

## 郵送自記式調査とWeb調査による自由回答の分析例

○渡會 隆・小山裕之（東京サーベイ・リサーチ）＊中谷吉孝（博報堂）

### 1) 市場調査環境の変化

各種調査環境が急変している。いわゆるサンプリングひとつとっても住民台帳を例にあげれば、地方自治体によっては住民台帳の「閲覧不可」あるいは「閲覧予約が相当先」になる、すぐには見られないところが多い。選挙人名簿もほぼ同様の状況にある。電話帳への非掲載も多く、さらに携帯電話の普及に従って世帯固定電話の番号簿は個人を特定するには用をなさないようになってきた。たとえ、ようやく世帯にたどりついて留守・不在/留守番電話と調査不能で回収率は年々厳しい状況に追い込まれつつある。いきおい街頭調査/集合会場調査/来店客・入場者調査で対象者に調査員が面接聴取するか、あるいは留置・郵送調査に頼ることが多いのが「従来調査法」の現状である。

### 2) web調査への過大な期待

2000年通信白書によればインターネット利用者2706万人/普及率にして21%に達したといわれる。(民間調査では1600～1900万人と差があるが) 利用者の性・年齢をみると圧倒的に20・30代の男性が多く、また地域的には県庁所在地などの都市部は24%だが町村部では14%とみられている。仮にweb調査の母集団と従来調査の住民台帳や選挙人名簿からのサンプリングあるいは街頭調査や来店客調査のランダム性と比べると全く異質な対象群を調査していることになる。インターネット・リサーチは「早い・安い・大量」をセールス・ポイントに売込みは激しい。しかし、母集団の性質の違い、調査法の差(質問票の質量、謝礼・インセンティブ、調査員の介在など) 回答の信頼性(謝礼目当て、なりすまし、複数回答、サンプルの偏り/性年齢・PCへの熟練度など) 等いまだ不明のまま取り扱われている。つまりデータの「正確さ」を曖昧にしたままweb調査へ過大な期待が先行しているように見受けられる。

### 3) 従来調査法とweb調査法の比較

web調査といっても「パネル型/モニター契約」「クローズ型/サンプリング」「オープン型/この指とまれ」に大別できる。従来調査法と比べ調査員が介在しないで、対象者自身が自記式で回答する、といった点では従来調査法の「郵送調査法」と類似している。今回の事例に示す実験調査は選択肢型の質問と自由回答型の両方でweb調査/パネル型と郵送調査/パネル型で比較することとし、特に自由回答文の部分をいかに客観的に解析比較できるかが検証のキーポイントとなる。始めに取得するデータそのものの等質性を保つために博報堂の生活者モニター「HABIT」からサンプリングし、調査票の構成も(画面と紙の違いはあるが) 同一にした。またweb調査のデータは書き込まれたテキストデータをそのまま取り込み、郵送調査で回収されたデータは手書きされた文言を訂正せずテキストデータ化したものを原データセットとした。

#### 4) 実験調査のねらいと調査設計メモ

実験調査のねらいは以下に示すように

- 1) 従来調査法である、いわゆる面接調査法とか郵送法などにweb調査法が取って替わりうるのか特に今回は郵送自記式調査とweb調査の比較を行った。
- 2) H A B I T DB という一元管理されたパネル型/モニター契約の対象サンプルの中で画面と紙の違いしかない調査票上に回答してどんな差違がみられるのか
  - a) 自由回答形式の場合/選択肢設問の場合にそれぞれどんな違いがあるか
  - b) 質問量で回答の(質量)に違いがあるか
  - c) 回収率はどうか
  - d) 実施上の時間・手間・コストはどうか
- 3) 解析ツール「InfoMiner」の機能(分かち書き・解析・クラスター化)のキレと操作性実験などである。こうしたテーマから(表1)の調査設計にした訳である。

< 表1 調査設計 >

	郵送調査	WEB調査
マスターサンプル	H A B I Tパネル	H A B I Tパネル
	15歳から69歳	電子パネル
	男女個人	男女個人
	3,740名	3,500名
スクリーニング	①パソコンの自宅使用者 & 2年以内購入者	20歳~39歳
	②パソコンの自宅非保有者 & 会社・学校での使用者	
発送数	576名	973名
回収数	351名	228名
調査法	郵送調査	e-mail + HP
調査時期	1998年12月	1998年12月

この調査は、統計数理研究所の大隅教授が開発したテキスト型データ(自由回答)解析ツール『InfoMiner』を使用し、博報堂の生活者データベースH A B I Tのパネルを利用し、実査および集計は東京サーベイ・リサーチが行うという三者の共同実験調査である。

#### 5) 調査結果から

調査票(実際の調査票は本報告時に提示)の質問構成は選択肢型質問と自由回答型質問の混合で、かなり膨大な質問量である。通常これほどの質問量で構成すること自体、調査への協力度・回収率の低下を予想させるので行われぬ。しかし、今回は質問量と回答傾向も実験のねらいでもあり、また対象者がパネル・モニター契約者でもあったので協力度・回収率も見込んで実施した。郵送調査発送数576/回収数351(回収率61%)、web調査発送数973/回収数228(23%)である。

今回はその中の21の自由回答型質問に限定し、その特徴的な傾向を述べる。サマリーは「21質問の単語頻度比較」(表2)にみる通りである。

- 1) 郵送調査の総単語数(分かち書き数)は総じてweb調査のそれよりも多い、しかし異なり単語率(同じ表記のものは1語とした率)ではweb調査の方が高率を示す。
- 2) 総単語数も異なり単語数も質問の順に低減傾向が見られる。特に同様質問の繰り返しで「回答欄(つまり回答数)非限定」時に、その傾向は強まる。回答をいくつまでと限定する方が異なり単語率も下がらない。
- 3) 質問文のワーディングに含まれるワードは回答にもリーディング・ワードとして頻度が多く出やすい。(この場合、分析に採用するかしないかの判断が必要)
- 4) 逆に出現頻度が極端に少ないものも分析への採否は要検討。単語の出現頻度分布/閾値分布で閾値1は全単語、閾値2は2回以上使用された単語の意味。閾値が高くなるに従って異なり単語率は急速に低減することに注意。(※本報告時に提示)
- 5) 同様繰り返し質問(例:SONY、NEC、APPLEのイメージ)は前の質問で回答した表現を比較のために繰り返し、それをさらに否定・肯定・強調条件として利用する傾向がある。
- 6) 選択肢型質問が先行し、その理由としての自由回答の場合(例:5段階選好度の理由)ある程度選択肢に沿った回答がえられるようだが、それでも表現される単語や表現の意味は選択肢の段階を横断的に混在するものもある。
- 7) クラスター化はサンプル数と異なり単語数のスパースな関係をInfoMinerで解析されるが、事前にサンプルや単語の吟味を精密に行えば、それなりに細分化も可能だが、仮にそうした事前の吟味作業を省略し、クラスター化を段階的(初動探索的)に繰り返しサンプルと単語の選別を行う客観性を選ぶかは調査質問の目的と回答の質量にもよる。
- 8) 郵送自記式調査とweb調査の自由回答は各クラスターへの分布からみてもサンプル特性を反映した構成をみせ、回答の仕方の本質は変わらないようである。
- 9) もっとも回収率は、郵便物で個人に確実に手渡される調査票と、web上で、まずメールでHPにアクセスを促し、当初は質問量も分からないアンケートに答えるという手順を考慮すると、回収率の差を縮めるには別の工夫がいると思われる。

例えば以下のようなことである。

- ・対象者の選び方・除外の仕方(サンプリングソース、スクリーニング基準、DB管理法)
- ・質問法と回答の傾向(質問量、ワーディング、スペース、回答限定数、ページ設定)
- ・謝礼・インセンティブの提示の仕方(ポイント、図書券、抽選・懸賞)
- ・対象者と調査者の信頼関係(パネル・モニター契約、DB管理程度、クローズ度)
- ・集計・解析の方法とデータ品質の点検(データ取得条件、いつ・どこで・誰に・方法)

(抄録作成者:渡會隆/小山裕之/e-mail:takashi.watarai@tsrn.co.jp,Hiroyuki.Koyama@tsrn.co.jp)

< 表2 21質問の単語頻度比較 >

			SONYに ついての 印象	NECにつ いての印 象	Appleに ついての 印象	SONY 商品、事 業認知	NEC商 品、事業 認知	Apple商 品、事業 認知
郵送 N=351	回答件数		345	339	335	348	342	329
	a総単語数	単語数	12630	9941	8134	2351	1753	1022
		平均単語数	36.0	28.3	23.2	6.7	5.0	2.9
	b異なり単語数	単語数	2186	1876	1718	439	411	289
		平均単語数	6.2	5.3	4.9	1.3	1.2	0.8
c異なり単語率	b÷a	17.31	18.87	21.12	18.67	23.45	28.28	
WEB N=228	回答件数		215	219	228	225	220	228
	a総単語数	単語数	6477	4754	4738	1563	1734	1481
		平均単語数	18.5	13.5	13.5	4.5	4.9	4.2
	b異なり単語数	単語数	1480	1219	1398	374	415	439
		平均単語数	4.2	3.5	4.0	1.1	1.2	1.3
c異なり単語率	b÷a	22.85	25.64	29.51	23.93	23.93	29.64	

  

			SONYを 言い表す 言葉(形 容詞)	NECを言 い表す言 葉(形容 詞)	Appleを 言い表す 言葉(形 容詞)	SONYの 長所	NECの長 所	Appleの 長所
郵送 N=351	回答件数		332	301	281	328	311	313
	a総単語数	単語数	3425	2621	2027	6034	4647	3785
		平均単語数	9.8	7.5	5.8	17.2	13.2	10.8
	b異なり単語数	単語数	891	914	744	1339	1190	1045
		平均単語数	2.5	2.6	2.1	3.8	3.4	3.0
c異なり単語率	b÷a	26.01	34.87	36.70	22.19	25.61	27.61	
WEB N=228	回答件数		221	227	225	227	226	226
	a総単語数	単語数	2363	2541	2391	3310	2898	3065
		平均単語数	6.7	7.2	6.8	9.4	8.3	8.7
	b異なり単語数	単語数	741	944	923	964	917	998
		平均単語数	2.1	2.7	2.6	2.7	2.6	2.8
c異なり単語率	b÷a	31.36	37.15	38.60	29.12	31.64	32.56	

  

			VAIOに ついての 印象	iMacに ついての 印象	VAIOを言 い表す言 葉(形容 詞)	iMacを 言い表す 言葉(形 容詞)	VAIOの 長所	iMacの 長所
郵送 N=351	回答件数		333	330	209	243	316	319
	a総単語数	単語数	5120	6653	1341	1666	3816	4140
		平均単語数	14.6	19.0	3.8	4.7	10.9	11.8
	b異なり単語数	単語数	1229	1505	503	591	962	1067
		平均単語数	3.5	4.3	1.4	1.7	2.7	3.0
c異なり単語率	b÷a	24.00	22.62	37.51	35.47	25.21	25.77	
WEB N=228	回答件数		226	225	223	222	226	226
	a総単語数	単語数	4413	4690	2042	2251	2992	2842
		平均単語数	12.6	13.4	5.8	6.4	8.5	8.1
	b異なり単語数	単語数	1238	1324	660	727	948	920
		平均単語数	3.5	3.8	1.9	2.1	2.7	2.6
c異なり単語率	b÷a	28.05	28.23	32.32	32.30	31.68	32.37	

  

			自宅PC の購入目 的と経緯	自宅PC の使用状 況	PCの運 人イメー ジ
郵送 N=351	回答件数		163	162	310
	a総単語数	単語数	5757	4974	5893
		平均単語数	16.4	14.2	16.8
	b異なり単語数	単語数	1261	1229	1329
		平均単語数	3.6	3.5	3.8
c異なり単語率	b÷a	21.90	24.71	22.55	
WEB N=228	回答件数		220	218	223
	a総単語数	単語数	8037	5474	4021
		平均単語数	22.9	15.6	11.5
	b異なり単語数	単語数	1721	1184	1072
		平均単語数	4.9	3.4	3.1
c異なり単語率	b÷a	21.41	21.63	26.66	