

〈研究ノート〉

人はどのような言葉で幸福を語るのか？*

— 幸福理由のテキスト・マイニング —

石 田 淳**

1 はじめに

本稿では、探索的に人々の幸福／不幸理由の典型的な「語り口」を浮かび上がらせることを目指す。具体的には、2005年に実施されたCOE幸福調査データを用いて、テキスト・マイニングの手法によってどのような言葉が幸福理由の自由回答記述に用いられたかを分析する。

COE幸福調査は、関西学院大学大学院社会学研究科21世紀COEプログラム『「人類の幸福に資する社会調査」の研究——文化的多様性を尊重する社会の構築」における研究活動の一環として実施された。本調査は、(株)マイボイスコム サービスを利用したネットリサーチであり、A票、B票の2つの調査票に対して、それぞれ500人の回答を得た。実施期間は2005年10月12日～14日である。調査対象者は、マイボイスコムの登録モニターであり、A、B票ともそれぞれ、マイボイスコム登録モニターから20～69歳の男女を男女比が均等になるように1,250人を無作為抽出し、それぞれの調査票について回答者の累計が500人を超えた時点で調査を終了した。

この調査の目的は、人々がいかに幸福を感じているか、またその背後にはどのようなメカニズムがあるかを探索的に検討することにあつた¹⁾。そこでまず、幸福感研究のもっとも基本的な変数である一般的幸福感の度数分布を確認しよう。図1は問7「全体的に言って、あなたはいまどの程度

幸福だとお感じですか」に対する回答の分布を示している²⁾。図1から、70%近くの人が「幸福」と感じていることが分かる。同様に図2は、問9「10年前と現在を比べて、あなたは幸福になったとお感じですか。それとも不幸になったとお感じですか」に対する回答の分布を示している。

これらの変数は、分析に際して基本的な従属変数となるものであり、他の変数との組み合わせによって単純なクロス表分析から、多変量解析まで様々な分析が可能になる。だが本稿では、それらの分析の前に、より探索的に「幸福理由」の自由

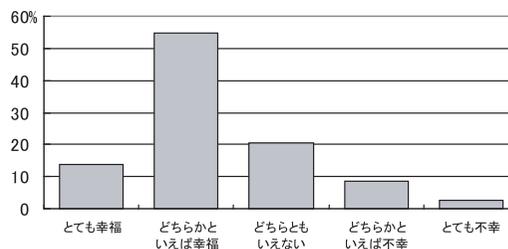


図1 一般的幸福感の度数分布 (A、B票)

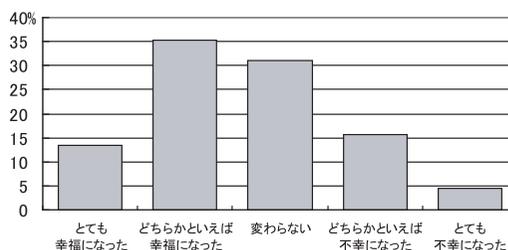


図2 10年前比較幸福感の度数分布 (A、B票)

*キーワード：一般的幸福感、テキスト・マイニング、数量化Ⅲ類

**日本学術振興会 特別研究員 PD

1) COE幸福調査に基づく研究成果として浜田(2006)がある。

2) ただし、A票では問7の選択肢中の「不幸」を「幸福ではない」と変更している。ワーディングの違いによる分布の大きな違いは見られなかった。

回答を用いた分析を行いたい。本調査では、問7、問9のそれぞれの直後に、「なぜそのようにお答えになりましたか。ささいなことでも結構です。理由を自由にお書きください」と回答理由を尋ねている。手書きよりもキーボード入力の方が簡単であるというネット調査の利点もあって、まとまった記述が多く集まった。ここでは、この自由回答記述をもとに、テキスト・マイニングによって「幸福についての語り」の分析を行いたい³⁾。

2 方法

調査では、さまざまな言葉づかいによって幸福もしくは不幸と答えた理由が語られた。いくつかの回答をピックアップしよう。

「食べるのに困らず、たまになら贅沢もできて、家族も皆健康だから」

「マイホームもあるし買いたいものも結構買えるし、時間もたっぷりあるし、主人は家庭思いなので」

「『みかん1コの幸せ』の言葉の通り、幸せは積極的に探せばいくらでも身近に転がっているものだ」

「体調が良くなく、仕事にストレスを感じている」

「持病が数多くあり医療費の自己負担の割合が増えたため金銭的な負担が多くなったため。おまけに不景気でボーナスが少ないこと」

これらの回答は、それぞれの回答者の事情が反映された個人的なものである。しかしながら、数多くの自由回答記述を集計することによって、どのような言葉や表現が幸福や不幸を語る際に多用されるのか、またある言葉がどのような言葉とともに用いられる傾向にあるのか、といった「語り口」の傾向を分析できるようになる。そのために、以下の手順に従って、自由回答記述を分析用

のテキスト・データへとリファインする作業を行った⁴⁾。

- ①形態素解析ソフトである「茶筌」(WinCha 2000)を利用して、各々の回答者の自由回答記述から「意味をもつ最小の言語単位」である形態素を抽出し、それらを活用のない基本形に直す。結果として、回答者ごとに「自由回答に用いた単語のリスト」が生成される。
- ②WordMiner1.1によってさらなるデータのリファインを行う。具体的には以下の通り。
 - 表記が混在している場合、カタカナ・ひらがな表記は漢字表記に統一する
 - 「周り／回り」など、誤字と思われるものは、コンコーダンス機能を用いて、文脈を確認の上、修正する
 - 「失業／失職」など、明らかに同様の意味で用いられており、ニュアンスの違いもないと判断できる場合は、同一語句として扱う
 - 記号、句読点、助詞を削除する
 - 助動詞、代名詞、接続詞、副詞、その他一語では意味をなさない言葉（「いう」、「思う」、「感じる」、「いる」、「ある」などの動詞、「ため」、「たち」、「もの」などの名詞）を削除する
 - 出現回数が10以下の単語を削除する

こうして作成されたデータをもとに、以下分析を行う。

3 単純集計結果

まずは、一般的幸福感の回答理由に用いられた言葉について見ていこう。表1は一般的幸福感の回答理由に用いられた言葉の出現回数の単純集計結果である。回答の中で出現した言葉として、「家族」がもっとも多く、次に多いのが「生活」、「健康」である。特に「家族」という言葉は、そ

3) フランスにおける全国調査をもとに、階級と幸福の語りとの関係をテキスト・マイニングによって分析したものととしてボードゥロ (2004) がある。

4) 利用したアプリケーション・ソフトの使用法を含むテキスト・マイニングの入門書として藤井ほか編 (2005) が有用である。

表1 幸福理由単純集計

語	度数	%	語	度数	%
1位 家族	181	8.14	11位 不幸	45	2.02
2位 生活	133	5.98	12位 幸福	44	1.98
3位 健康	114	5.13	13位 普通	42	1.89
4位 子供	98	4.41	14位 楽しい	41	1.84
5位 幸せ	89	4.00	15位 不自由	40	1.80
6位 自分	84	3.78	15位 暮らす	40	1.80
7位 人	68	3.06	17位 毎日	39	1.75
8位 経済	61	2.74	18位 困る	38	1.71
9位 仕事	59	2.65	19位 ストレス	33	1.48
10位 良い	57	2.56	20位 皆	31	1.39
			20位 元気	31	1.39
			計	2223	100.00

表2 幸福理由男女別単純集計（上位10語）

男性			女性		
語	度数	%	語	度数	%
1位 家族	61	7.50	1位 家族	120	8.51
2位 生活	55	6.77	2位 生活	78	5.53
3位 健康	39	4.80	3位 健康	75	5.32
4位 自分	36	4.43	4位 子供	69	4.89
5位 子供	29	3.57	5位 幸せ	64	4.54
6位 仕事	27	3.32	6位 人	51	3.62
6位 不幸	27	3.32	7位 自分	48	3.40
8位 幸せ	25	3.08	8位 良い	41	2.91
9位 経済	23	2.83	9位 経済	38	2.70
9位 幸福	23	2.83	10位 仕事	32	2.27
計	813	100.00	計	1410	100.00

れが含まれる「子供」を合わせるとかなりの割合で言及されていることが分かる。そのほか、8位に「経済」、9位に「仕事」といった経済的側面にかんする言葉が出現していることも注目される。

表2は一般的幸福感の回答理由に用いられた言葉の男女別の集計結果である。この結果を見る限り、性別によって使用される言葉が大きく異なるということはない。

次に、年代別の単純集計結果を見てみよう（表3）。ここでは年代による違いが見られる。まず、20代の最頻出単語は「自分」であるが、30代以降は「家族」になる。また、30代では「子供」

表3 幸福理由年代別単純集計（上位5語）

20代			30代		
語	度数	%	語	度数	%
1位 自分	30	6.61	1位 家族	61	6.91
2位 家族	27	5.95	2位 生活	59	6.68
2位 幸せ	27	5.95	3位 子供	53	6.00
4位 生活	22	4.85	4位 健康	50	5.66
5位 人	21	4.63	5位 幸せ	30	3.40
5位 不幸	21	4.63			
計	454	100.00	計	883	100.00
40代			50代以上		
語	度数	%	語	度数	%
1位 家族	63	11.23	1位 家族	30	9.23
2位 生活	32	5.70	2位 健康	24	7.38
3位 健康	26	4.63	3位 生活	20	6.15
4位 子供	22	3.92	4位 仕事	16	4.92
5位 幸せ	20	3.57	5位 経済	15	4.62
計	561	100.00	計	325	100.00

表4 幸福理由幸福感別単純集計（上位5語）

幸福			不幸		
語	度数	%	語	度数	%
1位 家族	165	9.62	1位 経済	8	6.06
2位 生活	107	6.24	1位 仕事	8	6.06
3位 健康	104	6.06	1位 ストレス	8	6.06
4位 子供	85	4.95	4位 自分	7	5.30
5位 自分	61	3.55	4位 良い	7	5.30
			4位 人	7	5.30
計	1716	100.00	計	883	100.00

が高い割合で出現している。さらに、興味深いのが30代以降で「健康」という言葉が頻出していることであろう。また、50代で「仕事」「経済」という言葉が上位に出現していることも興味深い。総じて、ライフ・コースに応じた言葉遣いの変化というものが見出せるだろう。

表4は、一般的幸福感の回答として「とても幸福」「どちらかといえば幸福」と答えた回答者と、「とても不幸」「どちらかといえば不幸」と答えた回答者ごとの頻出単語を示している。特に注目すべきは、不幸グループの最頻出単語が「経済」「仕事」「ストレス」であるということだ。大雑把に言えば、幸福は家族や健康といった生活領

表5 比較幸福理由単純集計

語	度数	%	語	度数	%
1位 子供	131	8.82	11位 仕事	41	2.76
2位 結婚	88	5.93	12位 変化	37	2.49
3位 変わる	81	5.45	13位 経済	36	2.42
4位 自分	72	4.85	14位 不幸	26	1.75
5位 幸せ	68	4.58	15位 自由	25	1.68
6位 家族	66	4.44	16位 昔	24	1.62
7位 生活	63	4.24	17位 楽しい	23	1.55
8位 人	47	3.16	17位 成長	23	1.55
9位 増える	45	3.03	19位 色々	22	1.48
10位 良い	44	2.96	19位 余裕	22	1.48
計				1485	100.00

表6 比較幸福理由男女別単純集計（上位10語）

男性			女性		
語	度数	%	語	度数	%
1位 子供	47	7.93	1位 子供	84	9.42
2位 結婚	31	5.23	2位 結婚	57	6.39
3位 変わる	29	4.89	3位 幸せ	55	6.17
4位 家族	27	4.55	4位 変わる	52	5.83
4位 生活	27	4.55	5位 自分	51	5.72
4位 増える	27	4.55	6位 家族	39	4.37
7位 人	22	3.71	7位 生活	36	4.04
7位 変化	22	3.71	8位 良い	28	3.14
9位 自分	21	3.54	9位 人	25	2.80
9位 仕事	21	3.54	10位 仕事	20	2.24
計	593	100.00	計	892	100.00

域、不幸は仕事や金銭問題など経済領域に関連して語られていると言えるだろう。

次に、10年前と比較した幸福感の変化についての自由回答記述の結果を見てみよう。表5は比較幸福感の回答理由に用いられた言葉の出現回数の単純集計結果である。最頻出単語は「子供」で、次に「結婚」、「変わる」となっている。子供の誕生や結婚といった家族形成イベントが、幸福感変化の理由として多く語られる傾向にあるようだ。

表6は比較幸福理由の男女別の集計結果である。ここでも、性別による使用単語の大きな差異は見られない。

表7は比較幸福理由の年代別の集計結果である。20代の最頻出単語は「自分」、30代、40代は

表7 比較幸福理由年代別単純集計（上位5語）

20代			30代		
語	度数	%	語	度数	%
1位 自分	26	7.74	1位 子供	68	11.31
2位 結婚	19	5.65	2位 結婚	52	8.65
2位 変わる	18	5.36	3位 幸せ	42	6.99
4位 良い	16	4.76	4位 家族	32	5.32
5位 人	15	4.46	5位 変わる	24	3.99
計	336	100.00	計	601	100.00
40代			50代以上		
語	度数	%	語	度数	%
1位 子供	41	11.23	1位 生活	17	9.29
2位 変わる	27	7.40	2位 変わる	12	6.56
3位 家族	20	5.48	2位 仕事	12	6.56
3位 生活	20	5.48	4位 自分	11	6.01
5位 変化	17	4.66	5位 子供	9	4.92
			5位 人	9	4.92
計	365	100.00	計	183	100.00

表8 比較幸福理由幸福感別単純集計（上位5語）

幸福になった			不幸になった		
語	度数	%	語	度数	%
1位 子供	112	14.07	1位 自分	14	5.79
2位 結婚	75	9.42	2位 仕事	12	4.96
3位 家族	52	6.53	3位 経済	11	4.55
4位 自分	45	5.65	4位 子供	9	3.72
5位 増える	38	4.77	4位 人	9	3.72
計	796	100.00	計	242	100.00

「子供」、50代は「生活」である。また、「結婚」という言葉が頻出するのは20代、30代である。ライフ・コースの各段階における家族形成イベントが、幸福感変化の主な理由として頻出していることが分かる。

最後に表8は、比較幸福感の回答として「とても幸福になった」「どちらかといえば幸福になった」と答えた回答者と、「とても不幸になった」「どちらかといえば不幸になった」と答えた回答者ごとの頻出単語を示している。やはりここでも、幸福は家族生活領域、不幸は経済領域に関連して語られている傾向が見られる。

4 数量化Ⅲ類の結果

次に、出現単語とカテゴリ変数のクロス表について数量化Ⅲ類による分析を行った。その結果を見てみよう。

図3は、幸福理由出現単語と一般的幸福感のクロス表に数量化Ⅲ類を行った結果の布置図である。単語やカテゴリ間の布置図上の距離がそれらの出現パタンの近似性を表している。たとえば、カテゴリ「とても幸福」の近くには、「好き」、「仲良く」、「一緒」などの言葉が散見されるが、これらの単語は「とても幸福」と答えた回答者がよく用いた単語で、また同時に用いられやすいことを示している。逆に、「どちらかといえば不幸」の周辺には「金銭」、「将来」、「ストレス」などといった言葉が見られる。図4は図3の「とても幸福」、「どちらかといえば幸福」周辺を拡大したものである。

次に、図5は比較幸福理由出現単語と比較幸福感のクロス表に、同様にして数量化Ⅲ類を行った結果の布置図である。「幸福」周辺には、「家族」、「子供」、「増える」、「結婚」などの言葉、「不幸」周辺には「離婚」、「収入」、「独身」、「ストレス」などの言葉を見ることができる。図6は図5の第四象限を拡大した図である。

5 おわりに

ここまで、幸福・不幸の理由、幸福感変化の理由についてのテキスト・マイニングによる探索分析を行ってきた。見出された傾向性としてもっとも注目されるものは、「幸福」は家族関係や健康といった生活領域に関連して、「不幸」は経済的領域に関連して語られる傾向にあるということだ。特に、幸福や幸福感変化の理由としては、ライフ・コースに対応した家族形成イベントが数多く言及されていた。もちろん回答者のそれぞれは、固有の経験にもとづいて幸福感の理由を語ってくれているのであるが、分析を通して見えてきたのはいわば「幸福の定型」のようなものが存在するということである。それはさらに具体的にはいかなるものだろうか。そして、他の文化や時代

による違いが存在するのだろうか。こうした問いに対して、今後はより精緻な分析や比較研究によって答えていく必要があるだろう。本研究は、その端緒の端緒に過ぎない。

【付記】

本稿はもともと、COE幸福調査報告書用に2006年に執筆されたものであるが、その後諸般の事情により報告書自体が出版されなかったために、これまで未出であったものである。

参考文献

- クリスチャン・ボードゥロ, 2004, 「幸福とは, 所有するものか, 状態か, 行うものか?」『先端社会研究』創刊号: 285-308.
- 浜田宏, 2006, 「客観的幸福と主観的幸福」『社会・経済システム』27: 71-84.
- 藤井美和・小杉考司・李政元編, 2005, 『福祉・心理・看護のテキストマイニング入門』中央法規.

A Text Mining Analysis of the Causes of Happiness and Unhappiness

ABSTRACT

In this research note, I conduct an exploratory analysis of the causes of people's happiness and unhappiness. I do so by employing data collected in the Kwansai Gakuin University 21st century COE program happiness survey of 2005. I specifically use the analytical method known as text mining, and analyze word frequencies and word relations that emerged in the open-ended questions related to happiness and unhappiness in the 2005 survey. As a result of this analysis, I find that the word "happiness" tends to be related concepts relevant to life situation such as "family," "child," and "health." On the other hand, unhappiness relates more strongly to the words relevant to economic situation.

Key Words: Subjective well-being, text mining, quantification theory III