The University of Osaka
Institutional Knowledge Archive

Title	インターネットに対する人々の意識 ： 自由回答の分析から
Author(s)	川端, 亮; 樋口, 耕一
Citation	大阪大学大学院人間科学研究科紀要. 2003, 29, p. 162-181
Version Type	VoR
URL	https://doi.org/10.18910/8256
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

「大阪大学大学院人間科学研究科紀要」第29巻, 2003年 3 月所収

インターネットに対する人々の意識
—— 自由回答の分析から ——

川　端　　亮
樋　口　耕　一

インターネットに対する人々の意識
―― 自由回答の分析から ――

川端　　亮
樋口　耕一

1．目　　的

1.1　インターネットの普及

1990年代の後半に入って、パソコンやインターネットは急速に普及してきた。パソコンやインターネットの普及率を正確に捉えることは難しいが、内閣府経済社会総合研究所が発表する世帯単位でのパソコン普及率では、1998年25.2%、2000年38.6%、2001年50.1%、2002年３月57.2%となっている。また、総務省による通信利用動向調査（2001年12月）では、インターネットの世帯普及率は、携帯電話による接続も含めて、60.5%になったと報告されている。これらの数値は、他の調査と比べると高い割合を示しているが、低い方の数値をあげてみても、インターネットの世帯普及率は、2002年３月で49.5%（情報通信総合研究所の「インターネットの普及予測等の調査」）、インターネットの人口普及率では、2001年12月で44.0%である（総務省［11］）。

このようなインターネットの普及にともなって、私たちの身の回りでもインターネットについて書かれたものを目にしたり、インターネットのことを話題にする機会が格段に増えた。表１は、最近の朝日新聞と毎日新聞で、「インターネット」という言葉を含む記事数を表にしたものである。93年までは、「インターネット」という言葉はほとんど新聞記事の中ではみられないが、95、96年に急増し、2000年の朝日新聞では、5,000件近くにまで達している。そして、記事数の急増にしたがって、おそらく記事内容も多様になっ

表１　「インターネット」を含む新聞記事の数

	朝日新聞[a]	毎日新聞[b]
1992年	0	2
1993年	1	2
1994年	84	37
1995年	467	283
1996年	1,411	1,080
1997年	1,621	1,057
1998年	1,700	1,098
1999年	3,036	1,666
2000年	4,920	2,807
2001年	3,509	1,864

a）朝日新聞については『CD HIASK』による検索結果で、東京、大阪、西部、名古屋の各本社版記事が含まれる。
b）毎日新聞については、東京版のみの記事数である。

てきていると推測される。同じような記事が繰り返される回数には限度があり、同じことの繰り返しだけでは、ニュースにならないからである。このような増加、多様化は、新聞記事だけでなく、雑誌や出版物においてもみられることであろうし、TV や映画などでもみられることであろう。さまざまなメディアに取り上げられることによって、「インターネット」という言葉は多くの人に知れ渡り、人々の日常会話においても「インターネット」という言葉が使われているだろう。

人々がどのように「インターネット」を意識するようになってきたかという変化も興味深いが、本稿においては、2001年の質問紙調査の自由回答項目の分析結果にもとづき、現在の人々の考え方を探る。なお、インターネットは単体で利用し得るものでなく、その利用にはパーソナル・コンピュータや携帯電話などの情報機器が必要となる。よって正確には、インターネットを中心としつつ、それらの機器を含む情報技術（IT）に対する意識が、本稿における分析の対象となる。

2．データと分析方法

2.1　データについて

インターネットなどの情報技術に対する意識を探るために用いるのは、「情報と生活に関する調査」[1)]留め置き票のＱ２である。この質問のワーディングを以下に示す。

Ｑ２．最近、「インターネット」や「情報技術（IT）」という言葉を耳にすることが増えたと思います。情報技術（IT）やインターネットについてあなたが考えるとき、どんなことが思い浮かびますか。思いつくものから順に３つまで、何でもご自由にお答え下さい。

回答欄については、約15.5㎝×2.5㎝の枠が３つ用意された。なお、記述内容が他の質問項目に影響されることを極力避けるために、この項目は質問紙の先頭近くに配置された。この質問より前にあるのは「Ｑ１.あなたは、インターネットやパソコン通信をご存じですか」という問いだけである。

2.2　分析方法

2.2.1　コーディングにおけるコンピュータの利用

本稿では、自由回答を計量的に分析する。自由回答のような質的データを計量的に分析するためには、データを数量化するためのコーディング作業が不可欠である。コーディングによっていったんデータを数量化することで、例えば、特定の事柄に言及した回答がいくつあったのか、またどういった人々が特定の事柄に言及したのかといった、計量的な分析が可能になる。

従来の多くの自由回答の分析では、このコーディング作業を人手で行っていた。つまり、複数のコーダーが決められたコーディング基準に従って、回答にコードを与えてい

くのである。しかしあらかじめ決めておくコーディング基準を完全なものにしておくことは、ほとんど不可能である。実際はコーディング作業を進める中で、基準にあわない回答が多数現れ、それらをコーディングできる基準を改めて作り直すという作業が不可欠である。この作業によって、コーディング基準はどんどん複雑になっていく。その結果、人間の頭では覚えきれないほどのコーディング基準ができあがり、誰がコーディングしても同じコーディング結果になることは、難しくなるのである。

そこでこの欠点を避けるために、本稿ではコーディング作業にコンピュータを用いる（川端［8］）。

コンピュータ・コーディングを行うメリットは、いくつかある。例えば、Seale［6］は、大量のデータを扱えること、信頼性があること、共同研究が可能となること、サンプルの選択に役立つことの４つをあげている。つまり、コーディングをコンピュータがするため、人間がコーディングしていたら何年もかかるような大量のデータであっても、わずかな時間でコーディング可能であり、またデータ量の多寡にかかわらず、コーディングの基準が揺らがない。このことは、データ量が大きくなればなるほど、メリットとなる。そして、その揺らがない基準を第三者に明示できるため、検証可能であるとともに、複数の研究者で共同研究することも可能となる。また、大量のデータの中から、ある条件を満たすデータを発見することも容易で、これによって、データの典型例や特異な例を抽出し、自分の行った分析の客観性を確認したり、仮説の変更や新たな仮説の発見に役立つ。

これら以外にも、コーディングによって、質的データを数値化することで、多変量解析の手法を応用することが可能となり、それによって、分析者の理論仮説や問題意識といったものに影響を受けない形でテキストデータの概要を把握、提示できること、各種の数値指標によって、特徴的な回答を発見できることなどがあげられる。こういったコンピュータ・コーディングの利点を活かすことで、単に分析の信頼性、客観性を向上し得るのみならず、テキストデータの探索的分析がより容易となり、仮説構成に大いに役立つのである。

本稿のコーディングは、筆者らの自作ソフトウェアである「KH Coder Ver. 1.20」を用いて行った[2)]。KH Coder はコーディングに必要な機能の大半を備えているが、多変量解析や検定などの統計計算機能を備えていない。よって、そういった処理を行う際には、統計ソフトウェアを利用した[3)]。

３．自由回答項目の分析

3.1　回答率と回答者の持つ傾向

記述内容の分析に入る前に、得られたデータの特徴を確認しておきたい。まず、３つの回答欄それぞれへの回答数、回答率などを表２に示す。表２から、１つ目の回答欄が

もっとも回答率が高く、２つ目、３つ目と順に回答率が減少していることがわかる。ただし、１つ目の回答欄にしてもその回答率は約68％と、必ずしも高い値ではない。

ここで2.1節でも触れたＱ１を用い、インターネットやパソコン通信に関する知識の有無と、自由回答への記入有無との関連を表３に示す。なお表３では、３つの回答欄のうち１つにでも記入されていれば「記入あり」として扱った。

表３から、インターネットやパソコン通信を知っている人ほど自由回答項目に記入する確率が高い。そして、Ｑ１で「知っている」と回答した人だけでなく、「言葉は知っている」とした人を併せた900人においても、自由回答記入者は651人で、回答率は約72.3％となる。これは、留置法で行った調査における自由回答項目としては、標準的な回答率といえるであろう（太郎丸［12］、佐藤［10］）。

すなわち、表２に示した全体での回答率が低く見えたのは、インターネットやパソコン通信を知らないという人々の回答率が著しく低かったためである。なお、インターネットやパソコンを通信を言葉だけでも知っているグループ、すなわち自由回答に記入が多かったグループは、50歳未満（98.5％）、高卒以上の学歴（96.9％）、ホワイトカラー（98.1％）、世帯年収が600万円以上（96.4％）、関東／近畿／東海に居住（94.3％）、人口10万を超える都市に居住（93.5％）している人々であった[4)]。カッコ内に示したのは、これらの属性を持つ人々のうち、インターネットやパソコン通信を言葉だけでも知っている人が占める割合である。

表２　得られた回答のあらまし

記述統計	回答欄		
	１つ目	２つ目	３つ目
回答数（回答率）	682（67.5％）	516（51.0％）	394（39.0％）
平均回答長（字）	12.75	12.72	15.25
回答長の標準偏差	13.99	13.06	21.59

カッコ内の数値は回答率。平均と標準偏差は回答のあったサンプルのみを使用して算出。

表３　知っているか否かと自由回答への記入

インターネット パソコン通信 を知っているか	自由回答（Ｑ２）に記入		合計
	なし	あり	
知っている	113（18.3％）	505（81.7％）	618
言葉は知っている	136（48.2％）	146（51.8％）	282
まったく知らない	69（77.5％）	20（22.5％）	89
合計	318（32.2％）	671（67.8％）	989

カッコ内は行のパーセント。（$\chi^2=171.89$　$df=2$　$p<.01$　*Cramer's V*＝.417）

3.2　回答の多い言葉

まず、得られた自由回答の例を挙げる。

例１：自分が知りたい情報が早く分かる

このような自由回答が、本調査では、1,592得られた。これは、異なる回答欄に記述されたものは、同じ回答者の回答であっても、別の回答として扱ったからである。この自由回答は、すべてテキスト形式で入力した。このデータをコーディング・プログラム KH Coder に投入すると、自動的に回答文を単語に区切り、どの回答者がどの単語を使用して回答しているかを示すデータを出力する。ただし、KH Coder は、例えば、「楽しかった」、「楽しく」、「楽しければ」などはすべて「楽しい」という１種類の語として抽出する。つまり、活用を持つ語は動詞、形容詞、形容動詞にかかわらず、すべて基本形として抽出している。例えば、例１の回答者は、「自分」、「知る」、「情報」、「早い」、「分かる」という単語を使用しているという結果が、出力されるのである。この結果を各単語ごとに集計した。データ中には1,576種類の語が含まれており、総単語数は11,298、異なり単語率は13.9%であった[5)]。

そして、どの単語を何人の人が使っているかを集計し、出現数が多かった140語を示した結果が、表４である。もっとも出現数が多かったのは、「情報」の295回であり、「情報収集」、「個人情報」、「情報交換」、「情報技術」などもあわせれば、情報に関わる言葉が非常によく使われている。また、Q２の質問文の中に「インターネット」という言葉を用いているので、「インターネット」という言葉が82回とかなり多く出現するのは当然であるが、「メール」と「ホームページ」という言葉も88回、71回と「インターネット」と同じくらい出現している。これら３つの言葉は、かなり広く浸透していることがうかがえる。それに対して「IT」という言葉は、Q２の質問文中に使われているが、その出現数は46回で、「インターネット」に比べるとおよそ半分ぐらいしか使われていない。「情報技術」も Q２の質問文中にあるが、出現数は11回で、情報技術（IT）という言葉が、インターネットやホームページほど多くの人には使いこなせない言葉であることがわかる。そのほかの出現数の多い語を見てみると、「出来る」が156回、「便利」が141回とインターネットを肯定的に評価する表現がみられる一方で、「犯罪」が47回とその悪い面に触れた表現もかなり多いことがわかる。

3.3　多変量解析によるデータ概要の把握

3.3.1　自己組織化マップの作成

次にコンピュータ・コーディングの第一段階として、表４に示した140の頻出語を用いて、出現パターンの似通った語ほど近くに布置されるような、単語の散布図を作成する。

コンピュータ・コーディングの最終的な目的は、得られたデータの一部である単語あるいは文字列にコードを与え、計量的に分析することにある。しかし、表４の集計結果

表４　回答中に頻出した140語

抽出語	出現数	抽出語	出現数	抽出語	出現数
情報	295	見る	16	問題	8
出来る	156	入る	15	氾濫	8
便利	141	増加	15	新聞	8
思う	127	時代	15	新しい	8
メール	88	検索	15	収集	8
インターネット	82	技術	15	見える	8
人	78	関係	15	興味	8
ホームページ	71	電話	14	感じる	8
知る	68	テレビ	14	楽しい	8
パソコン	60	コミュニケーション	14	ネットワーク	8
犯罪	47	オークション	14	料金	7
ＩＴ	46	調べる	13	友達	7
買物	43	人間	13	変化	7
早い	39	情報交換	13	世界中	7
自分	38	手段	13	進む	7
通信	37	売買	12	心	7
多い	36	大変	12	情報量	7
生活	34	経済	12	使える	7
社会	33	会社	12	産業	7
難しい	32	プライバシー	12	最新	7
世界	32	ウィルス	12	高い	7
得る	31	予約	11	購入	7
良い	30	情報技術	11	向上	7
悪用	30	広がる	11	見れる	7
利用	27	機械	11	可能性	7
情報収集	27	面	10	いろいろ	7
増える	26	販売	10	方法	6
必要	23	多く	10	変わる	6
使う	23	操作	10	幅広い	6
個人	23	進歩	10	発展	6
入手	22	出る	10	買える	6
携帯電話	21	使用	10	伝達	6
分かる	20	考える	10	他人	6
色々	20	言葉	10	相手	6
簡単	20	管理	10	少ない	6
ショッピング	20	悪い	10	商品	6
仕事	19	トラブル	10	作成	6
個人情報	19	普及	9	広い	6
ネット	19	若い	9	効率	6
家	18	持つ	9	恐れ	6
不安	17	事件	9	起きる	6
手	17	子供	9	機器	6
時間	17	今	9	易い	6
コンピューター	17	交流	9	やり取り	6
欲しい	16	居る	9	サービス	6
世の中	16	危険	9	お金	6
心配	16	旅行	8		

を見れば明らかなように、各語の出現度数はさほど大きくない。100以上の語が４語、50以上の語が10語、20以上の語が36語、10以上の語で84語である。したがって、このままで計量的な分析を行うことはかなり難しい。そこで各語を統合し、統合された語群にコードを与えていくことが必要となる。例えば「仕事」という言葉と「会社」という言葉を同義語として扱い、どちらにも「仕事」というコードを与えるというような処理が必要となるのである。すなわち、同義語を規定するある種の辞書を作成することで、きわめて多様に分布するデータを計量的に分析できる形に整理していく。こういった処理は、テキストデータの分析において特殊なことではない。本稿のもとづくデータは、異なり単語率が13.9%とどちらかといえば小さいが、それでもこの処理は必要であり、単語の統合は、コンピュータ・コーディングには、ほぼ必須のものといえるだろう。

この処理においては、例えば、「犯罪」と「プライバシー侵害」を同じ意味、もしくは同じカテゴリーに含まれる語として扱うのかどうか、さらには、「トラブル」も同じカテゴリーとして扱い得るのか、といった判断を分析者が数多く行わなければならない。この一連の判断の基準が何もなければ、分析者の持つ予断や暫定的な仮説や思いつきなどによって、恣意的に語の統合が行われる。もちろんその中には、正しい統合も数多くあるであろうが、誤った統合を行ってしまう可能性もある。そして、そのような恣意的な統合によって、集計結果に分析者の仮説が必然的に反映され、誤った仮説が成り立つという危険性がある。

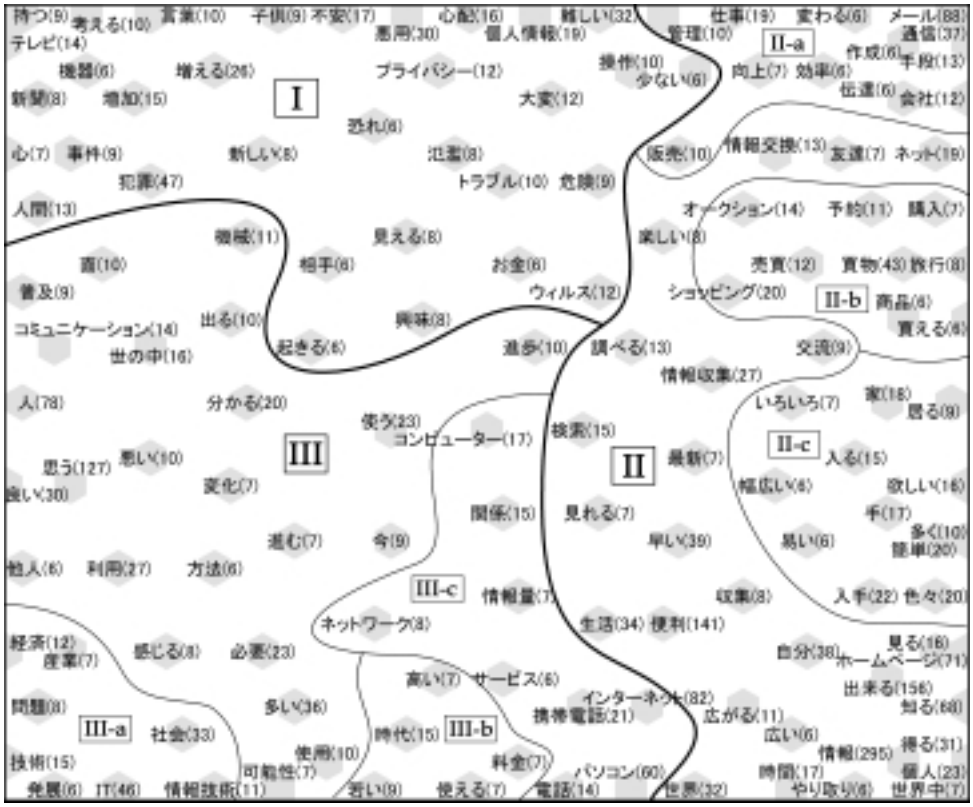
そこで、本稿ではそういった危険性を避けるために、語の統合に先だって、自己組織化マップという多変量解析の一手法を用い、統合以前の、表４に示した140語の関連を把握する。この手順を踏むことで、分析者にとっては関連ある語の把握が容易になり、3.4節で行う語の統合規則を記述する「辞書」の作成がよりスムーズに行える。また、分析結果を読む側にとっても、分析者がデータ全体の中からどの部分に焦点をあてて「辞書」を作成したのかがより明確になり、分析の客観性が向上する。

単語の散布図を描くための統計手法は多数あるが、今回は自己組織化マップ（Self-Organized Map）を用いることにした。自己組織化マップとはニューラル・ネットワークの一種であり、Kohonen［3］,［4］によって提案された、中間層を持たない２層型の教師無し競合学習モデルである。主成分分析、対応分析、多次元尺度法などの手法と異なる性質として、自己組織化マップによる散布図は多次元空間の線形写像によって得られるものではなく、またユークリッド距離空間でもない。このため、作成される図においては方向という概念が存在しないので軸の解釈はできないし[6]、５つ離れた語と８つ離れた語が同程度に異なるということが起こり得る。この性質がデメリットとなる場合もあるだろうが、むしろ、高次元空間の複雑で階層的な関係を２次元平面上に表現可能である（Kohonen［3］）という点を重視して、自己組織化マップを採用した。文書空間は明らかにそのような高次元空間であり（Doszkocs *et al*［2］）、自由回答データもそれに準じるものであろう。また、初動探査という意味合いが濃いこともあり、多数の次元ない

し成分を抽出しての詳細な分析が目的ではないということも、自己組織化マップを用いた理由の１つである。

頻出140語を用いて自己組織化マップを作成し、このマップにそれぞれの語とその出現数をラベルとして添付した[7)8)]。さらに、似通った語が集まったと思われる部分を線で区切ったものが図１である[9)]。また、線で区切ったそれぞれの部分にはローマ数字で番号を与えた。

以上のように、図１における単語の布置そのものは、異なるデータに対してもそのまま適用できる機械的な手順にのっとって行われている。その一方で、図１における線引きに関しては分析者による主観的な判断を含む作業であり、この点に注意が必要である。



884nodes, 31(columns) × 29(rows). Quantization error = .01543
（カッコ内は語の出現数。線引きは筆者による。）

図１　頻出140語のマッピング（Self-Organaizing Map）

3.3.2　マップの解釈

出現パターンの似通った語のグループとして、大きく分けるとＩからIIIまでの３グループが図１から見いだされた。以下、この３つのグループの詳細を順に見ていきたい。

Ⅰ　まずⅠの部分には、出現数の多い語として「犯罪」(出現数40)、「個人情報」(同19)などがあり、主として情報技術の悪用に関する語が集まっている。上述の「犯罪」と「個人情報」以外にも、情報技術がもたらし得る好ましくない事柄を表す「事件」、「トラブル」、「氾濫」、「ウィルス」といった語がこの部分に集まっている。また「不安」、「心配」、「恐れ」など、そういった好ましくない事柄に対する不安感を表す語もこのグループに含まれている。

その他にも、情報技術に対する否定的な印象を表す語が数多くこの部分に布置されている。実際の回答を挙げながら、見ていきたい。まず左端の下部には、「心」と「人間」の２語が近くに布置されており、実際の回答では「何か大事なもの（**人間**の**心**）が忘れられそうに思う（64歳　女性）」といった形で用いられていた。また、この部分の中央下部には「相手」と「見える」の２語が近くに布置されている。これらの語は「**見え**ない**相手**に対して不安に思う（68歳　女性）」といった用いられ方をしていた。少し下に目を移すと「興味」という語があり、「インターネット関連にはあまり**興味**なくしたがって全く接したことはありません（72歳　男性）」など、先ほどの「見える」と同様、実際の回答の中では否定された形で用いられていた。ここでは、KH Coder（茶筌）がすべての語を基本形になおして抽出している点に注意が必要である。最後に、この部分の右上を見ると「操作」と「難しい」の２語が近くに布置されており、「**操作**が**難し**そうで怖い（43歳　女性）」といった不安を示す回答も見受けられた。

Ⅱ　次に、Ⅱの部分を見ていこう。この部分はa、b、c、アルファベット無し、という４グループにさらに細かく分割してある。

Ⅱ−aの部分には「仕事」、「効率」、「会社」などの語が布置されており、職場における情報技術の利用を表す語が集まっている。実際の回答例としては「**仕事**の**効率**も上がる（57歳　男性）」などがあった。

少し下、Ⅱ−bの部分には「買物」、「オークション」、「予約」などの語が布置されており、主として「オンラインでの**買物**（20歳　男性）」に関わる語が集まっていることがわかるだろう。

さらに下に目を移してⅡ−cの部分を見ると「家」、「居る」、「入手」などの語があり、実際の回答としては「**欲しい**物が**家**にいて**手に入れる**事が出来る（43歳　男性）」といったものがあった。情報技術の利便性、なかでも、居ながらにしてさまざまな物を入手できるということを表す語が、この部分には集まったと考えられる。

アルファベット無しの部分は、上の３つと比べると、いくぶん雑多な語が布置されているという印象があるかもしれない。だが、以下のように確認していけば、この部分の解釈は難しくない。サ行変格活用の語を含む動詞に注目すると、出現数の多いものから「知る」、「得る」、「見る」、「情報収集」、「検索」などがある。次に形容詞、形容動詞に注目すると、出現数の多いものから「便利」、「早い」、「楽しい」などがある。

これらの語から、情報収集をはじめとする情報技術の有効利用、および、肯定的な評価を示す語がこの部分に集まっていることがわかるだろう。

以上より、職場における利用（Ⅱ−a）、買い物や予約（Ⅱ−b）、居ながらにして色々なものを入手できる利便性（Ⅱ−c）、そして情報収集ほか（アルファベット無し）といったように、総じて情報技術の有効利用を表す語がⅡの部分には集まっていたと解釈できる。

Ⅲ 最後のⅢの部分についても a、b、c、アルファベット無しという４グループにさらに細かく分割を行っている。

まずⅢ−a を見ると、「社会」、「情報技術」、「経済」などの語が布置されており、実際の回答例としては「**経済**発展が進む（76歳 男性）」といったものがあった。情報技術による社会経済や産業の発展を表す語が、ここに集まっている。

次にⅢ−b では「高い」、「料金」、「時代」、「若い」などの語が目に付く。これらの語は回答の中で「便利だが**料金**が**高い**（47歳 女性）」、「**時代**の流れ（65歳 男性）」、「（前略）**若い者**には必要。年寄りにはなじめない。（72歳 女性）」などの形で用いられていた。不安や恐れというほど明らかな否定的感情は見いだせないものの、情報技術からいくぶん距離を置いた回答といえよう。この部分には、情報技術を敬遠、傍観する表現に用いられるような語が集まったと思われる。

右に目を移すと、Ⅲ−c の部分には、「インターネット」、「パソコン」、「電話」、「コンピューター」などの語が布置されている。利用法や評価というよりも、情報技術そのものを表す一般的な語がこの部分には集まっている。

アルファベット無しの部分に関しては、いくぶん解釈が難しい。というのは、どんな回答の中にでも出現しているような、一般的な語が比較的多いからである。動詞に注目してみると、まず「思う」、「感じる」、「分かる」といった、単に考えを表現するためだけの語がある。さらに「利用」、「使用」、「使う」など、情報技術の利用を表すためだけの動詞もある。以上の語と比較して、より特徴的であると思われる動詞としては「変化」、「進歩」、「進む」、「普及」などがこの部分に布置されている。「長生きすればいろんな技術の**進歩**に驚く（79歳 男性）」、「人類生活の大きな**変化**（42歳 男性）」などの形で、これらの動詞は用いられていた。次に、形容詞を探してみると、左端付近に「良い」と「悪い」の２語が近くに布置されていることに気づく。これらの形容詞は「**良く**も**悪く**も世の中かわる（75歳 男性）」といった形で用いられることが多かった。なお、この回答の中に出現した「世の中」についても、このアルファベット無しの部分に布置されている語である。こうしてみると、情報技術によって引き起こされる、漠然とした世の中の変化や進歩を表現するための語が、アルファベット無しの部分には数多く集まったものと思われる。

以上より、情報技術による経済や産業の発展（Ⅲ−a）、情報技術の敬遠（Ⅲ−b）、

情報技術一般（Ⅲ–c）、さらには漠然とした世の中の変化（アルファベット無し）といったように、社会一般の変化に関するあまり具体的でない考えや印象を表すような語、および非常に一般的な語がⅢの部分には集まったといえよう。

得られた回答の概要　以上の自己組織化マップの解釈より、得られた自由回答データは主として以下に挙げる３つの主題から成っていたことがわかる。

まず１つ目は、図１上のⅠの部分から読みとれる、インターネットや情報技術を悪用しての犯罪やプライバシー侵害、さらには自ら利用する場合についても操作が難しい、などの点に対する不安や恐れである。

次に２つ目は、伝達や情報収集を行い仕事に役立てられることや、買い物や予約を行い日常生活にも役立てられることといった、インターネットや情報技術の便利さへの評価、ないしは便利さへの期待である。これは図１上のⅡの部分から読みとれる。

最後に３つ目は、図１上のⅢの部分から読みとれる、インターネットや情報技術によって社会一般が変化しつつあるという印象、ないしは変化への予感である。なお、ここで言う社会一般の変化には、技術・経済・産業などの進歩や発展といった比較的具体的なものから、単に進歩するとか変化するとかいった漠然としたものまで含まれる。

3.4　コーディング・ルールによる分析

3.4.1　コーディング・ルール作成

各単語の関連を自己組織化マップで把握した上で、コンピュータ・コーディングの次の段階に進む。ここでは統合する同義語を定義した「辞書」を作成し、それぞれの語群に KH Coder を用いてコードを与えた。具体的には、語の統合のルールである「辞書」をもとにして各回答が特定の事柄に言及しているかどうかを１／０の２値変数で表す。すなわち、言及していれば「1」、言及していなければ「0」という値を与える。本稿では、こうして割り当てられた値をコードと呼び、コーディングのための基準となる統合ルールである「辞書」のことをコーディング・ルールと呼ぶ。なお、分析者が行うのはコーディング・ルールの作成までであり、実際のコーディング作業については、与えられたコーディング・ルールにもとづいてパソコン（KH Coder）が自動的に行うものである。

こうした仕組みによって、コーディング・ルールという第三者に明示可能[10]な基準にもとづくコーディングが可能になった。また、いったん作成したコーディング・ルールは他のデータに対してそのまま用いることもできる。そして、結果として得られるコードを利用することで、計量的な分析を進めることができるのである。

KH Coder 用のコーディング・ルールとは、例えば、次のようなものである。

＊仕事
　仕事 *or* オフィス *or* 勤務 *or* 会社 *or* 求人 *or* 就職 *or* リストラ *or* リストラク

チュアリング *or*（会議 *and not* 井戸端）

一行目の「＊仕事」という部分は、このルールが「仕事」というコードを与えるためのものであることを示す。その際の条件が２行目以降に記述されており、このルールによって「仕事」、「オフィス」、「勤務」などの *or* で区切られた語を含む回答に「仕事」というコードが割り当てられることになる。最後の括弧の部分については、単に「会議」という語を含むだけでなく、同じ回答の中に「井戸端」という語が含まれていないこと、という条件を表している。これによって、多くの「会議」という言葉は「仕事」というコードに含まれるが、「井戸端会議」は除かれるのである。なお、KH Coder によるコーディングは回答を分類するのではなく、回答の中から要素を抽出するという考え方なので、１つの回答が複数のルールに合致すれば、当然、複数のコードが与えられることになる。

コーディング・ルール作成は3.3節で得られた自己組織化マップの解釈をもとにして行った。表５に作成した21のコードの一覧と、回答に各コードが与えられた回答者の数、およびその割合[11]を示す。表５においては、４つのグループにコードが分類されており、

表５　コーディング結果

コード名	コードが与えられた回答の例	回答者数	割合
メール	「メール」	114	16.7%
ホームページ	「ホームページ」	75	11.0%
パソコン	「パソコン」	74	10.8%
便利さ	「便利」	154	22.5%
日常生活－全般	「生活向上」	35	5.1%
日常生活－ショッピング	「買物」	121	17.7%
日常生活－その他	「旅行案内」、「遠隔医療」	39	5.7%
仕事	「仕事」	45	6.6%
情報収集	「情報収集」、「分からない事を調べる」	195	28.5%
スピード	「速い」、「早い」	82	12.0%
人間関係－コミュニケーション(＋)	「新しい人との出会い」	39	5.7%
人間関係－コミュニケーション(－)	「人のふれいあいが減る」	25	3.7%
犯罪・トラブル・悪用	「犯罪」、「ハッカー」	113	16.5%
個人情報・プライバシー	「個人情報が漏れやすい」	35	5.1%
格差	「機械音痴は取り残されそう」	34	5.0%
難しさ	「操作が難しそうで怖い」	49	7.2%
無関心	「興味がない」	32	4.7%
情報の多さ	「情報が多い」、「情報氾濫」	39	5.7%
世の中・社会	「情報社会」	49	7.2%
変化・進歩	「文化の進歩」	25	3.7%
経済・産業	「ＩＴ産業の不況」	26	3.8%

一番上のグループは情報技術を表す一般的な用語をコード化したものである。その下の３つのグループは順に、情報技術の便利さへの評価や期待、悪用への不安や懸念、社会一般の変化といった3.3節で発見された要素をもとに作成したコードである。

なお、作成したもののごく僅かしか出現していなかったコード、すなわち回答者数が25に満たないコードは表５に挙げておらず、以下の分析からも省いた。あまりに出現数の少ないコードは、以下で行うような統計的分析にたえないためである。こういったコードの例としては、「携帯電話」、「ブロードバンド」、「料金の高さ」などがあった。

3.4.2　回答欄によるコード出現率の変化

自由回答項目に用意された３つの回答欄には、回答者がもっとも強く、もしくは、はっきりと念頭においていた事柄から順に、記述されていったと考えるのが自然であろう。それでは、この３つの回答欄ごとに、記述内容に何らかの特徴や差異があったのだろうか。

この点を明らかにするために、表５に挙げた21のコードすべてについて、回答欄ごとの出現率に差があるどうかを確認した。すなわち、コードの出現有無と、回答欄の番号とを用いたクロス集計表を21作成した。そして、回答欄ごとに出現率が明らかに異なった（$p<.05$）コードのみをリストアップし、回答欄ごとの出現率を示したものが表６である。

表６　回答欄によるコード出現率の変化

コード	回答欄		
	１つ目	２つ目	３つ目
パソコン	6.60%	3.10%	3.81%
便利さ	16.13%	6.01%	6.09%
情報収集	18.91%	9.69%	7.11%
スピード	7.48%	4.65%	3.05%
犯罪・トラブル・悪用	3.96%	8.33%	14.21%
人間関係－コミュニケーション(－)	0.88%	0.97%	3.81%
個人情報・プライバシー	1.17%	2.52%	3.55%
日常生活－その他	1.32%	2.91%	4.57%

(*) 回答欄ごとの出現率に明らかな（$p<.05$）差異があったコードのみを記述。
また、パーセントは各回答欄の中でのコード出現率である。

表６を見ると、１つ目の回答欄でもっとも出現率が高く、２つ目、３つ目と出現率が減少しているコードとして「パソコン」、「便利さ」、「情報収集」、「スピード」があることがわかる。これらのコードのうち「パソコン」については情報技術を表す一般的な語であるが、「パソコン」以外の３つのコードは、情報技術の便利さへの評価や期待を表すものである。

逆に、１つ目の回答欄に比べて２つ目、３つ目と出現率が増加しているコードには「犯罪・トラブル・悪用」、「人間関係－コミュニケーション(－)」、「個人情報・プライバシー」、「日常生活－その他」がある。見てわかるとおり、「日常生活－その他」を除くすべてのコードは、悪用への不安をはじめとして、情報技術へのさまざまな懸念を表すものである。

以上より回答欄ごとの特徴として、便利さへの評価・期待といったポジティブな記述が後になるほど減少する一方で、悪用への不安といったネガティブな記述は逆に増加していることがわかった。回答者がもっとも強く、はっきりと念頭においていたであろうものとしては、情報技術の便利さへの評価・期待といったものが比較的多かったのに対して、その後２つ目３つ目となると、悪用への不安をはじめとする好ましくない事柄が増加していたのである。２つ目、３つ目の回答欄には文字通りの“ Second thought ”が記述されることが多かったと言えよう[12)]。

4．議　　論

4.1　インターネットに対する意識

以上の分析において、インターネットを中心とする情報技術に対する意識を、質問紙調査における自由回答項目を通じて探索してきた結果、次のような点が明らかになった。まず、分析者の予断を交えない形で、得られた回答の概要を把握することを試みたところ、大きく分けて（i）便利さへの評価・期待、（ii）悪用や犯罪への不安、（iii）漠然とした世の中や社会の変化、という３つの主な要素が見いだされた（3.3節）。さらに、こうしたデータ概要を把握した上で、それをもとにしてコーディングを行い、コーディング結果を用いてより詳細な分析を行った。その結果、インターネットや情報技術に関してまずはじめに想起されるのは便利さへの評価・期待であり、悪用や犯罪への不安については、その後から文字通りの“ Second thought ”、「考え直し」として想起されていた（3.4.2節）。

用意された選択肢にまるをつけるだけでよい通常の質問項目よりも、本稿で用いた自由回答型の質問項目は、答える際に自分なりの回答を考えなければならないという点で、負荷の高い質問項目である。さらに、まず最初に思いつくことを書けばよい１つ目の回答欄よりも、２つ目、３つ目の回答欄に記入する時の方が、自分なりの回答を考えねばならないという負担は大きかったであろう。このように負荷の高い状況で、インターネットないし情報技術に対する不安や懸念がもっとも多く現れた点には注意が必要である。

なぜなら、負荷の低い通常の質問項目を用いたとすると、インターネットや情報技術といった言葉からまず最初に想起される便利さへの評価・期待だけをすくい上げてしまい、悪用や犯罪をはじめとするさまざまな問題点への懸念を捉えそこなう危険がある。この場合、インターネットや情報技術に対する、実際以上に高い評価や期待が、結果と

して得られることになるであろう。

また、一般消費者の視点から見た場合、インターネットをはじめとする情報技術を自らの生活に取り入れるか否かを選択する際には、それにかかる費用のことを勘案するのが当然である。そうした費用ないしコストのことを考える時には、自由回答項目に記入する時以上に高い負荷がかかり、より慎重に「考え直し」がなされることであろう。よって一般消費者の視点から見るならば、上述の、実際以上に高い評価や期待が結果として得られるという危険性を、より重く受け止める必要がある。

回答欄を複数用意した自由回答項目の分析によって、インターネットや情報技術という言葉から人々が、まずは期待を、次に懸念を示すという順番を発見し、その結果として上述のような危険性を見いだせた点が、本稿における主な発見である。

付記

本研究は平成13年度科学研究費基盤研究 A(2)13301007「情報通信技術（IT）革命の文化的・社会的・心理的効果に関する調査研究」（研究代表者 : 直井優）の研究成果の一部です。

注

1）2001年10月に訪問面接法と留置法を併用して実施。日本全国の20歳から89歳までの男女を対象とし、サンプル抽出は層化二段無作為抽出法によって行われた。有効回収数は1,011（67.4%）。

2）KH Coder はフリーソフトウェアであり、現在 http : //hey.to/KO−ichi から入手できる。KH Coder は内部で形態素解析ソフトウェア「茶筌」（松本ほか［14］）ないし「JUNAM」（黒橋ほか［9］）を利用している。なお、分析手順・アプローチの詳細は異なるが、計量的かつ探索的なテキストデータ分析を目指すという点で、本稿で行う分析の方法と、大隅ら［7］の方法は似通っている。また、大隅らの方法を実現するためのソフトウェアとして、「WordMiner」が日本電子計算株式会社から発売されている。

3）KH Coder は統計ソフトウェア（SPSS）とデータをやり取りするための機能も備えているので、統計ソフトウェアを用いる際にも特殊なデータベース操作などは必要ない。

4）インターネットやパソコン通信を知っている人の割合が明らかに高い（$p<.05$）グループを列挙した。なお、ここでは男女差はみられなかった。

5）KH Coder（茶筌）が抽出するのは厳密には単語ではなく形態素であるが、本稿では便宜的にこれを語と見なす。なお、例えば「情報収集」という文字列が「情報」と「収集」という２語に分割されるなど、分割が細かすぎると思われる例も見受けられた。そこで、KH Coder に予備的な解析を行わせて分割が細かすぎる組み合わせを洗い出し、最終的な集計のための解析では、そのようには分割されないようにした。この手順の詳細については樋口［13］をご参照いただきたい。

6）もちろん、入力されたデータが２次元で近似できるものであった場合には、解釈可能な方向が見いだされることもある。

7）KH Coder による集計の結果は表７−(A)に示すような形のデータとして得られる。表７−(A)は、各回答の中（行、またはケース）に、それぞれの単語（列または変数）がいくつ出現しているかという

集計である。この形のデータをそのまま使用しても、それぞれの列（変数）の間の相関を計算すれば、各単語の出現パターンが似通っているかどうかを調べることができる。また、それをもとにしてマッピングを行うこともできる。しかし、以下に述べるような理由から、今回はそのような方法でのマッピングを行っていない。

一般に、あまりに多くの語を１つの図の中にマッピングすると、語が互いに重なりあって判読できなくなってしまう。よって、マッピングを行えるのは150程度の頻出語に限られ、出現数の少ない語は図から省かねばならない。仮に、表７−(A)に示した形状のデータから出現数の少ない語を省くとすると、出現数の少ない語の列（変数）を削除し、頻出語の列だけを残すという作業を行うことになる。この場合、出現数が少なくてマッピングされない語の出現位置情報が完全に失われてしまう。例えば表７−(A)で、出現数の少ない単語Ｃの列（変数）を削除するならば、当然、単語Ｃの出現パターンをマッピングに利用することができなくなってしまう。

表７　データの形状

（A）語の出現数（KH Coder による出力）

	単語 A	単語 B	単語 C
回答１	30	15	5
回答２	10	25	0
回答３	20	0	1

（B）各単語が出現するという条件での期待値（変形後）

	単語 A が出現するという条件（回答１〜３すべてが該当）	単語 B が出現するという条件（回答１と回答２が該当）	単語 C が出現するという条件（回答１と回答３が該当）
単語 A の期待値	(30+10+20)÷３	(30+10)÷２	(30+20)÷２
単語 B の期待値	(15+25+ 0)÷３	(15+25)÷２	(15+ 0)÷２
単語 C の期待値	(5+ 0+ 1)÷３	(5+ 0)÷２	(5+ 0)÷２

たとえマップ上に表示できるのは一部の頻出語であっても、マッピング実行時には出現数の少ない語の位置情報も利用した方が、より精緻なマッピングが行えるであろう。特に自由回答データの場合には、比較的長い新聞記事のようなデータとは異なり、３語、４語といった短い回答も多いので、そもそもマッピングに利用できる情報の量が少ない。したがって、もともと少ない情報量をさらに減らしてしまわないために、出現数の少ない語の情報をも、マッピング時には利用したい。また一般に、出現数が少ない語ほど、特殊な語、あるいは特徴的な語であることが多い点も、そういった語の出現パターン情報を利用するべき理由の１つである。これらの理由から、あらかじめ表７−(B)に示す形状にデータを加工しておくこととした。

表７−(B)は、「各単語が出現するという条件」（列、または変数）のもとで計算した、それぞれの単語の出現数の期待値（行、またはケース）である。表７の例では単語Ｃが出現しているのは回答１と回答３である。そこで、回答１と回答３だけで出現数の期待値を計算すると、次のようになる。

単語 A の期待値 : (30+20)÷２
単語 B の期待値 : (15+ 0)÷２
単語 C の期待値 : (5+ 1)÷２

この３つが「単語Ｃが出現するという条件」における各単語の期待値であり、表７−(B)の例における単語Ｃの列（変数）の値となる。

この形状のデータであれば、仮に単語Cの列（変数）を削除したとしても、単語Aの列（変数）と単語Bの列（変数）の相関を計算する際に、期待値として３行目に残っている単語Cの情報を利用できる。すなわち、出現数の少ない語の列（変数）を削除しても、それらの語の出現パターン情報は行（ケース）に残ることになる。よって、頻出語として残った語をマッピングする際、残らなかった語の出現パターン情報も活かすことができる。

以上のような理由から、KH Coderから出力されたデータを統計ソフトウェア上で表７-(B)のような形状に加工してから、自己組織化マップを作成することにした。

なお、「単語Aが出現するという条件」における各単語の期待値は、他にどのような単語と一緒に、単語Aが使われているのかを表している。すなわちこれは、単語Aが使われる際の文脈を表すものと言える。よって、今回の方法では、各単語が使われている文脈が似通っているかどうかにもとづいて、単語間の類似度を測定し、マッピングを行ったことになる。

８）ここでは助詞、助動詞なども含めてKH Coder（茶筌）によって抽出された語をすべて含む集計データを作成した。このデータを統計ソフトウェア「SPSS 10.1 for Windows」上に読み込み、１つの回答にしか含まれない語、および１つの語しか含まない（抽出されていない）回答をデータから省いた。その上で、表７-(B)に示した形状にデータを変換した。マッピングする頻出語の決定に当たっては、まずどのような文の中にでも出現するような一般的な語、すなわち助詞、助動詞、記号（句読点など）を排除した。また、ひらがなのみからなる語も一般的なものが多いので、これも排除した。こういった語にはKH Coderによって特別な品詞名が与えられるので、容易に一括して排除することができる。さらに今回は副詞を排除し、その上で頻出していた140語、表４に示した語をマッピングすることとした。なお、最終的に用いたデータの列（変数）数は、１つの回答にしか含まれない語を省いたので、633となった。

９）自己組織化マップの作成に当たっては専用ソフトウェア「Viscovery SOMine Plus ver.４」を用いた。このソフトウェアについては、Deboeck & Kohonen［1：252－254, 277－316］が詳しい。このソフトウェア上で、すべての変数に対してSigmoid変換を適用するという前処理を行った。また学習については、「学習率係数」を必要としないバッチSOMが用いられ、93回の「バッチ」からなる学習サイクルが24度繰り返された。

10）紙幅の都合で、本稿で分析に用いたKH Coder用のコーディング・ルールは添付できなかったが、これをご覧になりたい方は樋口〈ko-ichi@hus.osaka-u.ac.jp〉までご連絡いただきたい。

11）自由回答項目（Q２）に記述がある回答者の中での割合を算出した。

12）特に「犯罪・トラブル・悪用」コードについては、２つ目３つ目の回答欄で出現率が増加しているだけでなく、出現数そのものの増加も顕著である。

文献

［１］Deboeck, G., Kohonen, T. 1998 *Visual Explorations in Finance with Self-Organizing Maps* Springer-Verlag（徳高平蔵 田中雅博 監訳『金融経済問題における可視可情報探索―自己組織化マップの応用』シュプリンガー・フェアクラーク 1999）

［２］Doszkocs, T. E., Reggia, J., Lin, X. 1990 "Connectionist Models and Information Retrieval" *Annual Review of Information Science and Technology* vol. 25 pp. 209－260

［３］Kohonen, T. 1988 *Self-Organization and Associative Memory* 2nd edition. Springer-Verlag（中谷和

夫 監訳『自己組織化と連想記憶』シュプリンガー・フェアラーク 1993）

［4］Kohonen, T. 1995 *Self-Organizing Maps* Springer-Verlag（徳高平蔵 岸田悟 藤村喜久郎 訳『自己組織化マップ』シュプリンガー・フェアクラーク 1996）

［5］Ritter, T. & Kohonen, T. 1989 “Self-organizing semantic maps” *Biological Cybernetics* 61 pp. 241－254

［6］Seale, C. 2000 “Using Computers to Analyse Qualitative Data” Silverman, D. *Doing Qualitative Research : A Practical Handbook* London : Sage pp. 154－174

［7］大隅昇 Lebart, L. 2000「調査における自由回答データの解析―InfoMiner による探索的テキスト型データ解析」『統計数理』48(2) pp. 339－376

［8］川端亮（編著）1999『非定型データのコーディング・システムとその利用』平成８年度～10年度科学研究費補助金（基盤研究(A)(1)）（課題番号 08551003）研究成果報告書

［9］黒橋禎夫 長尾真 1998「日本語形態素解析システム JUMAN version 3.61」京都大学大学院情報学研究科

［10］佐藤裕 2000「質問紙調査の自由回答項目における『家族言説』」中河伸俊（編）『家族をめぐる言説の実証的研究』平成９年度～11年度科学研究費補助金（基盤研究(B)(1)）研究成果報告書

［11］総務省（編）2002『情報通信白書 平成14年版』ぎょうせい

［12］太郎丸博 1999「自由回答の回答率と回答の長さ―調査法および回答者の社会的属性が及ぼす影響」川端亮（編著）『非定型データのコーディング・システムとその利用』平成８年度～10年度科学研究費補助金（基盤研究(A)(1)）（課題番号 08551003）研究成果報告書 pp. 61－69

［13］樋口耕一 2003（近刊）「コンピュータ・コーディングの実践」『年報人間科学』第24号

［14］松本裕治，北内啓，山下達雄，平野善隆，松田寛，浅原正幸 1999「日本語形態素解析システム『茶筌』version 2.0 使用説明書 第二版」NAIST Technical Report NAIST-IS-TR 99012

Observations on what and how people think about the Internet : Analyzing open−ended questions

Akira KAWABATA and Koichi HIGUCHI

Today, propagation of the Internet and number of articles about it in newspapers are increasing very rapidly. It is certain that people are using and talking about the Internet more and more frequently. The purpose of this paper is to figure out what and how they think about the Internet.

The data were answers written for open ended questions in JIS2001 : the Japan survey on Information Society. And they were analyzed by means of “Computer Coding”, a relatively new automated approach to code and analyze qualitative data such as open ended questions.

At first, the authors tried to summarize the data without imposing interference of their own prejudice by following procedure. The computer was used to extract words from the data automatically and a SOM : Self−Organizing Map of those words was made. As a result, it was found that the answers could be classified roughly into (i) positive evaluations or hopes for its convenience, (ii) fears of crimes or troubles, and (iii) general impressions about changing society and economy. After the summary of the data was obtained, the answers were coded for closer quantitative analysis. The results of the analysis are as follows. When people think about the Internet, positive evaluations or thoughts of its convenience often come to their mind at first. After that, fears of crimes or troubles tend to come to their mind as second thoughts.

Thus the authors argue that when one researches what people think about the Internet, he should consider a possibility that he may get only the positive evaluations or hopes which come to their mind at first and he may fail to get second thoughts such as fears of crimes or troubles.