

QUE SAIS - JE ?

Les méthodes quantitatives

FANNY BUGEJA-BLOCH

MARIE-PAULE COUTO



Imprimé par un visiteur avec l'adresse IP 194.154.200.98 sur
www.cairn.info - 23/10/2017

Nous remercions particulièrement les membres du laboratoire de sociologie quantitative pour leurs conseils, ainsi que les enseignants-chercheurs des universités Paris 8 et Paris 10 pour leur collaboration.

À lire également en
« Que sais-je ? »

Serge Paugam (dir.), *Les 100 mots de la sociologie*, n° 3870.
Sophie Alami, Dominique Desjeux, Isabelle Garabuau-Moussaoui, *Les méthodes qualitatives*, n° 2591.
Chesnais, Jean-Claude, *La démographie*, n° 2546.

ISBN 978-2-13-063161-3
ISSN 0768-0066

Dépôt légal — 1^{re} édition : 2015, juin

© Presses Universitaires de France, 2015
6, avenue Reille, 75014 Paris

Imprimé par un visiteur avec l'adresse IP 194.154.200.98 sur
www.cairn.info - 23/10/2017

INTRODUCTION

Ce « Que sais-je ? » s'adresse en premier lieu à des étudiants cherchant à s'initier aux méthodes quantitatives appliquées aux sciences sociales et aux professeurs souhaitant les leur transmettre. Il concerne particulièrement les sociologues, mais il emprunte aussi des exemples à la démographie et à l'histoire. En outre, il présente trois originalités.

D'abord, les « méthodes quantitatives » ne sont pas entendues comme une simple panoplie d'outils techniques (calculs d'indicateurs et de proportions). Elles renvoient davantage à l'ensemble du processus de quantification. Les traitements statistiques au sens strict (chap. V) ne sont abordés qu'après avoir explicité l'ensemble des étapes nécessaires à la construction des données sur lesquelles on les applique. Dans cette perspective, une attention particulière est accordée à l'enquête par questionnaire (chap. III et IV), mais pas avant que n'ait été déterminé ce en quoi elle se distingue (ou non) des autres façons de recueillir des matériaux (chap. II). Le dernier chapitre est l'occasion, quant à lui, d'aborder la question de la restitution des résultats, partie intégrante du processus de quantification.

Si cet ouvrage considère l'ensemble des étapes relatives à la quantification, c'est qu'il propose une posture « réflexive » (chap. I). « Pour le sociologue, cette posture consiste à soumettre à une analyse critique non seulement sa propre pratique scientifique (opérations, outils et postulats), mais également les conditions sociales de

toute production intellectuelle¹. » Les chiffres obtenus dépendent du déroulement de différentes étapes. De plus, chacune d'elles peut faire l'objet d'une analyse propre. Les statistiques sont donc bien envisagées ici comme une construction sociale. Depuis un certain temps déjà, les sciences humaines ont quitté l'illusion du positivisme : celle d'atteindre un « chiffre vrai ». En revanche, les pages qui suivent défendent l'idée moins répandue que la réflexivité des étudiants à l'égard des résultats statistiques ne peut s'aiguiser qu'au contact de la pratique de la quantification. Elle passe à la fois par la rigueur nécessaire pour les produire et par le fait de rencontrer des obstacles et de devoir procéder aux ajustements qu'ils engendrent.

Au fil du texte, nous nous efforçons ainsi de mettre ces principes à exécution en appuyant nos exemples et nos analyses sur un dispositif pédagogique mené simultanément à l'université Vincennes-Saint-Denis (Paris 8) et à l'université Paris Ouest Nanterre La Défense (Paris 10). Au cours de l'année universitaire 2012-2013, les étudiants de deuxième année de licence ont construit, exploité et valorisé, ensemble, une enquête sur le logement et l'habitat des étudiants des deux universités.

1. R. Sandrine, « Réflexivité », in S. Paugam (dir.), *Les 100 mots de la sociologie*, Paris, Puf, « Que sais-je ? », p. 21-22.

Chapitre premier

L'USAGE DES STATISTIQUES EN SCIENCES SOCIALES

I. – Au fondement des sciences sociales

L'activité de quantifier, c'est-à-dire de « faire exister sous forme numérique ce qui, auparavant, était exprimé seulement par des mots et non par des nombres¹ », est au fondement de beaucoup de sciences sociales. Parmi ces disciplines, nous pouvons citer par exemple la démographie et la sociologie. Moins évident encore : l'histoire a aussi utilisé, à certaines périodes favorables, la quantification.

1. **La démographie.** – La démographie analyse la taille et la structure des populations humaines. Elle étudie également leurs évolutions. Elle est l'une des sciences sociales dont le développement reste étroitement lié à l'usage des outils statistiques et des sources sur lesquelles on a coutume de les mettre en œuvre.

En 1662, la première de ces sources fut les bulletins de mortalité de la ville de Londres, eux-mêmes extraits des registres paroissiaux. Au milieu du XVII^e siècle, John Graunt, drapier londonien, s'est emparé de la mort comme objet et a ainsi mis à mal un certain nombre de superstitions. Par exemple, on pensait « que Londres, la plus grande ville du monde, avait entre deux et six

1. A. Desrosières, S. Kott, « Quantifier », *Genèses*, n° 58, 2005, p. 2-3.

millions d'habitants ; que les épidémies de peste coïncidaient avec l'avènement des rois et que la population diminuait, etc. ¹ ». Au contraire, John Graunt a montré que malgré la peste, la population de Londres n'a cessé de croître et qu'à cette époque elle n'excédait pas les 384 000 habitants. Comment est-il parvenu à une telle démystification ? D'abord, John Graunt a collecté l'information sur les naissances, les décès et leurs causes dans les bulletins paroissiaux. Ensuite, il a classé les informations recueillies dans des tableaux synthétiques ². Enfin, à l'aide de différents indicateurs, il a analysé ces séries numériques pour proposer une estimation de la population londonienne. Plus que les résultats obtenus, c'est la démarche scientifique mise en œuvre qui est pertinente. C'est la raison pour laquelle les *Observations naturelles et politiques sur les bulletins de mortalité* ³ de John Graunt peuvent être considérées comme étant à l'origine de la démographie. En outre, déjà à cette date, l'auteur a soumis à la critique les matériaux qu'il avait collectés. Pour chaque chiffre utilisé, il s'est demandé « comment il a été établi, par qui, dans quelle intention ⁴ ».

2. La sociologie. – Beaucoup de manuels le rappellent, la sociologie française a, elle aussi, été fondée sur un certain usage du chiffre et sur un ouvrage

1. É. Vilquin, « Une édition critique en français de l'œuvre de John Graunt (1620-1674). Présentation d'un ouvrage hors collection de l'INED », *Population*, vol. 33, n° 2, 1978, p. 416.

2. Ces tableaux deviendront l'ancêtre de ce que l'on nomme aujourd'hui « tables de mortalité ».

3. J. Graunt, *Observations naturelles et politiques sur les bulletins de mortalité* [1662], Paris, INED, « Classiques de l'économie et de la population », 1977.

4. É. Vilquin, « Une édition critique en français de l'œuvre de John Graunt (1620-1674). Présentation d'un ouvrage hors collection de l'INED », art. cité, p. 419.

fondateur, *Le Suicide* de Durkheim¹. Ce philosophe de formation a proposé une définition de la sociologie en réponse à l'essor d'une autre discipline : la psychologie. Pour ce faire, il s'est emparé de l'un de ses objets d'étude et en a présenté une nouvelle analyse. Au premier abord, le suicide constitue un acte intime et individuel. Pourtant, on ne passe pas à l'acte par hasard. En effet, Durkheim a relevé certaines régularités : les hommes se suicident plus que les femmes ; le taux de suicide est moindre en temps de guerre et varie également en fonction de la religion et de la taille de la famille. Outre les motifs personnels, les variations du taux de suicide obéissent donc à une loi sociale.

Selon l'analyse de Durkheim, les variations dépendent du degré d'intégration des sociétés. « Le suicide varie en raison inverse du degré d'intégration des groupes sociaux dont fait partie l'individu². » Par là, Durkheim pose les jalons de la sociologie et d'une méthode qui lui est constitutive (mais non exclusive) : l'analyse statistique des variations concomitantes. Sa vision s'inscrit également dans une logique holiste de la société. Selon cette logique, les phénomènes sociaux ne sont pas la somme des actions individuelles ; ils sont extérieurs aux individus ; ils s'imposent à eux.

Selon Stéphane Beaud et Florence Weber, l'enquête statistique a durablement dominé la sociologie française au détriment de l'enquête ethnographique³. En France, ce rapport de force a évolué notamment sous l'effet de la critique des indicateurs et des catégories statistiques

1. Durkheim, *Le Suicide* [1897], Paris, Puf, « Quadrige », 2007.

2. *Ibid.*, p. 221.

3. S. Beaud, F. Weber, « Le raisonnement ethnographique », in S. Paugam (dir.), *L'Enquête sociologique*, Paris, Puf, « Quadrige Manuels », 2010, p. 229.

par l'ethnométhodologie (voir *infra*, « Réflexivité et posture du quantitatifiste »).

3. Un regain d'intérêt en histoire. – Durant les décennies 1960 et 1970, l'histoire quantitative, avant d'être vigoureusement critiquée, s'est imposée, sinon dans les pratiques, au moins comme un exemple. Claire Lemerrier et Claire Zalc ont décrit ce tournant positiviste dans les sciences sociales ¹. Les chercheurs tentaient de présenter des « vérités statistiques » sur le modèle des « sciences dures ». Leurs ambitions étaient alors de produire des séries longues à partir de sources historiques. Les critiques adressées à ces séries longues sont à l'origine du désaveu qui frappa les méthodes quantitatives dans cette discipline. Alain Desrosières s'est fait l'écho de ces critiques ² : pour compter, il faut identifier. Or, les identifications qui font sens aujourd'hui n'en avaient pas nécessairement il y a deux siècles. Par exemple, « l'idée même de distinguer, de définir et de mesurer une *population active* est liée à l'extension d'un marché du *travail salarié* ³ ». Les séries longues sont ainsi suspectées d'anachronismes.

Dès lors, durant la décennie 1980, au moment même où les moyens matériels pour produire des statistiques se diffusaient (ordinateurs et logiciels), les historiens ont délaissé ces méthodes. À partir des années 2000, les méthodes quantitatives semblent connaître un nouvel engouement. En outre, la critique des séries longues et le nouvel enthousiasme pour les statistiques vont de

1. C. Lemerrier, C. Zalc, *Les Méthodes quantitatives pour l'historien*, Paris, La Découverte, « Repères », 2008, p. 8-18.

2. A. Desrosières, « Séries longues et conventions d'équivalence », *Genèses*, n° 9, 1992, p. 92-97.

3. *Ibid.*, p. 95.

pair avec le développement d'une approche moins positiviste, et plus réflexive, de ces méthodes en sciences sociales.

II. – Réflexivité et posture du quantitatifiste

On l'a vu, en 1662 déjà, John Graunt interrogeait la qualité de ses sources. Cependant il ne faisait pas encore de leur critique un objet de connaissance. La discussion proposée dans son ouvrage sur le degré de confiance à accorder aux matériaux vise principalement à évaluer leur fiabilité et donc les éventuelles erreurs de mesure. Ainsi, il était attentif aux instruments qu'il utilisait dans l'espoir d'approcher le « nombre vrai » de Londoniens.

Or, dans une posture réflexive, l'analyse ne porte pas seulement sur les nombres, mais sur l'ensemble des étapes qui a amené le quantitatifiste à les produire. Ces étapes sont en soi un processus social au cours duquel de petites et de grandes décisions (souvent contraintes) ont été prises. En admettant que les chiffres sont socialement construits, le chercheur se dégage de l'illusion de vérité. Cette réflexivité n'est pas si éloignée de celle qui est mise en œuvre dans le raisonnement ethnographique.

1. La quantification, un processus social. Dominique Merllié clarifie un certain nombre de malentendus autour des critiques adressées aux catégories statistiques¹. Outre la suspicion d'anachronismes des séries longues, on reproche aux catégories statistiques d'agréger des situations très différentes ou de simplifier excessivement le réel (en regroupant

1. D. Merllié, « Le travail des catégories statistiques », *Sociétés contemporaines*, n° 14-15, 1993, p. 149-163.

par exemple sous une seule étiquette tous les « étudiants »). Au demeurant, les quantitativistes se laisseraient dicter leur sujet et leurs questions de recherche par des systèmes de classement, souvent établis par d'autres qu'eux. C'est oublier sans doute que les opérations d'agrégation, de simplification, d'élaboration des catégories (en bref : d'identification) peuvent aussi être objet d'analyse. Les questions soulevées seraient par exemple : « Pourquoi les “étudiants” ont-ils été regroupés sous cette étiquette ? Comment ce regroupement a-t-il été mis en œuvre ? »

Les travaux d'ethnométhodologie ont parfois été appréhendés comme une charge contre les enquêtes statistiques, alors même qu'ils nous invitent surtout à déplacer notre regard du seul chiffre produit, aussi rigoureusement que possible, à toute la chaîne de production des données. Ce glissement ne vise pas seulement à identifier la façon dont les outils de mesure influencent les résultats, mais aussi à faire que les difficultés rencontrées soient autant d'éléments pertinents pour l'analyse. De la même façon qu'une « sociologie fondée pour tout ou partie sur des archives ne peut se passer d'une sociologie des archives ¹ » (chap. II), une analyse à partir d'une enquête statistique (sociologie quantitative) ne peut se priver de l'analyse du processus social qu'est l'enquête (sociologie de la quantification). Alors que le quantitativiste est présenté parfois comme enfermé dans sa tour d'ivoire, loin des réalités du terrain, son terrain est précisément toute la chaîne de production des données chiffrées.

1. L. Israël, « L'usage des archives en sociologie », in S. Paugam (dir.), *L'Enquête sociologique*, *op. cit.*, p. 174.

2. **L'analyse secondaire de données.** – Sur un terrain ethnographique, c'est souvent des déconvenues et de l'inadéquation entre les attentes (socialement situées) du chercheur et celles de ses enquêtés que l'on tire des connaissances. Le quantitativiste peut faire la même expérience. De ce point de vue, l'analyse secondaire de données peut être un atout.

Faire de l'analyse secondaire de données, c'est analyser une base de données construite par un tiers, le plus fréquemment une institution. Elle peut être publique (l'Insee, l'Inserm¹, Pôle Emploi, différents observatoires, etc.), académique (enquête réalisée par un groupe d'étudiants, un collectif de chercheurs, un laboratoire de recherche), associative (le Credoc², ATD Quart Monde³), privée (instituts de sondage tels Ipsos, Ifop, CSA, etc.). En conséquence, les bases de données ne sont pas systématiquement élaborées par des chercheurs en vue de répondre précisément aux questions qu'ils se posent. C'est d'ailleurs régulièrement ce qui est reproché aux exploitations secondaires qui en sont faites. Pourtant, l'analyse émerge aussi des décalages entre les attentes du quantitativiste et celles de l'institution qui est à l'initiative de l'enquête. Selon Dominique Merllié, « le fait même qu'il s'agisse de tester des hypothèses en fonction desquelles l'enquête n'était pas construite permet de porter un regard différent sur les "variables" qu'elle mobilise⁴ ».

Prenant l'exemple des étudiants, on peut donc se demander « pourquoi *cette institution* trouve pertinent

1. Institut national de la santé et de la recherche médicale.

2. Le Centre de recherche pour l'étude et l'observation des conditions de vie (www.credoc.fr).

3. Voir www.atd-quartmonde.fr.

4. D. Merllié, « Le travail des catégories statistiques », art. cité, p. 154.

de regrouper tous les “étudiants” sous la même étiquette alors qu’il me semble *à moi* primordial de s’intéresser finement à cette population ». Ce à quoi il est possible de répondre : « J’appartiens moi-même à cette catégorie, j’ai donc une conscience aiguë des différences internes au groupe que n’ont pas les agents à l’initiative de la nomenclature, puisqu’ils sont salariés. »

Autre exemple : les classements permettant de quantifier le travail des femmes au XX^e siècle sont imprécis et changeants dans les données de la statistique publique¹. *A priori*, cela fait obstacle à l’analyse de leur activité dans le long terme, mais cela nous rappelle aussi que la reconnaissance de l’emploi féminin dans la société a été difficile et que les agents élaborant les catégories statistiques ont longtemps été des hommes.

En France, beaucoup d’analyses secondaires s’appuient sur des données de la statistique publique, elles-mêmes issues de grandes enquêtes par questionnaire. L’Insee rassemble et centralise une grande part de l’activité de statistique d’État. Elle se compose de spécialistes du recueil des données, de spécialistes des méthodes, d’équipes au niveau régional et de deux écoles destinées à former les statisticiens. Il est rare qu’une institution concentre à ce point toutes ces fonctions. Donner à une seule institution la responsabilité de la collecte et de l’analyse des statistiques, au service de l’État, peut produire certaines dérives ou, du moins, un manque de transparence. Un collectif d’auteurs anonymes, Lorraine Data, fonctionnaires de la statistique publique pour la plupart, illustrent les différentes formes de pressions gouvernementales auxquelles ils ont pu être

1. M. Maruani, M. Meron, *Un siècle de travail des femmes (1901-2011)*, Paris, La Découverte, 2012.

confrontés¹. L'une des manipulations consiste à choisir l'indicateur qui répond le mieux aux attentes gouvernementales. Un nouveau seuil de pauvreté, nettement plus flatteur que le seuil international habituellement utilisé, a ainsi été adopté². Pour parer à cette situation de monopole et ouvrir le débat économique et social, des instituts indépendants se sont constitués, tels que l'Ires³ et l'OFCE⁴.

3. La chaîne de production d'une enquête statistique. – On le voit bien, les statistiques, « loin d'être l'enregistrement neutre d'actes ou d'effectifs, [sont] en fait le produit de déterminations sociales complexes⁵ ». Dans les grandes enquêtes par questionnaire, Alain Desrosières distingue six phases au cours desquelles se posent des problèmes sociologiques (fig. 1). Ces phases représentent en quelque sorte la chaîne de production des données statistiques, terrain du quantitatif réflexif. À l'instar de l'ethnographe, dont la seule présence dans un univers social implique qu'il y participe, les pratiques du quantitatifs sont parties prenantes du processus de quantification qu'il analyse.

1. Lorraine Data, *Le Grand Truquage. Comment le gouvernement manipule les statistiques*, Paris, La Découverte, 2009. Voir aussi une synthèse : M. Selz, « Les statistiques, un service public détourné », *La Vie des idées*, 2009.

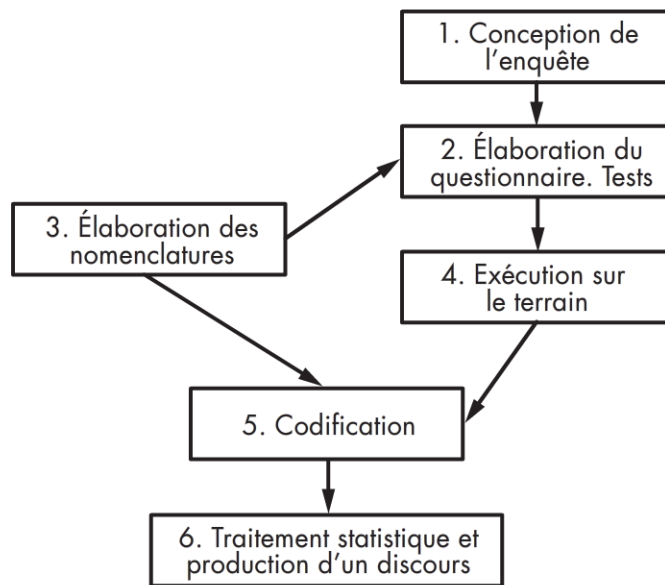
2. Lorraine Data, *Le Grand Truquage. Comment le gouvernement manipule les statistiques*, *op. cit.*, p. 103.

3. L'Institut de recherches économiques et sociales au service des organisations syndicales représentatives des travailleurs a pour fonction de répondre aux besoins exprimés par ces organisations dans le domaine de la recherche économique et sociale.

4. L'Observatoire français des conjonctures économiques est de surcroît ouvert à l'international.

5. J. Peneff, « La fabrication statistique ou le métier du père », *Sociologie du travail*, n° 2, 1984, p. 197.

Fig. 1. – Les six phases d’une enquête statistique



Source : A. Desrosière, « La portée sociologique des diverses phases du travail statistique », in J.-L. Besson (dir.), *Des mesures*, Lyon, PUL, 1986, p. 249.

La phase de la *conception de l'enquête* renvoie à la construction de l'objet d'étude. Un objet d'étude n'apparaît comme pertinent que si émerge une demande sociale, publique ou académique, d'informations. Durant cette étape, les différents acteurs (demandeurs et concepteurs d'enquête) interagissent. La conception de l'enquête est le résultat de ces interactions. Par exemple, l'enquête « Trajectoires et origines » (TeO) sur la diversité des populations en France doit son existence au croisement d'intérêts scientifiques, politiques et publics. En 2003, l'essor des questions relatives aux discriminations dans les débats publics a conduit le Premier ministre, Jean-Pierre Raffarin, à solliciter une grande enquête statistique sur l'intégration et les inégalités. Cette enquête a été menée par des chercheurs de l'Ined (Institut national d'études démographiques) avec

le soutien de l'Insee (Institut national de la statistique et des études économiques).

La phase d'*élaboration du questionnaire* est aux prises avec des « contraintes contradictoires : insérer le plus grand nombre de questions possibles, les adapter aux possibilités de réponses en contrôlant leur sens, en diminuer le nombre afin de parvenir à un temps de passation jugé possible¹ ». Toute la difficulté réside dans le fait que le questionnaire, unique, doit s'adapter à toutes les situations concrètes et à toutes les représentations. Ce questionnement suppose de transformer des situations souvent complexes en différents cas possibles et balisés à l'aide d'entretiens libres ou d'observations exploratoires. L'usage d'entretiens et d'observations exploratoires n'est pas anecdotique dans cette phase d'élaboration, mais constitue bien l'une des étapes déterminantes du processus de quantification (voir chap. III). La phase de test des questionnaires permet ensuite de vérifier l'adéquation entre les questions et les réponses, de faire des arbitrages (fermer certaines questions ouvertes au détriment de l'information afin de s'accommoder à des contraintes de coût ou de durée de passation des questionnaires). Le schéma ci-après montre que les versions tests et les réunions de travail ont été nombreuses au sein de l'Ined afin de finaliser le questionnaire de l'enquête TeO (fig. 2).

Fermer des questions revient à proposer une liste de réponses en mobilisant des nomenclatures, préexistantes ou créées pour les besoins spécifiques de l'enquête. C'est le cas, par exemple, des catégories socioprofessionnelles de l'Insee ou des échelles de satisfaction. D'autres, telles que les catégories dites « ethniques »,

1. A. Desrosières, « La portée sociologique des diverses phases du travail statistique », in J.-L. Besson (dir), *Des mesures*, Lyon, PUL, p. 252.

font en France l'objet de débats incessants. Georges Felouzis résume la controverse de la façon suivante : « Doit-on nommer au risque de renforcer une vision raciste et *racialisée* de la société ou s'abstenir de nommer et ainsi laisser dans l'ombre – ou la pénombre – des discriminations aussi structurantes et agissantes sur le destin des individus que celles liées à leur origine ethnique visible ¹ ? » Finalement, l'*élaboration des nomenclatures* consiste à créer une passerelle : d'un classement fondé sur des catégories indigènes ou spontanées, telles que les individus en parlent ordinairement, à un classement de catégories savantes, sociologiques ou conceptuelles. Il s'agit d'un exercice complexe. Lorsque Pierre Bourdieu disait : « La jeunesse n'est qu'un mot », il montrait déjà la difficulté de classer la jeunesse sur une échelle de classes d'âge. Par là, il voulait dire que la jeunesse ne signifie rien en tant que telle. Elle est polysémique et on ne peut lui assigner une réalité substantielle ². Catégorie juridique (on devient adulte à la majorité) lourde de représentations (la jeunesse n'est plus ce qu'elle était !), elle est pourtant « l'un des âges de la vie le plus rebelle au classement et aux dénombrements statistiques ³ ». De fait, la jeunesse est un processus qui se déroule dans la durée et qui évolue au fil des générations. C'est un parcours défini selon des calendriers : les itinéraires scolaires et professionnels, l'autonomie financière et

1. G. Felouzis, « L'usage des catégories ethniques en sociologie », *Revue française de sociologie*, vol. 49, n° 1, 2008, p. 128.

2. P. Bourdieu, « La "jeunesse" n'est qu'un mot », in *Questions de sociologie*, Paris, Minuit, 1984, p. 143-154.

3. L. Thévenot, « Une jeunesse difficile. Les fonctions sociales du flou et de la rigueur dans les classements », *Actes de la recherche en sciences sociales*, n° 26-27, 1979, p. 3-18.

résidentielle, la stabilisation dans l'emploi et la vie familiale, la poursuite ou la fin du célibat, la cohabitation et le mariage, le service militaire, etc.

Fig. 2. – **Élaboration du questionnaire de l'enquête TeO**

Mars 2006	Constitution d'un groupe de travail pour l'élaboration du questionnaire
Sept. 2006	1^{re} version du questionnaire (V1)
6-13-23 oct. et 10 nov. 2005	Nombreuses séances de travail pour finaliser cette version
Déc. 2006	Nouvelle version V1 du questionnaire testée en « focus-group » enquêteurs
<i>Séances de travail autour la version V1 :</i>	
12 déc. 2006	Bilan du « focus group » enquêteurs
15 déc. 2006	Réunion du Comité d'orientation
18 déc. 2006 et 9 janv. 2007	Réunions du groupe de travail
Févr. 2007	Premier test de la nouvelle version « papier » V2 du questionnaire (174 enquêtés en Île-de-France. Languedoc-Roussillon)
<i>Séances de travail autour de la version V2 :</i>	
6 mars 2007	Réunion du Comité d'orientation
12-13 mars 2007	Bilan du test
31 mars 2007	Groupe de discussion avec 30 enquêtés sur le module « image de soi : regard des autres »
Mars/avril 2007	Consultation de chercheurs a l'INED
Avril 2007	Livraison V3 du questionnaire aux équipes Blaise-Capi, pour mise au format CAPI (questionnaire sur ordinateur)

<i>Séances de travail autour de la version V3 :</i>	
	Séance du CNIS sur les questions sensibles (24 mai 2007) Comité du label (31 mai 2007) Réunion du Comité d'orientation (12 juin 2007) Courrier aux associations Séance du CNIS sur les questions sensibles (12 octobre 2007)
Nov. 2007	Version V4 du questionnaire CAPI
Sept./26 nov. 2007 au 7 janv. 2008	TEST du questionnaire V4 auprès de 600 enquêtes en Île-de-France, Aquitaine et Lorraine
Sept 2008 à déc. 2008	ENQUÊTE grandeur nature (24 000 personnes en France métropolitaine)

Légende :

Tests

Consultation

Production/collecte

Source : teo.site.ined.fr/fr/le_contenu_de_l_enquete/les_différentes_phases_de_l_elaboration_du_questionnaire

La phase d'*exécution de l'enquête sur le terrain* regroupe deux étapes : le plan de sondage et la passation (chap. IV). Au sein de la population de référence (celle qui nous intéresse compte tenu de l'objet d'étude), le plan de sondage définit la manière de sélectionner l'échantillon (la partie de la population réellement mesurée). La qualité d'un échantillon dépend de sa capacité à pouvoir généraliser les résultats obtenus sur celui-ci. Les modes de passation sont pluriels. Mettant en contact un enquêteur et un enquêté, ils ont en commun de créer un rapport de domination et de souligner la non-neutralité de la situation d'enquête.

La phase de *codification* est en constante évolution avec les progrès de la micro-informatique. Elle consiste

à produire une base de données (un tableau à deux dimensions) à partir des questionnaires : les enquêtés deviennent alors des lignes d'un tableau appelé base de données, les questions, des intitulés de colonnes, et les réponses, les cases du tableau (tableau 1). Avant d'analyser les données, il est souvent nécessaire de procéder à un apurement, c'est-à-dire à un nettoyage. Pour ce faire, on réalise des recodages, c'est-à-dire que l'on regroupe de façon encore différente les réponses des enquêtés.

Tableau 1. – **Base de données synoptique**

<i>Identifiant de l'enquêté</i>	<i>Âge</i>	<i>Diplôme</i>
Individu 1	20	Baccalauréat
Individu 2	22	Bac + 2
...

La phase de *traitement statistique* et de *production d'un discours* est celle abordée dans les deux derniers chapitres de cet ouvrage. Dans le cas d'une exploitation secondaire de données, c'est lors de cette phase que le chercheur intervient directement dans le processus de quantification. À ce stade, il doit réfléchir non plus seulement aux arbitrages pris par d'autres (concepteurs de l'enquête, enquêteurs, enquêtés, codeurs, etc.), mais également à ses propres pratiques. En la matière, il n'y a pas de mode d'emploi. Généralement, le chercheur débute avec un travail exploratoire. Il s'agit de s'approprier les données en réalisant des tris à plat, de découvrir leurs atouts et leurs limites en relation avec l'ensemble des étapes du processus de quantification. Vient ensuite le moment de tester les hypothèses posées lors de la construction de l'objet. Les hypothèses ne sont donc pas purement déductives puisqu'elles sont issues d'une connaissance préalable du sujet, notamment par

le biais des entretiens ou des observations exploratoires. En outre, en croisant certaines variables avec d'autres, on les reformule, on les ajuste, on les affine, parfois on les révisé. Au fil des traitements, le travail se confirme alors de plus en plus. C'est le temps de la validation ou de l'invalidation des hypothèses et de la rédaction d'un discours. À terme, analyser et présenter des données chiffrées est donc un exercice littéraire.

Chapitre II

LES APPROCHES QUANTITATIVES

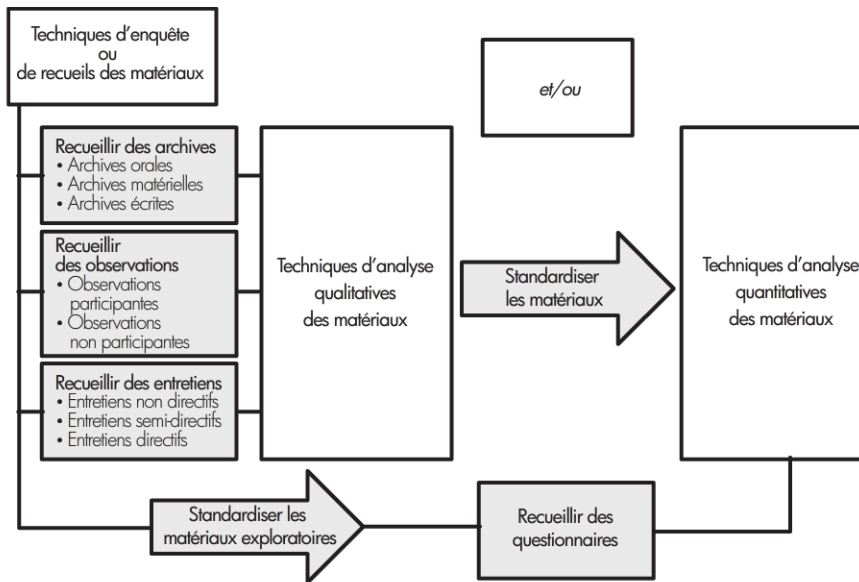
Articuler des approches quantitatives et qualitatives est souvent présenté comme un atout dans les sciences sociales. Chacune des deux méthodes apporterait donc des informations distinctes de la précédente et l'enrichirait. Afin de bien saisir l'intérêt spécifique des approches quantitatives, le présent chapitre les replace plus généralement dans le champ des méthodes en sciences sociales. Dans cette perspective, il est intéressant de différencier ce qui relève, d'une part, des techniques d'enquête et, d'autre part, des techniques d'analyse. Alors que les techniques d'enquête renvoient au *recueil des matériaux sur le terrain*, les techniques d'analyse sont mobilisées dans un deuxième temps, lors de l'exploitation ou du *traitement des données* recueillies. D'emblée nous pouvons signaler que la production de statistiques se distingue davantage comme technique d'analyse que comme techniques d'enquête ou de recueil des matériaux.

I. – Trois types de recueil des matériaux

Pour réaliser une étude ou une recherche en sciences humaines, il existe différents types de recueils des matériaux empiriques. Le recueil d'archives représente une première technique d'enquête. L'observation en constitue une deuxième. Le recueil d'entretiens non directifs,

semi-directifs ou directifs en est une autre (fig. 3). Dans tous les cas, ces opérations de recueil des matériaux consistent à négocier l'accès à un terrain et à y accumuler de l'information.

Fig. 3. – **Techniques d'enquête et techniques d'analyse des matériaux**



1. Recueillir des archives. – Liora Israël distingue plusieurs définitions de l'archive : « L'une, dans la vie quotidienne, désigne les documents divers rassemblés par une personne ou une famille, à des fins privées le plus généralement. Une deuxième définition, plus institutionnelle, désigne le stockage par une organisation des dossiers et des informations relatifs à son activité passée ¹. » L'organisation peut être une association, une

1. L. Israël, « L'usage des archives en sociologie », in S. Paugam (dir.), *L'Enquête sociologique*, op. cit., p. 168-169.

entreprise, une collectivité locale ou l'État. S'il s'agit le plus souvent de papiers, les archives peuvent prendre d'autres formes. Les chercheurs mobilisent des archives matérielles, telles que des œuvres d'art ou des objets du quotidien. Ils collectent également des archives orales, sous la forme d'enregistrements audiovisuels par exemple.

Pour ses travaux sur le devenir en métropole des Français rapatriés d'Algérie en 1962, Yann Scioldo-Zürcher a consulté des archives départementales, ministérielles et militaires. Il s'est intéressé à la correspondance privée de certains Français d'Algérie, mais aussi à la manière dont les pouvoirs publics les ont présentés dans les médias. En s'appuyant sur l'ensemble de ces pièces, il est parvenu à montrer la diversité de l'action de l'État envers les rapatriés¹.

La conservation et la communicabilité des archives publiques, c'est-à-dire relevant des missions de services publics, sont réglementées par la loi (loi du 15 juillet 2008 relative aux archives). Mais l'accès aux archives tient aussi beaucoup au travail de persuasions et à la relation de confiance que le chercheur instaure avec ceux qui les détiennent. Cela est également vrai pour les archives privées ou familiales. Ainsi, au cours d'une recherche doctorale sur les séjours de vacances que des descendants d'immigrés algériens ont passés au « bled », Jennifer Bidet a su négocier l'accès aux photographies de vacances de ses enquêtés². Dans cette perspective, les modalités d'obtentions des documents

1. Y. Scioldo-Zürcher, *Devenir métropolitain. Politique d'intégration et parcours de rapatriés d'Algérie en métropole (1954-2005)*, Paris, EHESS, « En temps & lieux », 2010.

2. J. Bidet, *Les Vacances au bled de descendants d'immigrés algériens. Trajectoires, pratiques, appartenance*, thèse de doctorat, université Lyon 2, 2013.

(sont-ils facilement accessibles ? par qui sont-ils détenus ?), les modes de sélection et de conservation des documents (où sont-ils rangés ? selon quels critères de classement ?), l'aspect ou encore le contenu de ces derniers (que voit-on ? que ne voit-on pas ?) sont autant d'informations pour le chercheur. De la même façon, les organisations (syndicats, caisses de retraite, entreprises, etc.) sélectionnent et mettent en forme les documents qu'elles conservent.

Autrement dit, avant l'arrivée du chercheur sur le terrain, un travail de filtrage a déjà eu lieu. Loin de constituer un problème, c'est ainsi que le recueil d'archives offre des informations sur le point de vue que les organisations portent sur elles-mêmes et sur leur fonctionnement. Outre les données issues des documents, les questions soulevées par le chercheur seront les suivantes : pourquoi précisément *ces* archives ont-elles été conservées par l'organisation étudiée ? Pourquoi l'ont-elles été sous cette forme ? Pourquoi m'autorise-t-on à y accéder ? La démarche est donc assez similaire à celle qui est mise en œuvre lors de l'accès à une base de données institutionnelle en vue d'une exploitation statistique de seconde main (chap. 1).

2. Recueillir des observations. – Si les archives offrent un regard (celui des institutions qui les produisent) sur une activité passée, le recueil d'observations permet quant à lui d'appréhender les actions en situation, c'est-à-dire en train de se dérouler dans un contexte donné. Dans cette technique d'enquête, héritée en sociologie de la première École de Chicago, on a coutume de distinguer l'observation non participante de l'observation participante. Dans un cas, l'enquêteur occupe une place qui n'existait pas auparavant dans le monde social étudié : celle d'observateur. Dans le

second cas, le chercheur observe, tout en se glissant dans un rôle ou dans des fonctions déjà présents sur le terrain avant son arrivée. Il se confronte ainsi directement aux exigences de ce dernier¹.

Damien Cartron a réalisé une observation participante *incognito* dans un fast-food². Il décrit la façon dont son engagement dans l'action en tant qu'équipier chez McDonald's a été riche d'informations. L'exercice d'un emploi non qualifié lui a permis, d'une part, de « se rendre compte de la difficulté à tenir ces postes déqualifiés³ » et, d'autre part, de saisir les ressorts de l'engagement des jeunes dans ce « petit boulot ». Observateur, en théorie dominant dans cet univers social, il est pourtant pris par l'excitation de la tâche. Il montre donc comment l'organisation du travail (objectifs à court terme, rapidité d'exécution, ambiance ludique) joue sur l'investissement des équipiers dans leur poste, au-delà des nécessités économiques. Ici, les informations recueillies renvoient non seulement à ce que le chercheur voit de l'organisation du travail, apprend des trajectoires des enquêtés, mais également de ce qu'il ressent lui-même sur le terrain (l'excitation) rapporté à sa position sociale (dominante).

En principe, le recueil d'observations ne fait pas intervenir d'autres filtres que les perceptions de l'observateur, toutefois, la seule présence sur le terrain d'un chercheur peut influencer les comportements des enquêtés. Il intègre alors ce phénomène à l'analyse. Là encore, la démarche présente des points communs

1. S. Chauvin, N. Jounin, « L'observation directe », in Serge Paugam (dir.), *L'Enquête sociologique, op. cit.*, p. 143-165.

2. D. Cartron, « Le sociologue pris en sandwich ! Retour sur une observation participante dans un fast-food », *Travail et emploi*, n° 94, 2003, p. 59-64.

3. *Ibid.*, p. 59.

avec celle d'un quantitativiste, d'une part, confrontant ses catégories de pensée à celles, par exemple, de la statistique publique et, d'autre part, explicitant comment ses propres pratiques influencent le processus de quantification.

3. Recueillir des entretiens. – Janine Barbot attribue deux filiations au recueil d'entretiens : la psychologie clinique et les enquêtes sociales, qui, à la fin du XIX^e siècle, portaient sur les pauvres dans les villes. Pourtant largement mobilisé par la première École de Chicago, son usage n'a été que tardivement formalisé en sociologie¹.

Par définition, le recueil d'entretiens revient à recueillir des discours. La manière dont le chercheur présente sa démarche et son sujet auprès des enquêtés trace un cadre de référence. Au sein de ce cadre, la parole est plus ou moins libre selon que l'enquêteur impose ou non les thèmes abordés. C'est pourquoi, on distingue souvent les entretiens non directifs, des entretiens semi-directifs ou directifs. Quelle que soit l'approche, ce type de recueil vise généralement à saisir le sens que les individus donnent à leurs actions. Ce faisant, on s'éloigne de la vision holiste de Durkheim pour s'inscrire dans une sociologie compréhensive, traditionnellement associée à l'un de ses contemporains allemand, Max Weber (1864-1920). Dans cette perspective, les actions sociales ne peuvent être comprises indépendamment de leurs significations subjectives, c'est-à-dire indépendamment des représentations des acteurs. Toutefois, là encore, des filtres interviennent. Comme pour l'observation, la situation d'entretien est

1. J. Barbot, « Mener un entretien de face à face », in S. Paugam (dir.), *L'Enquête sociologique*, *op. cit.*, p. 115-141.

une situation sociale à part entière. Elle est évidemment construite par le mode de sollicitation des enquêtés, le lieu de l'entretien, l'apparence ou encore les positions sociales respectives de l'intervieweur et de l'interviewé. Michel Pinçon et Monique Pinçon-Charlot décrivent les rapports de domination dans lesquels ils ont été pris lors d'entretiens semi-directifs avec des aristocrates et des grands bourgeois¹. Les chercheurs mentionnent le soin qu'ils ont été contraints d'apporter à leurs tenues vestimentaires².

Dans ce qu'on appelle les récits de vie, outre les enjeux de présentation de soi, interviennent des effets d'illusions rétrospectives, largement évoqués dans un article devenu classique de Pierre Bourdieu³. Plutôt que de chercher à tout prix à contrôler ces effets, il s'agit, ici aussi, d'en tirer parti. De fait, le travail de réorganisation, de simplification et d'unification des trajectoires opéré par les enquêtés au cours de leur narration n'est pas sans signification sociologique. La mémoire opère un tri dans le passé en fonction des exigences du présent. Dans son ouvrage sur l'expérience concentrationnaire, Michael Pollak analyse ce que les femmes enquêtées disent durant l'entretien, mais aussi ce qu'elles passent sous silence, leur vie avant et après le camp. « La disproportion de la place ici accordée aux différentes périodes de la vie indique la blessure durable infligée par la déportation⁴. » Ainsi, comme lorsque l'on s'intéresse

1. M. Pinçon, M. Pinçon-Charlot, « Pratiques d'enquête dans l'aristocratie et la grande bourgeoisie : distance sociale et conditions spécifiques de l'entretien semi-directif », *Genèses*, n° 3, 1991, p. 120-133.

2. *Ibid.*, p. 128.

3. P. Bourdieu, « L'illusion biographique », *Actes de la recherche en sciences sociales*, n° 62, 1986, p. 69-72.

4. M. Pollak, *L'Expérience concentrationnaire. Essai sur le maintien de l'identité sociale*, Paris, Métailié, « Sciences humaines », 2000, p. 232.

à la distribution des non-réponses dans une enquête d'opinion¹, c'est parfois dans les interstices de la situation d'entretien que l'on trouve des informations.

II. – Quantifier en sciences sociales

On vient de distinguer trois types de recueil des matériaux : archives, observations et entretiens. Pour l'instant, le cas particulier de l'enquête par questionnaire est resté en suspens. Aucun de ces recueils « n'impose ni n'exclut *a priori* la quantification² ». Si certains types de recueil s'y prêtent davantage, tous offrent l'occasion de faire du nombre. Un retour sur la manière dont la quantification est mise en œuvre dans chacun des cas révèle à la fois la pluralité des approches, leurs limites, mais également l'intérêt spécifique de la démarche.

1. Quantifier les différents matériaux. – L'usage des archives est souvent associé à la figure de l'historien, peu familier des chiffres. Or, les techniques d'analyse quantitatives des archives non seulement connaissent un regain d'intérêt chez les historiens, mais sont aussi de plus en plus investies par les sociologues. C'est ainsi qu'Alain Chenu quantifie l'apparition des personnes célèbres sur les couvertures du magazine *Paris Match*. Il analyse de cette manière la construction sociale de la célébrité et ses transformations de 1949 à 2005³. C'est également à partir d'un recueil d'archives

1. P. Bourdieu, « L'opinion publique n'existe pas », in *Questions de sociologie*, *op. cit.*, p. 222-235.

2. C. Lemerrier, C. Zalc, *Les Méthodes quantitatives pour l'historien*, *op. cit.*, p. 19.

3. A. Chenu, « Des sentiers de la gloire aux boulevards de la célébrité. Sociologie des couvertures de *Paris Match*, 1949-2005 », *Revue française de sociologie*, vol. 49, n° 1, 2008, p. 3-52.

que Joanie Cayouette-Remblière propose, selon ses termes, « une statistique ethnographique¹ » sur laquelle il est intéressant de s'attarder. Afin d'étudier les inégalités sociales dans un contexte de démocratisation scolaire, la sociologue a reconstitué la scolarité d'une cohorte d'élèves de la sixième au baccalauréat (s'il est obtenu), soit sur une période de sept à neuf ans. Pour ce faire, elle s'est appuyée sur les dossiers scolaires de ces élèves, archivés dans le dernier établissement qu'ils ont fréquenté. L'auteur a d'abord ouvert les dossiers avec l'œil de l'ethnographe, en observant par exemple la façon dont l'institution scolaire les a classés². Après avoir procédé à cette sociologie *des* archives, Joanie Cayouette-Remblière s'est appuyée sur la richesse des dossiers pour construire deux cent cinquante-sept variables renseignant l'origine sociale de l'élève, sa situation familiale, sa scolarité passée et sa scolarité année par année, ses vœux et les décisions d'orientation à la suite des conseils de classe de troisième et seconde, etc. Plus de cinq cents trajectoires scolaires ont ainsi été *standardisées* afin de donner lieu ensuite à une exploitation statistique. La *standardisation* des données est un préalable aux techniques d'analyse quantitatives (fig. 3). C'est d'ailleurs sur ce point qu'elles sont le plus souvent critiquées. La standardisation implique de soumettre les matériaux à une norme. Le chercheur renonce ainsi à certaines spécificités des données. Il abandonne une part de leur originalité pour faire entrer l'information disponible dans des standards, dans un nombre limité

1. J. Cayouette-Remblière, « Reconstituer une cohorte d'élèves à partir de dossiers scolaires. La construction d'une statistique ethnographique », *Genèses*, vol. 4, n° 85, 2011, p. 115-133.

2. Le fait que les documents soient rangés dans des dossiers individuels illustre le « processus d'individualisation inhérent à l'entreprise scolaire » (*ibid.*, p. 121).

de variables et, pour chacune des variables, de valeurs. De son propre avis, Joanie Cayouette-Remblière, qui a construit ces données sur la base des premiers dossiers consultés dans un lycée général et technologique, les a jugées insuffisantes à rendre compte des spécificités de l'enseignement professionnel, notamment de l'orientation en fin de CAP ou de BEP. Le processus de standardisation mis en œuvre est resté tributaire des informations mises à sa disposition par l'institution scolaire et de la manière selon laquelle elle a eu accès aux archives.

Selon Sébastien Chauvin et Nicolas Jounin, l'observation directe vise précisément à éliminer les dispositifs bureaucratiques s'interposant entre le chercheur et son objet¹. Mais, dans un même temps, ils signalent que « l'on croit encore trop souvent que la technique de l'observation est par essence "qualitative" débouchant sur des phrases et non des chiffres, sur de la prose et non de la statistique² ». Effectivement, les opérations de comptage peuvent être fréquentes lors d'un recueil d'observations : compter le nombre d'hommes et de femmes devant une école, le nombre de caméras de surveillance dans les rues de Paris, le nombre de touristes devant des monuments historiques, etc. Généralement, le comptage n'intervient qu'après un premier contact avec le terrain où le chercheur a laissé place à toutes ses impressions. Il implique de définir précisément une cible à l'aide d'une grille d'observations (pour les touristes, les passants avec un appareil photographique, accompagnés d'un guide ou parlant une langue étrangère, etc.), un laps

1. S. Chauvin, N. Jounin, « L'observation directe », in S. Paugam, *L'Enquête sociologique, op. cit.*, p. 144.

2. N. Jounin, *Voyage de classes. Des étudiants de Seine-Saint-Denis enquêtent dans les beaux quartiers*, Paris, La Découverte, « Cahiers libres », p. 76.

de temps et un périmètre (une file d'attente, entre 10 et 11 heures, puis entre 2 et 3 heures de l'après-midi). Ici, l'enquêteur n'utilise pas de catégories institutionnelles forgées par d'autres que lui¹, mais il trie et sélectionne tout de même l'information selon une norme qu'il s'est fixée.

Durant un entretien, les pratiques passent par le médium de la parole. La méthode est peu répandue, mais les mots se prêtent aussi à comptage. Toujours dans *L'Expérience concentrationnaire*, Michael Pollak s'intéresse aux occurrences des pronoms personnels « on » et « nous » dans les récits de femmes déportées. « En adoptant un registre plutôt qu'[un] autre, la personne révèle une certaine vérité d'elle-même² ». D'après son analyse, le pronom personnel « on », par définition indéfini, renvoie d'une part « non pas à des personnes identifiables, mais à des situations et structures » et d'autre part à un « groupe se transformant “en masse” [...] dont le locuteur se distancie »³. Aujourd'hui, la micro-informatique permet de systématiser et de complexifier les études lexicométriques.

2. Le cas particulier de l'enquête par questionnaire. – Durant une enquête par questionnaire, les questions sont en majorités fermées. En d'autres termes, l'interviewé choisit parmi une liste de réponses préconstruites. Pour le chercheur, cela suppose une compréhension en amont du « monde de référence » des enquêtés ou du « système de cohérence interne des informations

1. Une réserve cependant, on ne compte souvent que ce que l'on veut bien nous laisser voir. Pour exemple, les difficultés rencontrées par les étudiants de Nicolas Jounin à compter les banques privées, gérant d'importants patrimoines, dans le VIII^e arrondissement de Paris (*ibid.*, p. 77).

2. M. Pollak, *L'Expérience concentrationnaire*, *op. cit.*, p. 238.

3. *Ibid.*, p. 240.

recherchées¹ ». Un questionnaire est censé s'adapter au langage, aux représentations, aux modes de pensées des interviewés. Il requiert ainsi une connaissance certaine de l'objet et de la population enquêtée. Pour ce faire, une phase exploratoire est indispensable. Comme dans le recueil d'observations, l'observateur tâtonne et peut garder des traces de ces hésitations. Pour s'ajuster au mieux aux catégories de pensée des enquêtés, on réalise souvent plusieurs passations, qualifiées de passations tests, avant la passation finale. Celles-ci permettent de réviser, si besoin, la formulation de certaines questions ou de certaines réponses.

Pour comprendre l'expérience que firent du chômage des salariés licenciés de Moulinex, Manuella Roupnel-Fuentes s'est dans un premier temps imprégnée de son terrain par le biais d'entretiens informels et par l'observation (en participant aux réunions syndicales), pour construire et mener dans un second temps l'enquête par questionnaire². Dans cette étude, les questions, leurs modalités de réponse et les variables auxquelles elles donnent lieu s'appuient sur une connaissance intime du monde social étudié, de ses catégories indigènes. Ces dernières ont été élaborées par l'enquêtrice sans filtre institutionnel. Souvent décrit comme caractéristique des entretiens, son questionnaire comptait bon nombre de questions subjectives, sur les représentations des salariés licenciés³, ouvrant la voie à une sociologie

1. A. Blanchet, A. Gotman, *L'Enquête et ses méthodes. L'entretien*, Paris, Armand Colin, « 128 », 2007, p. 37.

2. M. Roupnel-Fuentes, *Les Chômeurs de Moulinex*, Paris, Puf, « Le Lien social », 2011.

3. Nous pouvons citer pour exemple, la question suivante : « Aujourd'hui, êtes-vous fière d'être Moulinex ? Non, je ne l'ai jamais été ; oui, toujours autant ; oui, mais un peu moins ; ou j'évite de le dire dans certaines circonstances. »

compréhensive. Notons que, de la même façon que les archives, la conservation des informations ainsi obtenues est soumise à un cadre juridique (loi informatique et libertés du 6 janvier 1978).

En quoi, finalement, l'enquête par questionnaire se distingue-t-elle des précédents types de recueil ? Tous les matériaux se prêtent à la quantification, mais la collecte de questionnaires lui est spécialement consacrée (fig. 3). C'est pourquoi le chercheur commence à standardiser les données avant le gros du recueil et non *a posteriori*. Une fois la standardisation achevée, le questionnaire n'est plus évolutif, comme l'est une grille d'entretiens au fil des rencontres. À ce stade, c'est-à-dire une fois le questionnaire finalisé, un retour en arrière devient presque impossible, d'où l'importance cruciale de la phase exploratoire.

3. Pourquoi quantifier en sciences sociales ? – Quelle que soit l'approche (archives, observations, entretiens, questionnaires), la standardisation des informations est au centre du processus de quantification. Elle implique de faire subir aux matériaux une transformation supplémentaire leur faisant perdre une part de leur richesse et de leur originalité. Pourquoi alors quantifier à tout prix ? Halbwachs disait : « La statistique permet d'atteindre les caractères d'un groupe, qui ont une réalité pour le groupe entier, mais qu'on ne découvrirait dans aucun membre de ce groupe pris à part et isolément¹. » Dans cette perspective, l'usage des statistiques ne représente pas un gain d'intérêt ; c'est l'objet de la recherche qui est différent. Le regard du quantitatif se porte sur les

1. M. Halbwachs, « La statistique en sociologie », in *La Statistique. Ses applications. Les problèmes qu'elles soulèvent (septième semaine de synthèse)*, Paris, Puf, 1944, p. 113-160.

agrégats, non sur les parties qui les composent. Quand on s'accorde aujourd'hui à dire que les techniques d'analyse quantitatives, et le raisonnement ethnographe par exemple, peuvent se croiser et interagir sur un même objet, c'est dans le contexte où l'un des atouts des statistiques est d'organiser l'abondance des informations. Face à une avalanche de données, l'œil du chercheur a tendance à être attiré par l'exceptionnel. La situation exceptionnelle, « extrême », nous en apprend toujours beaucoup d'un phénomène social. En effet, l'individu éloigné de la norme fournit nécessairement des éléments sur la manière dont cette norme a été construite. Cependant, encore faut-il savoir qu'un individu est exceptionnel. Pour ce faire, nous avons besoin de saisir les structures dans lesquelles il s'inscrit, d'évaluer « les grands traits du groupe dont nous l'extrayons ¹ ».

Les statistiques décrivent les phénomènes sociaux avec moins de finesse, mais elles permettent d'en analyser la fréquence : de leur caractère habituel, ce qui se passe en moyenne, ou extraordinaire, ce qui s'en écarte. Il s'agit d'un formidable outil pour déconstruire nos *a priori*. Le dispositif pédagogique proposé par Nicolas Jounin à ses étudiants de Seine-Saint-Denis en donne une bonne illustration. Menant une enquête par observations dans les « quartiers bourgeois » parisiens, ils ont été d'abord éblouis par le contraste social dont ils faisaient l'expérience. Leurs premières observations rendaient compte sans nuance de ce qu'ils s'attendaient à trouver dans les « beaux quartiers » (des « rues propres », de « belles voitures », des gens aux « habits neufs aux grandes marques ² »). Le détour par les sta-

1. C. Lemerrier, C. Zalc, *Les Méthodes quantitatives pour l'historien*, op. cit., p. 17.

2. N. Jounin, *Voyage de classes*, op. cit., p. 34.

tistiques leur a rappelé que, malgré la richesse du VIII^e arrondissement, on y constate une certaine diversité de populations. Les étudiants ont pu revenir sur le terrain mieux avertis, attentifs aux individus qui leur avaient jusque-là échappé. L'articulation des techniques d'analyse offre ainsi la possibilité de déplacer le questionnement. Pourquoi certaines catégories de population sont-elles restées de prime abord invisibles ? « Certains individus du quartier sont discrets car [précisément] subalternes¹. » De cette manière, les étudiants ont pu éviter l'écueil d'une généralisation abusive. Pour autant, cela ne signifie pas que les techniques quantitatives soient les seules en mesure de généraliser. Sans prétendre les quantifier, le chercheur qualitatif étudie sur un terrain spécifique des processus sociaux qui se produisent également ailleurs de ce cadre.

Quantifier permet donc non seulement de saisir avec précision les caractéristiques de la moyenne (ou de la norme) et l'homogénéité d'un groupe, mais également ce qui s'en écarte, les cas extrêmes ou l'hétérogénéité d'un groupe. De surcroît, les statistiques sont un outil particulièrement efficace pour comparer. Or, comparer des sortes de groupes sociaux constitue l'un des défis, si ce n'est pas le premier, de la sociologie. Étudier les inégalités salariales selon le sexe, les modes de vie des différents groupes sociaux, les discriminations ethniques sur le marché du travail suppose de mobiliser l'outil statistique. Ajoutons enfin (voir chap. V) que certaines méthodes statistiques offrent un autre regard. Les analyses de régression par exemple, qualifiées de méthodes « *toutes choses égales par ailleurs* », sont un « substitut à l'expérimentation² ».

1. *Ibid.*, p. 35.

2. *Ibid.*, p. 62.

III. – Transversal et longitudinal

Parmi les techniques d'enquête et d'analyse en sciences sociales, on distingue deux formes. « La métaphore de la photo et du film est souvent utilisée pour illustrer la différence entre les deux types d'enquête : une enquête en coupe enregistre une photo ponctuelle d'une situation économique ou sociale alors qu'un dispositif longitudinal de données permet de "filmer" sur un laps de temps plus ou moins long le devenir des individus ¹. » Parler de roman-photo, c'est-à-dire d'une succession d'images plus ou moins rapprochées, serait plus exact.

1. Les enquêtes transversales. – Les enquêtes transversales, qualifiées aussi en statistique d'enquêtes en coupe, interrogent une population à un moment donné. Cette méthode d'enquête est la plus répandue. Pourtant, presque tous les phénomènes sociaux, géographiques, épidémiologiques, démographiques sont dynamiques. Pour pallier ce problème, les démographes ont coutume de reconstituer à partir de données transversales des générations fictives.

Ainsi, l'espérance de vie à la naissance, calculée à une date t , représente la durée de vie moyenne d'une génération fictive qui serait soumise à chaque âge aux conditions de mortalité de l'année t considérée. Parmi les enquêtes de la statistique publique, celles de la DARES ² et du ministère du Travail et de l'Emploi sont, pour la plupart d'entre elles, transversales. Les études marketing se fondent, elles aussi, presque toutes sur ce

1. M. Safi, « L'analyse longitudinale données et méthodes », in A. Chenu (dir.), *La France dans les comparaisons internationales*, Paris, Presses de Sciences-Po, 2011, p. 161.

2. Direction de l'animation de la recherche, des études et des statistiques (DARES).

type d'enquête, puisqu'elles cherchent à répondre au besoin du client à l'instant t .

2. Les enquêtes longitudinales. – On parle d'enquête longitudinale dès que la dimension temporelle est présente dans le dispositif d'enquête, qu'il s'agisse d'entretiens ou de questionnaires. Les informations recueillies portent alors sur une durée. Le temps peut être intégré de deux manières différentes dans le dispositif d'enquête : soit en suivant un même individu sur une période de sa vie, soit en l'interrogeant plusieurs fois, soit une seule fois de manière rétrospective.

Lorsqu'on suit des individus, l'enquête prend alors la forme d'un panel. L'enquête Elfe (« Étude longitudinale française depuis l'enfance ») en est un exemple récent et relativement exceptionnel. Ce dispositif unique consiste à suivre vingt mille enfants, nés en France en 2011, de la naissance à l'âge adulte, soit sur une durée de vingt ans environ. L'enquête, associant médecins, épidémiologistes, démographes, sociologues, vise à comprendre comment l'environnement, la famille et plus généralement les conditions de vie et leurs changements au cours du cycle de vie influencent le développement de l'enfant et sa santé. Plus généralement, le panel consiste à choisir un échantillon, souvent représentatif d'une certaine population, et à interroger plusieurs fois ces mêmes personnes ou familles (fig. 4). Chaque nouvel entretien constitue une vague d'enquête. À l'échelle de l'Europe par exemple, l'enquête longitudinale portant sur les revenus et les conditions de vie, commanditée par Eurostat, a été inaugurée en 2003 et poursuivie par vague d'un an par chaque État membre ¹.

1. European Union Statistics on Income and Living Conditions (EU-SILC) (http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc)

Fig. 4. – Un exemple de dispositif longitudinal

Prévisions de collectes de données dans la cohorte Elfe

Observations permanentes	Âge de l'enfant	Enquêtes ponctuelles
Suivi dans l'échantillon démographique permanent (EDP) Collecte de données sur la consommation de soins (SNIIRAM)	(Grossesse) Naissance	Enquête en face à face (mère) + dossier médical + prélèvements biologiques + questionnaire auto-administré sur la nutrition et expositions pendant la grossesse + remise de « piège à poussière »
	6-8 semaines	Enquêtes par téléphone (mère et père)
Données géolocalisées sur l'environnement	3/5 mois 6/8 mois 9/12 mois	} Questionnaires nutrition
	1 an	
	2 ans	Enquêtes par téléphone (mère et père) + questionnaire médecin traitant
	3 ans (entrée en maternelle)	Enquêtes en face à face (mère) et par téléphone (père) + jeu avec l'enfant + « piège à poussière » + prélèvements biologiques
Suivi de la scolarité	5 ans	Enquêtes par téléphone (mère et père) + bilan compétences scolaires
	6 ans (entrée à l'école primaire : CP) ou 7 ans	Enquêtes en face à face (mère) et par téléphone (père) + test psychomoteur + examen médical et prélèvement sanguin
	8 ans	Enquêtes par téléphone (mère et père)
	11 ans (entrée en 6 ^e)	Enquêtes par téléphone (mère et père) + bilan compétences scolaires (+ examen médical ?)
	14 ans ?	Enquête en face à face

Source : L'équipe Elfe, « Le devenir de 20 000 enfants. Lancement de l'étude de cohorte Elfe », *Population & sociétés*, n° 475, 2011, p. 2.

L'intérêt d'une dimension temporelle dans l'étude des phénomènes sociaux est incontestable¹. Par exemple, la position sociale d'un individu ne se limite pas à sa situation actuelle. Un détour par son passé, sur son origine sociale, son niveau scolaire, sa profession précédente, sont autant d'éléments qui permettent en effet de saisir plus largement sa position sociale. La démarche pose néanmoins d'importantes difficultés. En effet, suivre les personnes d'une vague à l'autre constitue un dispositif coûteux et délicat. Certaines populations déménagent fréquemment sans penser à en informer l'enquêteur ; d'autres se lassent. Autrement dit, entre chaque vague, certains individus sortent du panel et l'échantillon se réduit. C'est le phénomène que l'on nomme l'*attrition*. Puisque ces sorties ne se font pas au hasard² et sont chargées de sens, au bout d'un certain temps, l'échantillon s'en retrouve biaisé. Pour fidéliser les enquêtés, des dispositifs originaux sont aujourd'hui instaurés. On peut citer l'ELIPSS (« Étude longitudinale par Internet pour les sciences sociales »). Ici, une tablette tactile et un abonnement Internet sont offerts à des panélistes, régulièrement sollicités pour répondre à des enquêtes. Il est possible aussi d'intéresser les panélistes à l'enquête en leur restituant régulièrement les résultats produits sous forme de lettres d'information. Un autre moyen de lutter contre l'attrition consiste à faire entrer des individus à chaque nouvelle vague afin que l'échantillon reste stable et conserve une forme de représentativité. On parle alors de *panel rotatif* ou de *panel glissant*. Ce principe

1. J. Cayouette-Remblière, N. Devèze, B. Geay, P. Lehingue, Appel à communication, Colloque pluridisciplinaire international : *Famille, éducation et politique : apports et enjeux des études longitudinales*, 2014, p. 2.

2. Par exemple, les étrangers, plus mobiles géographiquement que les nationaux, sont davantage touchés par l'attrition.

est mobilisé dans les enquêtes Emploi de l’Insee, où les individus sont suivis trois ans et l’échantillon renouvelé par tiers chaque année.

Les enquêtes rétrospectives, qualifiées aussi d’enquêtes biographiques, constituent un autre type de données longitudinales « qui consiste non pas à répéter les observations dans le temps de manière prospective, mais plutôt à interroger les individus sur leur passé de manière rétrospective, dans l’objectif de reconstruire de longues séquences d’événements ¹ ». Nombre d’enquêtes transversales contiennent un volet rétrospectif. Dans l’enquête Biographie et entourage de l’Ined (2000 et 2001) ce volet tient une place importante puisqu’il s’agit d’appréhender de manière détaillée l’ensemble des questions de mobilité chez les personnes nées entre 1930 et 1950 vivant en Île-de-France ². Les enquêtes par questionnaire rétrospectives rencontrent les mêmes obstacles que les entretiens biographiques, évoqués plus tôt et résumés sous l’expression de Pierre Bourdieu, qui parle d’*illusion biographique*.

3. Les enquêtes répétées. – La plupart des enquêtes (de la statistique publique et des instituts de sondage notamment) intègrent la dimension temporelle à moindre coût en se répétant dans le temps, mais sans interroger les mêmes individus. L’Eurobaromètre permet de suivre d’année en année l’opinion des citoyens européens. Les enquêtes Budget de familles de l’Insee répétées à intervalles de cinq ans plus ou moins réguliers, permettent d’analyser l’évolution de la structure des budgets

1. M. Safi, « L’analyse longitudinale données et méthodes », art. cité, p. 165.

2. C. Bonvalet, E. Lelièvre, *De la famille à l’entourage*, Paris, INED, « Grandes Enquêtes », 2012.

et les dépenses des ménages selon leurs caractéristiques sociodémographiques. Sans suivre les individus, ces enquêtes répétées permettent de suivre des générations. Grâce à la date d'enquête et à la date de naissance de l'enquêté, on reconstitue des groupes de personnes nées à une même date. On parle alors de pseudo-panel.

Après avoir réfléchi à l'usage du quantitatif et à sa place dans l'ensemble des méthodes quantitatives, il s'agit désormais d'entrer dans une phase plus pratique. Toujours avec un regard réflexif, il sera question d'établir les conditions qui mènent de l'élaboration du questionnaire à la restitution des résultats.

Chapitre III

L'ÉLABORATION D'UN QUESTIONNAIRE

On l'a vu : l'enquête par questionnaire est un mode de collecte des matériaux parmi d'autres. Sa particularité est d'avoir pour seul objectif de quantifier. Telle qu'elle a été décrite par Alain Desrosière et explicitée plus haut, l'action de *quantifier* est une action sociale au cours de laquelle on opère des choix, des compromis, des traductions et où l'on fait usage d'un certain nombre de conventions¹. Cette multitude de petites et de grandes décisions débute dès l'élaboration du questionnaire et, en amont de celui-ci, lors de la construction de l'objet d'étude. Afin de bien saisir la portée pratique et sociologique de ces choix, nous nous appuyerons désormais sur un dispositif pédagogique mené simultanément à l'université Paris 8 (Vincennes-Saint-Denis) et à l'université Paris 10 (Paris Ouest Nanterre La Défense). Durant l'année 2012-2013, les étudiants inscrits en deuxième année de licence de sociologie ont conçu, de toutes pièces, une enquête par questionnaire sur *le logement et l'habitat des étudiants* des deux universités. Ils en ont assuré la passation, l'exploitation et enfin la valorisation.

I. – La construction de l'objet

Faute de temps, l'enseignement des méthodes quantitatives à l'université fait parfois l'impasse

1. A. Desrosières, S. Kott, « Quantifier », art. cité, p. 2.

sur la phase de construction de l'objet d'étude, préférant se concentrer sur l'acquisition de savoir-faire techniques. Cette étape de définition d'un objet et d'une problématique n'est pas spécifique à cet enseignement. Dans le cas précis de l'enquête par questionnaire, elle semble d'autant plus décisive qu'elle comporte un caractère définitif. Une fois le questionnaire élaboré, imprimé et soumis aux enquêtés, il est en effet impossible de revenir sur ses pas et de redéfinir le sujet de l'étude ou la façon dont il a été appréhendé. Or, la manière dont le sujet a été construit a inévitablement des conséquences sur les résultats produits. En outre, les difficultés rencontrées à ce stade sont déjà des éléments pertinents pour l'analyse.

1. La problématique de l'enquête. – En sciences sociales, les enquêtes par questionnaires explorent une thématique resserrée selon une question de recherche. Ainsi, une enquête par questionnaire dispose déjà en soi d'une problématique. La problématique est une interrogation d'un genre particulier. D'abord, elle ne figure pas explicitement sur le questionnaire. Elle est propre au chercheur, les enquêtés n'ayant pas de réponse individuelle à y apporter. La problématique introduit une énigme sociologique. Cyril Lemieux décrit le processus selon lequel les sociologues mettent en énigme le monde social. Il faut « (1) s'emparer d'une croyance partagée ou d'un constat reconnu relatif à l'objet qu'on entend étudier ; (2) en tirer une série d'inférences logiques ou d'énoncés prédictifs ; (3) faire apparaître un ou plusieurs éléments *empiriques* qui contredisent les inférences logiques ou les prédictions qu'on vient de tirer ; (4) se demander comment, si les croyances partagées ou les constats

reconnus relatifs à l'objet sont *vrais*, ces éléments empiriques *peuvent* exister¹ ».

La problématique est donc, avant tout, une question introduisant une contradiction. Une formulation adéquate serait : *Alors qu'il est coutume de dire ou de voir cela, comment se fait-il que j'observe autre chose ?* Lors de l'année universitaire 2012-2013, la question de recherche proposée par les étudiants des universités Paris 8 et Paris 10 était : « *Alors que les étudiants sont parfois appréhendés comme un groupe homogène disposant d'un mode de vie unique (on parle de « la vie étudiante »), comment se fait-il que l'on observe une telle diversité dans leurs conditions objectives d'existence et dans leurs rapports à l'habitat ?* » Pour construire une telle question – Cyril Lemieux l'indiquait –, il faut disposer de premiers éléments empiriques sur l'objet étudié.

2. Les lectures et les terrains exploratoires.

Les premiers éléments empiriques indispensables à l'élaboration d'une problématique sont à trouver dans des lectures et sur des terrains dits exploratoires. Le terme « exploratoire » ne signifie pas qu'il faille en négliger l'importance. Au contraire, il s'agit bien là de la matière première de l'enquête à partir de laquelle va s'amorcer le processus de standardisation des données. Dans cette perspective, le travail bibliographique exploratoire consiste d'abord à lire des travaux sur le sujet étudié et sur des objets voisins. Ce travail se poursuit tout au long de l'enquête. Il permet de déterminer des sous-thèmes négligés dans la littérature et qu'il est parfois intéressant

1. C. Lemieux, « Problématiser », in S. Paugam, *L'Enquête sociologique*, *op. cit.*, p. 30.

d'approfondir. Cela offre l'occasion de clarifier les éventuels apports de l'enquête. À Paris 8 et Paris 10, ce travail, qui ne pouvait être exhaustif, a pris la forme d'une recension de plusieurs textes scientifiques, relativement accessibles. La lecture de ces travaux a été l'occasion de discuter les mots (confort, équipement, loyer, dépense, usage, etc.), d'évoquer leurs définitions et celles des concepts attachés au thème de l'enquête. C'est ainsi que les notions de « logement » et d'« habitat » ont pu être distinguées. « Le logement peut être *inhabité*, c'est-à-dire que l'on peut le concevoir pour lui-même, sans présence humaine. » En revanche, « l'habitat se conceptualise dans le logement et le logement accueille et délimite l'habitat »¹.

Ce travail bibliographique a été prolongé par un terrain exploratoire, en réalisant des entretiens non directifs auprès d'autres étudiants. Les entretiens devaient être conduits de préférence au domicile de l'enquêté afin d'allier l'observation à l'étude des discours. Ce moment a fait émerger la pluralité des modes d'habitat : vivre chez ses parents, être hébergé par un tiers ou vivre dans un logement indépendant en location, sous-location, avec ou sans aide financière familiale. Il a aussi permis d'approcher au plus près les catégories indigènes, c'est-à-dire faisant sens pour les enquêtés. Enfin, le terrain exploratoire a été l'occasion de confronter les représentations des intervieweurs à celles des interviewés faisant, par exemple, apparaître des différences dans l'appréciation d'un logement confortable. L'idée selon laquelle on évalue différemment le confort d'un logement selon la

1. F. Bugeja-Bloch, *Logement. La spirale des inégalités*, Paris, Puf, « Le Lien social », p. 4.

penne de la trajectoire résidentielle, largement étayée par les travaux de Pascal Dietrich-Ragon¹, a donc été avancée. Ainsi, le rôle de la phase exploratoire est bien crucial et multiple : préciser la problématique, remettre en cause ses *a priori*, saisir toutes les situations possibles en vue de la standardisation et faire émerger des premiers résultats. Forts des informations recueillies, les étudiants de Paris 8 et de Paris 10 ont rédigé en binôme une première version du questionnaire (V1).

3. **Les tests.** – Le travail d’exploration et de définition de l’objet se prolonge généralement par une phase de tests. À ce stade, il est encore possible de revenir sur ses pas, d’appréhender le sujet différemment, de reformuler, d’une part, la problématique de l’enquête, et, d’autre part, les questions et les réponses du questionnaire. Grâce aux tests, la longueur du questionnaire est ajustée, les formulations de question ambiguës, maladroitement ou inappropriées sont corrigées, les modalités de réponses oubliées sont ajoutées. Durant cette phase, on met donc à l’épreuve les normes que l’on va appliquer aux matériaux. Les étudiants de Paris 8 et Paris 10, eux aussi, ont mené plusieurs tests. La première version du questionnaire (V1), construite par binôme, a été soumise à un autre groupe d’étudiants lors d’une passation en face à face. Les corrections apportées ont donné lieu à une deuxième version du questionnaire (V2). Les deuxièmes versions ont été recueillies par les enseignants afin de construire un questionnaire commun aux deux universités (V3). La troisième version, commune à tous, a

1. P. Dietrich-Ragon, *Le Logement intolérable. Habitants et pouvoirs publics face à l’insalubrité*, Paris, Puf, « Le Lien social », 2011.

été testée en condition de passation « réelle », c'est-à-dire de la façon dont il est prévu de l'administrer aux enquêtés. Chaque étudiant a donc répondu seul en séance à une version papier du questionnaire. Leurs dernières remarques ont été intégrées dans une version finale (fig. 5).

II. – La structure du questionnaire

Pour élaborer un questionnaire, il n'y a pas de règles toutes faites, mais quelques grands principes ont fait leurs preuves. Certains d'entre eux concernent la structure du questionnaire, c'est-à-dire sa mise en page et sa cohérence interne. L'objectif que l'on se fixe, mais qui, en toute lucidité, ne sera jamais complètement atteint, est d'éviter les refus de passation et les abandons en cours de route.

1. Titre et présentation de l'enquête. – Les étudiants de Paris 8 et de Paris 10 ont eu d'emblée pour consigne de soigner la mise en page du questionnaire afin de susciter l'intérêt des enquêtés. Cela implique que ces derniers puissent prendre rapidement connaissance du sujet sur lequel ils sont interrogés et suppose donc la rédaction d'un titre sans ambiguïté. Un paragraphe introductif permet également de préciser les tenants et aboutissants de l'enquête. Celui-ci s'adresse à l'ensemble des enquêtés. La problématique, propre au chercheur, n'a pas à être mentionnée. Les termes trop techniques ou conceptuels sont à éviter. Au contraire, il s'agit de rendre la démarche intelligible et attractive au plus grand nombre. Toujours dans cette perspective figurent dans l'en-tête des remerciements et des garanties concernant l'anonymat des personnes.

Fig. 5. – Version finale du questionnaire

Date __/__/__

**Le logement et l'habitat des étudiants
des Universités Paris 8 et Paris 10**

Cette enquête vise à saisir la diversité des conditions de logement et des modes d'habitat des *étudiants de troisième année* des Universités de Paris 8 et Paris 10.

Quels étudiants pour quels logements ? Pourquoi certains quittent le domicile parental et d'autres pas ? Dans quelles conditions vivent-ils ? Quels usages les étudiants font de leur logement ? Comment y accèdent-ils et le financent-ils ?

Ce questionnaire est *strictement anonyme* et vos réponses serviront *exclusivement* à des analyses statistiques. Nous vous remercions de répondre à ce questionnaire avec le plus grand soin.

1. Vous êtes inscrit(e) en 2012-2013 à :
Paris 8 1 Paris 10 2

Et précisez la formation (ex : Droit, etc.) :
L3 de

2. Avez-vous obtenu le baccalauréat (ou équivalent étranger) ?
Oui 1 Non 2

Si oui, en quelle année ?

et dans quelle filière ?
Professionnelle 1
Technologique 2
Générale 3
Autre 4

3. Quel est votre sexe ?
Masculin 1 Féminin 2

4. Quelle est votre année de naissance ?
 1 9 . .

Tout d'abord, parlons de vos parents...

5. Vos parents ou tuteurs ont-ils poursuivi des études au-delà du baccalauréat ?

	Père/tuteur	Mère/tutrice
Oui	<input type="checkbox"/> 1	<input type="checkbox"/> 1
Non	<input type="checkbox"/> 2	<input type="checkbox"/> 2

6. Quelle est la situation actuelle de vos parents ou tuteurs ?

	Père/tuteur	Mère/tutrice
En emploi	<input type="checkbox"/> 1	<input type="checkbox"/> 1
Au chômage	<input type="checkbox"/> 2	<input type="checkbox"/> 2
Retraité(e)	<input type="checkbox"/> 3	<input type="checkbox"/> 3
Au foyer	<input type="checkbox"/> 4	<input type="checkbox"/> 4
Décédé(e)	<input type="checkbox"/> 5	<input type="checkbox"/> 5
Autre	<input type="checkbox"/> 6	<input type="checkbox"/> 6

7. Quelle est la profession actuelle de vos parents ou celle qu'ils exerçaient quand ils travaillaient ? (Soyez le plus précis possible)

Père/tuteur:
Mère/tutrice :

8. Exercent-ils ou exerçaient-ils leur profession comme :

	Père/tut.	Mère/tut.
Salarié(e) de l'État, d'une collectivité locale, d'un hôpital	<input type="checkbox"/> 1	<input type="checkbox"/> 1
Salarié(e) d'une entreprise publique	<input type="checkbox"/> 2	<input type="checkbox"/> 2
Salarié(e) du privé	<input type="checkbox"/> 3	<input type="checkbox"/> 3
À leur compte	<input type="checkbox"/> 4	<input type="checkbox"/> 4

9. Vos parents ou tuteurs sont :

	Père/tut.	Mère/tut.
Locataire d'un logement social (HLM)	<input type="checkbox"/> 1	<input type="checkbox"/> 1
Locataire d'un logement privé	<input type="checkbox"/> 2	<input type="checkbox"/> 2
Propriétaire	<input type="checkbox"/> 3	<input type="checkbox"/> 3
Autre :	<input type="checkbox"/> 4	<input type="checkbox"/> 4

10. Où résident vos parents ?

Père : Commune ou pays :
 Code postal :

Mère : Commune ou pays :
 Code postal :

Nous allons maintenant parler du logement dans lequel vous résidez en semaine...

11. En semaine, où résidez-vous ?

Commune :
 Code postal

12. Actuellement, en semaine, vivez-vous :

- Chez vos parents (ou tuteurs) 1
- Chez un tiers 2
- Dans un logement indépendant (seul, en couple ou en colocation) 3
- Autre : 4

13. Si vous ne vivez pas chez vos parents, êtes-vous :

- En résidence universitaire (CROUS) 1
- Locataire d'un logement social 2
- Locataire ou sous-locataire 3
- En colocation 4
- Propriétaire 5
- Logé gratuitement 6
- Autre : 7

14. Pour tous, combien de temps mettez-vous pour vous rendre à l'Université ?

..... heure(s).....minutes

15. Quels moyens de transport utilisez-vous pour vous rendre à l'Université (plusieurs réponses possibles) ?

- Transports en commun 1
- Véhicules motorisés 2
- Autres véhicules (vélo) 3
- À pied 4

Nous allons désormais parler du confort du logement dans lequel vous résidez en semaine...

16. Sur une échelle de 0 à 10, évaluer le confort de votre logement actuel (10 correspondant au degré maximum de confort) :

0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10

17. À l'intérieur du logement dans lequel vous résidez en semaine, disposez-vous des équipements suivants :

	Oui	Non
Toilettes	<input type="checkbox"/> 1	<input type="checkbox"/> 2
Douche/Baignoire	<input type="checkbox"/> 1	<input type="checkbox"/> 2
Réfrigérateur	<input type="checkbox"/> 1	<input type="checkbox"/> 2
Plaques de cuisson	<input type="checkbox"/> 1	<input type="checkbox"/> 2
Four/micro-ondes	<input type="checkbox"/> 1	<input type="checkbox"/> 2
Lave-vaisselle	<input type="checkbox"/> 1	<input type="checkbox"/> 2
Télévision	<input type="checkbox"/> 1	<input type="checkbox"/> 2
Ordinateur	<input type="checkbox"/> 1	<input type="checkbox"/> 2
Connexion Internet	<input type="checkbox"/> 1	<input type="checkbox"/> 2

18. Combien de pièces comprend votre logement (hors cuisine, sanitaires, garage et grenier) ?.....

19. Combien de personne(s) réside(nt) en semaine dans votre logement (vous y compris) ?.....

20. Si vous habitez dans un immeuble, combien d'étages compte-t-il ?.....

et à quel étage résidez-vous ?.....

et disposez-vous d'un ascenseur ?

Oui 1 Non 2

Nous allons désormais parler de l'usage que vous faites de votre logement...

21. Au cours du mois dernier, avez-vous reçu des membres de votre famille chez vous ?

0 fois 1

1 à 3 fois 2

4 à 7 fois 3

8 fois ou plus 4

22. Au cours du mois dernier, avez-vous reçu des amis chez vous ?

0 fois 1

1 à 3 fois 2

4 à 7 fois 3

8 fois ou plus 4

23. Au cours du mois dernier, où avez-vous effectué votre travail universitaire ? (plusieurs réponses possibles)

Toujours chez vous 1

Chez vous, la plupart du temps 2

Parfois chez vous, parfois ailleurs 3

Ailleurs, la plupart du temps 4

Toujours ailleurs 5

24. Depuis que vous êtes installé dans votre logement (ou chambre) y avez-vous fait ou fait faire des travaux (ex : peinture, tapisserie, ...)

Oui 1 Non 2

Nous allons maintenant parler de vos ressources...

25. Actuellement de quelles ressources disposez-vous (plusieurs réponses possibles) ?

Aide financière occasionnelle de votre famille 1

Aide financière régulière de votre famille 2

Bourse d'étude 3

Rémunération d'une activité professionnelle exercée l'été dernier 4

Rémunération d'une activité professionnelle exercée durant l'année universitaire 5

Ressources de votre conjoint(e) 6

Prêt bancaire aux étudiants 7

Allocation pour le logement 8

Autre(s) :..... 9

26. Si vous exercez une activité professionnelle durant l'année universitaire, de quoi s'agit-il (si vous avez plusieurs activités choisissez la principale) ?

.....

et êtes-vous ?

Salarié ou intérimaire 1

Stagiaire 2

Employé par un particulier 3

À votre compte (auto-entrepreneur) 4

Autre :..... 5

et est-ce une activité régulière ?

Oui 1 Non 2

et combien d'heures avez-vous travaillé au cours de la semaine dernière ?

.....

et quelle somme (en euros net) avez-vous gagné durant le mois dernier.....€

27. En comptant l'ensemble de vos ressources, de quel montant disposiez-vous le mois dernier (*en euros net*) ?.....€

Nous allons désormais parler de vos dépenses en matière de logement...

28. *Si vous vivez chez vos parents, tuteurs ou chez un tiers*, contribuez-vous au budget du foyer ?

Oui 1 Non 2

Si oui, combien le mois dernier :€

29. *Si vous habitez dans un logement indépendant*, quel montant avez-vous consacré à votre logement le mois dernier ?

Loyer mensuel (ou mensualité)	(ou charges) €
comprises		
Montant des APL	 €

et si APL à qui la CAF les verse-t-elle ?

À vous 1 Au propriétaire 2

et votre famille participe-t-elle à vos dépenses de logement ?

Non 1
En partie 2
En totalité 3

Nous allons maintenant parler de votre trajectoire résidentielle...

30. Depuis combien de temps vivez-vous dans votre logement actuel ?

.....années.....mois

31. Depuis que vous avez commencé vos études supérieures, avez-vous habité dans d'autres logements ?

Oui 1 Non 2

Si oui, combien :

32. Pourquoi avez-vous quitté ou quitteriez-vous le domicile parental (*plusieurs réponses possibles*) ?

Pour vous rapprocher de votre lieu d'étude	1
Pour vous rapprocher de votre lieu de travail	2
Pour vous rapprocher de vos amis ou de vos loisirs	3
Pour vous éloigner du contexte familial	4
Par envie d'indépendance, d'autonomie	5
Pour davantage de confort	6
Pour disposer d'un espace à soi	7
Pour vous mettre en couple	8
Autre :	9

33. *Si vous avez déjà cherché un logement*, sur une échelle de 0 à 10, décrivez l'ampleur des difficultés que vous avez rencontrées pour le trouver (*10 correspondant à un degré de difficulté maximal*) :

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

et, combien d'appartement(s) avez-vous visité ?

1 à 3	1
3 à 6	2
6 à 10	3
Plus de 10	4

et combien de temps a duré votre recherche ?

Moins d'un mois	1
1 à 3 mois	2
3 à 6 mois	3
6 mois ou plus	4

Merci d'avoir bien voulu remplir ce questionnaire

2. Connecteurs et filtres. – Lors de l'examen détaillé de différents questionnaires, tels que le bulletin individuel du recensement de la population, les étudiants de Paris 8 et de Paris 10 ont relevé la présence de connecteurs ou de filtres. Les connecteurs (encadrés en gras, fig. 3) permettent aux enquêtés de suivre la cohérence de l'ensemble du questionnaire. En maintenant leur attention sur le sens de la démarche, on cherche à éviter les abandons. Chaque section représente ainsi un nouveau sous-thème relatif à la problématique générale du logement des étudiants. Il s'agit en quelque sorte du plan du questionnaire. Nous avons finalement opté pour le vouvoiement de politesse, mais, compte tenu de l'âge du public auquel s'adressait l'enquête, la question du tutoiement a été soulevée en séance. Ce dernier ne faisant pas l'unanimité durant les tests (dynamisant l'enquête pour les uns, infantilisant et insultant l'enquêté pour les autres), l'idée a été abandonnée.

Outre les connecteurs, il arrive que les questionnaires présentent des filtres. Les filtres, comme leur nom l'indique, permettent de sélectionner une partie de la population enquêtée. D'emblée, ils annoncent aux enquêtés disposant de caractéristiques particulières qu'ils ne sont pas concernés par la ou les questions qui suivent. Lorsqu'ils sont clairs et bien apparents, leur rôle est d'éviter la lassitude de qui serait contraint de lire inutilement une ou plusieurs questions ne le concernant pas. Dans cette enquête, nous avons parfois manqué de clarté. Par exemple, l'intitulé de la question 13 spécifie (en italique et gras) que seuls les étudiants ne vivant *pas* chez leurs parents sont tenus de répondre. Pourtant, sur les sept cent soixante et un étudiants ayant déclaré à la question précédente (Q12) vivre actuellement, en semaine, chez leurs parents (ou tuteurs), quatorze ont, malgré tout, répondu à la question 13 (soit environ 2 %).

Six déclarent résider en résidence universitaire, quatre affirment être logés gratuitement, deux habitent dans un logement social et enfin deux étudiants sont dans une autre situation, sans plus de précision. Ce constat nous invite à revoir la formulation ou la visibilité de ce filtre, mais nous rappelle aussi que la frontière entre un logement à soi et celui des parents n'est pas toujours nette, qu'il existe des allées et venues, que les doubles résidences ne sont pas rares.

3. L'ordre des questions. – L'ordre des questions participe également de la cohérence globale d'un questionnaire. En principe, les enquêtés s'investissent davantage lorsqu'ils comprennent l'enchaînement des idées. Il est également conseillé de ne pas débiter avec des sujets vécus comme trop personnels. On préférera attendre qu'une relation de confiance s'instaure. Ajoutons que la manière dont les questions sont agencées exerce aussi une influence sur les réponses recueillies. « On gardera à l'esprit cet effet de l'"environnement" des questions non seulement lors de la conception du questionnaire, mais aussi lors de l'analyse des données ¹ ». C'est l'exercice auquel se prête Michel Gollac à partir de l'enquête « Conditions de travail » de l'Insee ². En 1984, l'une des questions concerne les charges lourdes. Elle était formulée de la façon suivante : « L'exécution de votre travail vous impose-t-elle de porter ou de déplacer des charges lour-

1. I. Parizot, « L'enquête par questionnaire », in S. Paugam, *L'Enquête sociologique, op. cit.*, p. 106. Cela est particulièrement vrai lorsqu'on compare les résultats obtenus à la même question énoncée dans des environnements différents.

2. M. Gollac, *Donner un sens aux données. L'exemple des enquêtes statistiques sur les conditions de travail*, Centre d'études de l'emploi, 1994.

des ? » La réponse attendue était « oui » ou « non ». Il s'agissait d'appréhender les charges réellement manipulées par les enquêtés ou la manière selon laquelle elles étaient ressenties. Mais, posée après d'autres portant explicitement sur les pénibilités au travail et notamment sur les postures « fatigantes à la longue », elle a pu être comprise comme une dénonciation de mauvaises conditions de travail (« porter des charges *trop* lourdes »).

Ainsi, le sens que les enquêtés accordent aux questions dépend du contexte de leur énonciation. Au demeurant, les enquêtés visent généralement à présenter une image cohérente d'eux-mêmes. Évitant de se contredire, ils donnent leurs réponses en fonction de celles apportées aux questions précédentes. Ainsi, la question 16 a volontairement été proposée avant d'autres dont l'objectif était d'approcher les caractéristiques dites objectives du logement (Q17, Q18, Q19). L'envie des étudiants de Paris 8 et Paris 10 était de recueillir une opinion spontanée des enquêtés et non une opinion énoncée au regard des équipements dont ils disposent. Un enquêté aurait sans doute eu plus de scrupule à donner à son logement une note en dessous de la moyenne après avoir affirmé posséder chez lui tous les équipements de la liste (toilettes, douche ou baignoire, réfrigérateur, lave-linge, télévision, etc.).

4. Le talon sociologique. – Outre les questions relatives à la thématique étudiée, un questionnaire contient toujours des interrogations destinées à approcher le capital économique, culturel et social des personnes interrogées, autrement dit leurs ressources. Cette série de questions constitue ce qu'on appelle le talon sociologique. On la trouve, en principe, au début ou à la fin du questionnaire. Elle ne change peu ou pas d'une enquête

à l'autre. Le questionnaire est aussi « l'art du plagiat¹ » ! Certaines questions ont déjà fait leurs preuves. Dans l'exemple qui nous occupe, les huit premières questions appartiennent au talon. Il faut être attentif à ne pas laisser l'enquêté, mais huit questions constituent un *minimum*. L'enquêté ne verra pas systématiquement le lien avec le sujet de l'enquête ; en revanche il s'agit pour le sociologue de collecter les principaux déterminants sociaux des phénomènes qu'il étudie : sexe, âge, diplôme et origine sociale. La catégorie socioprofessionnelle d'*ego* ou de ses parents, au sens de la nomenclature des PCS de l'Insee, est généralement admise comme indicateur de position sociale ou d'origine sociale². Or, nous y reviendrons, cette nomenclature est multidimensionnelle. En principe, pour la saisir, il faut poser pas moins de huit questions : (1) intitulé en clair de la profession ; (2) la position à l'égard du marché du travail (actif occupé, chômeur, retraité ou autre inactif) ; (3) le statut de la personne (salarié ou indépendant) ; (4) l'activité économique de l'établissement ; (5) le type d'employeur (privé ou public) ; (6) le grade ou l'échelon (pour les salariés de la fonction publique) ; (7) la classification professionnelle (au sens des conventions collectives) ; (8) le domaine de fonction (ingénieurs, technicien, etc.). Dans notre cas, nous n'avons ni l'espace dans le questionnaire (4 pages), ni le temps durant la passation (25 minutes) d'atteindre cet idéal. Seules quatre

1. J.-P. Grémy, « Questions et réponses. Quelques résultats sur les effets de la formulation des questions dans les sondages », *Sociétés contemporaines*, n° 16, 1993, p. 176.

2. D'après la définition de l'Insee, « La nomenclature des professions et catégories socioprofessionnelles dites PCS a remplacé, en 1982, la CSP. Elle classe la population selon une synthèse de la profession (ou de l'ancienne profession), de la position hiérarchique et du statut (salarié ou non) ».

questions subsistent pour déterminer l'origine sociale des enquêtés (Q5, Q6, Q7, Q8). Déjà intéressantes en soi, elles vont également jouer un rôle déterminant lors de la saisie des questionnaires (chap. IV).

III. – La formulation des questions et des réponses

Si l'architecture d'un questionnaire a une influence sur les refus de passation, les abandons de passation et les réponses des enquêtés, il en va de même de la formulation des questions et de leurs réponses. Il convient donc d'y être vigilant, tant lors de l'élaboration du questionnaire qu'au moment de l'exploitation des données. Il n'y a, ici encore, pas de règles établies, mais des astuces pour y faire face et en faire l'analyse. Elles concernent tous les types de questions : les questions fermées et ouvertes, les questions de fait et d'opinion¹.

1. Questions fermées et questions ouvertes. – Par définition, les questionnaires contiennent une majorité de questions fermées. Les enquêtés sont amenés à choisir leurs réponses dans une liste d'items prédéfinis. Dans la mesure où c'est là qu'intervient le plus gros du travail de standardisation des données, c'est précisément cette étape d'élaboration des items qui distingue l'enquête par questionnaire des autres techniques d'enquête. Seuls une connaissance profonde du terrain et de nombreux tests permettent de mener à bien cette phase.

« Mener à bien » l'élaboration de questions signifie, d'abord, trouver des formulations n'induisant ni jugements normatifs ni, trop fortement, certaines réponses.

1. Cette distinction est notamment faite par F. de Singly, *Le Questionnaire. L'enquête et ses méthodes*, Paris, Armand Colin, « 128 », 2012 (3^e éd.).

On sait par exemple, que l'ordre selon lequel on annonce les modalités de réponse induit des effets. Jean-Paul Grémy montre que « dans un questionnaire administré oralement par un enquêteur, il semble que le fait d'être cité en fin de liste augmente le taux de réponses. Dans un questionnaire présenté par écrit, c'est au contraire la position en tête de liste qui augmente la probabilité d'être choisi¹ ». Un moyen d'y faire face est d'invertir les modalités entre les enquêtés. En outre, dans la mesure du possible, les formulations doivent être sans ambiguïté, c'est-à-dire comprises par tous de la même façon. Pour ce faire, il faut employer des mots au plus près des catégories de pensée des enquêtés.

« Mener à bien » l'élaboration des items revient enfin à identifier l'ensemble des modalités de réponses probables. Tous les enquêtés doivent pouvoir se retrouver dans les réponses offertes. Pour les cas oubliés (parce que trop rares lors du terrain exploratoire) ou trop singuliers, il leur reste la possibilité d'indiquer « autre » et de préciser leur situation. Pour exemple, à la question 12, l'option « autre » a été choisie par trente-huit individus (soit 3 % des répondants ; voir fig. 3). Déclarant vivre chez une sœur, un oncle, une tante, de la famille, les précisions qu'ils apportent permettent de comprendre que l'expression « chez un tiers », sans doute maladroite de notre part, a été mal comprise par au moins huit d'entre eux.

Dans une enquête par questionnaire, les questions ouvertes sont (très) minoritaires. Avec elles, il n'y a pas de modalités de réponse imposées. On laisse les enquêtés s'exprimer librement. Dans le questionnaire de Paris 8 et de Paris 10, la question sur la profession des parents

1. J.-P. Grémy, « Questions et réponses. Quelques résultats sur les effets de la formulation des questions dans les sondages », *Sociétés contemporaines*, n° 16, 1993, p. 167.

(Q7) est un exemple (fig. 3). Des questions mixtes y figurent aussi, telles que la question 32, où il est proposé aux enquêtés de détailler une autre raison. Les réponses obtenues sont tout aussi diverses que « jouer de la musique » ou « découvrir Paris ». Cependant, des occurrences sont apparues plus fréquemment, telles que « voyage », « étudier à l'étranger », que nous aurions pu regrouper *a posteriori* sous une seule étiquette : « mobilité à l'étranger ». Ainsi, les questions ouvertes permettent d'accéder, davantage encore, aux catégories de pensée des individus et donnent parfois lieu à des analyses lexicales ou à des codages originaux (chap. IV). Nécessitant une standardisation *a posteriori* des réponses, leur exploitation reste plus longue et difficile. En outre, on les dit plus sensibles aux effets d'interactions avec l'enquêteur. Lors d'une enquête en face à face, on se confiera plus librement à un intervieweur affable ou, selon les sujets et les cas, à un homme plutôt qu'à une femme, etc. Les enquêteurs peuvent aussi noter les réponses avec plus ou moins de détails : aux mots près, avec les intonations de voix pour certains, avec approximation pour les autres.

2. Questions d'opinion et de faits. – Les questions fermées ou ouvertes sont susceptibles de porter sur des opinions ou sur des pratiques. Les questions d'opinion, souvent utilisées par les instituts de sondage, n'ont pas nécessairement bonne presse en sciences sociales. Bourdieu met en question trois postulats : « Toute enquête d'opinion suppose que tout le monde peut avoir une opinion ; ou, autrement dit, que la production d'une opinion est à la portée de tous [...]. Deuxième postulat : on suppose que toutes les opinions se valent. Je pense que l'on peut démontrer qu'il n'en est rien et que le fait de cumuler des opinions qui n'ont pas du tout

la même force réelle conduit à produire des artefacts dépourvus de sens. Troisième postulat implicite : dans le simple fait de poser la même question à tout le monde se trouve impliquée l'hypothèse qu'il y a un consensus sur les problèmes, autrement dit qu'il y a un accord sur les questions qui méritent d'être posées¹. » En outre, on les dit plus sensibles aux effets de formulation.

En sciences sociales, il est maintenant classique d'allier des questions d'opinion, ouvrant à une sociologie compréhensive, à des questions de fait portant sur les pratiques effectives des individus. Ainsi, dans le questionnaire des universités Paris 8 et Paris 10, tout l'enjeu de la question 33, sur l'appréciation subjective des problèmes rencontrés pour trouver un logement, est d'en croiser les résultats avec les propriétés sociales des répondants et avec les difficultés vécues en pratique (suite de la Q33). De prime abord, les questions de fait semblent plus tangibles que les questions d'opinion. Elles sont généralement plus fréquentes dans un questionnaire en sciences sociales. Toutefois, certaines d'entre elles demeurent sensibles². On sait, par exemple, que les Français peinent à répondre aux questions touchant à leurs revenus.

Le travail d'André Béjin sur l'enquête du groupe ACSF (Analyse des comportements sexuels en France) propose une analyse intéressante des réponses apportées

1. P. Bourdieu, « L'opinion publique n'existe pas », in *Questions de sociologie*, *op. cit.*, p. 222.

2. Nous limiterons ici aux questions sensibles du point de vue de l'enquêté, c'est-à-dire vécue par eux comme indiscrettes ou embarrassantes. Les questions sensibles peuvent l'être aussi du point de vue du droit. La loi informatique et libertés considère comme sensibles les « données à caractère personnel qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale des personnes, ou qui sont relatives à la santé ou à la vie sexuelle de celles-ci ».

à une question sensible et pose notamment le problème de la sous-déclaration¹. Dans l'enquête ACSF, deux questions portaient sur la pratique de la masturbation. La première est directe : « Vous êtes-vous masturbé(e) : souvent, parfois, assez rarement, jamais, non-réponse ? » Elle figurait au début du questionnaire. La seconde est indirecte : « Pouvez-vous me dire si, quand vous vous masturbez, vous parvenez à l'orgasme : toujours, plutôt facilement, plutôt difficilement, jamais, non concerné(e), non-réponse ? » Elle était placée en fin de questionnaire. Avec la formulation directe, seulement 42 % des femmes répondent s'être déjà masturbées. Avec la formulation indirecte, le pourcentage s'élève finalement à 51 %. Manifestement, on assiste à une sous-déclaration de la pratique dans le premier cas. L'examen des réponses des hommes aux deux mêmes questions ne montre pas, lui, de sous-déclaration importante. Trois effets interviennent ici : la position des questions dans le questionnaire, leur formulation, mais aussi le sexe des enquêtés. L'analyse des données tient compte de l'ensemble de ces éléments. Ici, il s'agit d'un cas de sous-déclaration, mais on assiste également à des cas de surdéclarations pour des pratiques valorisées socialement, telle la lecture. Soigner les formulations (avec des formes indirectes par exemple) et croiser plusieurs indicateurs sur un même sujet peut permettre d'identifier les déclarations abusives, et donc d'analyser les incohérences et éventuellement de les corriger ensuite. Dernière astuce, largement utilisée par les agents de la statistique publique dans les grandes enquêtes : circonscrire les questions dans le temps. Dans le questionnaire

1. A. Béjin, « La masturbation féminine en France. Un exemple d'estimation et d'analyse de la sous-déclaration d'une pratique », *Population*, 48 (5), 1993, p. 1437-1450.

des étudiants des universités Paris 8 et de Paris 10, cette stratégie a été adoptée pour la question 22 notamment : « *Au cours du mois dernier, avez-vous reçu des amis chez vous ?* » Pour les questions de fait, les formulations du type « généralement » et en « moyenne » sont à éviter à la faveur de questions portant sur une période donnée (« jours », « semaines », « mois »). En effet, dans le cas contraire, les enquêtés renseignent davantage la représentation qu'ils se font de leur pratique moyenne et non leurs pratiques effectives. Si dans la période considérée, la pratique a été exceptionnellement faible ou élevée, on peut renouveler la question (« Au cours de cette semaine... », puis « au cours de la précédente », etc.). En outre, le chercheur peut admettre, compte tenu du nombre de personnes interrogées, des situations marginales. Le problème est donc davantage de trouver le laps de temps approprié, qui fait sens pour les enquêtés. Dans l'exemple de Paris 8 et Paris 10, la période de la semaine, plus que celle du mois, aurait sans doute été plus adaptée. Le questionnaire élaboré, vient maintenant le temps de l'exécution sur le terrain.

Chapitre IV

L'EXÉCUTION SUR LE TERRAIN

Au cours du processus de quantification, le chercheur est en contact à plusieurs reprises avec les enquêtés, durant la phase exploratoire d'abord, au moment des tests ensuite. Il fait donc de nombreuses allées et venues entre ses catégories de pensée et celles des enquêtés. Fort de ces matériaux, il peut enfin finaliser le questionnaire. Une fois ce dernier achevé, il est inenvisageable de revenir en arrière.

Le processus de quantification ne s'arrête toutefois pas là. Au contraire, il s'agit désormais d'exécuter sur le terrain un protocole d'enquête défini par le chercheur. Ce protocole se compose généralement d'un plan de sondage, d'un mode de passation et se prolonge avec des consignes de saisie. Cette procédure sera inévitablement soumise à rude épreuve, malmenée au cours de nombreuses péripéties, celles-ci feront partie intégrante du processus de quantification.

I. – Procédures d'échantillonnage et inférence statistique

Les techniques d'analyse qualitatives et quantitatives visent toutes deux à éclairer des processus sociaux qui se passent ailleurs que sur la seule population observée. Dans un cas, le chercheur s'appuie sur la « qualité » des données, c'est-à-dire leur richesse. Dans

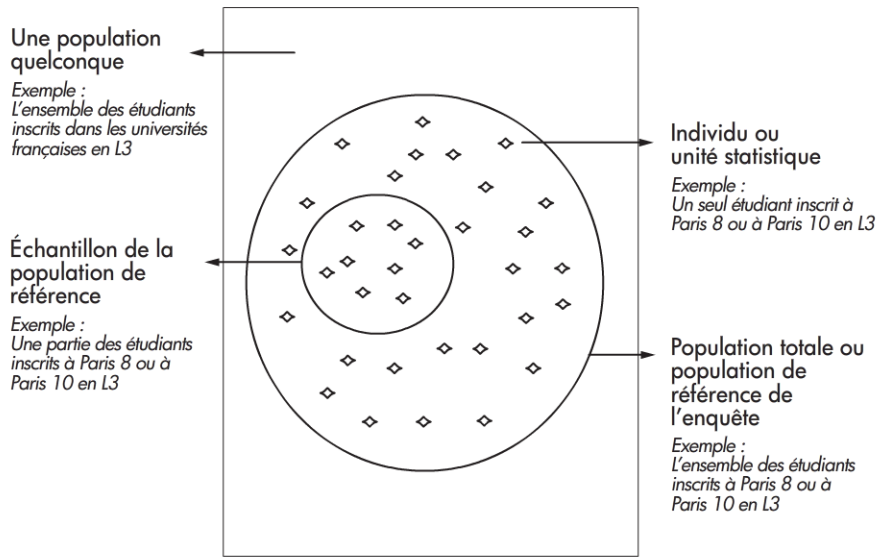
l'autre (en particulier lors d'une enquête par questionnaire), les matériaux sont standardisés et l'accent est mis sur la « quantité » des informations recueillies. Pour déterminer la « juste quantité » (ni trop, ni trop peu), les chercheurs en sciences sociales s'en remettent aux découvertes de statisticiens et à des notions comme l'« échantillon », la « représentativité » ou encore l'« inférence ». Il n'y a pas lieu ici d'entrer dans les détails mathématiques. Nous nous limiterons à présenter les grands principes de ces notions présentées notamment par Olivier Martin¹.

1. Qu'est-ce qu'un échantillon ? – L'exécution sur le terrain passe d'abord par l'élaboration d'un plan de sondage. On parle aussi de procédures d'échantillonnage. Pour construire un échantillon, il faut définir une population de référence. Dans l'exemple ci-dessous, une population quelconque est représentée (fig. 6). Il s'agit de l'ensemble des étudiants inscrits dans les universités françaises en troisième année de licence. Notre étude porte plus spécifiquement sur le logement et l'habitat des étudiants de deux universités franciliennes (Paris 8 et Paris 10). La population de référence de notre enquête est donc l'ensemble des étudiants inscrits en troisième année de licence dans ces deux universités.

Dans une enquête exhaustive, *tous* les individus de la population de référence sont interrogés. Ce type d'enquête coûteuse en temps et en argent est relativement rare. En France, on cite souvent le recensement de la population : le dernier recensement exhaustif de la population française date de 1999.

1. O. Martin, *L'Enquête et ses méthodes. L'analyse de données quantitatives*, Paris, Armand Colin, « 128 », 2012, p. 14-22.

Fig. 6. – Construction d'un échantillon



En effet, depuis 2004, son mode de collecte a changé et une partie seulement des résidents des communes de dix mille habitants ou plus sont interrogés. Il s'agit donc désormais d'un échantillon. Ainsi, dans une enquête par échantillon, seule une partie de la population de référence est observée. Dans notre exemple, une partie seulement des étudiants inscrits en troisième année de licence dans les deux universités devrait être interrogée (ou observée). Moins chères, les enquêtes par échantillon permettent aussi de mieux appréhender les conditions de passation réelle du questionnaire et leurs effets. Effectivement, il vaut mieux rester au plus près du processus de quantification en interrogeant un nombre relativement « restreint » d'individus que de solliciter beaucoup de personnes sans en maîtriser les conditions. D'autant qu'au-delà d'un certain seuil, il est probable que les nouvelles personnes interrogées ne nous en apprennent pas davantage sur le phénomène étudié. Pour définir ce seuil, les quantitativistes s'appuient sur la théorie des sondages.

La théorie des sondages indique comment obtenir le meilleur échantillon possible. La qualité d'un échantillon dépend de sa capacité à généraliser ce que l'on y observe à la population d'ensemble. Ce raisonnement porte un nom en statistique : l'*inférence*. En réalité, plus modestement, un échantillon correct doit permettre, d'une part, de minimiser les risques de se tromper et, d'autre part, de mesurer ce risque d'erreur. Cela tient notamment à la représentativité de l'échantillon.

Un échantillon est qualifié de représentatif quand il possède les mêmes caractéristiques (sociodémographiques) que la population d'ensemble. En conséquence, la notion de représentativité n'a de sens que si, d'une part, elle est rapportée à une population donnée (en l'occurrence représentative des étudiants de troisième année de Paris 8 et Paris 10) et, d'autre part, au sein de cette population, à certains critères (sexe, âge, origine sociale, etc.). La loi des grands nombres garantit généralement la représentativité de ces critères. Publiée par Jacques Bernoulli au début du XVIII^e siècle, elle précise que si l'on interroge suffisamment d'individus *au hasard*, on va s'approcher des caractéristiques de la population entière. Elle s'applique uniquement aux méthodes d'échantillonnage dites aléatoires.

2. Les différentes méthodes d'échantillonnage. – Il existe différentes méthodes d'échantillonnage, chacune comprenant des avantages et des inconvénients. D'abord, on distingue la famille des échantillons aléatoires (simple, stratifié, par grappe) de celle des échantillons non aléatoires (par quotas, notamment). Brièvement, la première méthode d'échantillonnage dite aléatoire simple (ou probabiliste) consiste à tirer des individus au hasard dans la population de référence de l'enquête. Les membres de la population de référence ont tous

une chance identique d'être tirés au sort, aucun critère sociodémographique ne présidant à leur sélection. Ce type d'échantillon ne nécessite donc pas une connaissance précise de la population de référence, notamment de sa structure par âge et par sexe. En revanche, pour le mettre en œuvre, il est indispensable de disposer d'une liste *exhaustive* de tous les membres de la population étudiée.

Disposer d'une telle liste est relativement rare. Pour les enquêtes représentatives de l'ensemble de la population française, seul l'Insee dispose d'une telle liste et est autorisé à en faire usage. Il s'agit de l'un de ses monopoles. C'est pourquoi nombre d'enquêtes menées dans d'autres institutions (Ined, Inserm) sont construites en partenariat avec l'Insee. L'échantillon aléatoire peut être stratifié. On découpe alors la population en strates et, au sein de ces strates, on tire au sort les individus. Dans le cas de l'enquête des étudiants de Paris 8 et de Paris 10, les strates auraient pu être les filières de l'université où les étudiants auraient été tirés au sort. L'intérêt ici est de s'assurer que l'on dispose d'un nombre suffisant d'étudiants dans chacune des filières pour y conduire des analyses fines. Un tel échantillon n'est pas toujours représentatif de l'ensemble de la population étudiée, mais seulement de chacune des strates. Une autre option est l'échantillon aléatoire par grappe. Dans ce cas de figure, on considère que les individus sont regroupés par grappe (un groupe de TD, par exemple). On va donc tirer au sort les grappes et, au sein de ces dernières, interroger tous les individus (tous les étudiants des TD tirés au sort). L'avantage ici est que seule la liste exhaustive des grappes (TD) et non des individus (étudiants) est indispensable. Notons que ces différents types d'échantillon peuvent se cumuler.

La seconde famille de méthodes est celle des échantillons non aléatoires. Dans ce cas de figure, les individus ne sont pas choisis au hasard. En d'autres termes, les chances d'appartenir à l'échantillon ne sont pas identiques pour tous les individus. L'exemple emblématique est l'enquête par quotas, largement utilisée par les instituts de sondage. Lors d'une enquête par quotas, l'enquêteur va chercher des enquêtés acceptant de répondre à l'enquête et respectant certains critères prédéterminés (par exemple : un étudiant de vingt-deux ans inscrit en troisième année de philosophie). En principe, les critères de sélection ont été établis pour que l'échantillon obtenu soit représentatif de la population d'ensemble. Il faut donc connaître en amont les caractéristiques socio-démographiques de la population que l'on souhaite étudier. L'échantillon ne sera représentatif qu'au sens de ces critères. En outre, l'approche se prête peu à la réflexivité, car on perd la trace des personnes qui n'ont pas souhaité répondre, disposant pourtant sans doute de certaines spécificités.

II. – Négocier l'accès au terrain et réaliser la passation

Les travaux en sciences humaines portant sur l'interaction sociale entre enquêteur et enquêté et sur la relation dissymétrique à laquelle elle renvoie sont nombreux. Ils ne sont pas spécifiques à l'enquête par questionnaire. Cependant, la division du travail à laquelle se prête ce type de recueil des données, au sein de la statistique publique notamment, conduit parfois le chercheur à perdre de vue tout ce qui se joue lors de l'administration du questionnaire aux enquêtés, c'est-à-dire lors de la passation. Pour en garder une trace, les étudiants des universités Paris 8 et Paris 10 ont rédigé

un compte rendu critique de passation. Il s'agissait de bien comprendre que le contact, même indirect, entre un enquêteur et des enquêtés constitue une situation sociale à part entière et, qu'à ce titre, il est impossible de le standardiser complètement. Le compte rendu de passation avait une visée descriptive et analytique. D'abord, les étudiants étaient chargés de décrire le mode d'accès des enquêtés, la relation entre enquêteur et enquêté et son contexte. Ensuite, ils devaient expliquer en quoi ces éléments ont pu intervenir sur le déroulement de la passation et la réception du questionnaire par les enquêtés.

1. Les différents modes de passation. – Il est commun de distinguer deux grands modes de passation. Les questionnaires peuvent être soumis par un enquêteur à un enquêté ou remplis seul par un répondant. Ici, on parlera de questionnaires autoadministrés. La présence d'un enquêteur auprès d'un répondant coûte cher, mais est souvent un atout. Il peut inciter l'enquêté à répondre à l'intégralité du questionnaire en l'encourageant et en lui apportant des précisions utiles. En outre, lorsque la passation a lieu en face à face, le chercheur maîtrise mieux le contexte de la collecte. Les enquêteurs peuvent adjoindre aux questionnaires quelques observations complémentaires, des remarques par exemple sur l'aspect extérieur du logement des enquêtés ou les effets de la présence d'un tiers durant l'entretien.

Avec un questionnaire autoadministré, dont l'un des exemples phares est l'enquête par Internet, le chercheur contrôle moins bien le profil des répondants et les conditions de la collecte. Récemment, l'enquête « Génération quoi ? La grande enquête sur les 18-34 ans », montée en partenariat par des journalistes et des sociologues, a remporté un franc succès auprès des

internautes¹. Cela nécessite un travail de rédaction du questionnaire important afin de le rendre attractif et de lutter contre les abandons. Au-delà du graphisme, l'option choisie par les concepteurs de l'enquête pour maintenir l'intérêt des enquêtés a été de leur restituer régulièrement en cours de passation les réponses données par les autres internautes. Chaque mode de passation produit des effets. Il faut donc choisir celui qui est le mieux adapté aux objectifs de l'enquête. Certains sujets par exemple se prêtent mieux à une enquête par téléphone, assurant une forme de distance entre enquêteur et enquêté, qu'à l'intrusion d'un enquêteur en face à face.

Concernant l'enquête des universités Paris 8 et Paris 10, le protocole était mixte. Sur une période de quinze jours, les étudiants devaient se rendre dans toutes les salles où des cours étaient dispensés aux étudiants de troisième année de licence, présenter brièvement l'enquête et ensuite seulement distribuer les questionnaires pour une passation autoadministrée. En théorie, ils devaient assister à cette dernière, se rendant disponibles pour d'éventuelles questions. Dans ce contexte, la problématique de l'accès au terrain et de ses conséquences a pu se poser, comme dans le cas d'une enquête ethnographique.

2. Négocier l'accès au terrain ou à une base de seconde main. – Pour administrer un questionnaire dans une salle de cours, et sur un laps de temps normalement consacré à l'enseignement, il faut évidemment négocier cette intrusion auprès du responsable de l'enseignement. Pour se faire, les apprentis enquêteurs disposaient d'une lettre-avis explicitant la démarche. Avec

1. Voir <http://generation-quoi.france2.fr/questions>.

un en-tête de l'université, la lettre devait leur assurer une certaine légitimité : ils ont malgré tout essuyé certains refus. Ils devaient alors revenir à la charge en proposant d'intervenir la semaine suivante. En outre, les enquêteurs devaient annoncer à l'enseignant, à l'avance, la durée de passation (20-25 minutes), quitte à la mésestimer sensiblement, ce qui ne va évidemment pas sans poser de problème déontologique. La bienveillance de l'enseignant qui les accueille selon sa sensibilisation à la démarche de l'enquête (différenciée selon les filières universitaires) ou les contraintes temporelles pesant sur le cours (en cas d'examen par exemple) ont été relevées dans les comptes rendus de passation comme autant d'éléments influençant l'attention portée par les enquêtés aux questionnaires.

Dans le cas d'une exploitation secondaire, il existe également des formes de négociation. L'objet des tractations est précisément d'accéder aux données appartenant par exemple à un syndicat, à une entreprise, à une association. La plus ou moins grande difficulté à les obtenir peut être un indicateur des enjeux que l'institution leur attribue. Les données issues de la statistique publique ou du monde académique sont, pour beaucoup d'entre elles, archivées et diffusées grâce à des centres d'archivages de données. Le Réseau Quételet donne notamment accès aux enquêtes de l'Insee par le truchement du Centre Maurice-Halbwachs¹, mais aussi à celles de l'Ined. D'autres sont par exemple archivées par le CDSP. Le Centre d'accès sécurisé distant aux données (CASD) fait également partie du réseau². Il

1. CMH-ADISP Centre Maurice-Halbwachs – Archives de données issues de la statistique publique (<http://www.cmh.ens.fr/>).

2. Le Centre d'accès sécurisé distant aux données (<http://casd.eu>) fait partie du Genes (Groupe des écoles nationales d'économie et statistique).

offre un équipement sécurisé conçu pour permettre aux chercheurs de travailler sur des données individuelles très détaillées, et donc particulièrement confidentielles. Les données de l'échantillon démographique permanent (EDP) ne sont accessibles qu'au travers de ce dispositif.

3. La passation, une situation sociale. – Selon Jean Peneff, les résistances rencontrées par l'analyse des statistiques comme faits sociaux tiennent précisément à la division des tâches dans les pratiques scientifiques et aux inégalités dans les conditions de travail : « Aux uns, sociologues, statisticiens qui étudient les représentations de la réalité : les tâches nobles, le prestige, de meilleures conditions d'emploi. Aux autres : enquêteurs, codeurs... les faibles rémunérations, la fragilité d'emploi, l'absence de prestige. Pourtant, sur ces derniers, repose une phase essentielle du travail scientifique ¹. » Rémy Caveng est allé au plus près de ces salariés déqualifiés dans les instituts de sondage ². La lecture de ses travaux par les étudiants de Paris 8 et Paris 10 a permis, plutôt que de les ignorer, d'objectiver sociologiquement la pratique dite de « bidonnage » dans les enquêtes par sondages. Le « bidonnage » renvoie à tous les manquements aux règles de passation prescrites par la hiérarchie de l'institut (dans le cas de Paris 8 et Paris 10, prescrites par les enseignants) qui impose en effet des contraintes difficilement tenables. Avant toute chose, c'est donc les contraintes pesant sur les enquêteurs qui engendrent les manquements aux règles de passation.

1. J. Peneff, « La fabrication statistique ou le métier du père », art. cité, p. 210-211.

2. R. Caveng, « La production des enquêtes quantitatives », *Revue d'anthropologie des connaissances*, vol. 6, n° 1, 2012, p. 65-88.

Bien que déqualifiée dans les instituts de sondage et dans ceux de la statistique publique¹, l'étape de passation est plutôt bien reçue par les étudiants dans le cadre d'un cours sur l'enquête par questionnaire. Malgré tout, les comptes rendus critiques de passation montrent que leurs pratiques sont très variables. Entre les étudiants respectant à la lettre les consignes fournies, ceux qu'un premier refus décourage et ceux qui recueillent les questionnaires au gré de leurs rencontres dans les couloirs de l'université, la différence est importante. Cependant, tous ont identifié un important biais de sélection. L'enquête prétendait interroger les étudiants inscrits en troisième année de licence. Or, seuls les étudiants effectivement présents sur le campus étaient enquêtés. Cet effet du mode de passation (à l'université) sur les données recueillies n'est évidemment pas négligeable. Les absents disposent à n'en pas douter de profils particuliers. Afin d'identifier les biais avec plus de précision, la structure de la population enquêtée (par âge, sexe, filière, etc.) a été confrontée avec celle, de l'année précédente, décrite par les données de l'OVE (Observatoire de la vie étudiante).

III. – Saisir les questionnaires

En principe, la phase d'exécution sur le terrain se prolonge par celle, plutôt austère, de la saisie des questionnaires, qui consiste à transformer les réponses en base de données, aujourd'hui informatisée. La saisie correspond à la phase de codification du questionnaire ou de son chiffrement. Face à la réalité de la saisie, ces consignes sont soumises à un certain nombre d'ajustements. À l'Insee,

1. Voir C. Bessière, F. Houseaux, « Suivre des enquêteurs », *Genèses*, n° 29, 1997, p. 100-114.

la déqualification des emplois de codeurs a longtemps été associée à une division sexuée du travail, les concepteurs des enquêtes étant massivement des hommes et les petites mains, les « chiffreuses », des femmes.

1. **Le principe de la saisie.** – En pratique, lors de la saisie, les enquêtés deviennent des lignes d’un tableau dit « base de données », les questions, des intitulés de colonnes, et les réponses, des valeurs remplissant les cases du tableau. Prenons l’enquête sur le logement et l’habitat des étudiants de l’université Paris 8 et de l’université Paris 10 et voyons comment apparaîtraient dans la base de données les réponses des étudiants Camille et Nathan à la question 12 :

Camille		Nathan	
12. Actuellement, en semaine, vivez-vous :		12. Actuellement, en semaine, vivez-vous :	
Chez vos parents (<i>ou tuteurs</i>)	×1	Chez vos parents (<i>ou tuteurs</i>)	1
Chez un tiers	2	Chez un tiers	2
Dans un logement indépendant (<i>seul, en couple ou en colocation</i>)	3	Dans un logement indépendant (<i>seul, en couple ou en colocation</i>)	3
Autre :.....	4	Autre :	×4
		Chez ma grand-mère	

Tableau 2. – **Base de données présentant la question 12**

Identifiant de l’enquêté	Mode d’habitat	Autre mode d’habitat
Camille	1	
Nathan	4	Chez ma grand-mère
...

En fait, la question 12 se décompose en deux sous-questions, la première porte sur le mode d'habitat, celui-ci se déclinant en quatre réponses possibles (chez vos parents, chez un tiers, dans un logement indépendant et autre) auxquels correspondent un numéro ou code (1, 2, 3, 4). Dans la statistique publique, certains codes respectent des conventions. Ainsi, les hommes sont systématiquement codés 1 et les femmes 2. Dans la question 12, la réponse « autre » est proposée afin de pouvoir s'assurer qu'aucune autre forme d'habitat n'a été omise. On suggère ainsi de renseigner cette autre forme lorsqu'elle est sélectionnée grâce aux pointillés. Lorsque la réponse « autre » est formulée, on renseigne alors non seulement la variable sur le « mode d'habitat » mais aussi celle sur l'« autre mode d'habitat ». Les deux sous-questions donnent ainsi lieu à deux variables, soit deux colonnes, dans la base de données. En résumé, la saisie consiste à transformer chaque question du questionnaire en une ou en plusieurs variables, et chaque réponse en valeur souvent représentée par un code.

Les progrès de la micro-informatique permettent désormais d'automatiser une grande partie de la saisie. Dans les grandes institutions comme l'Insee, les enquêteurs administrent même directement les questionnaires aux enquêtés au moyen d'un ordinateur et un logiciel opère seul la transformation des données. On parle de passation assistée par ordinateur ou de CAPI (Computer Assisted Personnel Interviewing). Le logiciel va également filtrer le questionnaire en ne présentant à l'enquêteur que les questions indispensables à poser selon les caractéristiques de l'enquêté. À Paris 8 et Paris 10, nous nous sommes inspirés de ce dispositif pour alléger le travail de saisie en l'automatisant en partie. Plutôt que de placer les

étudiants devant les lignes d'un tableur, nous avons utilisé un logiciel d'enquête. Dès lors, il a suffi aux étudiants de retranscrire les réponses apportées aux questionnaires papier, sur un questionnaire en ligne alimentant ainsi une base de données commune aux deux universités.

Malgré tout, avant d'analyser les données, il reste toujours à procéder à un apurement.

2. Le cas particulier de la PCS. – Parmi les codages *a posteriori*, il faut parfois plusieurs questions pour renseigner ou créer une seule variable de la base. C'est le cas des PCS. Pour la saisir, le codeur s'appuie sur une nomenclature disponible sur le site Internet de l'Insee. En théorie, il ne faut pas moins de huit informations pour placer les individus dans cette grille de positions sociales (voir chap. III). Cette grille a été conçue par l'Insee en 1954 (CSP) puis a été l'objet de plusieurs ajustements pour suivre les évolutions de la morphologie du marché du travail (1982¹ et 2003). Elle compte plusieurs niveaux d'agrégation emboîtés. Au niveau le plus agrégé, elle contient huit postes (tableau 3). Cet instrument n'est pas seulement déterminé par des considérations scientifiques. L'analyse menée par Luc Boltanski sur la formation du groupe des cadres est un bon exemple². Avant 1930, l'identité des cadres comme groupe social n'existe ni dans l'espace politique ni dans les statistiques. La reconnaissance de ce statut, et des droits associés, a supposé l'unification du groupe dans ces deux espaces.

1. Voir A. Desrosières, L. Thévenot, *Les Catégories socioprofessionnelles*, Paris, La Découverte, « Repères », 2002.

2. L. Boltanski, *Les Cadres. La formation d'un groupe social*, Paris, Minuit, « Le Sens commun », 1982.

Tableau 3. – PCS 2003 de niveau 1 :
liste des catégories socioprofessionnelles

<i>Code</i>	<i>Libellé</i>
1	Agriculteurs exploitants
2	Artisans, commerçants et chefs d'entreprise
3	Cadres et professions intellectuelles supérieures
4	Professions intermédiaires
5	Employés
6	Ouvriers
7	Retraités
8	Autres personnes sans activité professionnelle

3. De l'importance des non-réponses et des petits ajustements. – L'étape de la saisie permet de mettre au jour des erreurs commises durant l'enquête. Nous voudrions revenir ici sur deux maladroites qui nous rappellent l'importance des non-réponses et des petits ajustements auxquels le processus de quantification est soumis.

Selon Michel Gollac, « pour des raisons d'économie de saisie, les réponses "Non" a beaucoup de questions se traduisent par un "blanc" sur un fichier informatique. Elles se confondent alors avec les non-réponses à la question, ce qui est évidemment regrettable ¹ ». Ainsi, dans l'enquête sur le logement et l'habitat des étudiants de Paris 8 et de Paris 10, nous avons commis une première erreur lors de la rédaction de la question 32 sur les raisons, multiples, du départ du domicile parental (fig. 3). Ceux qui auraient souhaité apporter une réponse

1. M. Gollac, *Donner un sens aux données*, Noisy-le-Grand, Centre d'études de l'emploi, 1994, p. 29.

négative à chacune des propositions de réponse se confondent avec ceux qui n'ont pas souhaité répondre à cette question. C'est au moment de la saisie que l'erreur de formulation a été comprise, puisque cette étape permet de traduire les catégories de pensée des enquêtés en code ajusté à une base de données. Or, ne pas répondre à une question n'a évidemment pas la même signification que d'y apporter une réponse négative.

L'autre maladresse, commise plus tôt dans le processus de quantification, concerne la question 23 : « Au cours du mois dernier, où avez-vous effectué votre travail universitaire ? » Une précision était apportée à cette dernière : « Plusieurs réponses possibles ». Or, cette précision était inappropriée, les modalités de réponse devant être exclusives les unes des autres. Ainsi, malgré les nombreux tests et relectures, des imperfections échappent toujours au chercheur. Devant l'opportunité qui leur était offerte, certains enquêtés ont coché deux réponses. Dès lors, durant la saisie, nous avons été contraints de faire un choix entre deux possibilités. Ce choix a été appliqué uniformément, c'est-à-dire à chaque fois que la situation s'est présentée, mais il contenait une grande part d'arbitraire. Il est par conséquent de toute première importance de garder les traces de ces décisions qui, lors de l'exploitation, ne manqueront pas de peser sur le sens à accorder aux données.

Chapitre V

PANORAMA DES TECHNIQUES QUANTITATIVES

L'exploitation statistique de l'enquête n'a pas le charme de l'investigation de terrain. Le coût d'entrée dans la phase d'exploitation est effectivement plus élevé puisqu'elle nécessite l'utilisation d'un logiciel de traitement statistique et la maîtrise d'un certain nombre de techniques d'analyse. À condition d'acquérir plusieurs savoir-faire, cette étape peut se révéler stimulante dès lors que l'objectif est de répondre à une question de recherche.

I. – Les variables en question

Pour guider le novice, effrayé par l'ampleur de la tâche, on lui recommande de procéder en deux temps : construire d'abord un protocole de recherche et proposer ensuite les résultats de cette recherche. (1) Le protocole décrit l'objet d'étude ainsi que la problématique retenue pour l'exploitation des données. Cette dernière est plus resserrée que celle qui a été retenue pour la construction du questionnaire (voir chap. III). Il s'agit ensuite de construire une série d'hypothèses, c'est-à-dire une série d'affirmations qui seront à tester à partir des données de l'enquête. À chaque hypothèse, on associe la méthode qui sera appliquée pour tester sa validité. Par méthode, nous entendons les variables dont on veut se servir et la manière dont on souhaite les utiliser. (2) On passe ensuite à la production des résultats de recherche. Ils compilent l'ensemble des

traitements statistiques et des résultats associés à chaque hypothèse (validée ou non). Dès les premiers résultats, on est souvent amené à affiner ces dernières et à en tester de nouvelles qui ne semblaient pas initialement cruciales. Si la base de données porte sur un échantillon, il est nécessaire de réaliser, en plus, des tests statistiques, afin de s'assurer que les résultats sont généralisables.

Pour déterminer quel va être le traitement statistique le mieux adapté à l'hypothèse à tester, un premier élément simple à considérer est le type de variables que l'on souhaite utiliser. S'agit-il de variables nominales, numériques ou ordinales ? À chaque type correspond un traitement spécifique.

1. Les variables nominales, catégorielles ou qualitatives. – Les variables nominales, aussi appelées catégorielles ou qualitatives, permettent de nommer ou de dénommer une « qualité », c'est-à-dire d'associer une étiquette, une catégorie à un individu. Pour désigner les catégories d'une variable nominale, on parle de modalités de variable. Le sexe est une variable nominale à deux modalités : masculin et féminin. Les variables nominales, on va le voir, se prêtent à des tris à plat ou à des tris croisés, d'effectifs ou de pourcentages.

2. Les variables ordinales, discrètes ou d'échelle. Les variables ordinales ou d'échelle sont des variables nominales ordonnées. Il est donc possible de hiérarchiser leurs catégories. Les classes d'âge, découpées par intervalle de dix ans, sont un exemple de variable ordinale à sept modalités : « moins de 25 ans », « 25-34 ans », « 35-44 ans », « 45-54 ans », « 55-64 ans », « 65-74 ans » et « 75 ans et plus ». Les échelles de satisfaction en sont un autre. Mais cette hiérarchie ne s'impose pas toujours. Bien que le concept d'échelle sociale aille dans ce sens, il

est effectivement légitime de se demander si les catégories socioprofessionnelles forment une variable hiérarchique. Avec les variables d'échelle, les pourcentages cumulés constituent un autre outil intéressant.

3. Variables numériques ou quantitatives. – Les variables numériques ou quantitatives sont des variables dont la mesure a une signification réelle. Elles ont toujours une unité de mesure. Calculer le sexe moyen n'a aucun sens. En revanche, le calcul du revenu moyen des ménages (en euros) ou de la taille moyenne des logements (en mètres carrés) en a un. Outre la moyenne, avec ce type de variable, on peut établir des médianes et des écarts-types. L'âge, lorsqu'il est exprimé en continu (1 an, 2 ans, 3 ans, etc.) et non en tranche, est aussi une variable quantitative. Il est possible de créer des catégories ou des tranches à partir de ces nombres. On peut donc toujours transformer des variables numériques continues en variables d'échelle ou discrètes, mais la réciproque n'est pas vraie.

II. – Exploiter des variables numériques

Commençons par présenter les techniques statistiques adaptées aux variables numériques. Nous nous limiterons d'abord à des techniques relativement « simples », c'est-à-dire portant sur une seule variable (univariée) ou deux variables (bivariée). Si l'on souhaite décrire et analyser les groupes qui composent la société, objectif premier de la sociologie, ou les groupes de consommateurs, objectif premier des études de marché, il s'avère que les indicateurs de tendance centrale et de dispersion constituent de précieux outils.

1. Indicateurs de tendance centrale. – En statistique, la notion de tendance centrale désigne le centre autour duquel les différentes valeurs d'une variable « tendent »

à se regrouper. On parle aussi du centre de la distribution. Il en existe plusieurs mesures, la plus connue étant la moyenne. Les autres sont la médiane et le mode.

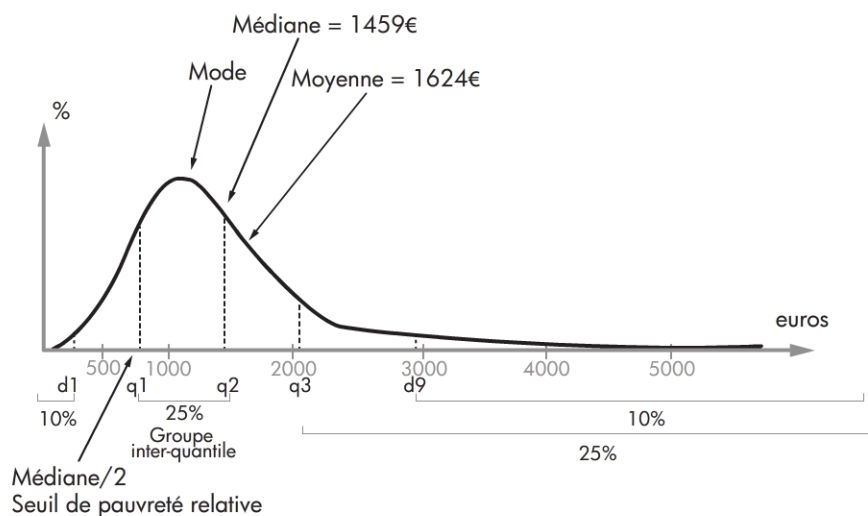
La *moyenne* est égale à la somme des données (des valeurs d'une variable numérique) divisée par leur nombre. D'après l'exemple du revenu salarial des individus (fig. 7), la moyenne est égale à la somme des revenus de chaque individu divisée par le nombre d'individus¹. On utilise aussi la moyenne pondérée en faisant la somme des différentes valeurs des revenus multipliés par leur poids. Le poids représente le nombre de fois où cette valeur apparaît dans la population totale. La moyenne présente un inconvénient. Elle est influencée par les valeurs extrêmes, c'est-à-dire rares mais très élevées, ou au contraire rares mais très faibles. La *médiane*, elle, n'est pas sensible aux valeurs extrêmes. Elle est plus représentative du *centre*. Par définition, elle partage la population étudiée en deux sous-populations de tailles égales. Autrement dit, c'est la valeur de la distribution d'après laquelle 50 % de l'échantillon se trouve au-dessus et 50 % en dessous.

En France, en 2010, le revenu salarial moyen mensuel des individus s'établissait à 1 624 € alors que la moitié des individus avait un salaire inférieur à 1 459 € par mois. Les très hauts revenus, visibles dans ce qu'on appelle la *queue de la distribution*, tirent bien la moyenne vers le haut. La moyenne, en tant qu'indicateur de tendance centrale, peut donc être trompeuse dans certains cas. Notamment lorsque la distribution est très dissymétrique, la moyenne diffère alors sensiblement de la médiane ; lorsque la distribution est très écrasée, de nombreux individus sont loin de la

1. Voir Insee, enquête « Revenus fiscaux 2004 », *Insee résultats*, n° 69, 2007.

moyenne. Un autre indicateur usuel de tendance centrale est le *mode*. Il correspond à la valeur prise par le plus grand nombre d'individus. Il s'utilise également sur des variables nominales. Dans ce dernier cas, le mode est alors la modalité présentant l'effectif le plus important. Le montant mensuel net du SMIC pour trente-cinq heures de travail par semaine s'établissait à 1 056 € en 2010, il est donc logiquement très proche du mode de la distribution du revenu salarial (fig. 7).

Fig. 7. – **Distribution du revenu salarial mensuel sur l'ensemble des salariés en 2010**



Champ : France métropolitaine, ensemble des salariés des secteurs public et privé hors salariés agricoles et apprentis-stagiaires.

Source : Données Insee, DADS 2010 définitif et SIASP, exploitation au 01/12.

2. Les indicateurs de dispersion. – Davantage que Durkheim, Halbwachs, son élève et neveu, a mis en évidence l'intérêt des indicateurs de dispersion. En effet, la recherche de régularité ne s'oppose pas à la recherche de l'hétérogénéité, et la moyenne doit de ce

fait être complétée par d'autres indicateurs pour refléter sincèrement la réalité sociale. Par opposition à la tendance centrale, la dispersion statistique reflète, elle, la variabilité des différentes valeurs prises par une variable numérique. Reprenons l'exemple du revenu salarial. L'hétérogénéité peut être étudiée à deux niveaux : à l'intérieur de la distribution, pour mettre en évidence des inégalités de positions entre le haut (les riches) et le bas (les pauvres) de l'échelle des revenus, mais aussi, par l'introduction d'une autre variable, le sexe ou le pays par exemple, pour comparer les revenus et mettre en évidence des disparités structurelles entre hommes et femmes sur le marché du travail, ou encore réaliser des comparaisons nationales.

Les *fractiles* permettent de définir les valeurs de telle sorte que la population se divise en un nombre de groupes de taille équivalent. La médiane est un fractile, elle divise la population en deux groupes de même taille. On parle de quartiles quand la population est divisée en 4, de quintiles lorsqu'elle est divisée en 5, de déciles, en 10 et enfin de centiles, lorsqu'on a 100 groupes de taille équivalente. Dans l'exemple, le premier décile de la distribution du revenu salarial, noté d1 (parfois DG1), s'élevait à 197 € nets par mois en 2010. Le premier décile correspond aux 10 % de la population aux salaires les plus faibles (compris entre le minimum d0 et d1). De l'autre côté de la distribution, les hauts revenus peuvent être analysés d'après leur appartenance au dernier groupe des déciles. Il s'agit des 10 % les plus riches. Ils se définissent alors par un revenu salarial supérieur à 2 883 €. Sur la courbe, la position respective des différents fractiles se prête à de nombreuses analyses. Si l'intervalle qui sépare les deux premiers quartiles (q2-q1) est faible, cela dresse le portrait d'une société en forme pyramidale, avec beaucoup de bas revenus et

peu de hauts. À l'inverse, si une large proportion de la population se concentre autour de la médiane (q_2) alors la société se définit par une large classe moyenne.

En effet, ces indicateurs permettent de créer des catégories. Il est fréquent de définir la classe moyenne non pas à l'aide d'un seuil de revenu figé, mais selon la position relative des revenus : entre la demi-médiane et deux fois la médiane. C'est, par exemple, sur ce type d'indicateur que l'on s'appuie pour dire que la société américaine est plus inégalitaire que la société suédoise. Cette dernière se définit par une très forte concentration de la population autour du niveau de vie médian, alors que la dispersion des revenus est nettement plus importante outre-Atlantique. La notion de dispersion peut également être abordée à l'aide du rapport interdéciles (d_9/d_1), qui met en évidence l'écart entre le haut et le bas de la distribution. Il constitue donc aussi une mesure d'inégalité de cette distribution. En France, ce rapport était de 15 en 2010 (fig. 7). Autrement dit, les 10 % les plus riches touchaient des salaires plus de quinze fois supérieurs aux 10 % les plus pauvres, ce qui est loin d'être négligeable.

L'*écart-type* et la *variance* sont, eux aussi, des indicateurs de dispersion. L'écart-type est la somme pondérée des écarts à la moyenne. La variance est seulement l'écart-type au carré. Ces deux indicateurs mesurent la dispersion ou l'étalement des valeurs de la distribution par rapport à la moyenne. Plus l'écart-type est grand, plus les valeurs autour de la moyenne sont dispersées. À coup sûr, l'écart-type de la distribution des revenus des Américains est plus grand que celui des Suédois.

3. Analyse bivariée. – Analyser isolément une variable numérique présente un intérêt, mais il est également enrichissant d'analyser conjointement deux variables

numériques ou de croiser une variable numérique avec une variable nominale. La notion de *corrélation* désigne la relation *linéaire* entre deux variables numériques. En principe, elle se borne à décrire des hausses ou des baisses produites par une variable sur une autre. Par exemple, les revenus des ménages et leurs dépenses en matière de logement sont corrélés positivement : les dépenses s'élèvent à mesure que s'élèvent les revenus. En revanche, plus les revenus sont importants, moins la *part* accordée au logement est élevée¹. Dans le dernier cas, la relation de corrélation est négative. Plusieurs indicateurs permettent de mesurer cette relation de corrélation. La covariance se définit comme le produit des écarts à la moyenne de deux variables numériques (une covariance élevée décrit un lien fort). Elle exprime la variation simultanée de deux variables. Le coefficient de corrélation linéaire, qui rapporte la covariance au produit des écarts-types des deux variables et qui mesure donc la covariation de X et Y compte tenue des variations propres de X et de Y, en est aussi un indicateur. Il est compris entre - 1 et 1 (égal à 1, il décrit une relation linéaire positive parfaite ; égal à - 1, une relation linéaire négative parfaite).

Il est également possible, et utile dans tous les champs des sciences humaines, d'analyser une variable numérique selon les modalités prises par une variable nominale. La *moyenne par classe* est couramment utilisée pour calculer par exemple la moyenne des revenus selon le sexe. À travers l'enquête réalisée en 2013 sur le logement et l'habitat étudiant, on révèle par exemple que les jeunes filles (en troisième année) vivant dans un logement indépendant ont deux ans de moins que les

1. F. Bugeja-Bloch, *Logement, la spirale des inégalités. Une nouvelle dimension de la structure sociale et générationnelle*, Paris, Puf, « Le Lien social », 2013.

jeunes hommes dans la même situation (tableau 4). Cette comparaison par catégories s'applique à tous les indicateurs de tendance centrale et de dispersion.

Tableau 4. – **Âge des étudiants selon leur lieu de résidence et leur sexe**

<i>Lieu de résidence en semaine</i>	<i>F</i>	<i>M</i>	<i>Ensemble</i>
Chez un tiers	23,7	25,4	24,3
Chez vos parents (ou tuteurs)	22,3	22,8	22,5
Dans un logement indépendant (seul, en couple ou en colocation)	24,0	25,8	24,5
Autre	26,1	24,0	25,6
Ensemble	23,1	24,0	23,4

Champ : Étudiants de licence 3 des universités Paris 8 et Paris 10 présents en 2012-2013.

Note : Les non-réponses sont exclues.

III. – Exploiter des variables nominales

Présentons désormais les techniques statistiques adaptées aux variables nominales ou catégorielles. Nous nous limiterons, là encore, à des techniques relativement « simples », c'est-à-dire portant sur une variable (univariée) ou deux variables (bivariées).

1. Tri à plat ou analyse univariée. – Pour étudier la distribution d'une variable nominale, on utilise le tri à plat. Il s'agit d'un tableau à une entrée donnant les effectifs ou les pourcentages pour chaque modalité de la variable. Parmi les 1 396 étudiants enquêtés en 2012, plus de la moitié (761, soit 54,5 %) vivait au domicile parental (tableau 5). Avec une variable ordinale, les pourcentages cumulés ont également de la pertinence. Par exemple, 21 % des étudiants n'avaient reçu

aucun ami au cours du mois précédant l'enquête et près de 70 % en avaient accueilli au moins à trois reprises (tableau 6). Ici, nous avons dû exclure les non-réponses, soit 58 individus (4,15 %), car elles ne s'ordonnent pas logiquement par rapport aux autres.

Tableau 5. – **Tri à plat de la variable du lieu de résidence hebdomadaire**¹

	<i>Effectif</i>	<i>Pourcentage</i>
Chez un tiers	78	5,6
Chez vos parents (ou tuteurs)	761	54,5
Dans un logement indépendant (seul, en couple ou en colocation)	519	37,2
Autre	11	0,8
Non-réponse	27	1,9
Ensemble	1 396	100

Tableau 6. – **Tri à plat de la variable de la réception d'amis**

	<i>Effectif</i>	<i>Pourcentage</i>	<i>Pourcentages cumulés</i>
0 fois	284	21,2	21,2
1 à 3 fois	645	48,2	69,4
4 à 7 fois	233	17,4	86,8
8 fois ou plus	176	13,2	100,0
Ensemble	1 338	100	

Note : les 58 « non-réponses » sont exclues.

Lors de l'exploitation de l'enquête, les étudiants de Paris 8 et de Paris 10 ont rapidement trouvé insuffisants

1. Après apurement de la base de données, c'est-à-dire après avoir attribué un mode d'habitat compte tenu des informations tirées des autres variables de la base.

les tris à plat pour répondre aux questions qu'ils se posaient. Apprentis sociologues, ils préféreraient croiser des variables pour étudier leur relation. Cependant, les tris à plat restent un outil indispensable. Avant de croiser des variables, étudier la distribution de chacune d'elles permet non seulement de s'assurer de leur cohérence, de traiter ou d'analyser les incohérences, de définir le sort des « non-réponses », « ne sait pas », etc., dont le sens n'est jamais anodin, et d'opérer des recodages, soit pour parer à la faiblesse des effectifs d'une catégorie, soit pour améliorer la lisibilité des résultats.

2. Tri croisé ou analyse bivariée. – Les tris croisés, tableaux croisés ou encore tableaux de contingence permettent d'analyser, cette fois, des relations entre variables. Ils ont permis aux étudiants d'assouvir leur curiosité. La plupart du temps, on cherche, à travers ces tableaux, à comprendre la relation entre deux phénomènes. Il s'agit souvent d'expliquer un phénomène précis par des déterminants sociaux, culturels ou économiques. On considère alors ces déterminants sociaux comme des variables *explicatives* (ou *indépendantes*) du phénomène étudié. Ils renvoient, le plus souvent, à l'âge, au sexe, à la catégorie socioprofessionnelle, au revenu ou encore au diplôme. À l'inverse, les variables *expliquées* (ou *dépendantes*), comme leur nom l'indique, sont les variables que l'on cherche à expliquer (que l'on croit dépendantes des précédentes). Il peut s'agir de la réussite scolaire, des choix résidentiels, des dépenses de loisirs, etc.

Avant de lire un tel tableau, quelques précisions de vocabulaire s'imposent. Un tableau de contingence est un tableau croisé entre deux variables nominales (ou ordinales) présentant des effectifs. Ces effectifs sont dits « observés » parce qu'il s'agit de ceux que l'on observe

dans l'échantillon ; le tableau 7 en est un exemple. Il interroge les liaisons possibles entre le mode d'habitat et l'établissement universitaire des étudiants. Les *marges* d'un tel tableau désignent les lignes et les colonnes « ensemble ». Ces distributions marginales sont importantes, elles nous renseignent sur plusieurs points. Dans l'exemple, on relève d'abord que les étudiants de Paris 10 sont plus nombreux que ceux de Paris 8 au sein de l'échantillon. On note ensuite qu'indépendamment de l'université d'accueil, plus de la moitié des étudiants habitent chez leurs parents et près du tiers vit dans un logement indépendant. Ainsi, les populations considérées sont de tailles différentes. Pour les comparer efficacement, il est nécessaire de les ramener à un dénominateur commun. Autrement dit, nous devons avoir recours aux pourcentages.

Tableau 7. – **Tri croisé entre le lieu de résidence et l'université de l'étudiant (effectif)**

	<i>Paris 10</i>	<i>Paris 8</i>	<i>Ensemble</i>
Chez un tiers	36	42	78
Chez vos parents (ou tuteurs)	510	251	761
Dans un logement indépendant (seul, en couple ou en colocation)	290	229	519
Ensemble	836	522	1358

Note : pour des raisons didactiques, les catégories « Autre » et « Non réponse » ont ici été supprimées. Elles représentaient à elles deux moins de 5 % de l'échantillon.

Plusieurs calculs de pourcentages sont possibles sur un tableau de contingence :

– **Les pourcentages du total** sont calculés à partir de l'effectif total (ici, 1 358). Ils donnent la proportion conjointe observée pour deux modalités. Dans notre échantillon, 37,5 % des répondants vivent

chez leurs parents (ou tuteurs) et étudient à Paris 10 : $(510/1358) \times 100 = 37,5$.

– **Les pourcentages en ligne** rapportent les effectifs de chacune des cellules à celui de la *marge en ligne* correspondante (tableau 8). Parmi les 761 étudiants habitant chez leurs parents (ou tuteurs), 67 % étudient à Paris 10 : $(510/761) \times 100 = 67$.

– **Les pourcentages en colonne** rapportent les effectifs de chacune des cellules à celui de la *marge en colonne* correspondante (tableau 9). Dans l'échantillon de répondants, 61 % des étudiants inscrits à Paris 10 vivent chez leurs parents ou tuteurs : $(510/836) \times 100 = 61$.

Tableau 8. – Répartition des étudiants des universités Paris 10 et Paris 8 selon le lieu de résidence (pourcentage en ligne du tableau 5)

	<i>Paris 10</i>	<i>Paris 8</i>	<i>Ensemble</i>
Chez un tiers	46	54	100
Chez vos parents (ou tuteurs)	67	33	100
Dans un logement indépendant (seul, en couple ou en colocation)	56	44	100
Ensemble	62	38	100

Tableau 9. – Répartition des lieux de résidence des étudiants en fonction de leur université (pourcentage en colonne du tableau 5)

	<i>Paris 10</i>	<i>Paris 8</i>	<i>Ensemble</i>
Chez un tiers	4	8	6
Chez vos parents (ou tuteurs)	61	48	56
Dans un logement indépendant (seul, en couple ou en colocation)	35	44	38
Ensemble	100	100	100

Les pourcentages en ligne et en colonne permettent de raisonner à effectifs égaux et de regarder comment se répartissent les différentes catégories. Dans le tableau 9, les pourcentages en colonne permettent de comparer les situations des étudiants de l'université Paris 8 à ceux de Paris 10. Dès lors, on observe que les étudiants de Paris 10 habitent davantage chez leurs parents (ou tuteurs) que les étudiants de Paris 8. Sur 100 étudiants inscrits en troisième année à l'université de Nanterre, environ 61 résident chez leurs parents, contre 48 seulement pour les étudiants de Saint-Denis. Ces derniers vivent donc plus souvent que les autres dans un logement indépendant. En effet, sur 100 étudiants inscrits à Paris 8, 44 résident dans ce type d'habitat contre 35 % des étudiants de Paris 10. En outre, il est vivement recommandé de comparer la situation de chacun de ces groupes avec la répartition moyenne, visible dans les marges du tableau. On identifiera alors des surreprésentations ou des sous-représentations par rapport à la situation moyenne, indépendante ici des établissements. L'habitat au domicile parental est surreprésenté à Paris 10 ($61 \% > 56 \%$, pourcentage marginal) et sous-représenté à Paris 8 ($48 \% < 56 \%$).

Le choix de recourir aux pourcentages en ligne plutôt qu'aux pourcentages en colonne ne répond pas à une logique mathématique. Il dépend de la question de recherche que l'on souhaite résoudre au travers du tableau. Face à ce choix, deux astuces mnémotechniques. (1) Les populations que l'on veut comparer sont celles que l'on doit ramener à un dénominateur commun, c'est-à-dire à une base de 100 individus. Pour comparer la situation des étudiants de Paris 8 à celle de ceux de Paris 10, il faut ramener ces deux populations à une taille équivalente (de 100 individus), on utilisera donc les pourcentages en colonne. (2) Les modalités de la variable dont on cherche

à mesurer les effets (variable explicative ou indépendante) doivent avoir un dénominateur commun. Dans l'exemple, les pourcentages en colonne permettent de mesurer les effets de l'établissement universitaire sur le mode de résidence (tableau 9) et les pourcentages en ligne, ceux du mode de résidence sur la faculté (tableau 8).

Cependant, une question demeure : peut-on se permettre de généraliser la portée des résultats observés dans notre échantillon à la population de référence de l'enquête ? Autrement dit, les effets que l'on observe ici sont-ils liés au hasard de la collecte ou sont-ils le reflet d'une tendance valable pour l'ensemble des étudiants présents dans les cours de troisième année des deux universités ?

3. Test du Chi-2. – Pour savoir si les différences constatées entre les deux établissements sont significatives, on a recours au test du Chi-2 (noté aussi Khi-2). Avec ce test, on complète l'analyse descriptive par l'analyse inférentielle. En d'autres termes, on teste la capacité de généralisation de nos résultats au-delà de notre échantillon. D'après l'exemple, on cherche à évaluer si les établissements ont un lien avec (ou une influence sur) les modalités de résidence des étudiants et s'il est possible de le généraliser. On se demande précisément si les répartitions des lieux de résidence à Paris 10 et Paris 8 diffèrent significativement de la répartition moyenne qui, elle, est indépendante de l'établissement universitaire. L'objectif est d'évaluer la distance qui sépare la situation réelle – celle réellement observée dans l'échantillon (tableau 10.1) – d'une situation théorique indépendante de l'université d'accueil – celle que l'on observerait si l'université n'avait pas d'effet sur les modalités de résidence.

Pour construire le tableau de ces *effectifs théoriques* (tableau 10.2), on applique aux effectifs totaux

observés à Paris 10 et Paris 8 la répartition moyenne, c'est-à-dire le pourcentage de la marge de chaque ligne (colonne « Ensemble » du tableau 9). Grâce aux deux tableaux, d'effectifs observés et d'effectifs théoriques, on peut mesurer les écarts entre la situation empirique et la situation théorique d'indépendance. Ces *écarts à l'indépendance* sont présentés dans le tableau 11. L'écart positif entre Paris 10 et l'habitat chez les parents décrit une surreprésentation (ou une attraction entre les deux modalités). Dans l'échantillon, il y a davantage d'étudiants de l'université Paris 10 vivant au domicile parental qu'il ne devrait y en avoir si leur répartition était identique dans les deux universités (si elle équivalait à la structure moyenne). L'écart négatif entre Paris 8 et le fait d'habiter chez les parents décrit à l'opposé une sous-représentation dans cet établissement.

La *distance Chi-2*, qui sépare la situation empirique de la situation théorique d'indépendance, est égale à la somme des contributions au Chi-2. Ces contributions, présentées dans le tableau se calculent case par case de la manière suivante :

$$\text{Contribution Chi2} = \frac{\text{Écart à l'indépendance}^2}{\text{Effectif théorique}} = \frac{(\text{Effectif observé} - \text{Effectif théorique})^2}{\text{Effectif théorique}}$$

Les cellules où la contribution est élevée désignent des modalités qui s'écartent nettement de la situation moyenne et contribuent donc fortement à accroître la distance Chi-2. C'est par exemple le cas des étudiants de Paris 8 habitant chez leurs parents. La distance Chi-2, qualifiée aussi de Chi-2 global, est ici égale à 24,48. Cette distance est-elle suffisamment grande pour conclure à un lien de dépendance entre mode d'habitat et établissement universitaire ? Dans quelle mesure peut-on généraliser ce lien ?

La table du Chi-2, accessible sur Internet, va nous permettre de répondre à cette question. Grâce à cette table, on va comparer le Chi-2 global à un Chi-2 seuil. Si la valeur du Chi-2 global est en deçà de la valeur du Chi-2 seuil, on accepte l'hypothèse d'indépendance entre les deux variables (ce qui n'est pas le résultat espéré) ; au-delà, on la rejette pour conclure alors qu'il y a bien un lien de dépendance entre les variables. Pour lire la valeur du Chi-2 seuil, il faut au préalable calculer le nombre de degré de liberté du tableau, noté ddl, il est égal à :

$$\text{ddl} = (\text{Nombre de modalités en ligne} - 1) \times (\text{Nombre de modalités en colonne} - 1)$$

Dans l'exemple, il vaut 2. On cherche alors dans la table du Chi-2 (tableau 13), la ligne associée à $\text{ddl} = 2$. En colonne, on choisit ensuite un risque d'erreur. On tolère au maximum une probabilité associée au Chi2 de 0,10, soit un risque de 10 % de se tromper lorsqu'on généralise les observations de l'échantillon à la population de référence dans son ensemble. Ici, pour un risque minimum de 0,1 % (soit une probabilité de 0,001) et un nombre de degré de liberté de 2, le Chi-2 lu dans la table est égal à 13,82. Le Chi-2 global, égal à 24,48, dépasse donc le Chi-2 seuil, ce qui nous permet d'affirmer (avec un risque de se tromper mince : de 0,1 %) que l'établissement universitaire et les modalités de résidence des étudiants sont liés¹. Une analyse plus poussée montre que ce lien s'explique avant tout par le lieu de résidence des parents des étudiants des deux facultés. Les parents des étudiants de Nanterre vivant en Île-de-France sont très surreprésentés. 64 % des pères des étudiants de Paris 10 sont Franciliens contre 44,6 % de ceux de Paris 8. À l'inverse, les parents des

1. Pour approfondir le test du Chi-2 ; voir notamment O. Martin, *L'Analyse de données quantitatives*, op. cit., p. 73-86.

étudiants de Saint-Denis vivent deux fois plus souvent à l'étranger (22 %) que ceux de Nanterre (11 %), ce qui les oblige à prendre un logement indépendant.

Tableau 10. – **Tableaux des effectifs observés et théoriques entre le lieu de résidence et l'université de l'étudiant**

<i>10.1 Effectifs observés</i>	<i>Paris 10</i>	<i>Paris 8</i>	<i>Ensemble</i>
Chez un tiers	36	42	78
Chez vos parents (ou tuteurs)	510	251	761
Dans un logement indépendant (seul, en couple ou en colocation)	290	229	519
Ensemble	836	522	1358
<i>10.2 Effectifs théoriques à l'indépendance</i>	<i>Paris 10</i>	<i>Paris 8</i>	<i>Ensemble</i>
Chez un tiers	48,0 (836 × 6 %)	30,0 (540 × 6 %)	78
Chez vos parents (ou tuteurs)	468,5 (836 × 56 %)	292,5 (540 × 56 %)	761
Dans un logement indépendant (seul, en couple ou en colocation)	319,5 (836 × 38 %)	199,5 (522 × 38 %)	519
Ensemble	856	522	1 358

Tableau 11. – **Écarts à l'indépendance du lieu de résidence selon l'établissement**

	<i>Paris 10</i>	<i>Paris 8</i>	<i>Ensemble</i>
Chez un tiers	- 12,0	12,0	0,0
Chez vos parents (ou tuteurs)	41,5	- 41,5	0,0
Dans un logement indépendant (seul, en couple ou en colocation)	- 29,5	29,5	0,0
Ensemble	0,0	0,0	0,0

**Tableau 12. – Contributions au Chi-2
du lieu de résidence selon l'établissement**

	<i>Paris 10</i>	<i>Paris 8</i>	<i>Ensemble</i>
Chez un tiers	3,0 (- 12,0) ² /48,0	4,8 12,0 ² /30,0	
Chez vos parents (ou tuteurs)	3,7	5,9	
Dans un logement indépendant (seul, en couple ou en colocation)	2,7	4,4	
Ensemble			

Tableau 13. – Table de distribution du Chi-2
Valeur limite de KHI2 en fonction des degrés de liberté
(en ligne) et du seuil de probabilité (en colonne)

	<i>Seuil de probabilité</i>				
ddl	<i>0,10</i>	<i>0,05</i>	<i>0,025</i>	<i>0,01</i>	<i>0,001</i>
1	2,71	3,84	5,02	6,63	10,83
2	4,61	5,99	7,38	9,21	13,82
3	6,25	7,81	9,35	11,34	16,27
4	7,78	9,49	11,14	13,28	18,47
5	6,24	11,07	12,83	15,09	20,52
6	10,64	12,59	14,45	16,81	22,46
7	12,02	14,07	16,01	18,48	24,32
8	13,36	15,51	17,53	20,09	26,12
9	14,68	16,92	19,02	21,67	27,88
10	15,99	18,31	20,48	23,21	29,59
11	17,28	19,68	21,92	24,72	31,26
12	18,55	21,03	23,34	26,22	32,91
13	19,81	22,36	24,74	27,69	34,53

14	21,06	23,68	26,12	29,14	36,12
15	22,31	25,00	27,49	30,58	37,70
16	23,54	26,30	28,85	32,00	39,25
17	24,77	27,59	30,19	33,41	40,79
18	25,99	28,87	31,53	34,81	42,31
19	27,20	30,14	32,85	36,19	43,82
20	28,41	31,41	34,17	37,57	45,31
21	29,62	32,67	35,48	38,93	46,80
22	30,81	33,92	36,78	40,29	48,27
23	32,01	35,17	38,08	41,64	49,73
24	33,20	36,42	39,36	42,98	51,18
25	34,38	37,65	40,65	44,31	52,62
26	35,56	38,89	41,92	45,64	64,05
27	36,74	40,11	43,19	46,96	55,48
28	37,92	41,34	44,46	48,28	56,89
29	39,09	42,56	45,72	49,59	58,30
30	40,26	43,77	46,98	50,89	59,70

IV. – Lire l'analyse factorielle et comprendre la régression

Jusqu'ici, les différentes techniques présentées se bornent à l'analyse de deux variables. Des méthodes plus complexes permettent pourtant des analyses multidimensionnelles. Elles sont l'objet de ce paragraphe. Nous présenterons d'abord l'analyse factorielle et ensuite les analyses de régression.

1. **Les analyses factorielles.** – Popularisées en France avec *La Distinction* de Bourdieu, les méthodes d'analyse factorielle sont particulièrement adaptées au

champ de la sociologie¹. Il en existe plusieurs types permettant aussi bien de traiter des variables numériques que des variables nominales. Dans cette méthode, toutes les variables ont le même rôle, sans distinction entre variables explicatives et expliquées. L'analyse factorielle permet de synthétiser dans un ou plusieurs espaces, à deux dimensions, un plan représenté par un graphique, un ensemble de variables ou de tableaux de contingence². De ce point de vue, cette méthode simplifie la réalité afin de la rendre immédiatement tangible par un schéma. La méthode donne de nouvelles variables, appelées facteurs ou axes, qui combinent et synthétisent l'information d'un grand nombre de variables. Une fois ces facteurs construits, il s'agit pour le sociologue de leur donner du sens en analysant les contributions des différentes dimensions mobilisées.

Prenons un exemple. Dans la lignée du célèbre espace des styles de vie de *La Distinction* de Bourdieu, un ouvrage collectif récent s'est donné pour objectif d'éprouver sa transposition à d'autres sociétés trente ans après³. Revisitant *La Distinction* pour la société britannique du XXI^e siècle⁴, des sociologues anglais ont réalisé une analyse des correspondances multiples (ACM), à partir de quarante et une variables, sur les pratiques et les goûts culturels, et proposé plusieurs espaces. Ils ont retenu des espaces construits à partir du croisement des

1. P. Bourdieu, *La Distinction. Critique sociale du jugement*, Paris, Minuit, 1979.

2. Voir M. Selz, F. Maillochon, *Le Raisonnement statistique en sociologie*, op. cit., p. 277, et J. Duval, « Analyser un espace social », in S. Paugam (dir.), *L'Enquête sociologique*, op. cit., p. 268-290.

3. P. Coulangeon, J. Duval, *30 ans après La Distinction de Pierre Bourdieu*, Paris, La Découverte, « Recherches », 2013.

4. T. Bennett, M. Gayo-Cal, B. Le Roux, M. Savage, E. Silva, A. Warde, D. Wright, « *La Distinction* revisitée : l'espace des styles de vie britanniques en 2003 », in P. Coulangeon, J. Duval, *30 ans après La Distinction de Pierre Bourdieu*, Paris, La Découverte, 2013, p. 179-205.

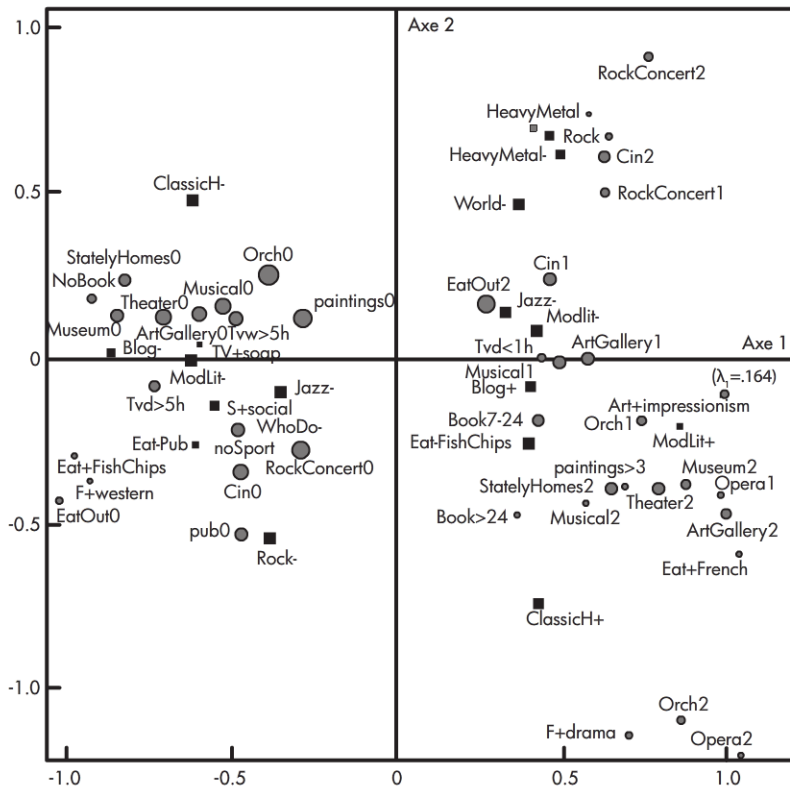
axes (ou facteurs) 1, 2, 3 à 4 puisque ces quatre nouvelles variables permettent, à elles seules, la restitution de plus de 80 % de l'ensemble des informations contenues dans les quarante et une variables initiales. Pour les besoins pédagogiques, on se focalisera ici sur le plan croisant les facteurs 1 et 2. Seules les variables qui contribuent le plus à expliquer l'axe 1 sont présentées dans la figure 8.

On constate que la majorité des modalités portant sur une pratique intense et un goût avéré se situent dans la partie droite de l'axe 1. S'y situent aussi bien ceux qui vont à un concert de rock au moins une fois par mois et ont un goût certain pour ce style musical que ceux qui se rendent à l'opéra ou dans des galeries d'art plusieurs fois par an et aiment la littérature moderne et l'impressionnisme. Les catégories décrivant une absence de pratique ou un dégoût se placent, elles, dans la partie gauche. On y trouve les individus qui ne vont ni au restaurant, ni au cinéma, ni au théâtre, et qui n'aiment ni le jazz, ni le rock, ni les séries noires, etc. Seuls font exception le temps passé devant la télévision et le goût pour les westerns. La proximité entre les modalités « Parfois à des concerts de rock » et « Souvent au cinéma » signifie qu'un grand nombre d'individus adoptent simultanément ces deux pratiques ou qu'ils ont une autre pratique commune fréquente. De ce constat, les auteurs déduisent que l'axe 1 présente le « degré d'engagement dans les pratiques culturelles ». En regardant ensuite la position des différentes catégories sociales sur cet axe (fig. 9), il apparaît qu'elles s'ordonnent du bas de la hiérarchie sociale, à l'extrémité gauche de l'axe, au sommet, avec les *higher professionals*, à l'extrémité droite. Ainsi, à l'opposition entre pratiques légitimes des catégories supérieures et pratiques illégitimes des classes populaires de *La Distinction* se substitue, dans cet exemple, une opposition entre (quasi-)absence de pratiques des

ouvriers et des employés et pratiques intenses des professions à responsabilités. On retrouve là l'antagonisme entre catégories favorisées aux goûts et pratiques variés (dites « omnivores ») et catégories défavorisées aux goûts et pratiques uniques (« univores »).

Fig. 8. – « L'axe du degré d'engagement dans les pratiques culturelles »

Interprétation de l'axe 1 ($\lambda_1=.164$) : les 59 modalités contribuant le plus à l'axe 1 (84 % de la variance), 35 modalités de participation (disques gris et italiques) et 24 de goût (carrés noirs).

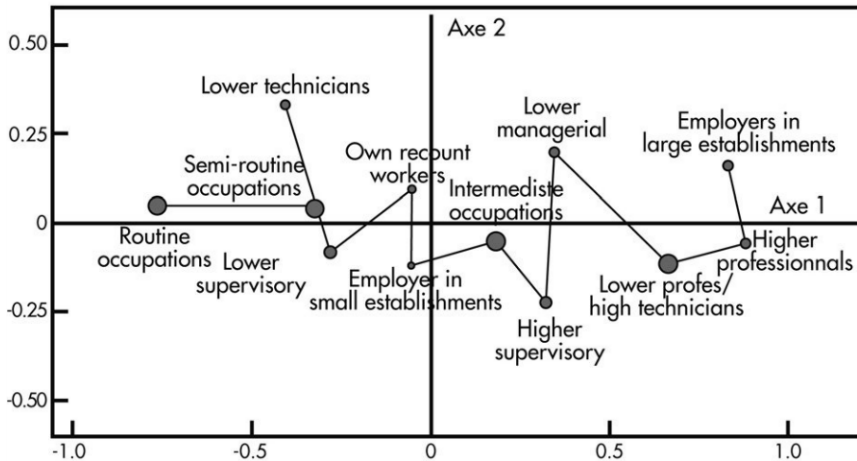


Source : Extrait de T. Bennett *et alii*, « La Distinction revisitée : l'espace des styles de vie britanniques en 2003 », in P. Coulangeon, J. Duval, *30 ans après La Distinction de Pierre Bourdieu, op. cit.*, p. 183.

Note de lecture : Pour les pratiques, 0 signifie « aucune » pratique, 1 « parfois » et 2 « souvent ». Pour les goûts, le signe + signifie « aimer », le signe – « ne pas aimer ».

Fig. 9. – La catégorie sociale
dans le plan factoriel 1-2

Nuage de modalités : les 12 modalités « catégories socioprofessionnelles » dans le plan 1-2 (variable supplémentaire).



Source : Extrait de Bennett *et alii*, « *La Distinction* revisitée : l'espace des styles de vie britanniques en 2003 », in P. Coulangeon, J. Duval, *30 ans après La Distinction de Pierre Bourdieu, op. cit.*, p 188.

Dans cet exemple, on a seulement présenté l'analyse des correspondances multiples (ACM). Elle constitue une forme d'analyse factorielle, spécifique à l'analyse d'un grand nombre de variables nominales. Selon les variables en jeu, les analyses varient, mais le principe demeure. Ajoutons que les méthodes d'analyse factorielle sont aujourd'hui particulièrement utilisées en marketing afin d'identifier les principales dimensions qui font l'image d'une marque ou pour établir des profils de consommateurs.

2. Les régressions logistiques. – Si l'analyse factorielle sert à décrire la manière selon laquelle un grand nombre de variables s'articulent, les régressions logistiques visent, au contraire, à démêler leurs effets respectifs.

L'enjeu est d'expliquer une variable (appelée variable « expliquée », « d'intérêt », « dépendante » ou « endogène ») par une ou plusieurs autres dites explicatives (« indépendantes », « exogènes » ou « covariables »). La régression indique des liaisons statistiques dont seule l'interprétation sociologique permet d'identifier ou non des relations de cause à effet. Ces méthodes sont également appelées méthodes « *toutes choses égales par ailleurs* » dans la mesure où elles servent à mesurer l'effet propre d'un facteur sur un phénomène indépendamment d'autres variables maintenues constantes. C'est donc une simulation de l'expérimentation « consistant à comparer le destin de deux populations identiques à un facteur près, celui dont on cherche à mesurer l'effet¹ ».

Cette méthode est très utile lorsque l'objectif est explicatif et que l'on souhaite distinguer l'effet propre de certains facteurs des effets de structure de population. C'est d'ailleurs dans cette perspective que Sybille Gollac l'utilise². La sociologue cherche à déterminer la place qu'occupe la fonction publique dans les trajectoires des individus issus de classes populaires. Dans le titre de l'article, elle formule une question : la fonction publique est-elle une voie de promotion sociale pour les enfants des classes populaires ? Afin d'analyser l'attrait du secteur public pour les enfants d'ouvriers, d'employés et d'agriculteurs, l'auteur souhaite comparer leur devenir à celui des salariés d'entreprises privées. Or, un problème se pose : les profils socioprofessionnels au sein de la fonction publique sont structurellement différents de ceux des entreprises. L'importance chez les

1. M. Selz, F. Maillachon, *Le Raisonnement statistique en sociologie*, *op. cit.*, p. 285.

2. S. Gollac « La fonction publique : une voie de promotion sociale pour les enfants des classes populaires ? », *Sociétés contemporaines*, vol. 2, n° 58, 2005, p. 41-64.

fonctionnaires du groupe des cadres et des professions intermédiaires, la faiblesse des ouvriers par rapport aux employés expliquent par exemple qu'ils soient nettement plus diplômés que les salariés du secteur privé. Les deux espaces (public et privé) et groupes (classes populaires et supérieures) sont donc difficilement comparables, à moins précisément de contrôler ces effets de structure de population et, notamment, la faiblesse du capital scolaire dans les milieux populaires. Ainsi, dans le tableau croisé présentant la part d'agents de l'État et des collectivités locales selon le groupe social du père, on note que seuls 20,7 % des enfants d'ouvriers sont fonctionnaires contre 27,6 % des enfants de cadres¹. Ce résultat *brut* est-il lié au peu d'attraits que les enfants d'ouvriers conçoivent pour le secteur public ou à leur moindre certification scolaire ? Finalement, ce résultat *brut* résiste-t-il lorsqu'on contrôle l'effet du niveau de diplôme des individus sur l'accès à la fonction publique ? Formulé différemment encore, y a-t-il un effet *propre* ou *net* de la catégorie sociale du père indépendamment du niveau de diplôme ?

Pour répondre à cette question, la régression logistique est heuristique. Elle se lit toujours par rapport à une situation de référence. Ici, l'auteur veut modéliser la probabilité d'être agent de l'État ou des collectivités locales ; elle a choisi pour ce faire la situation de référence suivante : un homme âgé de cinquante à cinquante-neuf ans sans diplôme dont le père était cadre salarié du secteur privé (non de l'État ou des collectivités ; voir tableau 14)². La régression va permettre d'éva-

1. *Ibid.*, p. 47.

2. Le choix de la situation de référence n'a aucune répercussion sur les résultats tirés de la régression, ces références dictant seulement le sens de la lecture.

luer l'effet propre que produit un facteur (par exemple ici, le sexe) sur la probabilité d'être fonctionnaire. Le raisonnement *toutes choses égales par ailleurs* signifie que les autres caractéristiques sont contrôlées. Ainsi, une femme âgée de 50 à 59 ans sans diplôme dont le père était cadre salarié du secteur privé, a plus de chances d'être fonctionnaire qu'un homme présentant les mêmes caractéristiques. Le signe du *coefficient*, positif, l'atteste. Pour évaluer dans quelle proportion ces chances sont plus élevées, il est d'usage d'utiliser l'*odds-ratio* ou rapport de chance, qui est égal à l'exponentielle du coefficient. Ici, à âge, diplôme, groupe social et statut du père identiques, les femmes ont 1,9 chance, soit presque deux fois plus de chance d'être fonctionnaires que les hommes.

Pour répondre à la question posée plus haut, lorsque le diplôme est contrôlé, on constate, désormais, que la probabilité d'être fonctionnaire des enfants d'ouvriers est supérieure à celle des enfants de cadres. *Toutes choses égales par ailleurs*, ils ont précisément 1,38 fois plus de chance d'être fonctionnaires que les cadres. « Une régression logistique sur la probabilité de devenir fonctionnaire montre [donc] qu'à diplôme égal le fait d'avoir un père agriculteur, employé ou ouvrier est plus favorable à l'appartenance à la fonction publique que le fait d'avoir un père cadre ¹ ». Ainsi, l'effet *net* dont rend compte ici la régression logistique ne va pas dans le même sens que l'effet *brut* décrit précédemment dans le tableau croisé. La dernière colonne du tableau nous indique, quant à elle, la significativité de ces résultats. On considère les résultats représentatifs jusqu'au seuil de 0,1 (10 %). Au-delà, ils sont non significatifs.

1. S. Gollac « La fonction publique : une voie de promotion sociale pour les enfants des classes populaires ? », art. cité, p. 48.

L'analyse de régressions logistique, ici heuristique, ne doit cependant pas faire oublier deux aspects : d'abord, que l'expression « *toutes choses égales par ailleurs* » est en partie abusive puisque seules sont contrôlées les variables introduites dans le modèle ; ensuite, que si elles démêlent les effets des différents facteurs, dans la réalité sociale ils demeurent fortement liés.

Tableau 14. – **Régression logistique sur la probabilité de devenir agent de l'État ou des collectivités locales**

Variables	Coefficient	Odds ratio	Seuil de représentativité
Sexe			
<i>Homme</i>	Référence	Référence	
Femme	<u>0,64</u>	1,90	< 0,0001
Âge			
15-24 ans	<u>-0,97</u>	0,38	< 0,0001
25-39 ans	<u>-0,54</u>	0,58	< 0,0001
40-49 ans	<u>-0,10</u>	0,90	< 0,0001
<i>50-59 ans</i>	Référence	Référence	
60 ans et plus	<u>-0,48</u>	0,62	< 0,0001
Diplôme			
Supérieur	<u>1,34</u>	3,82	< 0,0001
Baccalauréat +2	<u>0,89</u>	2,44	< 0,0001
Baccalauréat ou équivalent	<u>0,58</u>	1,79	< 0,0001
Diplôme inférieur au baccalauréat	<u>0,29</u>	1,34	< 0,0001
<i>Sans diplôme</i>	Référence	Référence	

Groupe social du père			
Agriculteurs	<u>0,26</u>	1,30	0,0002
Artisans, commerçants, chefs d'entreprise	<u>0,14</u>	1,15	0,043
Cadres	Référence	Référence	
Professions intermédiaires	<u>0,32</u>	1,38	< 0,0001
Employés	<u>0,35</u>	1,42	< 0,0001
Ouvriers	<u>0,32</u>	1,38	< 0,0001
Statut du père			
À son compte	- 0,08	0,92	0,2181
Salarié de l'État ou des collectivités locales	<u>0,54</u>	1,72	< 0,0001
Autre salarié	Référence	Référence	

Champs : Ensemble des actifs occupés de nationalité française regroupés dans la base de données « Individus » de l'enquête « Emploi 2001 » (sauf « fonctionnaires » déclarés agriculteurs, artisans, commerçants ou chefs d'entreprise).

Source : Tiré de l'Annexe p. 64 de S. Gollac « La fonction publique : une voie de promotion sociale pour les enfants des classes populaires ? », art. cité, p. 41-64.

Note : Les *odds-ratio* ont été calculés par nos soins.

Chapitre VI

VALORISATION ET RÉCEPTION DES RÉSULTATS

Les résultats statistiques étant, à ce stade de la démarche, produits, il s'agit désormais de les valoriser sous la forme d'un rapport ou, la plupart du temps, d'un article. Ce temps de la restitution, souvent négligé dans les manuels, occupe pourtant une place tout aussi centrale que la construction de l'objet ou la collecte des données. Selon le format de publication, le public visé, les attentes de l'institut commanditant l'étude ou plus généralement de l'institution (colloque, revue, etc.) qui va exercer la critique, quelques règles ou certains usages sont de rigueur. Les professionnels écrivent toujours « *en vue d'un public*, [...] ; [et] selon le public en vue duquel ils écrivent, ils ne s'obligent pas aux mêmes censures, ni aux mêmes formalismes ¹ ». Nous restituerons ici des conseils généraux pour présenter des données chiffrées, rédiger et interpréter.

I. – L'art et la manière de présenter des données chiffrées

Les conseils généraux concernant la valorisation de l'enquête et son exploitation ont largement été appliqués

1. C. Lemieux, « L'écriture sociologique », in S. Paugam, *L'Enquête sociologique*, *op. cit.*, p. 381.

par les étudiants de Paris 10 et de Paris 8. Par binôme, l'objectif fixé était de rédiger un document de quatre pages restituant leurs principaux résultats chiffrés et leurs analyses¹. Ce format, relativement court, implique un important travail de sélection des statistiques produites (tableau croisé sur variables qualitatives ou tableau de moyenne sur variables quantitatives). Mais quelle que soit la taille du document attendu, le chercheur procède toujours à un tri.

1. Une affaire de sélection. – Ainsi, valoriser l'exploitation d'une enquête est avant tout une affaire de sélection des résultats à mettre en valeur parmi tous ceux qui ont été produits en amont. Le quantitatifiste sélectionne les chiffres essentiels, c'est-à-dire les chiffres inédits ou importants pour le déroulement de sa démonstration ou pour le lecteur. En conséquence, il ne faut pas hésiter à faire une sélection assez drastique et qui a du sens : autant un résultat peut marquer ou éclairer le lecteur, autant une juxtaposition de résultats, apposés sans suivre le fil d'une démonstration, le fatiguera à coup sûr. On retient un résultat s'il raconte clairement une histoire, décrit un phénomène (ou, d'ailleurs, l'absence de phénomène), s'il constitue un point central de la démonstration, ou un point complémentaire que l'on souhaite toutefois préciser. Dans ce dernier cas, il est alors possible, selon le format attendu, de le placer en annexe ou en encadré. Au cours de cet exercice de sélection, il existe deux manières de restituer ses résultats, soit à l'aide de tableaux ou de graphiques, soit dans le corps du texte. Offrons, dans un premier temps,

1. Ce format de « quatre pages » est relativement classique dans la statistique publique. Pour en trouver des exemples, il suffit de consulter les revues *Insee Première* ou encore *Population et sociétés* de l'Ined.

quelques pistes pour la présentation des données sous forme de tableaux ou de graphiques.

2. Tableaux et graphiques. – « Tableaux et graphiques sont des illustrations, des compléments ou des “zooms” sur un point précis. Dans tous les cas, une figure est autonome : on doit pouvoir la comprendre sans avoir recours au texte de l’article. » Telles sont les recommandations formulées pour les publications de la statistique publique¹. Or, la tâche n’est pas aussi facile qu’il y paraît. Trouver la juste présentation, en particulier pour les graphiques, pose souvent des difficultés auxquelles on ne s’attend pas.

Reprenons l’exemple du chapitre précédent, sur le mode d’habitat des étudiants dans les universités parisiennes (Nanterre et Saint-Denis) et expliquons le cheminement qui nous conduit du tableau croisé d’effectifs (tableau 7) à un graphique fini. Le tableau est brut, tiré des résultats de recherche ; il n’a subi aucun recodage à l’exception des « non-réponses ». La première question porte toujours sur le traitement de ces « non-réponses » et ensuite sur la forme (ici, pourcentages en ligne ou en colonne) la plus adaptée à l’objectif du tableau. Dans l’optique de comparer les étudiants de Nanterre à ceux de Saint-Denis, on retient alors le tableau suivant (tableau 15) puisqu’il place les deux universités sous un même dénominateur. L’habitat au domicile parental apparaît être un trait caractéristique des étudiants de Paris 10. Ceux de Paris 8 se tournent, eux, presque autant vers un mode de vie dans un logement indépendant que chez leurs parents, et d’après

1. C. Benveniste, « Savoir compter, savoir conter. Épisode n° 3 – Des graphiques et des tableaux », *Courrier des statistiques*, n° 115, septembre 2005, p. 49.

le test du Chi-2, ces différences interuniversités sont significatives (voir chap. V).

Tableau 15. – Répartition des lieux de résidence des étudiants en fonction de leur université (pourcentage en colonne)

	<i>Paris 10</i>	<i>Paris 8</i>	<i>Ensemble</i>
Chez un tiers	4	8	6
Chez vos parents (ou tuteurs)	61	48	56
Dans un logement indépendant (seul, en couple ou en colocation)	35	44	38
Ensemble	100	100	100

Source : Enquête 2012-2013 sur le logement et l'habitat des étudiants de Paris 8 et Paris 10.

Champ : Les étudiants présents aux cours de L3 à la fin de 2012 de Paris 8 et Paris 10.

Note : Les non-réponses sont exclues.

Les chiffres étant désormais sélectionnés, la question de la forme de la représentation se pose alors : tableau ou graphique ? « Mieux vaut privilégier les graphiques », lit-on dans les recommandations du courrier des statistiques. Si les chiffres donnés dans les tableaux offrent une plus grande précision, la représentation graphique permet, elle, effectivement, grâce à une vision plus globale, de saisir instantanément les mouvements, les écarts, leur forme et leur ampleur. L'objectif premier d'un graphique est de présenter un résultat fort en un seul coup d'œil. Il ne faut pas pour autant bannir les tableaux, notamment s'ils sont courts, mais retenir qu'un tableau *se lit* alors qu'un graphique *se regarde*. Lorsque l'on souhaite mettre en avant un chiffre-clé, rien n'empêche d'insérer un chiffre dans un graphique ou d'utiliser des artifices de mise en page pour le souligner.

Lors de la réalisation d'un graphique, du choix de sa forme, il faut être honnête et transparent, et non tromper le lecteur. Pour ce faire, certaines règles doivent être respectées. Tout d'abord, quel type de graphique choisir pour quel type de série de données ? Dès que la série offre une certaine continuité : des tranches d'âges, des cohortes de naissance ou des périodes, le recours à un graphique sous forme de courbe se justifie, sinon il est plus cohérent de se tourner vers des représentations en histogramme (barres verticales) ou en barre (barres horizontales). Si les graphiques sous forme de secteurs (plus vulgairement appelés « camemberts ») ou d'aires peuvent être esthétiques, ils sont efficaces seulement dans certains cas de figure. Les aires sont intéressantes pour présenter des séquences, des successions d'étapes. Les secteurs ont un intérêt pour mettre en évidence la composition d'un unique phénomène, la structure du budget moyen des ménages par exemple. Mais attention, « la lecture de ce type de graphique présente un risque d'interprétation biaisée car l'œil distingue moins bien les différences entre secteurs (d'un camembert) qu'entre hauteurs (d'un histogramme) ¹ ».

En outre, un graphique, comme toute figure, ne peut se passer de certains éléments si l'on veut en faciliter la lecture tout en évitant des interprétations erronées. Un titre, le plus souvent énonciatif, afin de poser clairement ce dont il s'agit, est indispensable. Dans l'exemple, les titres : « Mode d'habitat selon l'établissement universitaire », énonciatif, ou « Diversité des modes d'habitat à Paris 8 et Paris 10 en 2012 », informatif, sont tous deux envisageables. La source (d'où proviennent les données ?) et le champ (sur quelle population

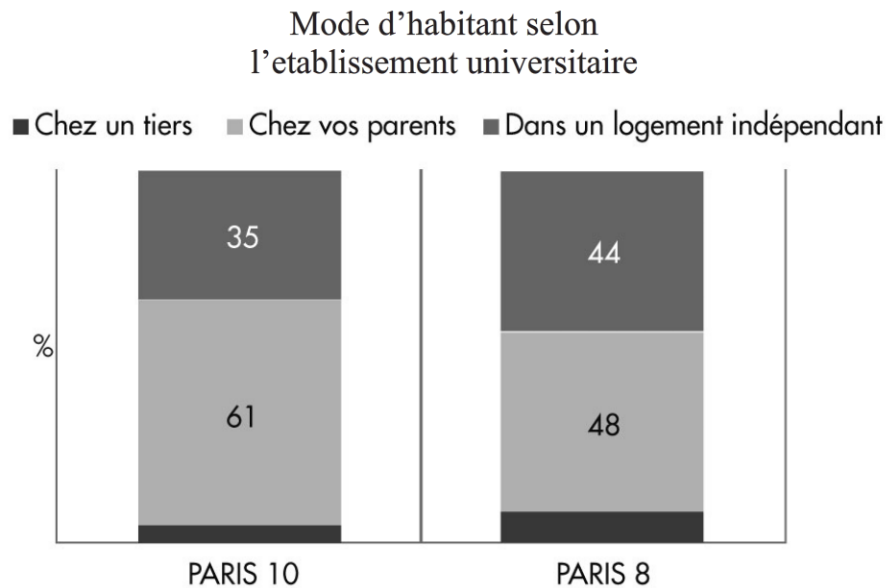
1. *Ibid.*, p. 52.

portent-elles ?) doivent être explicités. Ces informations apparaissent le plus souvent sous la figure, mais elles peuvent aussi être données une fois pour toutes dans le texte. L'unité doit être visible dans la figure, accompagnée communément d'une légende. Une note, placée sous la figure, peut aider la lecture en prenant l'exemple d'un point. La taille de la police et le choix des couleurs dépendent des contraintes liées à l'impression. Dès que l'on multiplie les graphiques dans le but de les comparer, il faut alors choisir une échelle et une origine communes pour ne pas fausser la lecture. Lorsqu'un graphique présente un grand nombre de modalités, les ordonner simplifie souvent le travail du lecteur.

Notre exemple porte sur des variables qualitatives, sans aucune logique de continuité : on opte donc pour une présentation en histogramme. Elle a l'avantage de ne pas hiérarchiser les données. Notre objectif est ici de comparer les deux universités et de pointer les différences entre les deux principaux modes d'habitat : soit chez les parents, soit en logement indépendant. Plusieurs options sont possibles. Discutons-en trois. La première, présentant un histogramme empilé à 100 %, s'accorde parfaitement avec le tableau. Il compare la distribution entre les deux universités et l'interprétation des chiffres ne saurait être ambiguë. Affichant clairement la base du pourcentage (chaque barre monte jusqu'à 100), on lit immédiatement qu'à Paris 10, 61 % des étudiants de L3 vivent chez leurs parents. On a choisi ici de rendre visibles les étiquettes, c'est-à-dire la valeur sur le graphique de chaque section (sauf pour l'habitat chez un tiers, moins central). Avec ce choix de graphique (option 1), c'est la proximité des modes d'habitat entre les deux facultés qui apparaît à première vue. Pour les options 2 et 3, on a retenu un histogramme regroupé : regroupé par université pour la

deuxième et par modalité d’habitat pour la troisième. Ces deux représentations sont équivalentes et mettent en relief, plus que notre premier choix, les divergences. L’option 2 souligne surtout les différences de composition entre les deux universités (le fait d’habiter chez ses parents est plus répandu que celui d’occuper un logement indépendant à Paris 10 ; la proximité entre ces deux modes de vie à Paris 8). L’option 3 marque davantage les différences pour chaque mode d’habitation (l’habitat au domicile parental est plus développé à Paris 10 qu’à Paris 8). La base 100 du pourcentage, l’université, se saisit plus immédiatement dans l’option 2 que dans la 3. Toutes ces options sont envisageables. Elles s’appuient sur les mêmes chiffres et répondent aux exigences d’un graphique, le choix dépendant du message que l’auteur souhaite délivrer et de la direction dans laquelle il veut amener son lecteur.

– **Option 1 :**



Source : Enquête 2012-2013 sur le logement et l’habitat des étudiants de Paris 8 et de Paris 10.

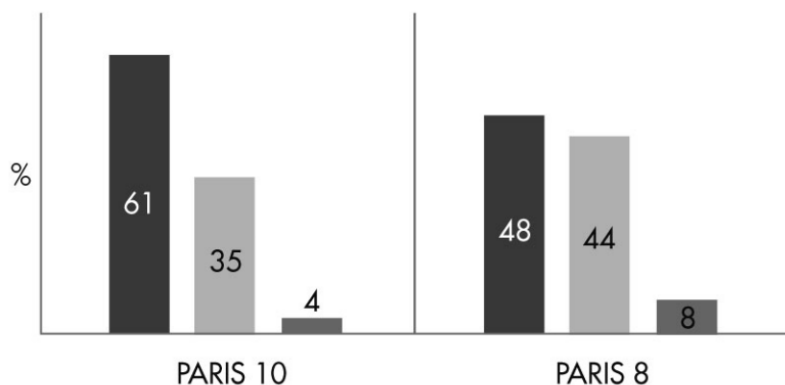
Champ : Les étudiants présents aux cours de L3 à la fin de 2012 de Paris 8 et de Paris 10.

Note de lecture : 61 % des étudiants de Paris 10 vivaient chez leurs parents.

– **Option 2 :**

Mode d'habitat selon
l'établissement universitaire

■ Chez vos parents ■ Dans un logement indépendant ■ Chez un tiers

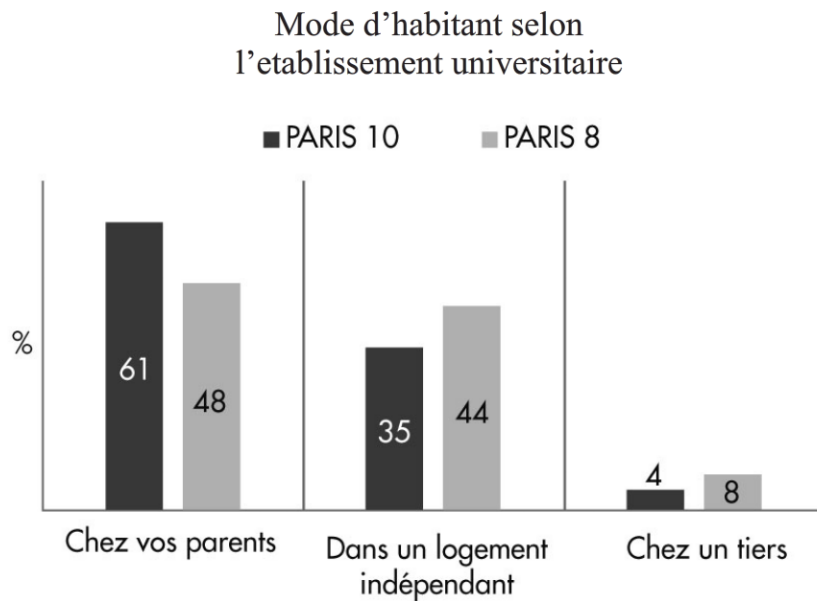


Source : Enquête 2012-2013 sur le logement et l'habitat des étudiants de Paris 8 et de Paris 10.

Champ : Les étudiants présents aux cours de L3 à la fin de 2012 de Paris 8 et de Paris 10.

Note de lecture : 61 % des étudiants de Paris 10 vivaient chez leurs parents.

– **Option 3 :**



Source : Enquête 2012-2013 sur le logement et l'habitat des étudiants de Paris 8 et de Paris 10.

Champ : Les étudiants présents aux cours de L3 à la fin 2012 de Paris 8 et de Paris 10.

Note de lecture : 61 % des étudiants de Paris 10 vivaient chez leurs parents.

II. – L'art et la manière de mobiliser des données chiffrées

Avec ces trois histogrammes, on voit bien que la mise en forme des résultats influence le message transmis. En outre, avec Howard Becker, rappelons que la phase d'écriture, loin d'être la simple transposition de la pensée, occupe au contraire un rôle actif, car réflexif, dans l'analyse. En effet, la mise en forme des idées les modifie en retour. Les ajustements que connaissent les pensées dépendent notamment du registre dans lequel l'auteur s'inscrit. Dès lors, il convient de l'explicitier.

1. Trois registres d'écriture. – On a montré l'importance de la phase de construction de l'objet et d'énoncé du problème. À cette occasion, on a cherché à mettre en évidence la façon dont l'objet, les matériaux empiriques et les statistiques s'articulent entre eux, en soulignant qu'ils ont en commun d'être des constructions sociales. Indispensable à la mise en ordre des preuves empiriques, l'écriture se doit de préciser clairement les processus de construction de l'objet ainsi que la ou les méthodes d'enquête. Ce faisant, l'auteur alterne entre différents registres : celui des matériaux empiriques, celui des concepts théoriques et enfin celui de leurs critiques.

Cyril Lemieux souligne avec justesse à quel point il est important de réaliser des allers-retours incessants entre les trois registres et de les mettre en évidence lors de la rédaction. Les données empiriques viennent ainsi appuyer le raisonnement et permettent tout au long du texte d'administrer la preuve. Il ne s'agit pas de simples éléments décoratifs. Bien au contraire, la démonstration de l'auteur prend appui sur ces dernières.

En mobilisant les différents registres, tout en prenant soin de les distinguer lors de la rédaction, on permet au lecteur, en toute honnêteté, de développer un regard critique à l'égard des résultats qui lui sont proposés. Le lecteur doit être en mesure d'identifier les avantages, mais aussi les inconvénients du protocole de recherche. De cette manière, on garantit la « vérificabilité » du discours. Il faut donc mettre à disposition simultanément ces données empiriques, les conditions dans lesquelles elles ont été produites et sa position théorique dans le champ. En outre, les allées et venues permettent à l'auteur de prendre position au sein de la littérature existante et d'anticiper les éventuelles remarques. L'objectif est simplement de viser l'honnêteté intellectuelle. Ainsi, lors de la restitution des résultats de notre enquête sur le

logement et l'habitat, un binôme dont le travail portait sur la sociabilité des étudiants a explicité sa démarche de la façon suivante : « L'enquête étant surtout orientée sur le logement et l'habitat des étudiants, les données recouvrant la sociabilité se limitent à la fréquence de réception des amis et de la famille. Ainsi, la sociabilité dont il sera question est la sociabilité chez soi. Sachant que les étudiants ont un mode de sociabilité basé en majorité sur des sorties et non sur la réception chez soi, il ne s'agit pas de dire que ceux qui ne reçoivent pas n'ont pas de sociabilité [...] mais plutôt de voir quelles caractéristiques objectives (équipement du logement proximité avec le lieu d'étude, nombre de pièces...) et subjectives (indice de confort) influencent la sociabilité chez soi¹. »

2. L'écriture en pratique. – Le choix du plan est souvent déterminant. Il doit être logique, et tout l'enjeu de l'écriture est de mettre en lumière cette logique. Bien que l'on cherche à tenir en haleine le lecteur, tenter de créer un effet d'attente est vain ; mieux vaut accompagner le lecteur et donner, tôt, les principaux résultats en expliquant la démarche qui y a conduit.

En outre, rédiger en mobilisant des données chiffrées suppose de déjouer certains écueils. Afin de ne pas alourdir le style ou pour éviter de lasser le lecteur, il convient de ne pas paraphraser les données d'un tableau ou d'un graphique. Textes et figures correspondent à des niveaux de lecture différents, il faut les faire dialoguer sans dupliquer les informations. « Tous les chiffres

1. F. Batamio, L. Kalle, « Les étudiants : quel logement pour quelle sociabilité ? » Nous restituons leurs propos tels que les étudiantes les ont écrits. Cet exercice universitaire et d'autres, issus de l'enquête 2012-2013, sont accessibles en ligne à l'adresse suivante : www2.univ-paris8.fr/sociologie/?p=5693.

figurant dans le texte doivent être parlants pour les lecteurs. Mieux vaut arrondir que de donner des chiffres trop précis. Les “près de”, “plus de” et autres expressions équivalentes sont plus nuisibles qu’utiles¹. ».

Cette quête de clarté et de simplification du texte passe souvent par un travail de réécriture, plus ou moins important selon les auteurs. Cette étape ultime consiste à supprimer tout ce qui n’apporte rien à la démonstration afin de mettre au jour les éléments essentiels pour répondre au problème posé. En pratique, ce travail d’autoclarification vise à trouver la formulation la plus claire, à simplifier au maximum le propos, à couper des phrases ou des paragraphes entiers, à rayer les mots inutiles : adjectifs, adverbes, etc.

III. – Médiatisation ou illusion

La diffusion des résultats constitue l’aboutissement de toutes pratiques scientifiques. Avec le développement de nouveaux supports, cette diffusion dépasse de plus en plus la seule sphère académique pour atteindre les médias et le grand public. Soit le scientifique cherche à diffuser son savoir, et les maisons d’édition à vendre leurs ouvrages, soit les médias entrent directement en contact avec l’auteur pour commenter l’actualité avec son regard d’expert. Or, cette étape ne va pas sans soulever différents problèmes et paradoxes.

1. Le discours scientifique dans les médias. – Le champ scientifique et le champ journalistique répondent à des enjeux contradictoires. En conséquence, et c’est

1. D. Temam, « Savoir compter, savoir conter. Épisode n° 1 – Des chiffres : point trop n’en faut », *Courrier des statistiques*, n° 112, décembre 2004, p. 57.

là un paradoxe, le discours des chercheurs est sollicité par les journalistes parce qu'il se définit comme scientifique. Or, ce discours perd précisément une part de son caractère scientifique lorsqu'il s'exprime dans les médias. D'un côté, lorsque le sociologue, par exemple, présente ses travaux dans un espace médiatique, espace par définition balisé et réduit, il le fait au prix d'une simplification telle que son discours perd tout ce qu'il a de scientifique. De l'autre, lorsqu'on lui demande de commenter, dans l'instant, une actualité, non seulement il a l'impression de ne pas être suffisamment expert de la question, ses sujets de recherche ayant un lien plus ou moins étroit avec le phénomène, mais encore il a besoin du recul, propre à la quête de neutralité du sociologue, et donc du temps que le journaliste ne peut lui accorder. En accordant cette interview, il délaisse encore une part de la scientificité de son discours et la spécificité de sa démarche. La diffusion au grand public suppose donc bien souvent de réduire la complexité des propos. Sont bien ici en cause les attentes, divergentes, entre sociologues et journalistes : « deux métiers de l'écrit souvent proches dans leurs centres d'intérêt mais presque opposés dans leurs temporalités ¹ ».

Les journalistes répondent à des enjeux commerciaux et financiers qui imposent un format, un temps et des catégories de pensées prédéfinies, souvent très éloignés des cadres disciplinaires revendiqués par le scientifique. L'entrée sur le devant de la scène médiatique soulève dans bien des cas une autre question déontologique. Appelé en tant qu'expert, l'auteur est de surcroît questionné sur sa position en tant qu'intellectuel. Dès lors, le scientifique peut-il, ou doit-il, s'engager et en faire part

1. C. Van de Velde, « Le sociologue et les médias », in S. Paugam, *L'Enquête sociologique*, *op. cit.*, p. 411.

dans l'espace public ? Rien ne l'en empêche *a priori* si, toutefois, il précise le registre. Donner son opinion en tant que citoyen ne peut être un problème à condition que cette parole s'exprime comme telle.

2. Quand les chiffres s'en mêlent. – Lorsque le discours s'appuie sur des chiffres, cette illusion de scientificité dans la sphère médiatique est encore plus nette. Les débats récurrents dans l'association Pénombre¹ illustrent à quel point le nombre est accaparé par les médias. Sans nombre, il n'y a pour ainsi dire pas d'informations. Mobiliser le nombre donne l'impression d'une connaissance : « L'illusion d'être allé y voir, de s'être intéressés à la réalité². » Cette illusion de connaissance oublie souvent ce que cet ouvrage a eu à cœur de répéter : le nombre est une construction sociale. Il est donc fondamental de connaître, et donc d'explicitier, les conditions dans lesquelles il a été produit pour le comprendre.

1. L'association Pénombre, www.penombre.org, est un espace de discussions et d'interrogations sur les usages du nombre, et de ses dérivés, dans l'espace public.

2. Association Pénombre, « Chiffres en folie. Petit abécédaire de l'usage des nombres dans le débat public et les médias », préface de P. Meyer, Paris, La Découverte, 1999.

CONCLUSION

Signalée dès l'introduction, l'expression « méthodes quantitatives », distincte de celle de « statistiques », désigne un processus social, celui de quantifier. Ce processus renvoie à toutes les opérations de standardisation des matériaux (archives, observations et entretiens), quelle que soit leur forme initiale. Une telle définition déplace sensiblement les termes d'un débat devenu récurrent, celui de l'articulation entre différentes méthodes.

Si l'*articulation* entre méthodes qualitatives et méthodes quantitatives est souvent valorisée, elle perd en partie son sens d'après la définition précédente. Articuler les méthodes suppose de les distinguer. Or, en quoi finalement l'approche de l'ethnographe, qui détaille tous les aspects de son terrain et les soumet à un questionnement, et celle du quantitatifiste, qui garde trace des moindres choix qui président à la quantification et qui les met à l'épreuve, s'opposent-elles ? Dès qu'elles sont menées avec réflexivité, il devient délicat de les opposer.

Le quantitatifiste, lui aussi, confronte ses catégories de pensée à celle des enquêtés ou de l'institution enquêtée. Dès lors, reprocher aux chercheurs travaillant à partir d'analyses secondaires de données l'inadéquation des matériaux utilisés avec leur objet est infondé dès que l'on aborde la quantification comme un processus social. Ainsi, dans l'analyse d'archives, il y a de nombreux enseignements à tirer des classements

qui ont été opérés par l'institution d'une part, et de ce qu'elle a bien voulu transmettre, d'autre part.

Si les méthodes qualitatives et quantitatives ne sont pas foncièrement différentes, il est souvent complexe de les mobiliser simultanément durant le travail d'écriture. Dès lors, on peut céder à la facilité qui consiste à utiliser les catégories statistiques comme simple élément de cadrage permettant de replacer dans leur contexte les entretiens, ou encore à mobiliser les entretiens afin de décorer ou de rendre plus vivantes les statistiques. Pour éviter cette paresse d'écriture, nous invitons le lecteur à poser une véritable question de recherche qui soit commune aux deux approches. Ainsi, l'analyse qualitative et l'analyse quantitative ne seront finalement que deux points de vue sur un même objet d'étude. Pour peu qu'ils apparaissent comme contradictoires, cela déplacera sensiblement la question de recherche. C'est de la contradiction de ces deux matériaux que Joanie Cayouette-Remblière a par exemple reconstruit sa problématique de recherche. D'un côté, les discours de parents d'élèves des quartiers populaires dévoilent qu'ils entretiennent de fortes ambitions scolaires pour leurs enfants (en CM2). De l'autre, les résultats statistiques rendent compte d'une orientation moins favorable pour les classes populaires (en fin de troisième). Elle montre alors non plus comment ces familles choisissent leurs orientations, mais comment leurs choix sont orientés au fil des trajectoires par l'institution scolaire¹.

1. J. Cayouette-Remblière, « Les classes populaires face à l'im-pératif scolaire. Orienter les choix dans un contexte de scolarisation totale », *Actes de la recherche en sciences sociales*, vol. 5, n° 205, 2014, p. 58-71.

BIBLIOGRAPHIE

- Association Pénombre, *Chiffres en folie. Petit abécédaire de l'usage des nombres dans le débat public et les médias*, préface de Philippe Meyer, Paris, La Découverte, 1999.
- Barbot J., « Mener un entretien de face-à-face », in Paugam S. (dir.), *L'Enquête sociologique*, Paris, Puf, « Quadrige Manuels », 2010, p. 115-141.
- Beaud S., Weber F., « Le raisonnement ethnographique », in Paugam S. (dir.), *L'Enquête sociologique*, Paris, Puf, « Quadrige Manuels », 2010, p. 225-246.
- Béjin A., « La masturbation féminine en France. Un exemple d'estimation et d'analyse de la sous-déclaration d'une pratique », *Population*, vol. 48, n° 5, 1993, p. 1437-1450.
- Bennett T., Gayo-Cal M., Le Roux B., Savage M., Silva E., Warde A., Wright D., « *La Distinction* revisitée : l'espace des styles de vie britanniques en 2003 », in Coulangeon P., Duval J., *30 ans après La Distinction de Pierre Bourdieu*, Paris, La Découverte, « Recherche », 2013.
- Benveniste C., « Savoir compter, savoir conter. Episode n° 3 – Des graphiques et des tableaux », *Courrier des statistiques*, n° 115, 2005, p. 13-17.
- Bessière C. et Houseaux F., « Suivre des enquêteurs », *Genèses*, n° 29, 1997, p. 100-114.
- Bidet J., *Les Vacances au bled de descendants d'immigrés algériens. Trajectoires, pratiques, appartenance*, thèse de doctorat, université Lyon 2, 2013.
- Blanchet A., Gotman A., *L'Enquête et ses méthodes. L'entretien*, Paris, Armand Colin, « 128 », 2007.
- Bonvalet C., Lelièvre E., *De la famille à l'entourage*, Paris, Ined, « Grandes Enquêtes », 2012.
- Bourdieu P., *La Distinction. Critique sociale du jugement*, Paris, Minuit, 1979.
- (dir.), *Questions de sociologie*, Paris, Minuit, 1984.
- , « L'illusion biographique », *Actes de la recherche en sciences sociales*, n° 62, 1986, p. 69-72.
- Bugeja-Bloch F., *Logement, la spirale des inégalités*, Paris, Puf, « Le Lien social », 2013.
- Bruno A.-S., *Les Chemins de la mobilité. Migrants de Tunisie et marché du travail parisien depuis 1956*, Paris, EHESS, « En temps & lieux », 2010.
- Caveng R., « La production des enquêtes quantitatives », *Revue d'anthropologie des connaissances*, vol. 6, n° 1, 2012, p. 65-88.
- Cartron D., « Le sociologue pris en sandwich ! Retour sur une observation participante dans un fast-food », *Travail et emploi*, n° 94, 2003, p. 59-64.
- Cayouette-Remblière J., « Reconstituer une cohorte d'élèves à partir de dossiers scolaires. La construction d'une statistique ethnographique », *Genèses*, vol. 4, n° 85, 2011, p. 115-133.
- , « Les classes populaires face à l'impératif scolaire. Orienter les choix dans un contexte de scolarisation totale », *Actes de la recherche en sciences sociales*, vol. 5, n° 205, 2014, p. 58-71

- Chenu A., « Des sentiers de la gloire aux boulevards de la célébrité. Sociologie des couvertures de *Paris Match*, 1949-2005 », *Revue française de sociologie*, vol. 49, n° 1, 2008, p. 3-52.
- Chauvin S., Jounin N., « L'observation directe », in Paugam S. (dir.), *L'Enquête sociologique*, Paris, Puf, « Quadrige Manuels », 2010, p. 143-165.
- Coulangeon P., Duval J., *30 ans après La Distinction de Pierre Bourdieu*, Paris, La Découverte, « Recherches », 2013.
- Desrosières A., « La portée sociologique des diverses phases du travail statistique », in Besson J.-L. (dir.), *Des mesures*, Lyon, PUL, 1986, p. 247-273.
- , « Séries longues et conventions d'équivalence », *Genèses*, n° 9, 1992, p. 92-97.
- Desrosières A., Thévenot L., *Les Catégories socioprofessionnelles*, Paris, La Découverte, « Repères », 2002.
- Desrosières A., Kott S., « Quantifier », *Genèses*, n° 58, 2005, p. 2-3.
- Dietrich-Ragon P., *Le Logement intolérable. Habitants et pouvoirs publics face à l'insalubrité*, Paris, Puf, « Le Lien social », 2011.
- Durkheim É., *Le Suicide* [1897], Paris, Puf, « Quadrige », 2007.
- Duval J., « Analyser un espace social », in Paugam S. (dir.), *L'Enquête sociologique*, Paris, Puf, « Quadrige Manuels », 2010, p. 267-290.
- Felouzis G., « L'usage des catégories ethniques en sociologie », *Revue française de sociologie*, vol. 49, n° 1, 2008, p. 127-132.
- Gollac M., *Données un sens aux données. L'exemple des enquêtes statistiques sur les conditions de travail*, Paris, Centre d'études de l'emploi, 1994.
- Gollac S., « La fonction publique : une voie de promotion sociale pour les enfants des classes populaires ? », *Sociétés contemporaines*, n° 58, 2005, p. 41-64.
- Graunt J., *Observations naturelles et politiques sur les bulletins de mortalité* [1662], Ined, « Classiques de l'économie et de la population », 1977.
- Grémy J.-P., « Questions et réponses. Quelques résultats sur les effets de la formulation des questions dans les sondages », *Sociétés contemporaines*, n° 16, 1993, p. 165-176.
- Halbwachs M., « La statistique en sociologie », in *La Statistique. Ses applications. Les problèmes qu'elles soulèvent (septième semaine de synthèse)*, Paris, Puf, 1944, p. 113-160.
- Israël L., « L'usage des archives en sociologie », in Paugam S. (dir.), *L'Enquête sociologique*, Paris, Puf, « Quadrige Manuels », 2010, p. 167-186.
- Jounin N., *Voyage de classes. Des étudiants de Seine-Saint-Denis enquêtent dans les beaux quartiers*, Paris, La Découverte, « Cahiers libres », 2014.
- Lemieux C., « L'écriture sociologique », in Paugam S. (dir.), *L'Enquête sociologique*, Paris, Puf, « Quadrige Manuels », 2010, p. 379-402.
- Lemieux C., « Problématiser », in Paugam S. (dir.), *L'Enquête sociologique*, Paris, Puf, « Quadrige Manuels », 2010, p. 27-51.
- Lemercier C., Zalc C., *Les Méthodes quantitatives pour l'historien*, Paris, La Découverte, « Repères », 2008.
- Lorraine Data, *Le Grand Truquage. Comment le gouvernement manipule les statistiques*, Paris, La Découverte, 2009.
- Martin Olivier, *L'Enquête et ses méthodes. L'analyse de données quantitatives*, Paris, Armand Colin, « 128 », p. 14-22.
- Maruani M., Meron M., *Un siècle de travail des femmes (1901-2011)*, Paris, La Découverte, 2012.
- Merllié D., « Le travail des catégories statistiques », *Sociétés contemporaines*, n° 14-15, 1993, p. 149-163.

- Parizot I., « L'enquête par questionnaire », in Paugam S. (dir.), *L'Enquête sociologique*, Paris, Puf, « Quadrige Manuels », 2010, p. 101-121.
- Pinçon M., Pinçon-Charlot M., « Pratiques d'enquête dans l'aristocratie et la grande bourgeoisie : distance sociale et conditions spécifiques de l'entretien semi-directif », *Genèses*, n° 3, 1991, p. 120-133.
- Peneff J., « La fabrication statistique ou le métier du père », *Sociologie du travail*, n° 2, 1984, p. 195-210.
- Pollak M., *L'Expérience concentrationnaire. Essai sur le maintien de l'identité sociale*, Paris, Métailié, « Sciences humaines », 2000.
- Rounel-Fuentes M., *Les Chômeurs de Moulinex*, Paris, Puf, « Le Lien social », 2011.
- Safi M., « L'analyse longitudinale données et méthodes », in Chenu A. (dir.), *La France dans les comparaisons internationales*, Paris, Presses de Sciences Po, 2011, p. 161-172.
- Scioldo-Zürcher Y., *Devenir métropolitain. Politique d'intégration et parcours de rapatriés d'Algérie en métropole (1954-2005)*, Paris, EHESS, « En temps & lieux », 2010.
- Selz M., Maillachon F., *Le Raisonnement statistique en sociologie*, Paris, Puf, « Licence socio », 2009.
- Singly F. de, *Le Questionnaire. L'enquête et ses méthodes*, Paris, Armand Colin, « 128 », 2012 (3^e éd.).
- Temam D., « Savoir compter, savoir conter. Épisode n° 1 – Des chiffres : point trop n'en faut », *Courrier des statistiques*, n° 112, 2004, p. 5-7.
- Thévenot L., « Une jeunesse difficile. Les fonctions sociales du flou et de la rigueur dans les classements », *Actes de la recherche en sciences sociales*, n° 26-27, 1979, p. 3-18.t
- Van de Velde C., « Le sociologue et les médias », in Paugam S. (dir.), *L'Enquête sociologique*, Paris, Puf, « Quadrige Manuels », 2010, p. 403-420.
- Vilquin É., « Une édition critique en français de l'œuvre de John Graunt (1620-1674). Présentation d'un ouvrage hors collection de l'Ined », *Population*, vol. 33, n° 2, 1978.

TABLE DES MATIÈRES

Introduction	3
Chapitre premier – L’usage des statistiques en sciences sociales	5
I. Au fondement des sciences sociales, 5 – II. Réflexivité et posture du quantitatifiste, 9.	
Chapitre II – Les approches quantitatives	21
I. Trois types de recueil des matériaux, 21 – II. Quantifier en sciences sociales, 28 – III. Transversal et longitudinal, 36.	
Chapitre III – L’élaboration d’un questionnaire	42
I. La construction de l’objet, 42 – II. La structure du questionnaire, 47 – III. La formulation des questions et des réponses, 56.	
Chapitre IV – L’exécution sur le terrain	62
I. Procédures d’échantillonnage et inférence statistique, 62 – II. Négocier l’accès au terrain et réaliser la passation, 67 – III. Saisir les questionnaires, 72.	
Chapitre V – Panorama des techniques quantitatives	78
I. Les variables en question, 78 – II. Exploiter des variables numériques, 80 – III. Exploiter des variables nominales, 86 – IV. Lire l’analyse factorielle et comprendre la régression, 97.	

Chapitre VI – Valorisation et réception des résultats	107
I. L’art et la manière de présenter des données chiffrées, 107 – II. L’art et la manière de mobiliser des données chiffrées, 115 – III. Médiatisation ou illusion, 118.	
Conclusion	121
Bibliographie	123

Cet ouvrage a été mis en pages et imprimé en France
par JOUVE
1, rue du Docteur-Sauvé – 53101 Mayenne
2202660M – Dépôt légal : juin 2015

Imprimé par un visiteur avec l'adresse IP 194.154.200.98 sur
www.cairn.info - 23/10/2017