意識に関する調査である．

3.1　データ概要

無記名自記式によるアンケート調査の対象者はつぎのとおりである．

・東京中学　21校　2456人(男子1285人，女子1171人)

・東京高校　25 校　2222 人(男子 1157 人，女子 1060 人，性別不明 5 人)
・大阪高校　10 校　1004 人（男子 472 人，女子 530 人，性別不明 2 人）
・山形中学　1 校　339 人(男子 161 人，女子 178 人)

調査項目は，以下に示す項目からなる．

Ⅰ「敬語についての意識」を問う調査項目 1 ～ 8
Ⅱ「敬語の使用」の具体的調査項目 9 ～15
Ⅲ「敬語についての意見」を尋ねる調査項目 16～19
Ⅳ フェースシート（東京以外に暮らしたか，一番長く住んだ所，両親の育った所，家の仕事，etc）

3.2 データ解析方針

形式概念によるデータ解析の有効性を示すことが目的なので，ここでは上のデータのうち東京中学，山形中学だけを分析対象とする．山形･東京合計約 2800 人のデータがあり，また調査項目の選択肢（ここでは，これらが属性として扱われる）が合計 140 個ほどである．これらの属性を一度に形式概念分析を適用することはコンピュータの処理能力上不可能である．また，敬語の具体的な使用に関わる場面や具体的敬語表現の使用に関する調査である質問 7 ～15 は，むしろ言語行動的･社会言語学的研究がふさわしいと思われるので，ここではこれらの質問を分析対象外とする．

質問 1 ～ 6 および質問 16～19 を分析対象とする（付録「ことばのアンケート（抜粋）」を参照）．前者は敬語についての意識，とくにその現状をどう評価し，感じているかについての調査項目であり，後者はそのような評価・感覚を抱いている生徒が敬語に対してどのような意見を有しているかを尋ねている．これらについて次の 2 点に注目し分析する．ひとつは，属性ごとに行う分析では明らかにしにくい属性値の組合せから捉えられる概念束の全体的構造を明らかにし，そこから全体的特徴点を闡明することである．いまひとつは，調査項目間にどのような関係等が存在するかを明らかにすることである．これによって，敬語に関する実際の評価意識と意見の間の関係が見出せる可能性がある．

分析対象項目以外では地域（山形・東京），性別（男子・女子）を考慮する．予備的分析で学年も考慮する意味があると予想されたが，東京のデータには 3 年生，1 年生がそれぞれ 4 件，2 件しか含まれておらず，山形の学年データと比較できないので，学年別分析は行わない．

3.3 データ解析結果

地域・男女別に行った形式概念分析をおこなった結果の概要を表 2 に示す．この表から，まず概念数が非常に多く Hasse 図に表すことは事実上不可能であること，また属性数に比し，含意規則が多いことがわかる．アトム数は異なる属性組合せの個数を表すが，対象数が異なりこのままでは比較できないので，アトム数を対象人数の平方根で割った値を示しておいた．概念束の全体的構造を把握する

表 2　形式概念分析結果

	人数(アトム内人数)	属性数	概念数	含意規則	アトム数	広がり[1)]
山形	男子　161　(160)	20	9961	1027	132	10.4
	女子　178　(177)	20	7197	864	109	8.2
東京	男子　1283（1237）	20	32889	1791	417	11.9
	女子　1170（1138）	20	28121	1542	369	10.9

1）広がり＝アトム数/$\sqrt{N}$

ため，まず 2.3 で述べた粗い概念束から全体的構造とその特徴を捉える分析を行い，ついで連関規則から属性間の関係を見出すことにする．

3.3.1　全体的構造と特徴

質問 1-6, 16-19 はいずれも二者択一の回答を求めているから，どちらかの回答の構成比は 50%以上である．したがって，構成比が 50%以上の概念だけをとりだせば，各質問項目単独で 50%以上の回答とそれらの回答どうしの絡み合いが把握できるはずである．しかし，構成比が 50%以上の概念を直接描いてもいまだ複雑であり，全体的構造が見えにくい可能性もあるので，構成比 50%以上の概念をさらに 3 つのレベル，すなわち 70%以上，70～60%，60～50%で粗い概念図を描くことにする．4 つのグループについて粗い概念図を描いた結果が図 4－図 7 である．これらの図はそれぞれ異なる様相を示し，何らかの特徴が表現しているように思われるので，これについて考察する．

⑴　男女の相違点

図 4－図 5 と図 6－図 7 から男女間の回答の傾向に以下の相違を読みとることができる．

1）男子は女子に比べ，関連する項目数が少ない，すなわち，男子では 70%以上（実際には 80%程度）の回答が集中するのは東京・山形とも 2 項目に限られるのに対し，女子では 4 ～ 6 項目が 70%以上である．

2）さらに 50%～70%についても，女子に比

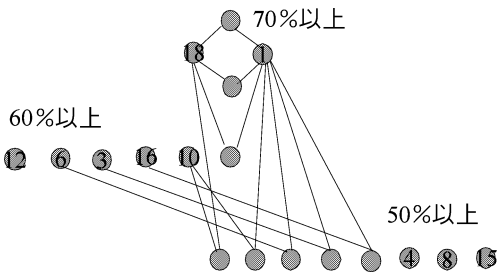

図 4　山形男子粗い概念図

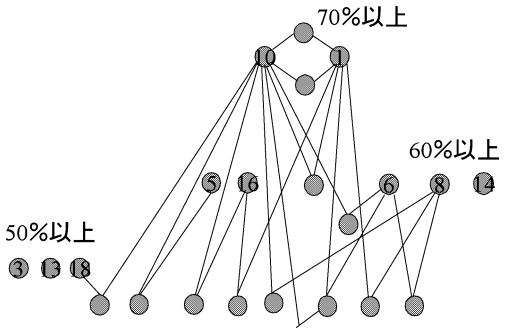

図 5　東京男子粗い概念図

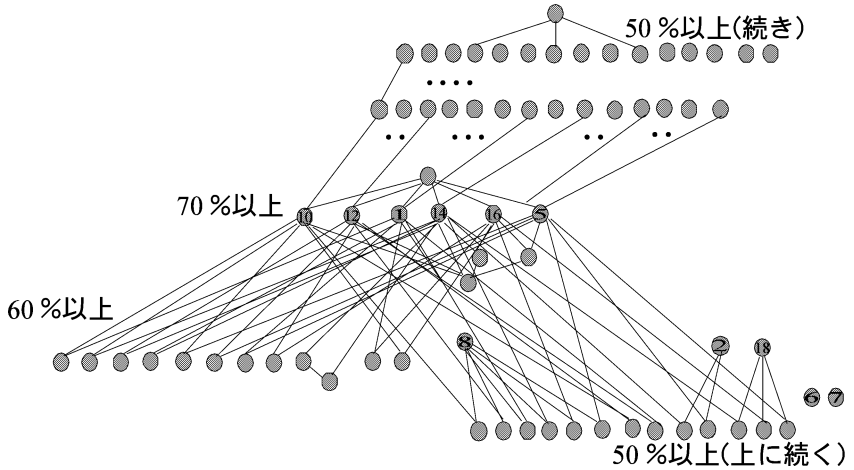

図 6　山形女子粗い概念図

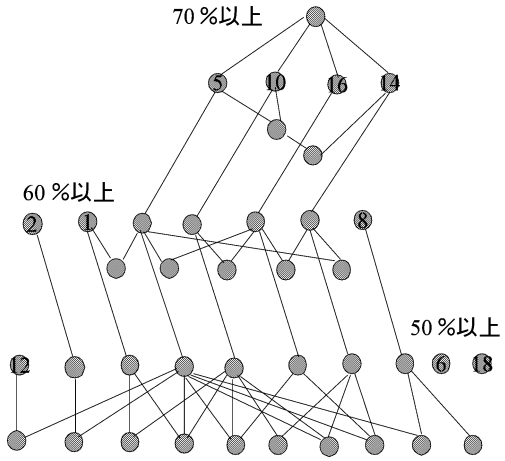

図7　東京女子粗い概念図

べ男子では関連する回答数が少ない．これは表2に掲げた概念数，アトム数からも説明することができる．すなわち，山形・東京とも女子に比べ男子のほうが，概念数およびアトム数が多い，つまり，少数派が多数いることを示している．

3）女子は男子に比べ関連する質問項目が多い，とくに山形女子はその傾向が顕著であり，東京女子に比べて50％～60％の概念数が多い．

4）男女･地域を問わず，質問18に対して「敬語は上下の規律が守れ，授業や部(クラブ)活動などの学校生活をするうえで欠かせないものだ（コード16)」という回答が60％以上あり，女子では70％にのぼり，とくに山形女子は80％を超える．

⑵　山形男子・東京男子の相違点

山形男子・東京男子の間には以下の相違を指摘することができる．

1）上述のように男子は50％以上の項目どうしが関連する概念が女子に比べ少ないが，山形男子は東京男子に比べ，さらにその数が少ない．山形男子は比較的少数の項目で特徴づけることができる．

2）山形男子の70％以上のものが，先生や上級生に対し，ていねい語や敬語を使うとよそよそしくなる（コード18）と思っている．

3）3つの質問，質問3（先生等に対することばづかい)，質問16(授業等に改まったことばづかい）および質問17（上級生等に対することばづかい）で山形と東京の過半数を超える回答が逆転する．すなわち，山形では3つの質問に対する回答がそれぞれ「あまり変わらない（コード4）」，「あらたまった，きちんとしたことばづかいがよい（コード12)」，「使わなくてもよい（コード15)」であるのに対し，東京はそれぞれ「変わると思う（コード5）」，「ふだんどおりの，ふつうのことばづかいでよい（コード13)」，「使うほうがよい（コード14)」である．

⑶　山形女子・東京女子の相違点

山形，東京の女子にはつぎの相違がある．

1）過半数を超える単独の回答項目は山形・東京で全く同じであり，その意味では両地域に差がないといえる．しかし詳しく見てゆくと，さらにつぎの相違点が見えてくる．

2）質問3，6，17, 18は山形･東京とも70％以上同じ回答をしているが，山形では質問1（ことばづかいが気になるか）と質問16（授業等に改まったことばづかい）に対し70％以上が「気にならない（コード1）」，「あらたまった，きちんとしたことばづかいがよい（コード12)」と回答しているのが目立つ．山形男子の場合も「ことばづかいが気になら」ず，東京に比し「あらたまった，きちんとしたことばづかいがよい」が多かったが，山形女子についても同様の傾向が見られることを意味する．

3）東京女子の場合，質問3（先生等に対することばづかいがかわるか）に対して「変わると思う（コード5）」という回答が構成比50％以上になる属性組合せに絡んでいる．

表 3　含意規則（支持度上位 10 個）

山形男子		東京男子		山形女子		東京女子	
含意規則	支持度	含意規則	支持度	含意規則	支持度	含意規則	支持度
3,9,18⇒1	0.19	3,4,10,16,18⇒1	0.023	5,6,10,12⇒1	0.27	3,5,6,8,12,16⇒10	0.059
3,14,18⇒1	0.13	3,4,6,13,15,17⇒1	0.016	3,8⇒1	0.27	0,2,7,8,14⇒5	0.048
3,7,9⇒1	0.11	3,4,6,8,13,15⇒1	0.014	6,10,12,14⇒1	0.26	2,7,10,12,14,16,18⇒5	0.044
3,6,8,13⇒4	0.11	3,4,6,13,15,18⇒1	0.014	6,10,14,18⇒1	0.23	3,6,8,12,16,19⇒10	0.042
3,7,11⇒1	0.11	3,4,8,10,13,16⇒1	0.013	5,6,8,12⇒1	0.23	0,6,12,19⇒16	0.041
3,5,10,18⇒1	0.11	1,5,7,16,19⇒14	0.013	6,8,12,14⇒1	0.22	7,11,19⇒5	0.041
3,5,7⇒1	0.11	3,4,6,8,13,16⇒1	0.013	3,6⇒1	0.21	3,5,6,8,16,18⇒10	0.040
5,9,14⇒16	0.11	0,8,10,12,1⇒16	0.013	6,8,14,18⇒1	0.20	0,2,7,8,12⇒5	0.038
1,9,13,15⇒18	0.11	3,4,6,8,13,19⇒1	0.013	5,6,8,18⇒1	0.20	0,2,7,13,14⇒5	0.037
6,13,16,18⇒1	0.10	0,12,14,19⇒16	0.012	1,7,12,14⇒16	0.17	0,9,10,14,19⇒16	0.035

4）山形女子では，質問 1（ことばづかいが気になるか－気にならない），質問 3（先生等に対することばづかいが変わるか－変わると思う）および質問 17（上級生等に対することばづかい－使うほうがよい）が構成比 50％以上になる属性組合せに絡んでいる．

3.3.2　連関規則からみた特徴

粗い概念束からアンケート結果の全体構造とその特徴点を捉えられることを示した．つぎに視点を変え，各質問項目の関連を分析して見ることにする．そのために 2.3 で述べた連関規則あるいは含意規則を調べることにする．表 2 に示したように含意規則はその数も多く，また該当する規則の支持度も小さい．参考までに表 3 に各グループの含意規則のうち支持度の高い上位 10 個を掲げる．

この表の支持度から判断する限り，含意規則は細かすぎ，調査項目間の全体的関係を把握するには適切ではないように思われる．まず，図 4－図 7 に示した粗い概念図の上で成り立つ連関規則を見出すことにし，含意規則については後述する（3.3.3）．2 つ以上の属性からなるノード（概念）を X⇒Y，X⋂Y＝ϕ に分解し，確信度 conf（X⇒Y）のできるだけ高いものを取り出せばよい．こうして求めた連関規則と信頼度を表 4 に示す．以下，各グループの質問項目間の特徴点を調べる．

⑴　男女差

表 4 を見てわかるとおり，男子にはとくに注目すべき連関規則は少ないのに対し，女子には興味深い連関規則が成立している．以下，グループ別に連関規則を検討する．

⑵　山形男子

「（言葉遣いが）気にならない」（コード 1）を含意する規則がほとんどである．これは 80％以上の生徒がコード 1 を回答しているためであって，敬語使用・意識の上で有意味な規則とは考えられない．唯一興味ある連関規則は 10⇒18 である．すなわち，「（学校での言葉遣いで）困った経験がない」（コード 10）生徒は，「（敬語はよそよそしいと）思う」（コード 18）傾向が強い．

⑶　東京男子

ここでも出現頻度の高い質問 6 の「（学校での言葉遣いで）困った経験がない」（コード 10）を含意する規則が多数を占め，つぎに多いのが質問 1 の「（言葉遣いが）気にならない」（コード 1）を含意する規則である．このうち注目すべき規則はつぎの 2 つである：

1）「（敬語は）欠かせないものだ」（コード 16）と考える生徒は，「（学校生活での言葉遣いで）困った経験がない」（コード 10）傾向が強い．

2）「（先生や上級生と話すとき言葉遣いが）変わると思う」（コード 5）生徒は，「（学校生活での言葉遣いで）困った経験がない」

（コード 10）と回答している．

なお，山形男子と異なり，東京男子ではコード 10 とコード 18 の関係が逆になっている．すなわち，「（敬語はよそよそしいと）思う」（コード 18）生徒は，「（学校での言葉遣いで）困った経験がない」（コード 10）と回答している．これは東京男子ではコード 10 が 80％以上の多数を占めているためである．

⑷ 山形女子

多くの連関規則があるが，コード 5 （「（先生や上級生と話すとき言葉遣いが）変わると思う」）とコード 16（「（敬語は）欠かせないものだ」）を含意する規則に注目すれば，その特徴を把握することができる．「（言葉遣いが）変わると思う」前提の中でとくに目立つのは，「（クラス討論や授業で）あらたまったきちんとした言葉遣いがよい」（コード 12），「（上級生や先輩などに）敬語を使うほうがよい」（コード 14）などである．「敬語が欠かせない」と思う生徒は，「（先生や上級生と話すとき言葉遣いが）変わると思う」，「（クラス討論や授業で）あらたまったきちんとした言葉遣いがよい」や「（上級生や先輩などに）敬語を使うほうがよい」などと回答する傾向が強い．

⑸ 東京女子

コード 5 （「（先生や上級生と話すとき言葉遣いが）変わると思う」）を含意する規則が極めて多く，その前提として敬語に対する意見である「（上級生や先輩などに）敬語を使うほうがよい」（コード 14），「（敬語は）欠かせないものだ」（コード 16），「（クラス討論や授業で）あらたまったきちんとした言葉遣いがよい」（コード 12）が明確に表れている．現実の言語遣いの上では「（学校生活での言葉遣いで）困った経験がな」く（コード 10），「（言葉遣いが）気にならない」（コード 1）という面も見られる．

女子には連関規則からつぎの地域差を読み取ることができる．東京女子は「敬語は欠かせないものだ」という敬語に対する規範意識がまずあって，その結果として「言葉遣いが変わると思う」などと回答しているのに対し，山形女子ではクラス討論・授業また先輩・上級生との言葉遣いで「敬語を使うほうがよいという意見の結果として「敬語は欠かせないものだ」と考えているように思われる．

山形と東京の女子生徒に認められる上記の相違をもたらす要因を探るにはさらに立ち入った調査が必要である．

3.3.2 含意規則による分析

表 3 に掲げたいずれの含意規則も支持度が小さい（適用範囲が狭い）．したがって，この調査データに関する限り全体的特徴を表現する含意規則は存在しないといってよい．しかし，これは含意規則が考察に値しないことを意味するものではない．狭い範囲で成立する有意味な対象グループの存在を否定するものではないからである．この観点から含意規則による分析を試みる．

いま「（敬語は）欠かせないものだ」（コード 16）と考える生徒（敬語支持派）はどのような言語行動あるいは敬語規範を有しているかを分析したいとする．簡単のため，上位 10 個の含意規則しか挙げていない表 3 を用いることにする．対象グループを東京女子に限ると，コード 16 を含意する規則は表 3 にはつぎの 2 個しかない．

0，6，12，19⇒16　　（支持度　0.041）

0，9，10，14，19⇒16（支持度　0.035）

この 2 つの含意規則から論理的につぎの含意規則を導くことができる．

0，19，（6∧12）∨（9∧10∧14）

⇒16

属性 0，6，9，10，12，14，16，19 を有する概念は存在しないので，支持度は 2 つの支持度の和 0.041＋0.035＝0.076 で与えられる。

この関係は，「（敬語は）欠かせないものだ」（コード 16）と考える東京女子の 7.6％（約 90

人弱）の生徒が「（言葉遣いが）気にな」り（コード 0），「（敬語は）よそよそしくない」（コード 19）と思っているという明確な言語行動と敬語規範を持っていることを表現している（ここでは，とりあえず残りの属性を無視した）．

このように有意味と思われる含意規則を取りあげ，これらを論理操作することによってある対象グループを特定し，その特徴を含意規則によって規定することが可能になる．

表 4　連関規則

山形男子		東京男子		山形女子		東京女子	
連関規則	確信度	連関規則	確信度	連関規則	確信度	連関規則	確信度
18⇒1	0.85	1⇒10	0.85	14⇒16	0.87	10⇒5	0.88
10⇒1	0.83	6⇒10	0.89	5⇒16	0.86	16⇒5	0.90
3⇒1	0.90	8⇒10	0.87	14⇒5	0.83	14⇒5	0.90
6⇒1	0.87	6⇒1	0.83	12⇒5	0.88	1⇒5	0.87
10⇒18	0.75	8⇒1	0.83	1⇒16	0.81	14⇒16	0.83
16⇒1	0.81	16⇒10	0.82	1⇒5	0.80	16⇒10	0.82
		5⇒10	0.82	12⇒16	0.86	14⇒10	0.82
		8⇒6	0.77	1⇒14	0.81	2⇒5	0.91
		1,6⇒10	0.90	10⇒1	0.84	1⇒10	0.84
		16⇒1	0.78	10⇒16	0.83	14,16⇒5	0.91
		18⇒10	0.85	10⇒5	0.83	8⇒5	0.89
				5,14⇒16	0.91	10,16⇒5	0.89
				12⇒14	0.81	10,14⇒5	0.90
				10⇒14	0.81	8⇒10	0.85
				12⇒1	0.80	12⇒5	0.91
				12,16⇒5	0.90	1,10⇒5	0.87
				8⇒16	0.86		
				8⇒5	0.85		
				1,14⇒16	0.88		
				1,5⇒16	0.87		
				12,14⇒16	0.93		
				8⇒1	0.84		
				10⇒12	0.74		
				12,14⇒5	0.91		
				8⇒14	0.80		
				5,10⇒16	0.86		
				1,14⇒5	0.83		
				10,14⇒16	0.88		
				1,12⇒5	0.89		
				1,12⇒16	0.88		
				10,16⇒1	0.85		
				8⇒10	0.77		
				5,10⇒1	0.84		
				18⇒14	0.80		
				5,12,14⇒16	0.93		
				5,8⇒16	0.90		
				2⇒5	0.83		
				8⇒12	0.76		
				18⇒5	0.79		
				10,14⇒5	0.84		
				10,14⇒1	0.84		
				18⇒16	0.78		
				2⇒14	0.81		

4．比較考察

ここで分析対象としたアンケート調査の詳細な報告書（国立国語研究所，2002）が出版されている．報告書中の質問1－6，16－19に関わる部分を参照しながら，形式概念分析から得られた結果を比較考察しよう．報告書のデータ分析手法は基本的に調査項目ごとの構成比の比較である．質問1－6，16－19に限っていえば，回答の構成比を比較し，男女差，地域差を考慮した考察を展開している．「得られた主な知見」としてまとめられた事項の中から関係部分の一部を引く（国立国語研究所，2002：p.133）．このまとめは本稿で扱わなかった東京高校および大阪高校の分析を含んでおり，中学生と高校生では構成比などで異なるものもあるが，全体的傾向としては変わらないと考えてよい．

「1）ふだん学校で自分自身の言葉遣いが「気になるほうだ」（コード0―引用者，以下同様）と回答した生徒は2～3割であった．言葉遣いをあまり気にせず学校生活を送っている生徒が多い．

2）先生や上級生に対する場面で自分の言葉遣いが「気になるほうだ」（コード2）と回答した生徒は5～6割いる．成人の社会と比べ複雑性の少ない学校社会においても，目上との人間関係の中では，約半数の生徒が言葉遣いを気にしながら学校生活を送っている．

3）先生や上級生に対する場面で自分の言葉遣いが「あまり変わらない」（コード4）と回答した生徒は2～4割にとどまり，6～8割の生徒は何らかの点で言葉遣いが「変わる」（コード5）と回答している．「変わる」の内訳で多くの割合を占めたのは，狭い意味での「敬語」のたぐいである（本稿ではこの分析省略――引用者）．」

上の引用から想像できるように，質問に対する回答をそれぞれ独立に分析し，得られた知見といってよい．これらの知見は何も形式概念分析に拠らずとも，10個の回答項目の構成比があれば展開可能である．もちろん，図4－図7からそれは可能である．試みに，図4山形男子を上記の引用の視点から対応する展開すればつぎのようになろう．

1）ふだん学校で自分自身の言葉遣いが「気になるほうだ」（コード1，「気にならない」が8割以上だから）と回答した生徒は2割以下であった．言葉遣いをあまり気にせず学校生活を送っている生徒が多い．

2）先生や上級生に対する場面で自分の言葉遣いが「気になるほうだ」（コード3，「気にならない」が6割以上だから）と回答した生徒は4割程度いる．成人の社会と比べ複雑性の少ない学校社会においても，目上との人間関係の中では，4割程度の生徒が言葉遣いを気にしながら学校生活を送っている．この数値は他のグループに比し，1～2割少ない．これは地域差と考えられる．

3）先生や上級生に対する場面で自分の言葉遣いが「あまり変わらない」と回答した生徒は5割以上おり，5割弱の生徒は何らかの点で言葉遣いが「変わる」と回答している．

以下同様，報告書にまとめられたような内容を展開することは可能であるが，形式概念分析の示しているものはこれにとどまらない．とくに粗い概念束（図4－図7）からでも回答項目の関連が読み取ることが可能である．ここでは図4－図7の中で最もシンプルな図4を例にその関連を読み取ってみよう．

1）言葉遣いが「気にならない」（コード1）し，敬語はよそよそしいと思う（コード18）生徒が65％近くいる（1かつ18）．

2）しかし，敬語は「欠かせないものだ」（コード16）と思うが，言葉遣いが「気にならな

い」（コード 1）生徒も 50％程度いる（1 かつ 16）．

3）上の 2）は先生や上級生と話すとき言葉遣いが「変わると思う」（コード 5）と答えた生徒が 40～50％いることと符合する．これは図 4 の「変わらない」（コード 4）と回答した生徒が 50～60％であることからわかる．

4）言葉遣いが「気にならない」（コード 1）ことは，具体的には先生や上級生と話すとき自分の言葉遣いが「気になら」（コード 3）ず，言葉遣いに困った「経験はない」（コード 6）生徒や，クラス討論などで言葉遣いに困った「経験はない」（コード 10）生徒が 50～60％程度いることに反映している．

5）クラス討論等で「改まった，きちんとした言葉遣いがよい」（コード 12）が 6 割程度いる．

他方，表 4 の連関規則はつぎのような情報を提供している．

6）質問 1－6（日常の敬語使用）と質問 16－19（敬語の規範意識）関係が連関規則として表現されるので，それを敬語意識とその言語行動の観点から分析すれば新たな展開が可能である．

7）具体的に挙げるなら，連関規則に見られるつぎの傾向はそのひとつである．東京男女とも「敬語は欠かせないものだ」という意識が日常の言語行動に現れる傾向がある．これに対し，山形では具体的な場面における敬語使用の必要性が「敬語は欠かせないものだ」という結果を導いているように思われる．

以上のように粗い概念束と連関規則を用いることによって報告書で示されている知見はもとより，さらに，3.3.2 で述べたように表 3 に示した含意規則を利用すれば，適用範囲が狭いけれども厳密な論理的含意関係から対象の一層明確な特徴を把握することが可能である．

5．結　論

形式概念分析は属性で記述された対象に内在するデータの特徴を抽出する方法である．すなわち，対象の有する属性間の関係だけから対象全体の特徴を記述する．この特徴がさまざまな分野，例えばデータ解析，情報検索，知識発見，知識表現，概念的クラスタリング，クラスの階層デザインと管理など広く応用される理由である．本稿では，実用規模のアンケート調査のデータ解析へ適用するために必要な手法を提示するとともに，その有効性を実証した．その方法は以下の特徴を有する．

1）形式概念分析は，数理的方法であって元のデータの情報を全く失わない解析方法である．

2）概念束を図に描く（Hasse 図）ことによって，対象の全体構造とその特徴を捉えることができる．しかしデータ数が多い場合，図を描くことが事実上不可能になるので，粗い概念束を描くことによって全体構造を把握し，必要に応じて細かく分解してゆくのが有効である．

3）属性間の含意規則あるいは連関規則から属性間の依存関係を把握することができる．

とくに，含意規則（連関規則）は社会情報解析にとって有効である．なぜなら，含意規則と論理操作を結合することによって対象部分集合間の相互連関，すなわち，対象集合に内在する，質的に異なるあるいは部分的に対立する対象を取り出すことを可能にするからである．これは論理と複数の価値システムからなる過程の分析を目指す社会情報解析へのひとつの足がかりを与えると思われる．

本稿のデータ解析に用いたツールは自前の小さな Java プログラムである．概念束の計算には Ganter の Next-Closure アルゴリズム（Ganter, 1984; Wille, 1999）を用い，含意規則として Duquenne-Guigues Basis を計算している．図を描く機能は現在開発中である．

形式概念分析が社会科学・人文科学のデータ解析方法として広く普及されるには，つぎの点に留意したツールの開発が必須である．

1）実用規模のデータに対するデータ形式を標準化する．

2）実用規模のデータの場合，概念数，含意規則数が膨大になるので，適切かつ標準的なデータ結果の表示法，とくに可視可法を確立する．

3）データ数が多い場合，膨大なコンピュータパワーを要するので，概念束，含意規則を計算する高速アルゴリズムを研究する（例えば，Stumme, 2002）．

注

1）O を集合 M における関係とする．O がつぎの性質をもつとき，O を順序関係という．

⑴ 反射律 xOx

⑵ 推移律 xOy，yOz ⇒ xOz

⑶ 反対称律 xOy，yOx ⇒ x＝y

ある集合 M において一つの順序関係 O が与えられたとき，（M，O）を順序集合という．たとえば，自然数の間の通常の大小関係≦は，自然数の集合 N における一つの順序関係であり，（N, ≦）と記す．

2）順序集合 L において，任意の二元 a, b に対して｛a，b｝の上限および下限がいつも L の中に存在するとき，L を束（Lattice）という．順序集合 L の空でない任意の部分集合 A に対してその上限 supA および下限 infA が存在するとき，L を完備束という．ここで，A（⊆L）の上限（最小上界）supA とはつぎの条件をみたす要素である：

⑴ a∈A ⇒ a≦supA

⑵ （a≦A ⇒ a≦x）⇒ supA≦x

supA と双対的に，すなわち上の二つの条件の≦を≧に置き換えて，A の下限（最大下界）infA が定義される．

謝辞

本稿のデータ解析に利用させていただいた「学校の中の敬語調査」を実施し，そのデータを公開し利用に供された国立国語研究所，とくに言語行動研究部第一研究室，杉戸清樹，尾崎喜光および塚田実知代の三氏に感謝の意を表します．方法を考察研究しようとするものにとって公開データは何ものにもかえがたい貴重なものでした．日頃議論し有益なコメントをくださいました大國充彦，高橋徹，田中一各先生に記して謝意を表します．また，貴重なコメントをしてくださいました査読者にお礼申し上げます．

参考文献

Davey & Priestley (2002): Introduction to Lattices and Order (2ndED.), Cambridge University Press

Ganter, B. (1984): Two Basic Algorithms in Concept Analysis, FB4-Preprint No.831, TH Darmstadt

Ganter, B., Stumme, Wille (eds) (2005): Formal Concept Analysis, *Lecture Notes in Computer Science* 3626, Springer

Ganter, B and Wille, R. (1999): Formal Concept Analysis, Springer

Kantardzic, M. (2003): Data Mining: Concepts, Models, Methods, and Algorithms, IEEE Press

国立国語研究所（2002）：『学校の中の敬語 1 ── アンケート調査編 ──』，三省堂

長田博泰（2004）：「形式概念にもとづく質的分析」，『社会情報』（札幌学院大学社会情報学部紀要），Vol.4 No.1，pp.19-37

大國充彦，鳥居喜代和，長田博泰，田中一（1999）：「社会情報解析 ── 判決文における論理情報過程と価値情報過程との相互連関について」，『社会情報学研究』，No.3, pp. 63-76，日本社会情報学会

Stumme, G. et al. (2001): Conceptual Clus-

tering with Iceberg Concept Lattices, *Proc. GI-Fachgruppentreffen Maschinelles Lernen '01*. October 2001 (2002): Computing Iceberg Concept Lattices with TITANIC, *Journal on Data and Knowledge Engineering*, Vol.42, issue 2, pp.189-222

田中一（1999）：「情報と情報過程の総合的考察」，『社会情報学研究』，No.3，pp.77-90，日本社会情報学会

Wille, R. (1982): Restructuring lattice theory: an approach based on hierarchies of concepts, *Ordered sets* (ed. I. Rival) pp. 445-470, Reidel, Dordrecht-Boston

Wolff, K. E. (1996): Comparison of graphical data analysis methods, *SoftStat '95 advances in statistical software 5* (ed. Faulbaum, F. & Bandilla, W.) pp.139-151, Lucius & Licius, Stuttgart

付録「ことばのアンケート」（関係部分）

［　］内の番号は図表中で引用するコードを表す―引用者．

Ⅰ．まずはじめに，毎日の学校でのことばづかいで感じていることを答えてください．「どちらかといえば」という程度でもかまいません．それぞれ選んでください．

1．ふだん，学校で，自分自身のことばづかいが気になるほうですか？
　1．気になるほうだ．［0］
　2．あまり気にならないほうだ．［1］

2．学校生活のなかでも，とくに先生や上級生と話すとき，自分自身のことばづかいが気になるほうですか？
　1．気になるほうだ．［2］
　2．あまり気にならないほうだ．［3］

3．先生や上級生と話すときと，親しい同級生と話すときとで，自分自身のことばづかいで変わるところがあると思いますか？
　1．あまり変わらない．［4］
　2．変わると思う．［5］
　→ことばづかいのどんなところですか？具体的に書いて下さい．

4．これまでに，先生や上級生へのことばづかいのことで困った経験はありますか？たとえば，どういうことばづかいをしたらよいかわからなかった，もっと別の言い方をしなければならないのにまちがえた，など．
　1．そういう経験はない．［6］
　2．そういう経験がある．［7］

5．先生や上級生・先輩（せんぱい）から，ことばづかいのことで注意されたり，教えられたりしたことがありますか？
　1．そういう経験はない．［8］
　2．そういう経験がある．［9］

6．クラス討論，生徒会活動，部（クラブ）活動などで，司会や議長をする場合や意見をみんなの前で発表する場合などに，ことばづかいのことで困った経験はありますか？
　1．そういう経験はない．［10］
　2．そういう経験がある．［11］

7－8．省略

Ⅱ．省略

Ⅲ．つぎに，ことばづかいや敬語についてのあなたの意見を聞きます．正しいとか，そうあるべきだとかいうのではなく，自分の考えを答えてください．

16．いまのあなたのクラスを考えて，クラス討論や授業での発言のときは，ふだんのことばづかいとは少しちがった，あらたまったことばを使うのがよいと思いますか，ふだんどおりのことばづかいでよいと思いますか？　あなたの意見に近いほうを選んで○をつけてください．
　1．あらたまった，きちんとしたことばづかいがよい．［12］
　2．ふだんどおりの，ふつうのことばづかいでよい．［13］

17. 学校のなかでは生徒同士であっても，上級生や部（クラブ）活動の先輩などには敬語(ていねいで，相手をうやまったことば)を使うほうがよいでしょうか，使わなくてもよいでしょうか？
 1．使うほうがよい． [14]
 2．使わなくてもよい． [15]
18. 現在の学校で使われている敬語について，次の二つの意見があります．あなたの意見に近いほうに○をつけてください．
 1．敬語は上下の規律（きりつ）が守れ，授業や部（クラブ）活動などの学校生活をするうえで欠かせないものだ． [16]
 2．敬語はかたくるしく面倒（めんどう）だから，学生生活のためにはかえって邪魔（じゃま）になる． [17]
19. 先生や上級生に対してていねいな敬語を使うと，どうしてもよそよそしくなって，親しい心の交流やざっくばらんな(気楽な)つきあいがしにくくなる，という意見があります．あなたは，この意見を……？
 1．そう思う． [18]
 2．そうは思わない． [19]