

XAVIER MARC

Les modalités de recueil des réponses libres en institut de sondage : le rôle de l'enquêteur, les consignes et les procédures de contrôle, les perspectives d'amélioration

Journal de la société française de statistique, tome 142, n° 4 (2001), p. 21-28

http://www.numdam.org/item?id=JSFS_2001__142_4_21_0

© Société française de statistique, 2001, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

LES MODALITÉS DE RECUEIL DES RÉPONSES LIBRES EN INSTITUT DE SONDAGE :

Le rôle de l'enquêteur, les consignes et les procédures de contrôle, les perspectives d'amélioration

Xavier MARC*

RÉSUMÉ

Cette note s'intéresse aux problèmes liés au recueil des données et à leur mise en forme : dans quelles mesures les procédures de recueil sur le « terrain » par les instituts de sondages sont-elles adaptées à l'analyse des réponses données à des questions ouvertes ? Pour apporter des réponses très empiriques à cette interrogation ainsi que de possibles pistes d'amélioration, il convient de revenir sur deux étapes dans la fabrication d'un sondage : l'interrogation et la codification des questions ouvertes.

Mots clés : Sondage, questions ouvertes, recueil des données, enquêteurs, qualité, normes ISO.

ABSTRACT

This article is about the problems related to data collection and codification. To what extent the procedures used by research institutions to collect data in the field are well adapted to the analysis of the responses given to open-ended questions ? A truly empirical response to this question along as well as possible ideas for improvement, results from the reconsideration of two important stages of the process adopted to carry out the survey : one concerns the manner in which the open-ended questions are asked and the second refers to how the responses are codified.

Keywords : Survey, open-end questions, collection of data, interviewers, quality, ISO standards.

1. Introduction

La réflexion ¹ qui suit porte sur la qualité du recueil de verbatims et n'est pas la première à soulever cette question de la responsabilité des enquêteurs ou de

* (GRETS / EDF R&D) 1 avenue du Général de Gaulle, 92141 CLAMART Cedex, E-mail : xavier-c.marc@edf.fr

1. Le présent article se base sur un travail bibliographique, sur des entretiens avec des responsables des principaux instituts BVA, CSA - TMO, IPSOS, ISL, Louis Harris, MV2 Conseil et SOFRES ainsi que sur notre expérience de terrain en instituts et chez l'annonceur.

celle du terrain d'un institut dans les réponses émises par les interviewés. Loïc Blondiaux (1998), Denis Duclos et Hélène-Y. Meynaud (1996), par exemple, se sont attardés sur ce thème à différentes reprises. Les JADT² ont également accueilli plusieurs interventions sur ce thème. On peut également se remémorer la position déjà ancienne de Jean-Pierre Pagès qui insistait sur les variations observables d'un institut à l'autre et qui en conséquence a toujours tenu à conserver le même institut pour réaliser les études d'Agoramétrie. S'interroger aujourd'hui sur les conditions du recueil de données d'enquêtes, plus encore de données ouvertes, revient à se poser des questions relativement déroutantes comme celles-ci : dans quelle proportion un sondage mesure-t-il une opinion ou au contraire la fabrique-t-il ? *A minima*, dans quelle mesure biaise-t-il les opinions recueillies ? Sachant que poser une question ouverte est, pour de multiples raisons, plus compliqué que poser un enchaînement de questions fermées³, il est légitime de se demander **quelle qualité de prestations avons-nous entre les mains, et quelle exigence de qualité pouvons-nous attendre ?** Pour mieux comprendre les enjeux de la qualité, nous rappellerons l'essor considérable des sociétés de télémarketing, des centres ou plates-formes d'appel (*'call center'*) qui, pour le dire de manière abrupte, ne tirent pas forcément la qualité vers le haut.

Revenir sur la définition du « terrain » et sur les phases d'interrogation et de codification permettra de s'interroger sur la culture du terrain et les fausses ressemblances entre questions ouvertes et questions fermées avant de conclure sur d'hypothétiques voies d'amélioration.

2. La production des données : rôle du terrain et des enquêteurs

Les rappels qui suivent constituent une première plongée dans la boîte noire de la fabrication d'une question ouverte et par extension d'un sondage⁴.

L'étape dite du « terrain » est fondamentale dans la production d'étude. Le terrain recoupe plusieurs acceptions terminologiques. Il désigne déjà une phase dans la production d'études : tout ce qui se rapporte au recueil de données, tout le travail de production qui n'est effectué ni par un directeur d'études ni par un chargé d'études relève de la compétence du terrain : **programmation informatique et mise en ligne du questionnaire, passation des interviews par les enquêteurs, saisie et codage des réponses correspondent ainsi à la phase de terrain.**

Par extension, le terrain renvoie également à **un lieu géographique**, à des locaux où se déroulent ces différentes phases de la production de données que les questions soient ouvertes ou fermées.

2. Informations sur les JADT depuis 1992 disponibles via www.cavi.univ-paris3.fr/lexicometrica/jadt/index.htm

3. L'enquêté est moins assisté ; il y a une rupture de rythme dans la passation, *etc.*

4. Le lecteur avisé trouvera toujours au cas par cas des différences d'un institut à l'autre. Ne sont données ici que des tendances générales, partagées par le plus grand nombre.

La population des enquêteurs est la plus importante en nombre sur le terrain. Il n'existe pas d'enquêteur type même si ces enquêteurs représentent **une population relativement instable**, constamment **renouvelée**, et **assez précaire**⁵. Il est d'ailleurs amusant de constater que c'est sur la population la plus instable et la plus précaire que repose la majeure partie de la qualité d'une enquête par sondage !⁶ Un enjeu majeur pour les responsables du terrain est d'ailleurs de fidéliser ses propres enquêteurs. Les perspectives d'évolution de carrière d'un enquêteur ne sont pas d'ordre salarial, mais plutôt en termes de missions différentes à remplir, missions présentées comme plus valorisantes : en encadrant à son tour d'autres enquêteurs ; en se détournant vers d'autres fonctions comme le codage de questions ouvertes, par exemple. Retenons que **les personnes qui codent les questions ouvertes aujourd'hui sont en partie celles qui assuraient la passation des questionnaires hier**.

3. L'étape essentielle de l'interrogation

Quelles sont les habitudes de fonctionnement, les modes de travail en vigueur sur les grands terrains d'études ? La quasi totalité des questionnaires téléphoniques est passée via un logiciel CATI (Computer assisted telephone interviewing). À chaque écran correspondent une question, les consignes et les réponses s'y rapportant. Chaque validation d'une réponse entraîne le passage à la question suivante, donc à l'écran suivant. Toutes les questions, *a fortiori* les questions ouvertes doivent être lues en mot à mot (ce mot à mot s'avère une règle érigée en dogme absolu !). Inutile de préciser que **la saisie intégrale est difficile à obtenir**. Le bon sens suffit pour comprendre que les enquêteurs n'ont pas de formation de sténotypiste et ne tapent pas sur un clavier d'ordinateur à la vitesse du débit oral. Une bonne astuce pour l'enquêteur, astuce fortement recommandée d'ailleurs, est de répéter à voix haute ce que le répondant déclare.

Les enquêteurs, s'ils sont à peu près libres de suivre ou de ne pas suivre cette règle, sont en revanche obligés de relancer l'interviewé.

Il existe plusieurs types de relances. On distingue principalement les relances sur le contenu et les relances extensives. **Les relances dites « sur le contenu » servent à creuser une réponse**, à l'enrichir. On se sert d'une partie des informations données par l'interviewé et on lui demande de préciser, d'expliquer ce qu'il vient de dire.

« Vous me dites 'joli', c'est-à-dire ? Pourquoi dites-vous 'joli' ? »

Les relances dites extensives servent à ajouter un autre aspect à la réponse donnée. *« Joli, d'accord. Quoi d'autre ? (...) Mais encore ? »* Ici, on privilégiera *« quoi d'autres ? »* *« Mais encore ? »* *« Mais aussi ? »* *« Et puis ? »* par rapport à *« c'est tout ? »* *« Pas d'autres raisons ? »* *« Vous ne voyez rien*

5. Ceci est plus flagrant sur les sites parisiens et dans les grandes villes. Des terrains basés par exemple à Saint-Brieuc ou Orléans échappent en partie à ces caractéristiques.

6. Voir notamment Duclos, D. et Meynaud H.Y. (1996).

d'autres ?». En effet, il ne faut pas laisser de refuges possibles à un enquêté, lui laisser l'occasion de se réfugier par une réponse réflexe dans la négative. Il convient d'ouvrir au maximum les relances.

4. Coder les réponses, une activité plus valorisante

Lorsque la phase d'interrogation est terminée, il s'agit de codifier les réponses ouvertes. « La codification est l'action qui permet de regrouper des verbatims portant sur un même sujet dans un même code réponse ».

Exemple 1 : si dans la liste des verbatims, on trouve « clarinette », « piano » et « violon », alors, on peut créer un code « Instruments de musique ». Dans le fichier des réponses toutes les personnes qui auront donné un instrument de musique particulier se retrouveront sous le même code générique « instrument de musique ».

Cette codification se fait manuellement⁷. Une personne du terrain propose une codification sur les réponses correspondant à, en général, 15 à 20% des individus, proposition de codification qu'il envoie à son responsable d'études qui doit la valider. Puis d'anciens enquêteurs codent proprement dit les réponses : le processus reste bien entendu itératif.

Pour mieux évaluer le matériau de première main sur lequel travaillent des enquêteurs, voici un contre exemple (cf. tableau 1) permettant de mettre en exergue les erreurs les plus courantes. Il s'agit d'un extrait d'un relevé à la question ouverte « *Pourquoi ?* » suivant la question fermée « *Resteriez-vous fidèle à EDF en cas d'ouverture du marché à la concurrence ?* ».

Nous ferons ici l'économie de remarques sur les déficiences en orthographe et en grammaire, ces erreurs pouvant être corrigées après relecture. De même, certaines abréviations (*tps* pour « temps », *pquoi* pour « pourquoi », *tjrs* pour « toujours », *dpd* pour « dépend », *quest* pour « question ») ainsi que certains sigles (EDF, ED F, E.D.F) peuvent là encore être corrigés *a posteriori*.

Quelques regrets peuvent être émis sur des consignes sans doute lacunaires : on remarque en effet que chaque répondant n'a proposé qu'une seule réponse⁸ ; visiblement, aucune relance n'a été effectuée. De l'information est irrémédiablement perdue.

Certaines erreurs sont à ce stade irréversibles.

Certains enquêteurs n'ont saisi qu'un seul mot : l'enquêté a-t-il répondu de manière si abrupte ? L'enquêteur a-t-il déjà procédé à sa propre codification ? D'autres enquêteurs ont retranscrit les réponses en tant que narrateurs. La première réponse illustre ce point avec un changement de pronom personnel. Enfin, la dernière réponse est quelque peu déroutante : la réponse, saisie à la

7. Elle se fait de plus en plus de manière automatique grâce à l'aide informatique de logiciels.

8. Graphiquement, une relance est généralement matérialisée par un « / » et la fin d'une réponse par « // » ou « CT » comme Commentaire terminé ou C'est tout suivant les habitudes.

LES MODALITÉS DE RECUEIL DES RÉPONSES LIBRES

TABLEAU 1. – Extrait d'un listing CATI de recueil d'une question ouverte

ELLE QUITTE EDF SI MOINS CHER AILLEURS ET A UNE AUSSI
BONNE QUALITE QU' EDF.
JE N' ADMET PAS LES MONOPOLES ET IL FAUDRAIT LE BRISER UN JOUR.
POUR CHANGER DE FOURNISSEURS
SANS COMMENTAIRE
CAR C' EST PLUS PRATIQUE DE RESTER CHEZ EDF
ELLE VERRAIT LES AVANTAGES DES AUTRES
SATISFAIS D'E D F
AUCUNE IDEE
IL VERRAIT LES AVANTAGES MAIS IL N'EN SAIS RIEN
SI IL Y A DES MEILLEURS PROPOSITIONS AILLEURS...
LE PROBLEME DES TARIFICATIONS ET CELUI DES SERVICES
EST PRIMORDIAL.
J'IRAI VOIR SE QUE PROPOSE LA CONCURRENCE
ETUDES DES OFFRES
IL FAIT TJRS CONFIANCE A E.D.F.
PRIX
CAR EDF EST UN BON RESEAU , FIABLE.
EDF N'A PAS DE CONCURRENT ET CE SERAI BIEN QU'ELLE EN AIT UN
CA DPD DES PRIX PROPOSES PAR LES CONCURRENTS
ELLE PRENDRA LE CONCURRENT QUI A LES TARIFS LES PLUS INTERESSANT.
FIABILITE
CELA DÉPENDRA DES PRIX ET DU SERVICE. CELUI D'EDF SERA
TOUT DE MÊME DIFFICILE ÉGALER
IL Y A UNE RÉELLE CONFIANCE DANS LES TECHNICIENS ED F.
JE NE VOIS PAS PQUOI JE CHANGERAIS
C'EST UNIQUEMENT UNE QUEST DE PRIX
JE SUIS FIDÉLE EDF DEPUIS LE TPS.
JE SUIS DANS L'ATTENTE DU MEILLEUR RAPORT QUALITÉ PRIX
ET C'EST NORMAL
JE NE PEUX AFFIRMER QUE JE RESTARAIS CHEZ EDF.
ON VERRA LES TARIFS
CONFIANT
EDF A UNE BONNE NOTORIÉTÉ ET EST TRÈS EFFICACE.
POURQUOI CHANGER ?
ON DOIT VOIR LES CONCURRENTS ET APRÈS ON POURRA CHOISIR
J'ANALYSERAI EN FONCTION DES BESOINS ET DES TARIFS
ET JE FERAIS UN CHOIX
IL EST INNONDE DE SONDAGE.

troisième personne du singulier, n'a rien à voir avec la question et ressemble davantage à un argument pour ne pas répondre à un questionnaire.

Retenons tout de même que cet exemple, bien que réel, n'est pas représentatif des relevés auxquels le chercheur est soumis à condition que celui-ci anticipe les questions de saisie.

Pour la codification proprement dite, quelques règles tacites existent. On veille notamment à conserver un poste « autre » inférieur à 5% ; on veille également à éviter les codes qui ne correspondraient qu'à 2 ou 3% des réponses. Avec cette méthode, on passe progressivement de verbatims (théoriquement) saisis intégralement à des codes fermés que l'on quantifie. On ne raisonne plus sur des mots, mais bien sur des regroupements de mots, sur des thèmes, bref sur des artefacts. Toute l'information purement lexicale est perdue. Les outils d'analyse textuelle sont dans ce cas inapplicables : on ne peut pas s'intéresser à la longueur de la phrase, au vocabulaire employé, à l'emploi de verbes modaux, *etc.* Nous obtenons donc des thèmes, des familles de mots et de phrases que l'on quantifie, que l'on hiérarchise.

De tout ce qui précède sur le terrain, l'interrogation, la saisie, la codification, il est possible de retenir un fait principal : **il existe une culture du terrain, culture dont les enquêteurs – codeurs – saisisseurs (il s'agit des mêmes individus) sont très imprégnés.**

Cette culture se retrouve dans le codage de questions ouvertes perçues trop souvent comme des questions qu'il faut fermer (*cf.* le plan de codification). Les enquêteurs lors du recueil même des verbatims sont tentés de déjà coder ce qui est dit. Si l'on n'y prend garde, et sans que l'enquêteur soit forcément blâmable, toute une description peut se réduire à un ou quelques mots, qui, s'ils sont cités en nombre suffisant, deviendront un item du même nom dans la codification ultérieure. Que d'informations perdues⁹ même si les réponses restent utilisables avec des visées davantage modestes.

5. Des premières pistes d'amélioration

Dans ce contexte, quelles voies d'améliorations peut-on proposer pour obtenir un corpus écrit riche et fidèle à ce qui a été énoncé ?

Nous écartérons d'emblée la **reconnaissance vocale** pour des raisons technologiques évidentes. Peut-être dans quelques années, dans quelques décennies ?

En revanche, pourquoi ne pas demander aux enquêteurs de noter les réponses au **stylo puis** de les **saisir** en fin de passation du questionnaire. Cela se pratique déjà un peu. Ce n'est pas non plus la solution idéale puisque cela comporte également des risques d'erreurs et cela allonge les délais (donc certainement les coûts ?).

9. La démonstration est encore plus nette dans le cas de questions dites semi-ouvertes.

LES MODALITÉS DE RECUEIL DES RÉPONSES LIBRES

Une solution plus fiable serait d'**enregistrer** les réponses données à des questions ouvertes et de les **saisir ensuite**. Nous observons toutefois les mêmes écueils de délais et de coûts supplémentaires. En outre, il n'est pas possible de faire l'économie de questions déontologiques : faut-il prévenir les enquêtes qu'ils sont enregistrés ? Quelle incidence cela a-t-il sur les taux de réponses ?

Bref, aucune solution idéale n'émerge. Il ressort de tout ceci qu'il faut déjà mettre en adéquation les objectifs et les moyens de l'étude. Si l'on se destine à utiliser des outils d'analyse textuelle, il convient de se poser toute une série de questions avant même de démarrer l'étude. Le **cahier des charges** doit être précis, y compris sur les aspects concernant les questions ouvertes, ce qui n'est quasiment jamais le cas ; les **consignes** écrites et orales sont importantes ; **encadrer** le terrain prestataire (lors du **briefing** ET lors d'**écoutes** nombreuses et régulières) sont autant d'aspects indispensables. Un **pilote** (c'est-à-dire la passation de quelques questionnaires avant le démarrage de l'étude afin de vérifier le bon enchaînement des questions et la bonne compréhension de celles-ci) permet également de profiter des remarques et de l'expérience des enquêteurs.

C'est pourquoi, au GRETS (EDF R&D), nous essayons de replacer toutes ces questions dans une démarche plus large, dans une réflexion autour de la qualité de terrain. Nous sommes conscients que des efforts sont faits par exemple avec l'introduction de normes ISO, normes et procédures toutefois insuffisantes et pas assez ad hoc. Nous avons donc des exigences particulières sur les questions ouvertes et nous essayons d'accompagner les instituts sur ces questions de qualité. Nous ne pouvons qu'encourager les efforts considérables déjà entrepris par plusieurs instituts de renom. Les investissements en matériel et en personnels d'encadrement se sont par exemple accentués ces dernières années. **Mais en fin de compte, la puissance du traitement sera toujours fonction des choix initiaux et de la qualité du recueil.**

Bibliographie

- ADETEM (2001), *La qualité des enquêtes quantitatives*, Revue française du marketing, cahier 185, pp. 3-72.
- ASU (Association pour la Statistique et ses utilisateurs) (1992). *La qualité de l'information dans les enquêtes*, Paris, Dunod.
- BLONDIAUX L. (1998), *La fabrique de l'opinion, une histoire sociale des sondages*, Paris, Seuil.
- DUCLOS D., MEYNAUD H.Y. (1996), *Les sondages d'opinion*, Paris, La Découverte.
- MICHALOWSKA A. (2002), « Ad'hoc quanti, un terrain toujours plus fiable », *Marketing Magazine*, 69, pp. 52-59.
- MAYER N. (2002), *La démocratie à l'épreuve, une nouvelle approche de l'opinion des Français*, Paris, Presses de Sciences Po, collection académique.
- RAY D. (2001), *Mesurer et développer la satisfaction clients*, Paris, Editions d'Organisation.
- SALOMON A.C., BENOIT A.M. (1999), *Les enquêtes d'opinion*, Voiron, La lettre du cadre territorial, Dossiers d'experts.

LES MODALITÉS DE RECUEIL DES RÉPONSES LIBRES

Pour en savoir plus, de rares émissions télévisées ont abordé le sujet du recueil des données de sondage ; en outre, quelques pages de sites web donnent des premières pistes techniques ou de déontologie sur ce recueil de données.

La Sept, Arte (1994). Bonjour, c'est pour un sondage!, soirée thématique.

France 2 (Mars 2002). Questions pour un sondage, Envoyé Spécial (Grenier, B., Cardoze, J.).

http://www.ipsos.fr/CanalIpsos/cnl_static.content.asp?rubId=35

http://www.sofres.fr/sofres/reso_enq.htm#activites

<http://www.esomar.nl>