

3

Techniques de constitution des données

LES DONNÉES INVOQUÉES

- 1 De la participation observante à l'observation participante
- 2 L'observation systématique
- 3 L'observation électronique
- 4 Le recueil des documents d'archives comme documents invoqués et les biographies

Les problèmes de constitution des données vont varier selon que les données sont invoquées, suscitées ou provoquées. Dans le cas des données invoquées, les deux problèmes majeurs concernent l'accessibilité des sources et la critique du témoignage (voir plus loin : la critique historique des sources). En effet, par rapport au problème envisagé, il convient de se demander si l'inventaire des sources de documentation est complet, de veiller à leur accès le plus direct et d'examiner la qualité du matériel découvert. Ce matériel est-il de première main, est-il le produit d'un observateur direct des événements étudiés, semble-t-il constituer un rapport sincère et objectif, quel contexte et quels partis pris ont pu orienter sa rédaction, l'enjoliver ou le réduire ? Dispose-t-on de l'intégralité du document, est-il authentique, est-il cohérent et déchiffrable ? Autant de questions à envisager pour valider le témoignage d'un matériel invoqué. D'autres questions s'y ajoutent dans le cas de la recherche-action intégrale et de l'observation participante où la qualité des données dépend aussi de la qualité de la prise de notes par l'observateur : quel était le contexte de l'observation, le rôle d'observateur était-il déclaré ou caché et, dès lors, quelles confidences et conduites suscitait-il, quelle était la familiarité de l'observateur avec le terrain, avec la culture du milieu, etc. ?

1 De la participation observante à l'observation participante¹

Les problèmes de l'observation, qu'elle soit participante ou systématique, peuvent s'aborder à partir d'une analogie avec le témoignage en justice. Au tribunal, le juge demande au témoin de dire la vérité, toute la vérité et rien que la vérité. Il demande au témoin de lui apporter le témoignage des faits, se réservant le droit de porter le jugement, après que les avocats auront passé une bonne partie de leur interrogatoire à essayer de vérifier, sinon de miner, la crédibilité du témoin. En recherche, l'observateur doit rencontrer cette exigence de vérité imposée au témoin par le juge. Mais quelques problèmes s'y ajoutent. D'abord, l'observateur

1 Pour plus de détail, voir :

Laperrière A., *L'observation directe*. Ch.10 dans Gauthier B., *Recherche sociale*, Sillery : PUQ, 1984, p. 227-246.

Chauchat H., *L'enquête en psychosociologie*. Paris : PUF, 1985. Plus particulièrement les chapitres III, *L'observation* (p. 87-98) et IV, *L'observation participante* (p. 99-120).

Postic M. et J.M. De Ketele, *Observer les situations éducatives*, Paris : PUF, 1988. Particulièrement, partie 2, le Ch.III, *L'observation expérientielle* (p. 129-174) et la section « observation en cours de processus » traitant de l'observation interactive (p. 222-226).

n'est pas le témoin accidentel des faits; c'est lui qui se rend sur le terrain pour chercher des faits à observer et il se trouve donc être un témoin intentionnel. La difficulté réside alors à mettre entre parenthèses son intention, pour n'être qu'un témoin capable de rapporter les faits, tous, et entiers, mais rien que les faits. Ensuite, le chercheur ne sera pas confronté comme le témoin à deux avocats qui l'interrogent et le confrontent pour vérifier la crédibilité de son témoignage : le chercheur devra lui-même assurer la crédibilité de ce qu'il rapporte en utilisant plusieurs techniques, dont le recoupement de ses observations, de ses sources d'information et de ses... interprétations! Car pour amplifier la difficulté, le chercheur doit aussi être juge de la valeur et de la signification apportées par les faits d'observation qu'il a lui-même produits. Or, la communauté des chercheurs attend bien de lui que non seulement ses données soient crédibles, mais aussi que ses interprétations soient impartiales, qu'elles ne soient pas contaminées par ses préjugés, ses espérances, ses croyances.

Le jeu de mots du titre de cette section veut souligner que la principale variation de cette technique tient au rôle du chercheur observateur. L'observation participante (ou expérientielle) consiste à se rendre sur le terrain où le problème trouve son contexte, à s'immerger dans la situation et à participer aux activités quotidiennes des acteurs, selon différents degrés d'implication. Cette immersion engagée a pour objectif de pouvoir observer *in situ* et de décrire l'ensemble de la situation problématique à partir de la perception impliquée, vécue, qu'en a le chercheur en relation avec les acteurs. La question méthodologique tient donc à la relation entre, d'une part, l'immersion et l'implication du chercheur, c'est-à-dire à sa participation, et, d'autre part, sa capacité d'effectuer une observation vraisemblable et pertinente.

À l'une des extrémités, le chercheur est essentiellement participant, engagé le plus complètement possible et n'ayant, à la limite, qu'à s'observer lui-même, car il s'identifie aux autres participants de la situation. Alors, on peut se demander dans quelle mesure il est capable de s'extraire, de prendre un recul suffisant pour observer, c'est-à-dire pour fournir un reflet médiatisé, instrumenté, de ce qu'il vit. À l'autre extrême, bien que sur le terrain et participant aux activités, le chercheur est surtout observateur. Dès lors, se pose la question de savoir si ce qu'on lui donne à observer, à ce titre, est bien ce qui se passerait en dehors de sa présence. Le chercheur doit donc soigneusement choisir son degré d'implication en fonction de ce qu'il privilégie. L'implication maximale favorise la pertinence : on observe bien le problème à observer. Mais elle diminue la vraisemblance et l'objectivité :

les préconceptions du chercheur interfèrent avec l'observation par manque de recul. À l'inverse, par suite de la mise à distance, les participants peuvent montrer au chercheur autre chose que ce qu'il souhaite voir : il y a donc diminution de la pertinence. En contrepartie, l'observateur peut mieux contrôler ses biais, ce qui favorise une augmentation de l'objectivité, et observer plusieurs acteurs à la fois dans une situation, ce qui peut accroître la vraisemblance.

L'implication sur le terrain a, par ailleurs, une incidence immédiate sur la possibilité matérielle d'enregistrer les observations. Intégralement immergé, quand le chercheur va-t-il pouvoir enregistrer ce qu'il a perçu pour en faire une observation ? Si le délai est trop grand, ce qu'il enregistre ne correspond plus à sa perception immédiate mais à un souvenir. Or, le souvenir sélectionne nécessairement les événements les plus émotionnellement chargés ! L'observation dans ce cas apporte-t-elle plus d'informations valides que des entrevues ? S'il veut enregistrer des perceptions encore fraîches, le chercheur doit pouvoir se retirer, ne fut-ce que quelques minutes mais régulièrement, pour procéder à ses enregistrements. Or, ce retrait de la participation, non seulement neutralise momentanément son implication, ce qu'il ne pourra corriger qu'en assumant toutes les décisions et les actions prises par les autres pendant son retrait. Mais ce retrait, qui doit être effectif si le chercheur veut cacher qu'il enregistre afin de garder, aux yeux des autres, son statut de participant, l'empêche de percevoir ce qui se passe pendant que son attention est prise par son enregistrement.

D'autres problèmes sont aussi à signaler.

Si, en cas d'implication maximale, le chercheur est invité par les acteurs à collaborer à la solution des problèmes qu'ils vivent (cas de recherche-action avec contrat ouvert²), chacun devient «co-chercheur» et «co-acteur» pour l'autre. L'implication du «co-chercheur» comme «co-acteur» pose au moins deux questions : celle du rapport entre l'action et la recherche et celle de la propriété des résultats. Qui, dans ce cas, est l'auteur ? Qui a fait, qui a cherché et qui peut publier quoi au nom de qui ?

Un autre problème réside dans le rapport entre la disponibilité du terrain (son ouverture au chercheur) et la visibilité des problèmes que le chercheur voudrait analyser. Pour caricaturer le dilemme, disons qu'une école où il n'y a pas de problèmes évidents, dont les membres ne se sentent

2 Voir le texte de A. Morin : Réflexions sur la recherche-action à partir d'une expérience d'animation. *Prospectives*, XX, 1-2, 1984, p. 49-54.

pas sous tension ni menacés, sera probablement plus ouverte à la collaboration avec des chercheurs qu'une école en crise où chacun tient à se réfugier dans son petit domaine. Par ailleurs, la déontologie interdit au chercheur de s'intégrer dans une école qui va bien pour y susciter des problèmes qu'il pourrait ensuite observer. Le dilemme disponibilité-visibilité conduit à un autre, celui de la négociation plus ou moins ouverte, plus ou moins dissimulée de l'entrée du chercheur. Le rôle que les acteurs du terrain acceptent de reconnaître au chercheur et que le chercheur leur manifeste aura, comme son niveau d'implication, une incidence sur la pertinence et la vraisemblance des données.

Une des difficultés supplémentaires que devra surmonter le chercheur qui utilise l'observation participante résulte de sa richesse : il va rapidement accumuler une très grande quantité de matériel à traiter. De plus, ce matériel ne sera pas homogène : données topologiques, observations d'actions, recueil de récits, enregistrement de discussions, d'échanges de biens matériels, d'échanges symboliques, etc. D'où des problèmes de codage, de classification, de systématisation qui rendent difficiles certains types d'analyse et de traitement.

Par suite de ces difficultés et de la subjectivité dont est nécessairement entachée ce genre d'observation, les «objectivistes» préfèrent travailler par observation systématique. Notons, avant de l'aborder, que rien n'interdit à un chercheur de pratiquer, parmi d'autres méthodes de constitution des données, l'observation participante et l'observation systématique. Pour peu que les phases de collecte soient décalées ou que la recherche se fasse en équipe, aucune méthode n'est exclusive des autres.

2 L'observation systématique³

Dans son introduction à *La méthodologie de l'observation* (page 7), De Ketele écrit :

3 Pour plus de détail, voir :

Simon A. et G. Boyer, *Mirrors of Behavior*. Philadelphie, CIN, 1970.

Dusseault G., et coll., *L'analyse de l'enseignement*. PUQ, 1973.

Strayer F.F. et R. Gauthier, L'approche éthologique de l'observation du comportement, *Apprentissage et socialisation*, 1982, 5, 1, p. 12-23 (numéro spécial sur l'éthologie humaine).

Gauthier R., Techniques d'enregistrement : du papier crayon au microprocesseur, *Apprentissage et socialisation*, 1982, 5, 1, p. 48-58.

Chauchat H., *L'enquête en psychosociologie*, Paris : PUF, 1985. Plus particulièrement le chapitre 5 : L'observation directe, p. 121-142. .../...

Les phénomènes les plus banals sont les plus méconnus et les plus galvaudés. L'homme de la rue croit bien connaître ce qu'est l'observation, au contraire de l'homme de science qui sait combien ce processus si fondamental est en fait bien complexe ... Il est curieux de constater que la plupart des spécialistes qui ont largement utilisé l'observation ou ont écrit à son propos n'aient pas tenté de donner une définition de l'observation !

Après cette citation, on notera seulement que le principal problème de cette technique provient de ce que bien des jeunes chercheurs croient savoir naturellement observer, qu'ils s'y lancent sans préparation, et que leurs résultats sont dramatiquement pauvres, peu fiables, sans grande valeur. Les défauts les plus courants de ces prétendues observations viennent de leur contamination par les préjugés, les émotions, l'anecdote. Cela aboutit, non pas à des rapports pertinents, complets et factuels, mais à des jugements et à des opinions basés sur les parties sensationnelles des événements.

Paradoxalement, c'est en voulant améliorer la fiabilité, la fidélité et la validité des observations que les chercheurs ont découvert la difficulté d'obtenir des données valables à partir de l'observation. Pour surmonter cette difficulté, on construit des grilles d'observations, c'est-à-dire des listes d'éléments à observer dont les définitions sont les plus précises et les plus opératoires possibles. Afin d'établir leur validité, les observations faites à partir de ces définitions, sont comparées aux inscriptions obtenues grâce à d'autres systèmes inscripteurs conçus selon les mêmes concepts que la grille. On compare aussi les observations obtenues par plusieurs personnes utilisant la même grille pour décrire en même temps le même événement : cela dans l'espoir d'établir la fidélité de ces grilles.

L'observation systématique tente d'établir sa crédibilité en introduisant, entre le chercheur et les événements, un intermédiaire technique (une méthode) dans le but de produire le recul nécessaire à l'objectivité. De plus, cette technique peut être utilisée par plusieurs observateurs, ou par le même observateur dans plusieurs situations, et cela de manière systématique, ce qui permet de garantir la valeur des faits rapportés. L'introduction de cet intermédiaire méthodique est même, aux yeux de certains chercheurs, censée produire elle-même les données, le chercheur n'agissant plus que comme opérateur, ce qui donnerait aux données une fiabilité instrumentale certaine. En effet, utilisant une grille,

.../... Postic M., et J.M. De Ketele, *Observer les situations éducatives*, Paris : PUF, 1988. Particulièrement, partie 2, le Ch. II, L'observation systématique (p. 105-128).

De Ketele J.M., *Méthodologie de l'observation*. Bruxelles : De Boeck-Wesmael, 1988.

l'observateur ne peut voir que les faits prévus dans la grille, et tous les faits qu'il observe ne peuvent être rapportés que dans les termes de la grille. Dans la mesure où la grille a été élaborée par un autre chercheur que l'observateur, et dans la mesure où ce dernier fonctionne comme un système technique et intelligent, le résultat de l'observation devrait correspondre aux faits, il devrait être une simple inscription.

Mais, l'observateur, même armé d'une grille, même encadré par un système d'observation, est toujours un témoin intentionnel : il est bien des cas où l'observateur est le créateur de la grille, et quand cela n'est pas, il n'est pas socialement naïf dans son rapport au créateur ou au chercheur. De plus, l'observation systématique est maintenue dans la classe des techniques produisant des données invoquées, parce que les comportements sont produits sans contrainte par les sujets observés et parce que les chercheurs, malgré la technique, ont le problème de faire correspondre les données observées avec les catégories de la grille.

L'existence de grilles (*Mirrors of Behavior* en présente une centaine) prétendent valider et être fidèles à la suite des procédures de leur construction, ne garantissent pas encore que leur utilisation par les observateurs maintienne ces qualités. En effet, les observateurs peuvent interpréter différemment une même définition, si bien qu'il devient important d'entraîner les observateurs à l'utilisation de la grille envisagée. Même alors, si les observateurs s'entendent sur ce que signifie la définition des comportements à observer, c'est-à-dire sur la correspondance entre une définition et un ensemble de comportements, encore faut-il que, dans l'action, ils fassent correspondre tel comportement particulier à une même définition. Cela n'est pas évident, surtout quand le comportement perçu ne fait pas partie de la gamme des comportements strictement associés à la liste des définitions de la grille : il y a des comportements ambigus, imprévus, qui se perçoivent différemment selon le contexte, la culture, etc. De plus, une observation n'est jamais, comme toute mesure, qu'une approximation de l'événement observé. Il faudrait donc collecter plusieurs observations du même événement, puis les comparer pour évaluer la précision de la description. À cette fin, l'observation devrait pouvoir se faire au même instant par plusieurs observateurs, ce qui pose le problème de leur synchronisation. La technique du bruiteur relié à un chronomètre ne garantit pas encore la simultanéité des observations, car l'un peut être distrait, le second peut enregistrer l'événement qui se produit juste après le top, alors qu'un troisième finit par anticiper le top et enregistre l'événement qui se produit pendant celui-ci. En plus des problèmes de construction des grilles (précises, opératoires, sans ambiguïté), se pose donc celui de la formation des observateurs.

Une manière de réduire le problème consiste à construire la grille tout en formant les observateurs, ce qui produirait un assez grand accord sur la traduction des définitions en comportements et sur le classement des comportements dans les catégories de définition. C'est la stratégie suivie par ceux qui utilisent l'observation éthologique en éducation ou en psychologie (observation des jeunes enfants, des malades, etc.). Leur perspective méthodologique consiste à réduire l'observation à celle de gestes (moteurs) dont on dresse la liste par observation naïve, informelle dans une situation typique de celles qui serviront pour la constitution des données. Une fois tous les « patrons moteurs » (les gestes) relevés pour le contexte visé, on les organise en une liste descriptive, soumise ensuite à une vérification sur le terrain. Celle-ci porte sur la précision des définitions et sur leur pertinence pour le contexte de l'observation. Enfin, on réduit le nombre de catégories afin qu'elles soient manipulables (maximum 15 à 20) de manière automatique (sans demander des temps de réflexion qui introduiraient des périodes sans observations ou des confusions), et l'on entraîne les observateurs à la manipulation des claviers qui permettront l'enregistrement des données. Ces étapes franchies, on vérifie si la grille est fonctionnelle et si les observateurs sont bien entraînés, en testant la fidélité inter-observateurs et intra-observateurs, et en reproduisant ces tests sur divers échantillons.

Pour intéressante que puisse être cette stratégie qui jumelle la formation des chercheurs avec la préparation de la grille, elle pose encore des problèmes. D'abord, elle comporte un présupposé réductionniste du psychique au biologique : elle observe des gestes moteurs et consiste essentiellement en une analyse biologique des comportements. En conséquence, même si, par rapport aux grilles classiques, elle gagne en précision dans l'enregistrement des inscriptions, elle perd en signification, car une fois les inscriptions obtenues, il faut encore les interpréter, leur donner du sens. C'est alors que le problème du classement des comportements dans des catégories de significations sociales et culturelles devient difficile. Or, on ne peut procéder aux analyses des séquences des comportements si on ne tient pas compte de leur signification : les comportements humains sont psychosociaux et n'ont de sens que dans l'intersubjectivité. Pour savoir si tel geste du bébé est une réponse au gargouillis de sa mère ou si c'est un réflexe de détente, le chercheur doit réintroduire sa subjectivité. Or, la mise à distance de cette subjectivité, à la suite de la réduction biologique et de la procédure de constitution de la grille éthologique, est telle que les débats interprétatifs ne trouvent pas

souvent de solution. Entre eux, les membres d'une équipe qui ont participé à la recherche s'entendent et semblent convaincus du sens de leurs observations, mais les chercheurs extérieurs à la recherche ont souvent bien de la peine à admettre leur interprétation et à se laisser convaincre. La construction éthologique des grilles produit des observations dont la validité est interne à l'équipe de recherche, mais elle n'est guère transférable.

Enfin, l'observation éthologique pose aussi le problème du nombre d'éléments différents observés. Pour être quasi automatisé, ce nombre est réduit à une quinzaine, ce qui est peu devant la complexité des situations à observer. Son utilisation semble donc limitée. De plus, l'analyse des comportements humains ne peut isoler une conduite de son contexte. Le sens du geste d'un des partenaires d'une interaction ne peut se comprendre qu'en tenant compte du sens qu'il attribue aux gestes des autres. Cela pose, comme l'a montré Watzlawick⁴, le problème de l'endroit où il faut placer les bornes qui indiquent où commence et où finit une interaction, problème essentiel pour interpréter un comportement. Or, ce problème n'est pas résolu par l'éthologie; elle le néglige.

L'introduction d'une méthodologie systématique, soit la construction et l'utilisation de grilles, ne donne donc pas aux données d'observation le statut d'inscriptions strictement représentatives des faits à observer. De plus, les limites inhérentes à leur emploi par l'observateur humain (c.-à-d. le nombre restreint de catégories mémorisables et automatiquement utilisables) en restreignent l'usage à l'observation simultanée d'une partie très réduite de l'ensemble des événements qui se produisent en un instant dans un échange comme une séquence pédagogique. L'observation systématique ne peut être efficace que pour constituer des données nombreuses sur des problèmes limités où n'intervient qu'un nombre réduit de facteurs. De plus, leur usage impose la présence de «corps étrangers» à la situation, les observateurs avec leurs grilles ne pouvant pas se fondre dans le paysage humain, puisqu'ils ne peuvent être, dans ce cas, des participants. L'observation systématique, qu'elle soit d'inspiration éthologique ou non, n'est jamais écologique : elle perturbe toujours les rapports entre les partenaires «naturels» de la situation. Elle introduit un artifice dont il faudra tenir compte lors de l'interprétation; même lorsque l'observation est indirecte grâce à l'électronique.

4 Watzlawick P., Helmick-Beavin J. et D.D. Jackson, *Une logique de la communication*. Paris, Seuil, 1972.

3 L'observation électronique

On ne peut, en cette fin de XX^e siècle, parler d'observation sans envisager les possibilités de l'observation électronique : microphones, caméras et caméscopes. Outils pratiques, aides précieuses devenues indispensables mais qui ne sont pas sans poser de problèmes. Au début, il semble qu'on y ait surtout pensé pour décharger le chercheur de l'enregistrement manuel des faits à noter ou pour permettre une observation à distance. Jointes à l'informatique, ces outils pourront, espère-t-on parfois, rendre l'inscription des faits aussi objective qu'elle le serait dans un laboratoire... de physique. Mais ils ne sont pas sans problèmes.

D'abord, à moins de procéder comme dans les films d'espionnage en cachant tout, en dissimulant le matériel et en contrôlant l'enregistrement à distance - ce qui coûte encore cher -, l'introduction des appareils contamine la situation. Le caméscope et le magnétophone ne passent pas inaperçus dans une classe, d'autant plus qu'ils ne fonctionnent pas seuls : ils sont accompagnés d'un opérateur. À moins d'être dissimulée dans un plafond et télécommandée, la caméra devient la spectatrice privilégiée ou, au contraire, l'intruse crainte. Certains sujets manifestent même des réactions de défense : les uns ne supportent pas l'enregistrement (refus de l'image, du reflet, refus de se reconnaître), d'autres au contraire s'exhibent, caricaturent, font le clown. Lorsqu'on dissimule la caméra, le bruit des moteurs (faisant tourner les objectifs et les rotules du socle) attire l'attention et signale qu'il y a un espion. Lorsqu'elle n'est pas dissimulée, les mouvements de l'opérateur constituent des stimuli supplémentaires. Même un simple micro modifie la situation : on parle fort, distinctement pour le micro, et l'on doit s'interrompre pour que l'opérateur puisse retourner la cassette ou la changer. Donner aux élèves, comme cela fut suggéré par plus d'un, l'occasion de manipuler ces appareils pour leur permettre de les apprivoiser ne fait que réduire l'impact de leur introduction. L'utilisation d'une aide électronique pour l'enregistrement modifie donc toujours la situation à observer (cela vaut aussi pour les entrevues).

De plus, les procédés techniques ne fournissent pas plus que l'observateur humain une reproduction fidèle. L'image et le son sont déformés par les capteurs (les optiques et les filtres) et par les amplificateurs qui ne les restituent qu'en partie (perte de certaines fréquences, des couleurs, du contraste, etc.). En outre, ces procédés ne prennent pas tout : ils sélectionnent. Le champ d'une caméra est plus réduit que le champ visuel d'un humain, ses capacités d'adaptation et de mouvement sont plus lentes

et sont focalisées, ce qui, comparativement à l'humain, réduit de beaucoup la quantité et la qualité de matériel enregistré. Les micros aussi ont leurs problèmes de déformation et de sélection : manque de discrimination entre les voix qui se superposent, prise de son directionnelle, etc. De plus, ils sont très sensibles aux parasites (sonnerie du téléphone, bruit d'un avion, d'une voiture... ou même parasites techniques, comme le bruit du moteur du magnétophone) qui, dans certains cas, peuvent masquer et rendre inaudible ce qui devait être enregistré.

Enfin, ces moyens techniques sont sujets aux pannes et aux erreurs de manipulation. Quel qu'en soit le responsable, la frustration est la même lorsque l'on constate, rentré au laboratoire, que la partie la plus intéressante des événements, que l'on a bien vue, n'a pas été enregistrée ou qu'elle est inaudible, invisible.

Ces techniques d'observation électroniques ont, malgré leurs défauts, des avantages certains. Lorsqu'on peut les utiliser en limitant raisonnablement leur contamination de la situation, elles permettent des reprises de l'observation et donc des contrôles. Il est étonnant de voir combien de choses nouvelles on remarque à chaque visionnement d'un même enregistrement. De ce fait, ces techniques deviennent des outils majeurs pour la formation, l'entraînement et le contrôle des observateurs. De plus, elles captent la trace du climat des événements, elles en restituent le contexte, grâce au caractère analogique de l'enregistrement qu'elles permettent, ce que ne peut faire une inscription manuelle. En ce qui concerne l'observation proprement dite, si elles soulagent l'observateur d'une tension, celui-ci ne peut totalement se fier à elles. Il doit toujours continuer sa prise de note manuscrite pour les éléments essentiels de l'observation. En effet, le retour sur un enregistrement et son déchiffrement sont des opérations très onéreuses, et le chercheur se rendra rapidement compte qu'elles alourdissent son travail. Lorsque c'est possible, il est plus efficace d'enregistrer manuellement un maximum d'observations, et d'utiliser la trace électronique pour compléter ce qui manquerait et pour retrouver le climat et le contexte de la situation. En les considérant comme des compléments et des outils de contrôle, le chercheur se garantit de ne pas perdre l'information qu'il cherchait et, qu'en vertu de la loi de la frustration universelle, ces techniques auront justement manqué.

4 Le recueil des documents d'archives comme documents invoqués et les biographies⁵

Il n'est pas nécessaire de faire de la recherche sur un lointain passé pour être confronté avec les problèmes de la recherche historique. La compréhension d'une situation qui se développe exige souvent le recours à l'analyse de documents quasi contemporains, qui viennent d'être produits et qui l'ont été dans le but de servir de références, de ressources, d'indicateurs pour ceux qui, peu après leur production, pourraient ou devraient s'en servir dans l'action. Ainsi en est-il des statistiques nationales, des bulletins publiés par les ministères et les institutions, des lois et des décrets. Des archives de toutes sortes empiètent les comptes-rendus des assemblées, des conseils, des comités ainsi que tous les documents, mémoires ou simples mémos qui y sont déposés. Les archives contiennent aussi des documents comme des registres de présence, les livres comptables avec les listes de paie, les originaux des bulletins, les listes de graduation et d'autres documents attestant de la vie ou de la production des institutions : journaux de classe des élèves ou des professeurs, cahiers d'examen des élèves, production pour des concours, etc. Tous ces documents contiennent des traces de l'activité des membres de ces institutions et fournissent donc un matériel, ou des inscriptions, que l'on peut invoquer comme témoin des événements que l'on souhaite étudier.

L'examen de tels documents invoqués pour chercher à comprendre les origines et le contexte d'une situation pose immédiatement deux problèmes à l'égard de leur statut.

Le premier problème est lié à la diversité des conditions ou des contextes dans lesquels ces documents ont été produits. Chaque document résulte d'une intention de son auteur et comme tel, s'il est le gardien d'une

5 Pour plus de détail, voir :

Léon A., *Introduction à l'histoire des faits éducatifs*. Paris : PUF, 1980.

Best J.W., *Research in Education*. Englewoods Cliffs, N.J.: Prentice-Hall Inc., 1959. Particulièrement le chapitre 10, Historical Research, p. 340-360.

Bourdé G., et H. Martin, *Les écoles historiques*. Paris : Seuil, 1983.

Burgess R. (Éd), *Strategies of educational research*. Philadelphie, P.A. : The Falmer Press, Taylor & Francis Inc., 1985. Particulièrement le chapitre 6 de Alison Andrew, In Pursuit of the Past : Some Problems in the Collection, Analysis and Use of Historical Documentary Evidence, p. 153 -178, et le chapitre 7 de June Purvis, Reflections upon Doing Historical Documentary Research from a Feminist Perspective, p.179-205.

Simmel G., *Les problèmes de la philosophie de l'histoire*. Paris : PUF, 1984.

trace des événements, il n'est pas le témoin de l'ensemble des traces de l'événement. Chaque auteur, même contemporain de l'événement, ne parle que de son point de vue, c'est-à-dire de la position qu'il occupe, compte tenu de la position occupée par ses interlocuteurs. Dès lors, le recours à des documents d'archives implique, quasi nécessairement, que chaque document soit confronté à des documents d'un autre type, afin de compléter l'information et de construire progressivement le noeud de l'événement. Le crédit accordé à cette reconstruction sera relatif aux croisements des sources et des types de documents que l'on aura pu faire. À cette fin, on distingue au moins cinq types de matériaux d'archives :

- 1. *Les matériaux oraux* : en général, des entrevues d'administrateurs, de professeurs, d'employés, de parents, d'élèves, etc. (Dans certains cas, ces entrevues sont construites pour la recherche et pendant la recherche; il s'agit alors de données suscitées telles qu'envisagées plus loin sous le titre histoire de vie ou récits de pratiques. Elles rencontrent alors les problèmes de statut propre aux données suscitées.)
- 2. *Les matériaux officiels* tels les lois, statistiques, règlements, programmes, annuaires, horaires, bulletins, journaux, etc.;
- 3. *Les écrits non publiés*, tels les rapports internes, notes de service, correspondances, etc.;
- 4. *Les reliques* comme les bâtiments, les plans, les peintures, les photographies, les équipements, les manuels, les examens, les travaux écrits, etc.;
- 5. *Les récits indirects* comme les romans, les fictions d'époque, les films, etc.

Le deuxième problème est typique des travaux historiques : il s'agit de la critique du témoignage apporté par les documents invoqués. D'abord, on tente d'identifier la source du document, de préciser sa date, le lieu de sa production et son contexte. Ensuite, on pose la question de l'authenticité et de l'intégrité. Est-il authentique, a-t-on affaire à l'original, à une copie ou à un faux construit pour égarer (cela n'arrive pas que dans les romans), pour détourner l'attention ou pour se l'accaparer, a-t-il été produit par celui qui le signe ou par un de ses subalternes ? Est-il complet et intégral, ou bien a-t-il été censuré, en a-t-on modifié certaines parties, a-t-il été accidentellement ou volontairement amputé ? Le chercheur doit aussi évaluer la signification à attribuer au document à partir d'au moins cinq questions :

- 1° L'étendue du témoignage et le point de vue de l'auteur (c.-à-d. quelle perspective a-t-il sur les événements qu'il relate, a-t-il été témoin impliqué dans ce qu'il relate ?);
- 2° L'objectivité du témoignage (c.-à-d. l'effet des biais de l'auteur, de ses options philosophiques et politiques, de ses appartenances sociales et culturelles sur son témoignage, soit les déformations, les sélections et les négations induites par ces biais);
- 3° La sincérité du témoignage (c.-à-d. cache-t-il des choses, dit-il tout ce qu'il sait);
- 4° La cohérence du témoignage (c.-à-d. se contredit-il, ou les apparentes contradictions sont-elles dues à la complexité des événements; encore qu'une trop grande linéarité pourrait n'être qu'une construction de l'auteur qui enlève du témoignage ce qui lui apparaît dissonant ?)
- 5° La clarté du témoignage (c.-à-d. la possibilité de le déchiffrer, de le comprendre sans ambiguïté, ce qui impliquerait que le chercheur doive compléter la compréhension, mais à partir de quoi...?)

Ces deux problèmes, dus aux matériaux, se reflètent dans le travail du chercheur lui-même, qui devra témoigner d'un minimum de probité, de courage et de lucidité dans le dépouillement des matériaux. Ainsi, il doit éviter de combler les trous, les manques du matériel, en y ajoutant des choses de son propre cru et en les faisant passer pour originales. Il doit aussi accepter d'interpréter tous les matériaux, il ne peut pas éliminer ceux qui ne collent pas avec ses connaissances antérieures, avec la lecture qu'il a faite des autres documents collectés. Il doit accepter les évidences produites par certains documents, même si elles ne rencontrent pas ses hypothèses, ses anticipations, ses préconceptions, ou si elles choquent ses appartenances et ses options.

Une attention particulière doit être apportée à deux types de documents d'archives parfois utilisés comme témoignages invoqués : les biographies et les rapports de recherche. En effet, les biographies et les autobiographies sont le plus souvent apologétiques : leur auteur ne déteste pas son sujet, que ce soit lui-même ou un de ses héros. Même s'il est honnête et sincère, la construction de la biographie sera marquée de la nécessité d'en produire une image intéressante pour la postérité. Les rapports de recherche sont aussi marqués par le risque de l'auto-justification : le chercheur qui l'écrit est toujours soucieux de sa crédibilité. Dès lors, la rédaction contient parfois des rationalisations *post hoc* sur les choix méthodologiques et interprétatifs, des négligences dans le récit des

manipulations, des oublis qui permettent au rapport de maintenir sa cohérence. De plus, alors que la recherche est un processus souvent hésitant, plein de faux pas, de détours et de marches en arrière, parfois complètement cyclique, la rédaction du rapport linéarise le tout et donne l'image d'une démarche rectiligne, assurée, et tendue logiquement et stratégiquement vers un objectif.

Ces nécessités rédactionnelles compliquent la critique historique du témoignage apporté par les biographies et les rapports de recherche, car il est bien difficile, à moins d'avoir été un observateur direct, ou d'en avoir un sous la main (le héros de la biographie lui-même ou le rédacteur du rapport), de départager les parties du récit qui sont enjolivées, ou qui sont déplacées, de celles qui sont restées brutes. Ces documents doivent donc, au mieux et dans la majorité des cas, être considérés comme n'apportant pas un témoignage de première main, et comme nécessairement biaisés par le contexte et les nécessités de leur production.

LES DONNÉES SUSCITÉES

- 1 Les problèmes généraux
 - 2 L'histoire vivante : les histoires de vie et les récits de pratique
 - 3 L'entrevue (libre, semi-structurée et structurée)
 - 4 L'entrevue clinique
 - 5 L'examen systématique
 - 6 La sélection des sources et l'échantillonnage
-

1 Les problèmes généraux

Les données issues d'interactions entre le chercheur et l'informateur posent différents problèmes, dont celui de la maîtrise de la technique de l'entrevue. Il importe de signaler ici qu'elle diffère s'il s'agit d'une entrevue individuelle ou de groupe. Un problème général peu examiné est celui la préparation du contact avec les informateurs.

Il faut soigner la préparation du contact. On ne le soigne pas assez. Trop souvent, l'enquêteur déambule en quête d'un sujet, il aperçoit un étudiant qui lit à la cafétéria, il lui soumet ses questions, puis le remercie et court derrière un autre qu'il vient de voir s'asseoir devant son café. Il ne sait pas quel a été l'impact de son entrée en contact, de ses questions, comme il ignore dans quel état d'esprit le sujet se trouvait. Le sujet est reparti, et l'enquêteur ne sait même pas dans quelle disposition le sujet s'est trouvé à l'égard de ses questions. Était-il seulement prêt à répondre ? A-t-il répondu pour se désennuyer, pour passer le temps, ou s'est-il senti pressé par le temps ? Il est donc important de préparer le contact avec les informateurs de façon à créer un contexte favorable à la production d'une information vraisemblable et pertinente. La préparation du contact comporte quatre éléments.

Le premier élément semble aller de soi, mais l'expérience montre qu'il n'est pas toujours respecté : il s'agit de la formation du chercheur aux techniques d'entrevue. Le chercheur doit prendre le temps de se former et de ce point de vue, les livres, même très bien faits, sont insuffisants. Il existe des documents audiovisuels qui sont d'une aide certaine, mais il ne faut pas négliger des essais sous contrôle d'un superviseur expert. Un entraînement, sous forme d'essais enregistrés sur magnéto-scope puis analysés, de préférence avec un comparse ou un superviseur, devrait faire partie de toute recherche utilisant des données suscitées. En effet, dès que le thème, le contexte et l'échantillon d'informateur changent, la technique d'entrevue doit s'adapter et une mise au point de l'enquêteur est à reprendre.

Le deuxième élément est le *contact préliminaire*. Le chercheur a pensé à tel type d'informateurs. Mais la personne à laquelle il s'adresse fait-elle vraiment partie de cette catégorie ? Par exemple, dans une recherche sur les pratiques de recherches en éducation, il fallait tenir compte du fait que le domaine a été contaminé par beaucoup de monde : les psychologues et autres spécialistes des sciences humaines. Donc, la quête de sujets risque fort de conduire à des personnes qui, tout en étant en éducation, font de la recherche psychologique à propos de l'éducation. Ce n'est peut-être pas

idéal pour l'échantillon d'une telle recherche. Il est donc important de savoir si, de fait, la personne à qui on s'adresse dispose de l'information et est disposée à la donner; sinon on perd son temps. Lorsqu'on s'est assuré d'avoir un informateur potentiel, il faut lui expliquer ce que l'on cherche, les objectifs de la recherche, le type d'entrevue, le rôle qu'il y tiendra, la durée probable de l'entrevue (important pour qu'il puisse organiser ses activités), le nombre d'entrevues, et le rassurer sur la confidentialité du protocole. On prend le rendez-vous et on annonce qu'une confirmation par écrit suivra, ainsi qu'un rappel téléphonique la veille de l'entrevue.

Une fois ce contact préliminaire établi, on envoie un *document explicatif* qui est une *maquette* de la recherche. Il permettra à l'informateur de se préparer. Si on veut qu'il nous donne de l'information, il est important qu'il puisse ramasser sa documentation et ainsi se préparer. Il est important de ne pas le prendre à brûle-pourpoint. Comme le disait un jour un sujet surpris : «C'est gênant quand on te prend à l'improviste et qu'on te demande d'improviser; puis tu risques plus d'improviser pour sauver la face que de chercher dans ta mémoire la réponse dont l'autre aurait besoin.» Un informateur préparé est dans de meilleures dispositions pour fournir une information pertinente. On envoie donc une maquette de la recherche dans laquelle on rappelle à quel genre de questions le sujet peut s'attendre, le but de la recherche, ses retombées, sa pertinence sociale ou professionnelle. Bref, cette maquette détaille ce qu'on a dit lors du premier contact, et on garantit à nouveau la confidentialité.

Quelque temps avant la date fixée pour le rendez-vous, on *rappelle* l'informateur potentiel pour savoir s'il a bien reçu le document, s'il a eu le temps de le lire, s'il est toujours d'accord et, si oui, pour *confirmer* précisément la date, l'heure et le lieu du rendez-vous.

2 L'histoire vivante : les histoires de vie et les récits de pratique¹

À l'heure actuelle, on entend généralement par histoires de vie ou récits de pratique des récits construits pour fin de la recherche, à la demande

1 Pour plus de détail, voir :

Chalifoux A., Les histoires de vie. Chapitre 12 dans Gauthier B., *Recherche sociale*. Sillery, QC : PUQ, 1984, p. 277-291.

Pineau G. et Marie-Michèle, *Produire sa vie : autoformation et autobiographie*. Montréal, QC : Albert St-Martin, 1983. Particulièrement la 2e partie, «Les histoires de vie» (p. 117-190) et la 3e partie «Marie-Michèle» qui donne une excellente illustration du travail de constitution et d'analyse (p. 193-381).../...

du chercheur, et portant spécifiquement sur des périodes précises de la vie ou sur des activités particulières. Ce sont en quelque sorte des morceaux d'autobiographies dirigées, produites sur demande, afin de permettre au chercheur de mettre en évidence la dynamique d'une évolution et de l'interaction entre le sujet, ses rôles, son milieu, sa culture et les institutions.

On rencontre aussi quelques cas d'analyse (voir R. Lourau) où le matériel a été produit, comme dans les premiers usages du terme, en dehors du contexte de la recherche. Il s'agit alors, soit de correspondance décrivant une période de vie, de journaux plus ou moins intimes, soit de carnets de notes, de journaux de bord, toutes sortes de cahiers écrits en tant que mémoire externe de l'acteur pour lui-même, en prévision d'un avenir où il aurait peut-être besoin d'y recourir, de retrouver le souvenir. Cette situation se rencontre chez des écrivains, chez des chercheurs professionnels, chez ceux qui ont besoin de noter quasi sur le vif les impressions, les intuitions, les observations, les réflexions peut-être non pertinentes au moment même où elles se sont produites, mais qui pourraient être utiles plus tard.

Lorsqu'il s'agit de la construction d'un récit dirigé, on rencontre principalement trois manières d'effectuer le recueil du matériel.

Le récit peut s'enregistrer dans un cadre détendu, devant un auditoire familial comprenant le chercheur, que celui-ci soit ou non un membre de la famille, un ami, un invité ou un collègue. Une fois le récit lancé, il se poursuit dans une interaction familière comme une conversation de veillée où les différentes personnes de l'auditoire peuvent parfois interrompre pour mieux, ou plus, en apprendre du narrateur.

Le récit peut aussi être produit dans un cadre plus isolé, le narrateur étant seul devant sa feuille ou devant son micro, pour répondre à la demande du chercheur. C'est le plus souvent le cas des journaux de bord que des volontaires acceptent de remplir tous les jours ou toutes les semaines, et que les chercheurs rencontrent épisodiquement pour les motiver, pour les aider à continuer et pour évaluer la pertinence de poursuivre.

Enfin, le récit peut être complètement construit en interaction entre le chercheur et l'informateur. Cependant l'interaction doit être contrôlée et limitée aux interventions nécessaires à la compréhension du sens des mots et du récit. En effet, le problème de ce type de constitution des données

.../... Goodson I., *History, Context and Qualitative Methods in the Study of Curriculum*. Chapitre 5 dans Burgess R. (Éd), *Strategies of educational research*. Philadelphie, P.A. : The Falmer Press, Taylor & Francis Inc., 1985, p. 121-151.

Lourau R., *Le journal de recherche*, Matériaux d'une théorie de l'implication. Paris : Méridiens Klincksieck, 1988.

reste toujours d'obtenir l'information sous la forme d'un «récit de X» et non pas d'un «récit pour Y».

En effet, les «récits construits pour» le chercheur n'ont pas le même statut que les récits de l'informateur. Gilles Houle disait un jour que les meilleures histoires de vie avaient été les premières qui furent analysées, soit la correspondance entre des immigrants polonais en Amérique et leur famille restée là-bas, car, disait-il, au moins les chercheurs ne s'étaient pas mêlés de leur rédaction. Le problème évident de ces histoires de vie idéales, c'est qu'elles ne sont ni fréquentes ni facilement disponibles (protection de l'intimité). Dès lors, si les chercheurs veulent obtenir le matériel nécessaire à leur recherche, ils doivent susciter sa production chez des bénévoles. Le risque en ce cas est semblable aux effets Rosenthal : le narrateur est tenté de produire un matériel en fonction des attentes du chercheur, il est attentif à la moindre de ses interventions, de façon à réunir des indices sur ce que le chercheur a comme hypothèse, sur les informations qui lui sont pertinentes. De plus, si le chercheur donne des renforcements par des encouragements et par des questions qui insistent sur des thèmes semblables (c'est toujours les mêmes questions qui reviennent !), le narrateur produit une histoire qui, sans être une fiction, devient un arrangement destiné à produire un effet de satisfaction chez le chercheur. La matériel s'écarte alors de ce que le chercheur devrait obtenir.

Il faut cependant noter que les histoires de vie et les récits de pratique sont toujours un récit construit en fonction de quelqu'un : l'histoire de vie est un intertexte, texte qui se situe à la fois entre la perception que l'auteur a de son souvenir, entre la cohérence que l'auteur a besoin de donner au fil de son existence, entre l'image de son existence qu'il souhaite projeter, et les justifications de ses actes qu'il désire présenter aux autres, le tout dans un contexte donné. La construction d'une histoire de vie est très dépendante de facteurs psychosociaux. Pour reprendre l'exemple de l'immigration, il est clair que la famille restée en terre natale perçoit souvent les émigrants comme des chanceux (ils ont pu partir...) qui, en même temps, l'ont abandonnée (quand vont-ils revenir pour reconstituer la famille ?); et dans ses échanges, la famille le leur fait sentir. La perception de cette culpabilité, que la famille projette, marque le contenu des récits des immigrants : ils accentuent la nécessaire réussite de l'installation au nouveau pays, mais en même temps, ils atténuent tout ce qui pourrait renforcer les plaintes de la famille abandonnée. Quelle signification attribuer alors à de tels récits, même si les chercheurs n'interviennent pas dans leur constitution ? Ces récits ne sont à l'évidence pas les témoins fidèles de ce qui a été vécu par leurs auteurs.

À la limite, mais à condition que le chercheur soit plus attentif à percevoir les ruptures du récit de l'auteur qu'à chercher dans ce récit les réponses à son problème, la construction du récit dans une interaction informateur-chercheur produit un matériel qui résiste mieux à la critique du témoignage. Dans cette perspective, la procédure consiste à recueillir d'abord sous forme orale une première version du récit. Ensuite, une fois la narration transcrite, le chercheur et le narrateur font une co-lecture du récit de façon à compléter ce qui, tant aux yeux du narrateur que du chercheur, peut constituer une lacune rendant difficile une compréhension non ambiguë du texte. C'est essentiellement dans cette deuxième phase que se place l'interaction, et celle-ci se centre essentiellement sur les ruptures du récit (les phrases inachevées, les sauts brusques d'un sujet à un autre, etc.), ce qui permet de combler les lacunes introduites lors de la première formulation orale.

Lorsque la narration est ainsi complétée, l'interprétation peut débiter. Deux manières de faire se rencontrent : dans la première, le chercheur interprète le récit et soumet ensuite son interprétation au narrateur, afin que celui-ci en évalue la validité phénoménologique et suggère les corrections nécessaires de son point de vue; la seconde revient à prolonger l'interaction narrateur-chercheur pour la phase interprétative. La construction de l'analyse est alors dialectique par «co-réflexion» ce qui devrait garantir, selon Pineau, une meilleure pertinence de la grille et des concepts interprétatifs à l'égard de l'univers du narrateur. Mais cette pertinence dépend du rapport entre le narrateur et le chercheur, des capacités de persuasion de ce dernier, ainsi que des connaissances psychosociologiques et des capacités critiques du narrateur tant à l'égard de son propre récit que des interprétations proposées par l'analyste.

En plus de la question du statut du récit, deux problèmes techniques se présentent. D'abord, le choix d'un narrateur informé pose un problème de dépistage et de sélection. Ce genre de recherche procède généralement avec quelques récits, sinon le matériel devient vite trop abondant. Comme ces recherches visent à obtenir un matériel exemplaire, au sens de pertinent par rapport au problème, il est essentiel que les narrateurs soient choisis en fonction de leur participation à des situations en rapport avec l'objet de la recherche. La question s'élargit alors : non seulement les narrateurs doivent avoir vécu cette expérience, être capables de s'en souvenir et d'en construire un récit, mais le chercheur doit d'abord avoir évalué et sélectionné des situations stimuli pertinentes qui conduisent, de fait, les narrateurs à fournir un matériel utilisable pour la recherche. Ensuite, l'édition des histoires de

vie est confrontée au problème rédactionnel du respect de l'anonymat, non seulement du sujet auteur de la narration, mais aussi de son contexte avec les personnages qui y ont été impliqués.

3 L'entrevue (libre, semi-structurée et structurée)²

Si les histoires de vie visent la reconstruction d'une tranche de vie afin de dégager le sens de certains événements, l'entrevue vise plutôt à obtenir des informations sur les perceptions, les états affectifs, les jugements, les opinions, les représentations des individus, à partir de leur cadre personnel de référence et par rapport à des situations actuelles. Alors que la recherche du sens et la narration imposent à l'histoire de vie le passage de l'analogique (simultané et discontinu) au linéaire, l'entrevue libre tente de préserver l'expression des contradictions, des tensions, des conflits, des cahots, des ruptures et des circularités qui font partie de l'expérience humaine. Cela impose au chercheur une attitude d'écoute plus que la maîtrise de techniques, une manifestation de réceptivité afin de stimuler l'engagement et l'expression.

En fonction du genre d'information visée et du degré d'implication que cette information exige de la part de l'informateur, le style de l'entrevue devra se situer entre deux extrêmes qui balisent cette méthode. Comme technique de constitution des données d'une recherche, l'entrevue va se situer quelque part entre la conversation et le questionnaire. Plus l'information recherchée est personnelle ou intime, plus le style de l'entrevue se rapproche de la conversation : c'est l'entrevue libre. Plus l'information

2 Pour plus de détail, voir :

Nahoum C., *L'entretien psychologique*. Paris : PUF, 1963.

Daunais J.P., L'entretien non-directif. Dans Benoît Gauthier, *Recherche sociale*. Sillery, QC : PUQ, 1984, Ch. 11, p. 247-275. Voir aussi l'ensemble de 11 cassettes vidéo réalisées en collaboration avec le service audiovisuel de l'Université de Montréal.

Chauchat H., *L'enquête en psychosociologie*. Paris : PUF, 1985. Particulièrement le chapitre VI, L'entretien de recherche, p. 143-178.

Guillaumin J., *La dynamique de l'examen psychologique*. Paris : Dunod, 1965.

Ghiglione R., *L'homme communiquant*. Paris : Armand Colin-collection U, 1986.

Postic M. et J.M. De Ketele, *Observer les situations éducatives*. Paris : PUF, 1988. Voir particulièrement la section : Méthodes d'observation clinique, p. 130-138.

Ska B., De l'entrevue clinique pour fin de diagnostic. *Revue des sciences de l'éducation*, IX, 2, 1983, p. 267-277.

Van der Maren J.M., Comment conduire un groupe sans le brutaliser. *Information*, XVI, 7, 1977, p. 14-19.

recherchée se rapproche des opinions quasi publiques, plus l'entrevue peut ressembler à un questionnaire et devenir une entrevue structurée. Entre l'entrevue libre, appelée aussi non-directive, et l'entrevue structurée, qui ressemble à une liste de questions demandant des réponses ouvertes, à développement, on trouve diverses variantes comme l'entrevue clinique (le stimulus est une tâche à effectuer, un problème à résoudre), la méthode des incidents critiques (le stimulus est un scénario interrompu et qu'il faut poursuivre), les jeux de rôle et les formes adaptées à la recherche de la méthode des cas.

Comme déjà évoqué, lorsqu'on a insisté sur la préparation du contact avec les informateurs, les entrevues doivent être soigneusement préparées. Outre le contact préliminaire, le matériel produit ne sera pertinent que si trois conditions préalables sont respectées :

- 1° Le chercheur doit se donner un entraînement aux techniques de l'entrevue, en particulier au style de l'entrevue qu'il veut utiliser, afin d'éviter les erreurs qui contaminent les données;
- 2° Il doit sélectionner attentivement un terrain et des informateurs disposant de l'information recherchée et disposés à la transmettre;
- 3° Il doit se familiariser, s'acclimater, aux caractéristiques du terrain, du milieu où il va faire ses entrevues; il doit apprendre son environnement socioculturel : organisations géographiques et sociales, coutumes, manières de faire et dialectes locaux.

Parmi les techniques de *conduite des entrevues*, quelques points essentiels doivent être soulignés. Mais les connaître ne suffit pas, il s'agit d'habiletés à développer par des exercices pratiques. On ne peut assez insister sur l'efficacité de mini-entrevues enregistrées et que l'on peut revoir à plusieurs reprises avec un superviseur, pour apprendre à manifester avec naturel ces habiletés.

A. Les informateurs ne seront bien disposés à l'entrevue que s'ils ne se sentent pas pris dans un entretien semblable à l'interrogatoire policier. Cela implique *trois attitudes*. Le chercheur doit indiquer complètement et le plus clairement possible le thème et les objectifs avant l'entrevue et les rappeler en début d'entrevue. Ensuite, le chercheur doit respecter le rythme d'expression et de réflexion de l'informateur : ne pas le brusquer, respecter les silences nécessaires à la réflexion, écouter et manifester son écoute en utilisant ce qu'a dit l'informateur lors des résumés. Quand il relance l'entrevue, il doit poser ses questions à partir des réponses antérieures, de façon à ce que l'entretien suive la pensée de l'informateur sans le forcer à

de brusques changements de thèmes. Si une réorientation de l'entrevue est nécessaire, il faut prévenir l'informateur, fermer le thème antérieur, introduire le nouveau thème, et lui donner le temps de le prendre à son rythme. Enfin, ne pas manifester une attitude de réserve méfiante en essayant de coincer l'informateur dans des questions pièges, par des reprises de ce qu'il a dit sous forme de confrontation, et par des sauts incessants d'un thème à l'autre, qui donnent à l'informateur l'impression qu'on veut lui faire perdre la tête pour mieux le surprendre.

B. Dans l'*entrevue libre*, la recherche de l'information appropriée nécessite que l'intervieweur ait en tête un schéma général de l'ensemble des thèmes qu'il souhaite aborder. Mais la souplesse nécessaire au respect de la démarche de l'informateur et du climat de l'entrevue exige, le plus souvent, le sacrifice de quelques thèmes qui resteront inabordés, sinon oubliés. Il faudra les reprendre dans la seconde entrevue. En effet, lors d'une recherche par entrevue, il est important de prévoir deux entretiens : le premier, pour la collecte principale du matériel, est d'un style plus ouvert que le second qui, plus structuré, a pour fonction de compléter les thèmes qui n'ont pas été abordés, de combler les lacunes, d'éclaircir les ambiguïtés et de prendre des informations générales (âge, formation, etc.). Ces informations techniques ne s'obtiennent qu'à la suite de questions exigeant des réponses précises. Aussi on évite de les poser, surtout au début de la première entrevue, car elles induiraient un style fermé, un jeu question-réponse brève, pour la suite de l'entretien.

C. Les *entrevues de groupe* constituent une variante fort intéressante lorsque l'information recherchée est une *représentation sociale*, partagée par plusieurs personnes. Elle est fort utile lorsque l'on souhaite contrôler par le groupe les représentations sociales qu'on a inférées à partir d'entrevues individuelles, et lorsqu'on veut évaluer leur poids dans la dynamique d'un groupe. Mais l'entrevue de groupe produit autre chose que la moyenne d'un ensemble d'entrevues individuelles, et elle ne peut pas être utilisée pour faire l'économie de ces dernières. Comme pour une entrevue individuelle, la préparation du contact est essentielle. Mais la technique de la conduite d'une telle entrevue diverge sur quelques points. Nous en relevons deux qui sont déroutants pour les chercheurs entraînés à l'entrevue individuelle; ils doivent perdre ces deux automatismes d'animation pour réussir une entrevue de groupe.

Dans l'entrevue individuelle, il est important que l'intervieweur manifeste son écoute par des reprises, par des «hum hum» et autres gestes d'encouragement. Au contraire, en entrevue de groupe, il importe que

l'animateur intervieweur ne manifeste pas ces formes d'encouragement. Une fois le problème posé, il doit laisser aux membres du groupe le soin de manifester leur accord ou leur désaccord avec ce qui se dit et, pour cela, il lui faut éviter de se manifester trop souvent. En effet, s'il se manifeste par des interventions automatiques et naturelles en conversation de soutien et d'écoute, la personne qui parle aura tendance à s'adresser à lui, le chercheur, et non aux autres; et les autres auront aussi tendance à répondre au chercheur et non pas à la personne qui vient de parler. Non seulement cela ralentit le rythme de la discussion, mais plus encore cela met l'animateur intervieweur au centre du groupe. Or, lorsqu'il est dans une telle position, il est, malgré lui, amené à prendre parti, le groupe lui demandant implicitement ou directement son avis, l'approbation ou la désapprobation de ce qui vient d'être dit. Lorsque cela arrive, la réunion perd de son intérêt, puisque son but est de mettre en évidence, non pas l'accord de l'animateur avec un ou plusieurs membres, mais bien les positions partagées par le groupe. C'est aux membres du groupe, et non à l'animateur intervieweur, de valider les opinions émises.

Lorsqu'un membre du groupe, ou l'ensemble du groupe, demande à l'animateur de prendre parti, celui-ci peut toujours répondre par des réactions *écho* (il renvoie la question à celui qui la pose), *relais* (il refile la question à un autre membre) ou *miroir* (en renvoyant la question à l'ensemble du groupe). Mais ces techniques ne sont utilisables qu'une fois ou deux au cours d'une entrevue de groupe, sinon elles font artificielles, et les membres refusent rapidement de continuer à jouer un jeu qu'ils estiment faussé. Il est donc préférable que l'animateur intervieweur perde cette habileté typique des entrevues individuelles et qu'il se montre très discret afin que les validations et les contrôles proviennent du groupe lui-même.

Cependant, l'animateur intervieweur des entrevues de groupe doit intervenir de temps à autre, notamment pour effectuer des bilans qui relancent la discussion, et pour effectuer des synthèses qui explorent les confrontations en soutenant les minoritaires et les timides (on a dit blanc, mais, si on n'a pas dit noir, on a aussi dit bleu, alors...?). Cela exige de lui qu'il perçoive les divergences et les nuances dans les opinions partagées par le groupe. À cette fin, une technique essentielle consiste à ne pas regarder celui qui parle (qui sentirait cela comme un encouragement, ce qui nous renvoie au problème précédent). L'animateur intervieweur se contente d'*écouter celui qui parle*, et d'enregistrer, par des mots clefs, l'essentiel de ce qu'il dit, afin de pouvoir *voir les réactions des autres* à ce qui se dit. La plupart du temps, ces réactions, non verbales, ne peuvent être vues. Or,

si l'animateur ne les perçoit pas, il risque de ne jamais les enregistrer, car ceux qui les manifestent les auront peut-être oubliées avant d'avoir leur tour pour s'exprimer. L'animateur intervieweur doit donc balayer du regard les membres du groupe qui ne parlent pas, afin d'anticiper qui il devra encourager, ensuite, du regard afin d'obtenir une réaction à ce qui a été dit.

Malgré leur richesse, les techniques d'entrevues (individuelle ou de groupe) ne recueillent un matériel pertinent que si le chercheur prend garde à ne pas introduire des thématiques personnelles par les questions et par le vocabulaire utilisé dans ses interventions. Or, l'informateur a tendance à suivre ces suggestions parce qu'il désire être un bon informateur, c'est-à-dire celui qui donne à l'autre les informations que ce dernier a plaisir à entendre. De plus, afin de permettre à l'informateur de maintenir un rapport d'égalité avec le chercheur, l'intervieweur doit donner à son informateur l'occasion de valider chacune de ses reformulations, synthèses, explicitations ou bilans. Puisqu'il s'agit d'une quête d'information auprès d'un informateur, cette information ne peut être considérée comme valide que dans la mesure où l'informateur donne lui-même son accord au contenu et à la forme de ce que l'intervieweur en retient. Cette validation par l'informateur est une exigence essentielle pour garantir la validité phénoménologique de l'information, et pour assurer l'informateur de l'importance que l'intervieweur accorde à ce qu'il sait et à ce qu'il veut bien lui dire. Si l'informateur perd le sentiment de cette importance, il réduit le flot, la qualité et la pertinence de son information, pour ne plus chercher qu'à satisfaire le narcissisme de son intervieweur et à terminer la discussion au plus vite. Dans la préparation et la réalisation des entrevues, le chercheur rencontre des embûches qui conditionnent la qualité et le statut du matériel. Le chercheur doit ne pas se méprendre sur la valeur de ce qu'il a obtenu, alors qu'il en est lui-même le principal responsable.

4 L'entrevue clinique³

Cette méthode a plusieurs appellations, toujours accolées du terme clinique, qui dépendent de ses objectifs et de son contexte : entrevue clinique

3 Pour plus de détail, voir :

Ska B., Quelques précisions sur l'entrevue clinique pour fin de diagnostic. *Revue des sciences de l'éducation*, IX, 2, 1983, p. 267-277.

Postic M., et J.M. De Ketele, *Observer les situations éducatives*, Paris : PUF, 1988. Voir particulièrement la section «Méthodes d'observation clinique» p. 130-138.

Berg D.N. et K.K. Smith, *Exploring Clinical Methods for Social Research*. Beverly Hill, CA. : Sage, 1985 (surtout pour une étude critique des méthodes cliniques).

(en recherche avec des sujets capables de verbaliser leur réflexion), observation clinique (en recherche avec des jeunes sujets ou des sujets qui verbalisent difficilement), et examen clinique (dans la pratique évaluative ou diagnostique avant le traitement). L'idée à la base de cette technique se résume à demander au sujet d'exprimer les perceptions qu'il a de son action ou de son état, de verbaliser son action pendant qu'il effectue une tâche particulière, et à observer les conduites du sujet confronté à des nécessités d'action. L'origine de cette technique, sinon de son appellation, est médicale (grec «klinê» = lit) : pour mieux connaître l'état du malade, le médecin se rend à son chevet et l'interroge sur ce qu'il ressent. Les possibilités de cette technique en psychologie et en éducation résultent d'une analogie évidente : pour savoir comment l'enfant construit une notion ou résout un problème, on lui demande de réaliser une tâche, de résoudre un problème, et on «discute intelligemment» avec lui pendant qu'il s'exécute. L'entrevue clinique a donc pour objectif de mettre en évidence des processus par la mise en acte et par la réflexion sur ces actions dans la réalisation d'une tâche.

Les difficultés sont immédiates : disposer d'un inventaire de tâches pertinentes au problème de la recherche, et être capable de discuter de ce qu'il fait avec un sujet occupé par une tâche, et cela de manière «intelligente» aux yeux du sujet ! Cette capacité de discuter intelligemment avec le sujet dépend strictement des habiletés cognitives et relationnelles du chercheur. Il n'est pas facile, que le sujet soit adulte ou enfant, de lui poser des questions qui l'incitent à réfléchir sans le ridiculiser, sans l'inférioriser, et sans lui suggérer ou lui montrer (enseigner) les actions qu'il pourrait faire pour satisfaire les hypothèses. Aussi, l'examen attentif des protocoles complets (action-réaction) d'entrevue clinique doit avoir, entre autres, comme objectif de vérifier si l'intervieweur n'a pas induit les réactions du sujet et s'il a pu maintenir une discussion intelligente et intéressante pour son sujet. Sans quoi, on ne peut avoir aucune garantie à propos du statut à accorder au matériel obtenu.

La construction d'un inventaire de tâches pertinentes à l'exploration d'un processus nécessite une maîtrise du domaine de recherche au plan des théories et des pratiques, parce que l'entrevue clinique suit une stratégie de cascade d'hypothèses et de vérifications successives à propos des processus activés par le sujet.

L'entrevue se déroule comme suit :

- 1° Le chercheur propose au sujet une tâche, dont la réalisation implique (pense-t-on à partir de la théorie ou des pratiques éducatives) que le sujet active un processus (celui qu'on cherche à mieux connaître);

- 2° Le chercheur observe l'activité du sujet et suscite, quand c'est possible, sa réflexion sur cette activité;
- 3° Le chercheur induit de ce que le sujet fait, ou de ce qu'il dit, une hypothèse sur le processus ou le mécanisme activé;
- 4° Le chercheur cherche à vérifier immédiatement cette hypothèse en soumettant au sujet une nouvelle tâche, susceptible de produire des indices pertinents en regard de l'hypothèse. Cela exige que les actions posées dans cette nouvelle tâche et les réflexions soient telles que le chercheur sera capable de confirmer ou d'infirmer l'hypothèse;
- 5° En cas d'infirmerie de l'hypothèse, ou bien le chercheur dispose d'une hypothèse de rechange qu'il éprouve par une nouvelle tâche pertinente, ou bien il soumet au sujet une tâche exploratoire, espérant que son comportement permettra d'induire une nouvelle hypothèse;

Et ainsi de suite dans une cascade en chaîne comportant tâche, induction d'hypothèse, vérification dans une tâche, induction, jusqu'à ce que le chercheur pense avoir réussi à mettre en évidence les processus utilisés par le sujet.

Il faut donc que le chercheur puisse planifier l'entrevue en sachant, au moins d'une manière hypothétique, quelles actions sont des bons indices de quels processus, et quelles tâches suscitent les actions qui témoignent des processus à propos desquels le chercheur émet des hypothèses. Autrement dit, le chercheur doit disposer d'un modèle du fonctionnement du sujet, c'est-à-dire d'un modèle théorique (hypothétique) des processus impliqués dans la réalisation des tâches, et des indices de l'activation de ces processus. Comme il s'agit d'une entrevue clinique, le chercheur doit non seulement connaître l'ensemble de ces relations «tâches-processus-actions», il doit en plus en avoir une maîtrise telle qu'il puisse, en cours d'entrevue, passer d'une tâche à l'autre sans avoir besoin de demander au sujet de lui accorder une pause pour y réfléchir. Cela implique donc que les chercheurs doivent acquérir une connaissance et une expérience théorique et pratique du modèle «tâche-processus-action» du domaine de recherche, avant de se lancer dans l'utilisation d'une telle technique. Pour ceux dont l'expérience est insuffisante, il existe une solution de rechange, c'est l'examen systématique.

5 L'examen systématique

L'examen systématique est un type d'examen exploratoire que le chercheur, tenté par l'entrevue clinique, devrait envisager s'il dispose d'un

modèle théorique des relations entre tâches-processus-actions, et s'il n'a pas la maîtrise suffisante des habiletés de l'entrevue et de la cascade en chaîne des tâches du modèle. Comme l'entrevue clinique, l'idée est issue de pratiques diagnostiques. Par exemple, en neuropsychologie, les chercheurs ont construit des modèles du fonctionnement cérébral qui impliquent qu'une lésion produit plusieurs symptômes et qu'un symptôme peut provenir de plusieurs lésions. Dès lors, afin de pouvoir diagnostiquer une lésion, il importe de procéder à l'examen systématique d'un ensemble de tâches pour observer quels symptômes sont présents parmi les symptômes possibles.

Le raisonnement de l'examen systématique peut être schématisé comme suit. S'il y a trois lésions connues, A, B, C, telles que la lésion A produit les symptômes a, b, c, que la lésion B produit les symptômes b, c, d, et que la lésion C produit les symptômes c, d, a, alors il faut et il suffit d'examiner toutes les possibilités de combinaison des symptômes a, b, c, d, pour identifier correctement une des lésions A, B ou C. Dans la mesure où l'observation des symptômes peut être simplifiée et réduite à l'observation de quelques signes facilement reconnaissables, alors cet examen systématique peut être effectué par des techniciens ou même par des machines (comme les analyses de sang), ce qui débarrasse le médecin de la nécessité de faire l'examen clinique tout en lui donnant la garantie d'un diagnostic objectif et technique. D'où l'éclosion d'une pratique médicale où les examens de laboratoire se sont multipliés malgré leur coût.

L'examen systématique, construit selon l'analogie ci-dessus, peut aussi être envisagé dans la recherche éducationnelle. Si on dispose d'un modèle théorique (même hypothétique et provisoire) des relations entre tâches, processus et actions, on peut construire un examen dans lequel un ensemble planifié de tâches (stimuli d'actions indices des processus) est soumis aux sujets. Si le modèle est pertinent et si l'ensemble des tâches stimuli peut être présenté aux sujets, alors on peut construire un examen tel qu'à la fin de cet examen, le profil des actions (réussites et échecs) réalisées devrait permettre d'identifier les processus utilisés par les sujets. L'examen systématique constitue, à la fois, un outil d'exploration des fonctions du sujet, et un outil de vérification de la pertinence du modèle théorique, si l'on peut comparer le diagnostic qu'il produit à un autre diagnostic (valide) obtenu par ailleurs.

Mais cette technique est relativement lourde, car le nombre de tâches à réaliser dans un examen systématique correspond au moins au nombre d'hypothèses à vérifier dans la cascade clinique pour s'assurer de l'activation

des processus examinés. De plus, comme le chercheur souhaite pouvoir compenser la perte des contrôles multiples d'une hypothèse que permet la cascade en chaîne de l'entrevue clinique, il est tenté de multiplier les tâches d'un examen systématique afin de pouvoir confronter les données (les trianguler) et ainsi de contre-valider les hypothèses qui résistent. Dès lors, il faut souvent réduire le nombre des processus à évaluer pour que l'examen ne soit pas trop long, car si c'est le cas, on risque de collecter du matériel parasite (fatigue, distraction) parfois difficile à distinguer des indices de l'activation des processus.

Dans certaines recherches, afin que les productions de différents sujets puissent être comparées, l'examen systématique est standardisé et transformé en test : on propose aux sujets des listes d'actions entre lesquelles il devra choisir pour accomplir chaque tâche. Dans ce cas, les données ne sont plus suscitées, mais provoquées, et on n'est pas certain que la réponse du sujet correspond bien à ce qu'il aurait fait si on n'avait pas limité son choix. On peut seulement dire que la réponse choisie est probablement celle qui est la plus proche de ce qu'il aurait fait, à condition qu'il ait bien compris le sens des éléments de la liste proposée. Voilà posé un des problèmes majeurs des données provoquées. La construction d'un test permet au chercheur de savoir exactement ce qu'il a proposé au sujet et quel sens il attribue aux réactions du sujet. Mais rien ne garantit au chercheur que le sujet a compris les consignes comme le chercheur, et que le sens donné par le sujet à la réaction choisie correspond exactement au sens que le chercheur lui avait accordé.

6 La sélection des sources et l'échantillonnage⁴

Les problèmes de la sélection des sites, des événements et des sujets, pour les techniques envisagées jusqu'ici, concernaient surtout le dépistage de situations ou d'informateurs pertinents, c'est-à-dire susceptibles de fournir l'information recherchée. Comme on le verra ci-dessous, si on veut utiliser le mot échantillonnage dans ces cas, on doit l'utiliser en association avec les attributs «théorique» ou «rationnel», dans la mesure où les sujets ne sont pas retenus pour leur représentativité, mais pour leur capacité de fournir un matériel inducteur d'hypothèses.

4 Pour plus de détail, voir :

Beaud J.B., Les techniques d'échantillonnage. Chapitre 8 dans B. Gauthier (Éd.), *Recherche sociale*. Sillery, QC : PUQ, 1984, p. 175-200.

Reuchlin M., *Précis de statistique*. Paris : PUF, 1976. Voir spécialement le chapitre 7, Échantillon et population, L'inférence statistique, p. 181-219.

Plutôt que de s'intéresser aux dimensions d'un problème ou aux composantes d'un processus à partir d'un nombre très réduit de sujets, on peut vouloir connaître comment des traits, des habiletés ou des opinions se répartissent dans une population et comment cette population se caractérise par rapport à un éventail de traits, d'habiletés, d'opinions. Dans de telles recherches, on est confronté au problème de l'échantillonnage des sujets, des situations, des questions à examiner parmi l'ensemble des sujets, des situations et des questions possibles. En effet, dans beaucoup de cas, l'ensemble des individus à propos desquels on veut tenir un discours est hors d'atteinte, soit parce qu'ils sont trop nombreux et qu'on n'a ni les moyens ni le temps de tous les interroger, soit parce qu'ils ne sont pas disponibles pour diverses raisons. Si l'ensemble des individus (que ce soient des sujets, des situations ou des questions) est hors d'atteinte, on doit se résoudre à travailler avec une portion d'entre eux, c'est-à-dire avec un échantillon, que l'on souhaite représentatif de la population visée. Notons que le terme population ne désigne pas nécessairement un grand nombre de sujets. Si le chercheur veut déterminer les traits spécifiques et différentiels des 15 enfants d'une classe spéciale et qu'il a la possibilité de les rencontrer tous les 15, il travaille bien sur une population. On ne parlera d'échantillon que lorsque l'on veut étendre ou généraliser ce que l'on a observé sur quelques-uns à un ensemble d'individus plus nombreux, dont font partie les quelques individus (l'échantillon) que l'on a, de fait, observés.

Deux principes peuvent fonder un échantillonnage : la maquette et le hasard. Selon le principe de la maquette, ou du modèle réduit, on tente, à partir de différents critères arbitraires, de postulats ou de règles déduites d'une théorie, de sélectionner un nombre d'individus tel que, si l'on pouvait grossir l'image qu'ils projettent, on obtiendrait une image semblable à celle de la population. Selon le principe du hasard, on postule que les variations pour les différents traits qui caractérisent les individus dans une population se répartissent au hasard. À la suite de cela, on présuppose que, si on extrait au hasard (c.-à-d. sans qu'aucun principe ou régularité n'intervienne) des individus de cette population, on retrouvera dans cet échantillon la même répartition au hasard des variations pour les traits que celle qui caractérise les individus de la population. Le principe de l'échantillonnage au hasard repose donc sur deux postulats qui restent souvent à vérifier : 1° la répartition des variations au hasard dans la population; et 2° l'identité de la répartition dans l'échantillon avec celle de la population comme résultat d'une extraction au hasard. Or, en admettant que le premier postulat puisse tenir pour certains traits, cela ne valide pas nécessairement le second. En conséquence, nous devons savoir que la validité du principe de la maquette

repose sur la validité de la théorie dont sont déduites les règles de réduction, alors que l'échantillonnage au hasard repose sur deux postulats faisant appel au hasard et qui sont souvent invérifiables.

Les *échantillonnages en maquette* s'effectuent le plus souvent selon trois méthodes :

A. *L'échantillonnage raisonné* (ou théorique ou rationnel) en boule de neige, dans lequel on construit un modèle de l'individu en se basant sur des idées *a priori* (une théorie, un postulat) concernant les caractéristiques que devraient posséder les individus de l'échantillon. Ayant repéré quelques individus types, on y joint ceux qui sont en relation avec eux (selon le principe «qui se ressemble s'assemble»);

B. *L'échantillonnage en cascade contrastée* dans lequel on délimite, sur la base d'un modèle théorique ou de postulats, les caractéristiques fixes que doivent avoir tous les individus de l'échantillon et les caractéristiques qui devraient varier. On sélectionne alors un premier individu ayant les caractéristiques fixes et on lui demande de désigner un autre individu, possédant les caractéristiques fixes, mais manifestant des positions contraires ou différentes quant aux caractéristiques qui peuvent varier. Et ainsi de suite d'individu à individu;

C. *L'échantillonnage par quotas* pour lequel on détermine, au préalable et selon une théorie ou un postulat, quel pourcentage d'individus de la population manifesterait chacune des caractéristiques pertinentes : c'est la détermination du poids de chacune des strates, ou des composantes, de la population. Par exemple, on détermine, à partir des recensements, le pourcentage de femmes, hommes, enfants, adultes, vieillards, riches, pauvres, urbains, ruraux, etc. Ensuite, on compose l'échantillon en y incluant les individus selon le pourcentage de ceux qui doivent manifester les caractéristiques pertinentes : c'est la reproduction de la pondération des strates dans l'échantillon. L'échantillon comportera le même pourcentage de femmes, hommes, enfants, adultes, etc., que dans la population.

Ces trois méthodes d'échantillonnage, selon le principe de la maquette, sont sensibles à différents aspects de la description d'une population. On choisit idéalement la méthode en fonction de l'aspect que la recherche souhaite privilégier. Ainsi, l'échantillonnage par quotas est sensible à la proportionnalité de la réduction sur chacune des caractéristiques. Cela ne va pas toujours sans altérer l'image projetée de la population sur l'échantillon. Par exemple, si l'échantillon est une réduction au 100^e (échelle de 1 / 100), alors toute strate de la population composée de

moins de 100 individus disparaît de l'échantillon. Cela implique parfois de choisir des échelles particulières, avec compensation, pour maintenir la présence dans l'échantillon de toutes les strates de la population. L'échantillonnage raisonné en boule de neige est sensible aux tendances dominantes que manifestent les premiers individus retenus pour constituer l'échantillon : les individus qui manifestent des divergences ont tendance à disparaître. C'est donc un échantillonnage sensible aux effets de tendance centrale. Par contre l'échantillonnage en cascade contrastée est sensible aux effets de dispersion : les individus appartenant à la tendance dominante n'y sont normalement pas plus représentés que les divergents, et l'éventail des divergences y est habituellement visible.

Cette sensibilité à la proportionnalité, à la tendance centrale ou à la dispersion peut toujours être biaisée par quelques facteurs accidentels, parfois incontrôlables au moment de la constitution de l'échantillon, mais observables à l'analyse. Par exemple, en faisant une cascade contrastée, on indique généralement aux sujets sur quelles caractéristiques le contraste doit s'effectuer. Mais il se peut que tout en respectant cette base de contraste, les sujets indiquent les autres individus à l'intérieur d'une caractéristique, non prévue par le chercheur, qui les rassemble et à partir de laquelle ils se ressemblent beaucoup, alors que leurs différences sur les caractéristiques prévues pour le contraste sont mineures. Par exemple, cela pourraient être des enseignants qui, tout en enseignant des matières semblables selon des didactiques différentes, font partie d'une même secte qui leur impose des attitudes semblables. Il peut ainsi en être de l'appartenance à un groupe plus ou moins informel dont le chercheur ignore l'existence mais qui biaise, parfois jusqu'à l'annuler, le contraste entre certains individus. La qualité recherchée d'un échantillonnage de type maquette doit donc toujours se vérifier sinon au moment de sa constitution, au moins lors de l'analyse des données.

Les *échantillonnages au hasard* s'effectuent aussi selon diverses modalités :

A. L'*échantillonnage aléatoire simple* est effectué strictement au hasard, en respectant le principe selon lequel chaque individu a une chance connue, non nulle et égale d'être choisi. Cela implique que la liste des individus soit connue et que chaque tirage soit indépendant (c.-à-d. on remet dans l'urne le billet tiré après chaque tirage, de telle sorte que le nombre de billets dans l'urne reste constant);

B. Dans l'*échantillonnage systématique*, les individus sont tirés à intervalle fixe dans une liste dont le premier est tiré au hasard. Plus rapide et économique, ce tirage présente deux exigences afin de rester un peu au hasard : 1° la liste sur laquelle s'effectue le tirage ne peut pas être ordonnée (sinon seul le premier est tiré au hasard); 2° l'intervalle avec lequel on tire les suivants ne peut pas correspondre à un cycle qui affecterait la liste (exemple de cycle : on tire tous les 10 individus, mais pour composer la liste on a pris, pendant 10 jours, chaque jour, les 10 premiers qui se présentaient !);

C. L'*échantillonnage aréolaire* est une méthode topographique utilisable lorsqu'on ne dispose pas d'une liste des individus. On divise un espace en blocs. Chacun de ces blocs est considéré comme une des unités constituant une liste. Ensuite, on effectue un tirage entre les unités de cette liste (de blocs) par une des méthodes précédentes. Par exemple, sans devoir connaître les habitants d'une agglomération, on divise la carte de la ville en petites zones et on tire au hasard n zones parmi l'ensemble des zones;

E. Dans l'*échantillonnage en grappes ou par faisceaux*, on tire d'abord au hasard, non pas des individus, mais des groupes; ensuite, dans chacun des groupes retenus, on tire au hasard des individus (échantillon au 2° degré) ou des sous-groupes, pour lesquels un tirage subséquent produirait un échantillon au 3° degré d'individus, etc.;

F. L'*échantillonnage stratifié* combine l'échantillonnage théorique avec le hasard. On divise préalablement la population en strates selon des caractéristiques pertinentes au plan théorique ou selon quelques principes pragmatiques. Ensuite, on effectue dans chaque strate le tirage au hasard d'un échantillon. Comme dans le cas des échantillonnages par quotas, on peut vouloir compenser l'importance de groupes particuliers en augmentant le poids de leur représentation dans les échantillons. On parle alors d'échantillonnage stratifié non proportionnel.

Les échantillonnages aréolaires et en grappes présentent des faiblesses à vérifier au moment de la constitution des échantillons ou lors de l'analyse. Dans le cas des échantillons aréolaires, il est possible, sinon probable, que les blocs aient des densités relativement différentes, ce qui aboutit à une sur-représentation de certaines parties de la population au détriment d'autres. Par exemple, si certains blocs correspondent à des quartiers résidentiels riches, où les résidences sont dispersées et le territoire peu peuplé, à côté de blocs plutôt pauvres et surpeuplés, les pauvres seront probablement sous-représentés dans l'échantillon, et les riches sur-représentés. Une pondération

de la représentation en fonction de la densité des blocs doit alors être envisagée, même si les densités relatives, et donc les pondérations, sont difficiles à estimer. Les échantillonnages en grappes comportent la possibilité d'erreurs importantes, due à l'homogénéité renforcée des caractéristiques à l'intérieur des groupes. Si le principe «qui se ressemble s'assemble» joue lors de la constitution des groupes, et si ceux qui présentent des traits minoritaires dans un groupe, sont aussi minoritaires dans les autres groupes, ces minorités risquent de ne jamais voir tirer un de leur représentant. Ce tirage est donc peu favorable aux minorités dispersées : il produit une représentation des tendances dominantes dans les groupes, ce qui peut conduire à une image très erronée de la population.

Enfin, deux modes de constitution d'échantillons, relativement répandus, sont à éviter lorsque c'est possible :

L'échantillonnage accidentel, «au petit bonheur» ou «cafétéria», manque de représentativité. Il est biaisé par le temps, par le lieu, par les occupations, etc. Par exemple, on risque de rencontrer beaucoup plus d'étudiants de 2^e ou 3^e cycle à la bibliothèque que de jeunes étudiants du 1^{er} cycle. Malgré ses apparences, l'échantillonnage accidentel est sujet à de très forts effets de sélection.

L'échantillonnage volontaire est aussi biaisé par un fort effet de sélection. Rosenthal⁵ a montré que le volontaire est toujours quelqu'un d'intéressé, sinon directement par les effets attendus du projet de recherche, du moins par une valorisation sociale. Le sujet volontaire semble témoigner d'une psychologie particulière : moins conventionnel, plus jeune, moins autoritaire, plus sociable, ayant un grand besoin d'approbation sociale, etc.

Le moins bon des échantillons consiste évidemment à jumeler l'échantillonnage accidentel et volontaire. Lorsqu'on ne peut pas se passer de volontaires, ce qui malheureusement est assez souvent le cas, puisque le chercheur est fréquemment à la recherche de sujets, le mieux à faire est de contrôler leurs caractéristiques de façon à se rapprocher d'un échantillonnage par quotas.

5 Rosenthal R., La participation volontaire. Dans Lemaine G. et J.M. Lemaine, *Psychologie sociale et expérimentation*. Paris : Mouton-Bordas, 1969, p. 71-79.

LES DONNÉES PROVOQUÉES

- 1 Le sondage
 - 1.1 *Les problèmes généraux*
 - 1.2 *Les techniques de sondage*
 - 1.3 *Les plans de sondage*
- 2 Les questionnaires et les échelles
- 3 Les tests
 - 3.1 *Les caractéristiques générales*
 - 3.2 *Les qualités exigées des tests*
 - 3.2.1 La fidélité
 - 3.2.2 La validité
 - 3.2.3 L'étalonnage ou la référence
 - 3.2.4 La hiérarchisation
 - 3.2.5 La discrimination
 - 3.2.6 La standardisation
- 4 La manipulation des contingences
 - 4.1 *La formulation des consignes*
 - 4.2 *Les consignes explicites et implicites*

D'une manière générale, la constitution des données provoquées doit respecter des principes élaborés dans le cadre des méthodologies quantitatives, particulièrement dans les théories de la mesure, même si, au sens strict, les réponses aux questionnaires, aux échelles et aux tests sont qualitatives. Cependant, le problème de la formulation des consignes est semblable à celui rencontré pour les données suscitées. Mais, lorsque les données provoquées sont de type qualitatif et ne réclament pas une théorie de la mesure, elles exigent au moins une théorie de l'objet de recherche. Celle-ci servira de base à la détermination des dimensions à interroger ou des listes d'items à produire.

1 Le sondage¹

1.1 Les problèmes généraux

Les sondages ont de moins en moins bonne presse dans la mesure où les hommes politiques, qui semblent en être les principaux utilisateurs, montrent qu'ils ne savent pas s'en servir. En effet, bien des politiciens prennent les résultats des sondages pour un portrait fidèle de la volonté populaire et ne décident qu'en fonction des tendances dominantes qui y apparaissent. Or, comme le dit bien le titre du livre de Stoetzel et Girard, le sondage ne fournit qu'une image fugitive de l'opinion publique élaborée dans le contexte des consignes et du climat dans lequel les gens sont sondés. L'opinion livrée par les personnes interrogées n'est qu'une opinion publique : ce que chacun pense qu'il est bon de dire en public à la suite d'une question posée à un moment donné. Il y aurait donc un abus de langage à parler de sondage au sens de «sonder», c'est-à-dire d'effectuer un examen minutieux, approfondi et méthodique d'une situation ou des pensées d'une personne.

Le sondage d'opinion publique prend un sens figuré inverse : il s'agit d'une investigation discrète et rapide, contextualisée, fournissant une information peu raffinée et sans nuance sur l'opinion ou les attitudes publiques des gens en un moment donné. Bien conçu, il a son utilité : donner rapidement une image du climat public à propos d'un fait, d'un projet, d'une politique. Mais cette image doit être interprétée et elle ne peut pas être prise au pied de la lettre. Elle ne peut pas servir directement à une prise

1 Pour plus de détail, voir :

Stoetzel J. et A. Girard., *Les sondages d'opinion publique*. Paris : PUF, 1973.

Blais A., Le sondage. Chapitre 14 dans B. Gauthier (Éd.), *Recherche sociale*. Sillery, QC : PUQ, 1984, p. 317-357.

de décision avant un examen attentif des facteurs qui ont influencé les réponses obtenues. Cela veut dire que les résultats d'un sondage ne peuvent pas être fiables si on ne tient pas compte de la technique de sondage, du plan de sondage et des caractéristiques des informateurs.

Le deuxième problème sous-estimé par la plupart des politiciens tient aux aspects statistiques du sondage. Si les maisons de sondage ont pris l'habitude de donner un résultat ainsi qu'une marge d'erreur, elles n'insistent pas beaucoup sur la probabilité qu'elles ont de voir la réalité tout à fait différente de leurs prédictions. En effet, chaque prédiction d'un résultat est effectuée en comportant ce que les statisticiens appellent les limites de confiance de l'estimation, ce que les journalistes appellent une fourchette. Mais une probabilité d'erreur correspond toujours à cette *limite de confiance*, donnée par les deux «bornes» d'une prédiction. Or, pour un échantillon de taille semblable, plus la prédiction est précise, plus la probabilité d'erreur est grande. Ainsi, en tenant constant un échantillon, si l'on veut une prédiction certaine, il faudra accepter une fourchette assez large. Par exemple, on peut prédire, avec une faible probabilité d'erreur, que le résultat d'un parti à une élection se situera entre 40 et 70 % ($55\% \pm 15$) des voix. Si l'on souhaite plus de précision, c'est-à-dire réduire l'ampleur de la fourchette ou les limites de confiance, on devra accepter une possibilité plus grande d'erreur. Par exemple, on peut prédire que le résultat à la même élection se situera entre 48 et 54% ($51\% \pm 3$), mais en sachant que l'on a 5 chances sur 100 (les journalistes disent souvent : 1 sur 20) de voir cette prédiction tout à fait erronée. Or, cette probabilité d'effectuer une prédiction erronée est souvent tue, les journalistes insistant surtout sur les limites de confiance présentées comme la possibilité d'erreur de la prédiction, ce qui n'est pas le cas. Enfin, il faut se rappeler, bien sûr, que plus l'échantillon est grand, c'est-à-dire qu'il se rapproche de la population pour laquelle l'estimation est faite, plus la précision et la probabilité d'erreur de la prédiction se réduisent.

Notons au passage qu'en éducation, on effectue souvent les sondages auprès de l'ensemble du groupe. Dans ce cas, on vise la rapidité avec laquelle on peut obtenir une réaction précise à un nombre limité de questions de tous les élèves d'une classe ou d'un niveau, de tous les enseignants ou de tous les parents. Alors les calculs inférentiels n'ont plus aucun sens, puisque c'est l'ensemble de la population qui a répondu. Cela ne signifie pas que le portrait obtenu soit sans interprétation possible : la formulation et le plan de sondage auront certainement contaminé les réponses et il faudra en tenir compte.

De ce qui vient d'être évoqué rapidement, on retient que le sondage pose au moins trois problèmes :

- 1° Le problème statistique de l'inférence et de l'échantillonnage lorsque l'on veut prédire les résultats pour une population à partir d'un échantillon extrait de cette population;
- 2° Le problème du plan de sondage, c'est-à-dire de l'effet, sur le contenu des réponses, du moment et de la manière dont sont administrées les questions;
- 3° La formulation même des questions.

Le premier problème relève d'ouvrages de statistiques comme le précis édité par M. Reuchlin, le troisième problème sera traité dans la section suivante puisqu'il fait partie des difficultés associées aux questionnaires. La question du plan de sondage peut être décomposée en deux éléments : la manière de soumettre les questions et le moment.

1.2 Les techniques de sondage

Trois techniques de sondage sont assez courantes, chacune avec ses avantages et ses inconvénients : d'où un problème de choix.

A. Le *questionnaire écrit*, envoyé par la poste, a l'avantage de laisser aux informateurs le temps de ramasser du matériel et des idées. Il est idéal lorsque l'on veut sonder sur des habitudes qui se déroulent dans le temps (par exemple, les questionnaires Nielsen se rapportant à l'écoute des chaînes de télévision pendant une semaine). Ce type de questionnaire comporte quelques inconvénients. D'abord les consignes doivent être très claires, sans ambiguïté, puisqu'on ne pourra pas discuter avec les informateurs pour vérifier leur compréhension. De ce point de vue, elles doivent absolument avoir été mises à l'essai avant le sondage. Ensuite, la longueur des questionnaires crée un problème de disponibilité : l'informateur ne trouve pas le temps de répondre, il oublie de le faire, ou encore il doit s'interrompre et ne poursuit pas sa tâche. Enfin, ces éléments et des aspects motivationnels non contrôlés (l'intérêt des informateurs pour le thème du sondage) se conjuguent souvent pour aboutir à des taux de réponse assez faibles. Il n'est pas rare qu'un sondage postal n'aboutisse pas à plus de 10 % de réponses. Si l'on veut maximiser le taux de réponse, il faut choisir le face à face.

B. Le sondage peut être effectué en *face à face* : l'intervieweur se rend sur les lieux où il va trouver ses informateurs et leur pose ses questions.

C'est la forme la plus polyvalente d'administration des questions, car elle permet tous les formats de présentation (oraux, écrits, graphiques, avec des listes courtes ou longues). Elle permet d'ajuster directement la séquence des questions en fonction des réponses fournies et des cas de figure présentés par l'informateur et en fonction du temps de disponibilité. Par exemple, si l'informateur est pressé, on ne pose que les questions importantes; l'intervieweur peut aussi sauter à la question «X» après une réponse «non», c'est-à-dire qu'il suit l'embranchement prévu au plan du questionnaire sans que l'informateur s'en aperçoive. Enfin, ce mode d'administration permet d'adapter le vocabulaire et la syntaxe des questions au niveau de langage des informateurs, de leur donner des exemples, de préciser des définitions, si cela s'avère nécessaire et si cela n'interfère pas avec la standardisation des questions. De plus, une telle situation permet à l'intervieweur de motiver son informateur et d'entretenir son intérêt tout au long du sondage. Enfin, cette technique produit généralement un taux assez élevé de réponses : un intervieweur bien formé à l'abordage d'informateurs éventuels peut être pratiquement irrésistible. Mais elle coûte en général assez cher : le déplacement des intervieweurs, la durée des entretiens, le temps d'attente entre chaque informateur sont des éléments qui se payent à un taux horaire d'autant plus élevé que l'intervieweur est expérimenté et que le contenu du sondage est important et délicat à mener.

C. *Le sondage téléphonique* coûte moins cher que le face à face (pas de frais de déplacement) et permet un contrôle direct des enquêteurs par le chercheur (car dans le face à face, il faut faire confiance à l'intervieweur lorsqu'il rapporte, par exemple, qu'il a dû attendre après ses informateurs). Mais les informateurs potentiels peuvent plus facilement refuser de répondre, et la durée du sondage par téléphone ne doit pas durer plus de 20 à 30 minutes si l'on ne veut pas se faire interrompre brutalement. Si le sondage comporte plus que quelques questions, il importe de prévenir l'informateur afin qu'il puisse s'installer confortablement et qu'il puisse vous dire de le rappeler s'il ne dispose pas immédiatement du temps nécessaire.

1.3 Les plans de sondage

Outre le choix de la technique de sondage, le chercheur doit aussi s'interroger sur le plan de sondage autant que le géologue doit le faire : l'endroit et le moment d'extraction d'une carotte de terre influencent évidemment son contenu. Le contexte sociopolitique et historique d'un sondage conditionne l'état d'esprit des informateurs. Leur accès à l'information et leur présence sur le terrain des événements influencent

leur capacité et la richesse de leurs réponses. Leur implication dans ces événements peut biaiser leurs perceptions, et leur rôle peut altérer leur niveau de conscience, même si les questions portent sur des faits plutôt que sur des jugements de valeurs. Enfin, le moment même où s'effectue le sondage n'est pas sans jouer sur la disponibilité et la coopération des informateurs : a-t-on interrompu une tâche importante, les a-t-on pris pendant un temps de loisir ou de liberté, ou encore les retient-on pendant un repas alors que leur humeur à l'égard du sondage se désagrège au même rythme que le mets refroidit dans leur assiette ?

L'examen des effets du plan de sondage doit aussi envisager l'impact des caractéristiques des différents informateurs sur leurs réponses. Par exemple, comme Rosenthal l'a montré à propos des échantillons de volontaires, on sait que certaines personnes sont plus sujettes que d'autres à la *désirabilité sociale*, alors qu'à l'inverse, on rencontre toujours quelques individus qui profitent de ces occasions pour déverser leur bile et prendre le sondeur comme témoin de leurs accusations devant un tribunal. Dès lors, avant d'interpréter les réponses, il est utile de s'interroger sur la proportion d'informateurs qui manifestent à l'avance des propensions, des biais à l'égard de certaines réponses. Afin d'en faire une estimation, le sondeur peut avoir avantage à incorporer quelques questions qui visent à mesurer le désir de conformité et la sincérité des informateurs.

2 Les questionnaires et les échelles²

Les questionnaires et les échelles constituent des techniques en apparence assez simples, utilisées pour presque tout, et qui donnent... ce qu'on y a mis ! À la différence des entrevues, les questions d'un questionnaire ou d'une échelle proposent aux sujets interrogés de répondre selon un format déterminé. Il s'agit donc d'instruments produisant des données provoquées.

2 Pour plus de détail, voir :

Blais A., Construction et administration du questionnaire. Chapitre 14 dans B. Gauthier (Éd.), *Recherche sociale*. Sillery, QC : PUQ, 1984. p. 340-357.

Ferland F., La mesure des attitudes. Chapitre 15 dans B. Gauthier (Éd.), *Recherche sociale*. Sillery, QC : PUQ, 1984.

Selltiz C., Wrightsman L.S et S.W. Cook, *Les méthodes de recherche en sciences sociales*. Traduction D. Bélanger, Montréal, QC : HRW, 1977. Voir particulièrement le chapitre 9 : La récolte des données II, questionnaires et interviews.

Kornhauser A. et P.B. Sheatsley, L'élaboration des questionnaires et les techniques d'interviews, Appendice B dans Selltiz C., Wrightsman L.S et S.W. Cook, *Les méthodes de recherche en sciences sociales*. Traduction D. Bélanger, Montréal, QC : HRW, 1977.

Le format de réponse proposé peut prendre plusieurs formes, parmi lesquelles on rencontre les suivantes (cette liste n'est pas exhaustive) :

- L'alternative : oui-non, vrai-faux;
- Le choix parmi une liste plus ou moins longue, comme dans les questions à choix multiples des examens;
- Le choix parmi une liste d'énoncés ordonnés selon un principe (par exemple, selon une chronologie ou allant du général au particulier);
- Le choix d'un niveau sur une échelle de valeur (par exemple : excellent, très bien, bon, passable, insuffisant, nul) ou de fréquence (par exemple : jamais, parfois, souvent, très souvent, toujours);
- Le choix d'une position spatiale sur une échelle bi-polaire; par exemple :

ouvert					X			fermé
chaud		X						froid

- Le classement d'un certain nombre d'éléments (par exemple, des énoncés, des photos) dans un certain nombre de piles dont les effectifs sont déterminés à l'avance. La distribution des effectifs dans les piles peut prendre une forme que le chercheur détermine en fonction de ses objectifs (technique appelée *Q sort*). Par exemple, classer 25 énoncés, selon qu'ils correspondent tout à fait ou pas du tout à la perception du sujet, mais l'obligeant à les répartir de telle sorte que le nombre d'énoncés placés dans les cases corresponde au schéma suivant (courbe normale) :

tout à fait :	1	3	7	12	7	3	1	: pas du tout
---------------	---	---	---	----	---	---	---	---------------

Cela force le sujet à isoler les éléments qui sont vraiment représentatifs de sa perception, puisqu'il ne peut en mettre qu'un à chaque extrémité. À l'opposé, lorsque l'on souhaite forcer le sujet à opérer une dichotomie en l'empêchant de se réfugier dans une position centrale, on peut le contraindre à un schéma du genre (courbe en U) :

tout à fait :	8	4	2	1	2	4	8	: pas du tout
---------------	---	---	---	---	---	---	---	---------------

Ces exemples de format imposé aux réponses indiquent que la signification des réponses dépendra en partie du format, qui constitue en lui-même une situation ou un contexte psychosocial contraignant le sujet.

Le format de la réponse sera choisi en fonction de l'information recherchée, des contraintes que l'on veut imposer aux sujets et du traitement que l'on veut appliquer aux données, puisque le format imposé aux réponses rend possible certain type de calculs. Ainsi, étant donné que le premier exemple de technique *Q sort*, illustré ci-dessus, impose une forme de distribution normale aux réponses, cette forme permet d'effectuer, avec plus de légitimité, les opérations qui présupposent une telle distribution, comme certains tests statistiques paramétriques.

Mais la signification attribuable aux réponses dépend aussi de la forme et du contenu de chaque question. En ce qui concerne la forme, le chercheur devra être attentif à ce que le niveau de vocabulaire et la complexité syntaxique de ses questions permettent une compréhension non équivoque : ils devront donc être ajustés au niveau culturel et au type de langage coutumier des informateurs potentiels. Cela implique que la formulation des questions soit mise à l'essai, non seulement auprès d'experts, mais aussi auprès d'individus typiques de la population-cible. Le contenu des questions dépendra évidemment de l'objet de la recherche, des indices reconnus comme pertinents par rapport à l'information recherchée, de la volonté de vérifier par des recoupements la stabilité des réponses des sujets (ce qui implique de poser plusieurs fois des questions portant sur un même contenu), et de contrôler les biais possibles comme la désirabilité sociale, le conformisme, la sincérité, le mensonge, etc.

Enfin, il ne faudra pas négliger la mise en page du questionnaire selon deux points de vue qui appellent parfois des compromis : celui du sujet qui doit répondre et celui du codeur qui devra traiter les réponses.

Du point de vue du sujet, il faudra imaginer une présentation

- qui réduise les ambiguïtés (par exemple, en regroupant les questions qui se rapportent à un même secteur d'activité, qui touchent le même contexte, et en indiquant clairement les changements de contexte ou de sens d'une situation);
- qui simplifie sa tâche (par exemple, ne pas l'obliger à faire des calculs en lui demandant de situer sa réponse dans une tranche d'âge, mais lui demander sa date de naissance... le codeur effectuera la transformation);
- qui ne surcharge pas sa mémoire (par exemple, en reprenant et en maintenant sous ses yeux la liste des éléments entre lesquels il doit choisir, même si cette liste se répète pour plusieurs questions);

- qui ne lui demande pas des mouvements inutiles et parasites, comme de parcourir une liste et de devoir remonter en haut de la page pour inscrire sa réponse ou de devoir tourner la feuille pour l'inscrire au verso.

Du point de vue du codeur, il faudra prévoir une marge à droite et un alignement de l'emplacement des codes avec la position des réponses de façon à éviter les erreurs de codage. Grâce au traitement informatisé par lecteur optique, on peut éviter de devoir faire effectuer par le codeur les renvois de page pour regrouper les réponses aux questions qui se croisent (contrôle des biais) en insérant, dans le programme, des instructions qui réalisent automatiquement ces recoupements. De la même manière, on peut plus aisément demander aux sujets les informations d'identification en fin de questionnaire (on y répond plus facilement lorsque le gros de la tâche est terminé) et faire transférer, par le programme, ces informations en tête de fichier.

3 Les tests³

3.1 Les caractéristiques générales

Deux grands types de tests peuvent être considérés : les *tests critériés* et les *tests étalonnés*. Leur différence réside dans le type de référence à laquelle on va confronter les réponses du sujet qui a passé le test, afin de leur attribuer une valeur. Dans le cas des tests critériés, la référence est déterminée, le plus souvent, par consensus, et correspond aux exigences, estimées minimales ou optimales, pour accomplir une tâche, pour remplir une fonction ou pour accéder à une situation éducative ou professionnelle.

3 Pour plus de détail, voir :

Cronbach L.J., *Essentials of psychological testing*. New-York : Harper and Row. 4e édition, 1984. (Un classique parmi de trop nombreux ouvrages).

Selltiz C., Wrightsman L.S et S.W. Cook, *Les méthodes de recherche en sciences sociales*. Traduction D. Bélanger, Montréal, QC : HRW, 1977. Voir particulièrement le chapitre 10 : La récolte des données III, les méthodes projectives et autres méthodes indirectes.

Buros Institute of Mental Measurement : *The Ninth Mental Measurement Yearbook*, Lincoln, Nebraska, 1985; *The Supplement of the Ninth Mental Measurement Yearbook*, Lincoln, Nebraska, 1988. (Ces ouvrages fournissent une présentation, une revue critique des qualités psychométriques et une bibliographie pour les principaux tests) *Test Critiques Compendium* : Reviews of Major Tests from the Test Critiques Series. Kansas City, Mo : Test Corporation of America, 1987. (Fournit une évaluation de l'usage des tests les plus populaires par des professionnels).

Dans le cas des tests étalonnés, la référence est constituée par les performances d'un échantillon représentatif de la population-cible, par rapport à laquelle on veut situer les réponses du ou des sujets testés. Dans les deux cas, la référence n'est jamais définie une fois pour toute : les exigences pour accéder à un enseignement ou à une fonction peuvent changer avec la mutation des pratiques professionnelles, et les caractéristiques d'une population suivent les changements socio-culturels qui l'affectent.

Qu'il soit critérié ou non, un test est toujours une *situation simulée* (un modèle réduit en complexité, en durée, en réalisme). Il est composé d'une série de consignes provoquant des activités. La réalisation de celles-ci permet la comparaison des variations comportementales du sujet ou des sujets, confrontés à cette situation simulée, avec

- a) les comportements indices du critère de référence dans le cas des tests critériés, ou
- b) les variations comportementales du groupe de référence choisi (groupe étalon) obtenues dans la même situation simulée, dans le cas des tests étalonnés.

Comme instrument de mesure, le test doit être une situation simulée parce que l'opération de mesure implique

- un choix d'items parmi l'ensemble de ceux qui sont des indices pertinents de la caractéristique que l'on veut évaluer à travers ce modèle réduit de la situation réelle. En effet, on utilise un test parce que la situation réelle est inobservable dans son intégralité par manque de temps, par manque de moyens, pour des raisons déontologiques ou parce que l'observation complète en perturberait le déroulement;
- un contrôle des variables parasites et des variables non pertinentes, c'est-à-dire des autres caractéristiques de la situation ou du sujet qui peuvent interférer ou cacher la caractéristique que l'on veut évaluer;
- la constance des conditions dans lesquelles les mesures sont effectuées, car aucune comparaison n'est possible sans cette constance des conditions. De plus, la constance des conditions de l'application du test est un préalable à l'homogénéité des composantes de la mesure, homogénéité sans laquelle aucun calcul arithmétique ne peut être effectué. Or, c'est surtout par des calculs que les comparaisons pourront aboutir à des résultats manifestant, avec une certaine évidence, les relations que la recherche tente de découvrir.

Dès lors, l'interprétation des résultats, obtenus à partir de tests, doit tenir compte que ces données ont été obtenues à partir d'une situation simulée, c'est-à-dire artificielle, et non à partir d'une situation réelle, c'est-à-dire non manipulée et non réduite. Cela signifie qu'un écart subsistera toujours entre ce qui est possible de dire sur les résultats du test, et ce que l'on souhaite pouvoir dire à propos de la situation réelle que le test simule. Il en découle aussi que l'utilisateur des tests doit être attentif à certaines exigences méthodologiques que les tests doivent respecter.

3.2 Les qualités exigées des tests

Classiquement, on attend d'un test de qualité qu'il réponde à six exigences : il doit être fidèle, valide, étalonné ou critérié (selon le cas), fiable, hiérarchisé, discriminant et standardisé. Les concepts de fidélité, de validité et de fiabilité touchent toutes les données empiriques; elles ont été abordées au chapitre cinq. En ce qui concerne les tests, le développement de la psychométrie leur a conféré des significations particulières, techniques, dues aux méthodes de calcul des indices de ces caractéristiques d'un bon test. C'est seulement la signification méthodologique de cet aspect métrique qu'on examinera ici. Les développements plus récents à partir d'élaborations de la théorie de la mesure fondées sur les statistiques ne peuvent être abordés dans ce cadre; des publications spécialisées leurs sont consacrées⁴.

3.2.1 La fidélité

Sous le terme de *fidélité* on rencontre habituellement au moins quatre types d'indices qui n'ont pas du tout la même signification. De plus, cette notion a reçu une extension, tant sur le plan technique que conceptuel, avec le concept de *généralisabilité*.

A. Les indices de *stabilité* ou de *constance* indiquent, selon leur nom, si une mesure reste stable avec le temps. Lorsque la variable mesurée est censée ne pas évoluer avec la maturation, la stabilité est une caractéristique

4 Pour les *développements récents du testing*, voir :

Bejar I.I., *Achievement Testing*. Recent Advances. Newbury Park, CA : Sage, 1983.

Blais J-G et D. Ajar, Théorie des réponses aux items et modélisation. *Mesure et évaluation en éducation*, XIV, n°4, 1992, p.5-18.

Cardinet J. et Y. Tourneur, *Assurer la mesure : guide pour les études de généralisabilité*. Berne : Petre Lang, 1985.

Cardinet J., *Évaluation scolaire et mesure*. Bruxelles : De Boeck, 1986.

Weiss D.J. (éd.) *New Horizons in Testing*. Latent Trait Test Theory and Computerized Adaptive Testing. New-York : Academic Press, 1983.

importante. En effet, la mesure ne peut ni se dégrader ni s'amplifier par elle-même, si elle doit servir à mesurer l'effet d'une intervention sur la caractéristique mesurée. Les résultats à un tel test devront donc résister au temps (extinction et maturation) et à la reprise (apprentissage du test). C'est pourquoi le calcul de la stabilité s'effectue en comparant les résultats à un test et à un retest des mêmes sujets avec la même épreuve dans les mêmes conditions.

B. Les indices d'*équivalence* sont obtenus selon un principe semblable à celui de calcul de fidélité inter-observateur. Comme on s'attend à ce que l'observation du même phénomène par deux observateurs produise des rapports d'observation semblables, on s'attend aussi à ce que la même opération mentale appliquée sur des données différentes mais considérées comme équivalentes (de même type) produise des résultats semblables. Autrement dit, si un sujet quelconque maîtrise l'habileté à effectuer une opération telle qu'une multiplication sur des données comme 2 et 7 ($2 \times 7 = 14$), on s'attend à ce qu'il soit capable d'effectuer la même opération sur des données semblables comme 2 et 6 ou 2 et 8 et à fournir la réponse exacte. La méthode des coefficients d'équivalence consiste donc à construire des tests parallèles, soit des épreuves qui comportent les mêmes opérations appliquées à des données équivalentes et à comparer (souvent par corrélation) les résultats obtenus. Par analogie avec les plans de recherche (chapitre 9), on dira que la stabilité implique la similitude des résultats dans le temps, alors que l'équivalence implique la similitude des résultats dans l'espace.

C. Les coefficients d'*homogénéité* et de *consistance interne* se préoccupent de la similitude des résultats obtenus à l'intérieur d'un même test. En effet, si un test vise, par une série d'activités, à obtenir une mesure approximative d'une habileté déterminée, on doit s'attendre à ce que les résultats, obtenus dans chacun des sous-ensembles d'activités qui constituent le test, présentent une certaine convergence. Sinon, on peut craindre que les sous-ensembles dont les résultats divergent, mesurent en fait d'autres habiletés ou soient contaminés par des variables parasites indésirables; auquel cas la mesure fournie par l'ensemble du test comporte une part d'erreur due à ce sous-ensemble divergent.

Deux techniques ont surtout été utilisées. La première consiste à diviser le test en deux parties et à comparer les résultats obtenus. Le partage s'est le plus souvent fait par moitiés, par items pairs et impairs, ou par hasard. L'autre technique procède en comparant les variances mesurées pour chaque item entre elles et avec la variance calculée sur la totalité du

test. En effet, on peut logiquement s'attendre à ce que la position de chaque sujet ayant passé le test par rapport à l'ensemble du groupe des sujets reste semblable d'item à item si chaque item mesure la même opération. Autrement dit, s'il s'agit de la même habileté, un sujet faible restera faible d'item à item et l'écart entre sa faiblesse et la performance d'un sujet habile devrait aussi rester semblable d'item à item. Savoir si le test a une bonne consistance interne, revient à savoir si chacun des items qui le constituent contribue bien à une approximation de la même habileté, de la même opération ou de la même caractéristique.

D. *La généralisabilité* constitue un raffinement technique de l'approximation des erreurs de la mesure basé sur l'analyse de la variance. À partir d'un plan expérimental, on essaie d'isoler, pour un test donné, la part des variations (confondues, dans ce cas, avec les erreurs) attribuable aux diverses sources possibles d'erreurs, comme le test lui-même, le sujet testeur, le sujet testé, les conditions d'application du test, etc. L'hypothèse à la base de cette conception de la généralisabilité est que, si l'on connaît mieux quelle est la part de l'erreur due à chacune des sources, on peut mieux connaître ce que serait la mesure lorsque l'une ou plusieurs de ces sources d'erreurs sont contrôlées. En postulant que la part d'erreur due au testeur puisse être vraiment estimée, on pourrait déterminer quelle serait la mesure si l'influence du testeur n'existait pas. D'où l'idée de généralisation : si toutes les sources d'erreurs pouvaient être isolées et contrôlées, on pourrait avoir une vraie mesure (idéale). Resterait à discuter de la réalité de l'estimation d'une mesure ainsi épurée.

3.2.2 La validité

La validité des tests est une exigence que certains auteurs considèrent plus importante et englobant la fidélité, mais on ne dispose pas d'un développement métrique aussi raffiné pour cette notion. L'estimation de la validité reste une affaire de jugement auquel bien peu d'indices psychométriques apportent des fondements. En effet, si la fidélité est une estimation du rapport entre la mesure et le réel dont la mesure constitue un indice, la validité est une estimation du rapport entre cet indice d'un événement du réel et les concepts avec lesquels on parle de cet événement. Si le rapport entre un indice observable et un événement du réel observable est de fait mesurable, il en est tout autrement du rapport entre un indice observable et une idée, un concept; ce dernier échappant par définition à toute observation. Quatre types de validité sont classiquement envisagés : les validités (a) de contenu, (b) de concordance, (c) prédictive et (d) conceptuelle.

A. La *validité de contenu* ou d'apparence est estimée en considérant la parenté ou la similitude observable entre les opérations exigées par le test et les opérations supposément exigées par la situation que le test simule. Par exemple, on considère sur cette base qu'un test comportant des opérations arithmétiques est valide comme test d'arithmétique. Cela peut paraître évident lorsque les consignes du test ne comportent expressément que des opérations (du genre : $7 + 3 = ?$), mais cela l'est beaucoup moins lorsque les opérations à effectuer sont commandées par la résolution d'un problème exposé, non plus par des seuls symboles arithmétiques, mais à travers une histoire, un récit censé représenter une situation problématique. Dans ce cas, si la solution du problème exige bien l'habileté à exécuter les opérations arithmétiques, elle exige aussi l'habileté à comprendre le récit, à en dégager les propositions principales et à les traduire en symboles arithmétiques utilisables dans l'exécution d'une opération ou de plusieurs opérations. La validité de contenu n'est pas toujours aussi évidente que cela, loin s'en faut.

B. La *validité de concordance* (dite aussi validité immédiate) résulte de la comparaison des résultats du test avec le jugement obtenu par d'autres moyens, comme un examen scolaire, une entrevue, une observation clinique, un autre test dont on a déjà pu apprécier la validité. Tout dépend alors de la corrélation entre le test et le jugement, de la confiance attribuable au jugement obtenu par cet autre moyen et de l'indépendance entre le testeur et l'autre juge. Une étude célèbre conduite par Delay, Pichot et Perse⁵ à propos de la validité du test de Szondi, a bien montré que si la personne qui voit le sujet pour lui appliquer le test est celle qui fait aussi l'interprétation du test (situation fréquente dans la pratique), sa sensibilité clinique, lors du test, peut être suffisante pour contaminer son interprétation en faveur de la validité de concordance du test. Mais si le spécialiste qui applique le test diffère à la fois de celui qui interprète les résultats du test et de celui qui a effectué l'entrevue permettant de classer le sujet, alors les biais perceptifs sont contrôlés. Dans le cas en question, une validation en *double aveugle* a bien montré que le test n'était pas valide. Les épreuves de validation par concordance doivent donc être menées en aveugle; c'est une exigence minimale peu souvent rencontrée.

C. La *validité prédictive* consiste à mettre en relation les résultats au test avec la performance en situation réelle que le test est censé prédire. Ce type de validité est évidemment beaucoup plus puissant que les précédents,

5 Delay J., Pichot P. et J. Perse, *Méthodes psychométriques en clinique*. Paris : Masson, 1966. Voir en particulier le chapitre XVI : La validité du test de Szondi en clinique, p. 297-313.

mais il est plus difficile encore à mettre en oeuvre. En effet, dans certains cas, la situation réelle n'est que partiellement accessible à l'observation alors que, dans d'autre cas, la situation est d'une telle complexité que les critères d'observation de l'habileté que l'on veut identifier sont difficiles à préciser et à isoler.

D. La *validité conceptuelle* (ou de construction, en référence au constructivisme conceptuel) consiste en une évaluation simultanée du test et de la théorie (concept) qui a permis de construire le test et dont le test représente une opérationnalisation. Bien que la plus appréciée par les théoriciens, la validité conceptuelle paraît aux empiristes comme une validation tautologique : si la théorie du concept est cohérente et consistante, et si le test mesurant le concept est strictement fondé (déduit, construit à partir de la théorie), alors toutes les mesures effectuées avec le test dans des conditions différentes doivent correspondre aux prédictions que la théorie fournit pour ces conditions. L'échec des prédictions déduites de la théorie signifie que, ou bien la théorie a des problèmes de cohérence ou de consistance, ou bien le test n'est pas une opérationnalisation valide des concepts de la théorie, ou bien tant le test que la théorie ont des défauts. La validation conceptuelle s'effectue par la comparaison de tous les classements ou de toutes les prédictions que l'on peut faire en se basant sur l'ensemble des propositions théoriques se rapportant au concept envisagé. Techniquement, elle se décompose en deux volets : la convergence et la différenciation. Lors d'une validation conceptuelle par convergence, les différentes mesures du même concept doivent donner des résultats similaires. Lors d'une validation par différenciation, toutes les mesures du concept doivent permettre de le différencier des autres concepts.

3.2.3 L'étalonnage ou la référence

L'étalonnage (anglais : standardisation) consiste à construire l'échelle de référence à laquelle toutes les mesures individuelles seront comparées. Cette exigence est semblable pour les tests critériés. Trois problèmes principaux affectent l'étalonnage; ils permettent de longues discussions et ne se résolvent que par des consensus provisoires. Ce sont

1. La nature et l'origine de la base de référence (groupe pour un étalonnage, théorie pour l'établissement de critères);
2. La représentativité de l'échantillon fournissant les paramètres de l'étalon ou l'exhaustivité de la théorie précisant les critères;

3. La relativité spatio-temporelle des critères et des étalons, puisque les normes à partir desquelles les groupes sociaux et professionnels se définissent évoluent avec les mutations politiques et culturelles.

3.2.4 La hiérarchisation

La hiérarchisation des items d'un test résulte d'une exigence pratique et déontologique autant que théorique. Si les items sont hiérarchisés de telle sorte que les plus faciles précèdent les plus difficiles, on peut penser que si le sujet échoue à une série d'items consécutifs, il n'est pas nécessaire de lui faire perdre son temps et son estime de soi, en le confrontant à des items auxquels on peut prévoir qu'il échouera. Sur le plan métrique, la hiérarchisation a aussi des conséquences intéressantes. Si le degré de difficulté des items est organisé de telle sorte qu'on puisse observer entre eux une variation systématique, alors on peut construire une échelle de mesure au moins ordinale, comportant des rangs. De plus, si la connaissance des écarts entre les items sur l'échelle de difficulté est précise, on peut pondérer la valeur accordée à chaque item, de façon à rendre les écarts entre items égaux par rapport aux variations du concept que l'on mesure. De la sorte, l'échelle obtenue est considérée comme ayant des intervalles égaux. Enfin, la connaissance du degré de difficulté des items permet aussi des classements qualitativement plus justes, parce que l'évaluateur peut tenir compte de la valeur relative de chaque item dans le jugement global qu'il porte sur une performance. À cette fin, différents indices de difficultés ont été proposés; la plupart se basent sur la proportion des sujets qui réussissent les items.

3.2.5 La discrimination

La capacité de discrimination s'intéresse à la valeur du classement effectué pour le test. Par exemple, si un test est censé identifier un état mental comme la névrose d'échec, la question sera de savoir si la mesure produite par le test permet de bien classer ceux qui de fait, à partir d'un critère admis, ont une névrose et ceux qui n'en ont pas. La figure la plus simple pour examiner la discrimination se résume à une table à double entrée, comme dans le tableau 15.1. En poursuivant l'exemple, les FV du tableau 15.1 sont les sujets qui ne manifestent pas de névrose d'échec mais que le test classe comme névrotique, alors que les VF sont les sujets névrotiques qui ne sont pas identifiés par le test.

		Classement obtenu à partir du test	
		Vrai	Faux
Classement de fait	Vrai	VV	VF
	Faux	FV	FF

Tableau 15.1 : Table de discrimination des classements.

Le dilemme réside dans le nombre et le genre de faux classements que l'on accepte de tolérer, car la qualité discriminative se paie. En effet, comme il est dans la plupart des cas impossible d'obtenir une discrimination parfaite (VF et $FV = 0$), on doit se résoudre à décider quelle erreur on peut tolérer, soit les VF soit les FV , et de retravailler en conséquence la construction du test. Mais la tolérance des VF ou des FV est une question éthique, sociale et parfois économique.

Admettons un test de maturité scolaire ayant la prétention de prédire le succès en première année primaire et, en contrepartie, d'identifier les enfants qui auraient besoin d'une année préparatoire supplémentaire. Quel type d'erreur va-t-on tolérer ? Si l'on accepte des VF (matures, immatures au test), on se résout à faire perdre un an à des enfants qui sont prêts pour l'école mais que le test consacre comme immatures, alors que si l'on accepte des FV (immatures, matures au test), on fait monter en première année des enfants qui ne sont pas prêts, qui vont échouer et qui auraient bénéficié d'une année supplémentaire de préparation. Il y a là un choix de valeur qu'il faut trancher et pour lequel il n'y a pas de réponse toute faite, indépendamment des circonstances, de la variable mesurée et de l'impact humain et social des décisions qui seront prises.

Les deux cas pris en exemple simplifient le problème dans la mesure où le classement était dichotomique. Dans la plupart des cas, il s'agit de déterminer un seuil sur une échelle, tant du côté du test que du côté du critère. De plus, l'analyse porte aussi sur la capacité discriminative de chacun des items qui composent le test. En effet, si la contribution d'un item à la capacité discriminative du test (un examen scolaire, par exemple) est faible ou même négative (les bons et les mauvais réussissent de manière aléatoire), il faut alors se poser des questions sur le rôle de cet item dans cet test. D'autres analyses (analyses factorielles par exemple), en complément avec l'examen de l'homogénéité et de la consistance interne, pourront apporter des pistes d'interprétations si l'on veut conserver l'item et mieux comprendre ce qu'il mesure exactement.

3.2.6 La standardisation

Cette sixième caractéristique d'un bon test répond plus à des exigences méthodologiques de base qu'elle ne correspond à un paramètre : pour qu'un instrument produise une mesure fiable, fidèle, valide, etc., il est indispensable que cet instrument soit toujours appliqué dans des conditions identiques. Quand il s'agit d'un test, cela signifie que les modalités d'application (passation), de correction et d'interprétation doivent faire l'objet d'une convention contraignante qui définit de manière précise, sans ambiguïté, toutes les opérations qui seront effectuées à chacune de ces étapes.

4 La manipulation des contingences⁶

La manipulation des contingences peut être définie comme la modification planifiée des conditions de l'environnement physique, social ou psychologique dans lesquelles un ou des sujets sont amenés à poser une action.

C'est souvent dans le cadre de chapitres sur les plans expérimentaux que le problème de la manipulation des contingences est traité. Or, il nous semble important de distinguer la technique de constitution des données, soit la manipulation des contingences, des méthodes d'organisation de la constitution, qui ont un tout autre objectif, celui de mettre en place une argumentation permettant de formuler ou de vérifier des hypothèses. Par ailleurs, les plans de recherche peuvent très bien fonctionner sans qu'il y ait manipulation, car un bon nombre de facteurs faisant l'objet de recherche, interviennent naturellement, hors du contrôle des chercheurs. Réciproquement, les techniques de manipulation des contingences ne sont pas liées à l'un ou l'autre type de plan de recherche : on peut manipuler les contingences sans qu'il y ait expérimentation. Nous examinerons ici seulement le problème de la manipulation des contingences en rappelant que le chapitre 9 a été consacré aux plans de recherche et aux stratégies de conviction.

⁶ Pour plus de détails, voir :

Léon A., Cambon J., Lombroso M. et F. Winnikainen, *Manuel de psychopédagogie expérimentale*. Paris, PUF, 1977.

Robert M. et A. Fortin, Plans de recherche classiques, quasi expérimentaux et à cas uniques. Chapitres 5, 6 et 7 dans Robert M. (éd.), *Fondements et étapes de la recherche scientifique en psychologie*. Montréal, QC : Chenelière et Stanké, 1982.

Par manipulation des contingences, on entend toutes les techniques utilisées par les chercheurs pour introduire artificiellement un ou plusieurs facteurs dans l'environnement des sujets, de manière à observer l'effet de la présence de ces facteurs sur leurs conduites. Il s'agit d'une introduction artificielle, dans la mesure où l'apparition du facteur n'est pas spontanément dépendante de l'environnement des sujets. Cela veut dire que tout changement auquel les sujets sont soumis, qui a été décidé et planifié par d'autres personnes, est considéré comme une manipulation des contingences. Dans le cas des manipulations expérimentales, l'intervention est décidée et planifiée par le chercheur; dans le cas des manipulations non expérimentales, l'intervention peut être décidée à un autre niveau, politique par exemple. On trouve une illustration de ce genre de recherche dans les travaux portant sur l'effet d'une décision administrative, telle un changement de grille horaire ou une réduction budgétaire : les sujets (enseignants, élèves) y subissent une modification de leur environnement qui a été décidée et planifiée par les administrateurs ou les politiques, et leurs réactions à cette manipulation des contingences de leurs actions peuvent être étudiées par les chercheurs.

La manipulation des contingences pose au moins deux problèmes : celui des consignes par lesquelles on demande aux sujets d'agir et celui du rapport entre la modification projetée des contingences et sa perception par les sujets.

Le problème des consignes est double : la formulation et le contenu explicite d'une part, les consignes implicites d'autre part. On trouvera évidemment banal d'insister sur le niveau de vocabulaire utilisé dans la consigne : il importe que les sujets comprennent exactement ce qu'on leur demande de faire. Les chercheurs doivent se méfier des termes techniques qui leur paraissent évidents mais qui ne veulent rien dire ou, pire, qui veulent dire autre chose pour les sujets. Les malentendus existent aussi entre sujets et chercheurs dans les manipulations, et certains résultats très curieux s'expliquent parfois par le fait que le sujet a très bien compris... autre chose que ce qu'on voulait lui faire comprendre.

4.1 La formulation des consignes

Certains travaux, dont celui de Ska (1983)⁷, donnent des indications précieuses au sujet de la formulation des consignes. Ainsi elle note, comme on pouvait s'y attendre, que plus le texte de la consigne est long, plus le

7 Ska B., Mémoire des consignes et performance à la tâche. *Psychologica Belgica*, XXIII-1, 1983, p. 57-71.

temps de lecture total est long. Mais elle observe que le temps de lecture augmente lorsque le nombre de propositions explicitées diminue, c'est-à-dire lorsque tout n'est pas dit explicitement dans la consigne. Le fait que tout ne soit pas dit conduit le sujet à effectuer des inférences à propos de ce qui est resté implicite. Or, ce travail d'explicitation que le sujet doit faire lui-même, par inférence, pour compléter le texte, lui demande un surcroît de temps qu'il faut lui donner si l'on veut qu'il soit efficace. De plus, lorsque les consignes sont complètes, c'est-à-dire lorsqu'elles disent au sujet tout ce qu'il doit savoir, l'exécution de la tâche est plus rapide : la représentation que les sujets se font de la tâche est donc moins efficace lorsque les consignes ne sont pas complètes et explicites.

Par ailleurs, la formulation syntaxique a aussi un effet sur la tâche. Ska définit la formulation syntaxique simple comme composée de phrases ne comportant qu'une seule proposition, alors que dans la formulation complexe chaque phrase comporte plusieurs propositions reliées. Ainsi la phrase suivante est complexe : «Le jeu que vous allez faire consiste à deviner le dessin de deux figures que vous ne connaissez pas.» La formulation simple du même énoncé exige quatre phrases : «Vous allez faire un jeu. Le jeu consiste à deviner un dessin. Vous devinerez le dessin de deux figures. Vous ne connaissez pas ces figures.»

Cette distinction faite, Ska constate que la formulation syntaxique simple entraîne des temps de lecture plus longs, car elle exige plus de temps pour construire une représentation complète de la tâche. Mais de plus, les lecteurs des consignes syntaxiquement simples font plus d'erreurs dans l'exécution de la tâche. Les séquences, plus nombreuses et non reliées entre elles au niveau de la présentation des instructions, entraînent des essais plus nombreux, qu'en outre les sujets ne relient pas entre eux dans l'exécution de la tâche. Un trop grand morcellement de la consigne entraîne une représentation de la tâche elle aussi morcelée, au point que certains sujets ne retiennent même pas les erreurs qu'ils ont déjà faites et qu'ils les reproduisent. «Ce que l'on dit et la façon dont on le dit ont donc une influence sur l'exécution de la tâche prescrite» et la performance des sujets dans une recherche ne dépend pas seulement de leurs capacités, mais aussi de la formulation des consignes que le chercheur leur a données.

4.2 Les consignes explicites et implicites

Mais, en plus de l'effet des manières de formuler les consignes, le chercheur doit aussi être attentif aux consignes implicites qu'il fournit aux sujets, entre autres par le ton de sa voix, par ses attitudes, par la manière

dont il installe le sujet et présente la recherche, et par le dispositif d'observation ou d'enregistrement des comportements du sujet. En effet, tout sujet engagé dans une recherche tente de comprendre ce que le chercheur veut réellement qu'il fasse, que ce soit pour bien le faire, lorsqu'il est en position de collaboration, ou que ce soit pour saboter la recherche, lorsqu'il se sent exploité et qu'il veut résister. Le sujet tente donc de savoir ce que le chercheur a derrière la tête au-delà des mots utilisés, et cette attitude est à la source du problème des consignes implicites. Une recherche rapportée par Ghiglione (1988)⁸ illustre la dépendance des sujets par rapport aux intonations du chercheur. Dans une situation d'entretien non-directif classique, lorsque l'intervieweur répète la dernière proposition de l'interviewé en respectant toujours la même forme syntaxique (sujet-verbe-complément), mais en déplaçant l'intonation sur le sujet, le verbe ou le complément, on constate que l'interviewé modifie le thème de ce qu'il dit ensuite. Ainsi, rapporte Ghiglione, si l'intonation est placée sur le sujet de la proposition, ce terme prendra une place accrue dans le discours qui suit, alors que si l'accent tonique est placé sur le verbe, le discours subséquent accordera plus d'importance au sujet et au complément. Mais si l'accent tombe sur le complément, l'interviewé place le complément en position de sujet (changement de mode : actif-passif) pour lui donner plus d'importance. Enfin, lorsque l'intervieweur parvient à répéter la phrase sans intonation, l'interviewé témoigne d'un certain désarroi, il hésite et son discours devient chaotique : il ne sait que dire, n'ayant pas perçu d'indice quant à la demande d'information de l'intervieweur.

L'importance des consignes implicites, telle qu'illustrée par cette recherche de Ghiglione, pose le problème de savoir à quelle consigne le sujet de la recherche a réagi. D'où une stratégie utilisée par plusieurs : l'entrevue post-expérience dans laquelle on demande aux sujets de dire quelle a été leur perception de la demande formulée par le chercheur et quelles pourraient être les hypothèses que le chercheur voulait explorer par la tâche qu'ils viennent de réaliser.

Le contenu de la tâche qui sert de support ou de prétexte à la manipulation des contingences et dont l'exécution sera observée et analysée pour y déceler l'effet de la manipulation, doit aussi présenter certaines caractéristiques pour être efficace. La tâche doit évidemment avoir un

8 Ghiglione R., Une méthode d'analyse de contenu : l'analyse propositionnelle du discours. Dans Painchaud G. et M. Anadon, Conceptions et pratiques de l'Analyse de contenu. *Revue de l'Association pour la recherche qualitative*, Vol.1, 1, 1988, p. 82-117.

minimum de rapport avec les situations réelles dont elle constitue une simulation, lorsque l'observation ne porte pas sur des tâches naturelles. Elle doit aussi être sensible à la modification des contingences, et d'un niveau de difficulté tel que cette sensibilité pourra se manifester. En effet, pour pouvoir observer l'effet d'une modification dans les conditions d'exécution d'une tâche, il importe que l'exécution de la tâche puisse être affectée par la modification des conditions. Or, il est deux situations typiques où la tâche paraît insensible : lorsqu'elle est trop difficile pour les sujets et lorsqu'elle est trop facile. Dans le cas où la tâche est trop difficile, les sujets ne sont pas capables d'y manifester une performance suffisante pour pouvoir observer les différences dues à la manipulation. On se trouve alors devant un *effet plancher*. À l'inverse, lorsque la tâche est trop facile, la performance est parfaite dans toutes les conditions et l'on ne peut pas plus observer les effets de la manipulation; on se trouve alors dans le cas d'un *effet plafond*.

Enfin, on rencontre aussi des situations où une condition parasite l'expérience et rend la tâche résistante à la manipulation. C'est particulièrement le cas lorsqu'une consigne implicite cache aux sujets la manipulation de recherche et suscite une modification des conditions non voulue par le chercheur. Ainsi, dans une recherche portant sur l'apprentissage de la compensation chez des enfants handicapés mentaux, le chercheur voulait induire un apprentissage par des jeux variés de consignes et de feedback sans s'apercevoir que, chez ces enfants-là, la relation nouvelle créée entre eux et le chercheur était plus importante que l'apprentissage de la tâche. En conséquence, puisque les enfants se concentraient sur la relation nouvelle à l'expérimentateur et non sur l'apprentissage, le chercheur ne put percevoir l'effet de la manipulation. Les enfants du groupe expérimental, qui avaient semblé apprendre lors du cycle d'apprentissage, ne présentèrent pas les effets attendus lors du post-test. Par contre, les enfants du groupe contrôle qui découvraient l'expérimentateur à cette occasion, y manifestèrent des performances non attendues, mais dues à la relation qu'ils venaient d'entreprendre.

Enfin, la constitution des données par manipulation des contingences exige que soit clairement identifiée la variable mise en jeu dans la manipulation. En effet, dans certaines manipulations, la modification effectuée au niveau observable ne réalise pas nécessairement ce qui était censé l'être. Ainsi, dans une évaluation de l'effet du développement du rythme (variable indépendante) sur l'apprentissage de la lecture (variable dépendante), la manipulation effectuée était l'introduction de séances

d'exercices rythmiques. Les chercheurs postulaient que la réalisation de ces exercices suffirait à provoquer un développement du rythme. Or, l'analyse des protocoles a montré que l'introduction de ces exercices, tout en ayant apporté un bien-être aux enfants, n'avait en rien contribué au développement du rythme. L'analyse montrait aussi que les quelques effets, assez faibles, observés sur l'apprentissage de la lecture, étaient dus au changement de climat social, aux relations entre enfants et enseignants et à une contextualisation de leur apprentissage, plutôt qu'à un développement hypothétique du rythme, non observé par ailleurs. La manipulation avait dans ce cas introduit d'autres facteurs que ceux sur lesquels les chercheurs pensaient travailler. Ce problème est toujours délicat lorsque les variables indépendantes sont plutôt des variables intermédiaires, c'est-à-dire qu'elles ne sont pas observables, mais seulement inférées à partir d'indices indirects qui entretiennent avec la variable cible des relations discutables.

Les remarques que l'on vient d'esquisser à propos des consignes implicites, et du rapport entre la manipulation effectuée et les variables identifiées, posent encore une fois le problème du statut que le chercheur va accorder au matériel recueilli. C'est pour cette raison qu'avant d'entreprendre des analyses sophistiquées et longues sur les données, il convient de les examiner de façon à bien déterminer quelle crédibilité on peut leur accorder. Avant l'analyse et le traitement des données, il convient d'examiner quelle est leur qualité relative sur les plans de la vraisemblance, de la pertinence et de la représentativité.