15/09

Partie 1 : Les méthodes quantitatives en sciences sociales

Préambule :

Les **méthodes quantitatives** (où les données sont exprimés en chiffre) sont **inséparables** des **sociétés** dans lesquelles elles s'insèrent. La société va **influencer** la manière dont on met en œuvre ces méthodes et le sens des résultats que l'on peut donner aux produits. Insertion sociale des méthodes est introduite par **Alain Desrosières** dans la politique des grands nombres. Il montre **un lien très étroit** entre 3 choses :

- Manière dont on **décrit** la société (statistique, entretien, observation)
- La façon de **penser** la société
- Et modalité dont on **agit** au sein de la société

Deux objectifs historiques qui sont mêlés :

- Les statistiques pour connaître la société à des visées scientifiques (objectif de connaissance → s'autonomise au XIXème avec les universités)
- Un objectif pragmatique → les statistiques vont aider à la décision publique

Les deux sources historiques des statistiques

1) La statistique, instrument d'Etat (mesure de puissance)

LES statistiques remontent à l'antiquité : Empire romain sous auguste

LA statistique ne prend son sens au **XVIIème** = relatif à l'Etat du latin status (ancien) ou staticus (moderne) = désigne une collecte de données relative à la puissance des Etats.

Pourquoi une apparition au XVII ? Les premiers **grands états** apparaissent et donc on recourt au **recensement**. La statistique est donc essentiellement **descriptive** avec un objectif de **synthétiser** des données (c'est-à-dire simplifier des faits pour qu'ils soient plus facile à utiliser pour le **gouvernement** et plus facile à **enseigner**). Des acteurs **universitaires** = école descriptive allemande, université Rénale. Ils vont fonder la **démarche quantifiée**, une des première à représenter les données non pas sous forme littéraire mais sous forme de tableau croisé (appliquer une grille commune à différents états). Elle établit l'un des présupposés de la statistique = **« mise en équivalence » (Alain Desrosières)** = efface la singularité, cela suppose une mesure commune, les phénomènes mesurées sont devenus homogènes dans les faits réels (mesure reconnue, droit)

Remarque : pour que cette mesure ait un sens, il y a une homogénéisation. Cela implique DONC UNE **CONSTRUCTION POLITIQUE** DE L'ETAT, il porte cette homogénéisation dans les territoires.

2) La statistique comme activité scientifique autonome

Au **XVII**, en **Angleterre**, dans une visée scientifique, l'**« école des arithméticiens politiques** » est fondée. Il ne s'agit plus simplement des données pour puissance des états MAIS un **ensemble de procédés** qui vont permettre de **traiter** et **classer** l'information (prémices de différents instruments)

Pourquoi en Angleterre ? Dû aux spécificités de son état (faiblesse des recensements, Etat + faible, peu de données exhaustives)

Trois **spécifiés** de cette école :

- **Origine sociale** de ceux qui porte cette démarche (en Allemagne se sont des universitaires ici il s'agit plutôt du **MONDE DES AFFAIRES** ex : John Graunt, William Petty)
- Méthode statistiques : de nouvelles techniques de calcul et d'évaluation (ex : le multiplicateur de population)
- **Objectifs** des statistiques : dégager des **régularités** pour tenter de faire des **prévisions** (plus pragmatique : comprendre la logique des faits sociaux pour pouvoir prédire alors que la méthode de l'Allemagne consistait à synthétiser pour choisir)

Ex : travail de John Graunt sur la table de mortalité avec les registres (décès, mariage, baptême) avant la transition démographique = déduction // de la cohorte, on va savoir s'il est intéressant d'acheter, fondement de la future industrie des assurances

La table de mortalité de John Graunt (Londres, 1660)

Sur 100 personnes nées en 1570, N n'atteignent pas :	N	Sur 100 personnes nées en 1570, n décèdent :	n
0 ans	0	Entre 0 et 6 ans	34
6 ans	34	Entre 6 et 16 ans	26
16 ans	60	Entre 16 et 26 ans	15
26 ans	75	Entre 26 et 36 ans	9
36 ans	84	Entre 36 et 46 ans	6
46 ans	90	Entre 46 et 56 ans	4
56 ans	94	Entre 56 et 66 ans	3
66 ans	97	Entre 66 et 76 ans	2
76 ans	99	Après 76 ans	1

John Graunt (1620-1674), riche drapier londonien est connu comme un des **précurseurs de la démographie.**
Les **bulletins de mortalité** existaient à Londres depuis l'épidémie de peste en 1592, et y devinrent systématiques à partir de 1603.
Graunt consacre plusieurs années à **étudier** ces bulletins de mortalité, dressant méthodiquement de véritables statistiques, où il étudie la mortalité, la fécondité, la nuptialité, la migration, ainsi que des **observations épidémiologiques et économiques qui en découlent.**
Il met en évidence des **phénomènes démographiques** jusque-là inconnus (comme le fait qu'il naît plus de garçons que de filles) et **contredit** certaines croyances sur l'apparition des épidémies. Il réalise les premières tables de mortalité.
Ce travail lui vaut d'être nommé membre de la Royal Society (1664).
Les observations établies par Graunt serviront aussi de base aux premiers travaux statistiques sur l'espérance de vie, comme ceux de **Huygens.**

A partir de la **fin du XVIII** = diversification dans les domaines d'études (plus seulement la démographie)

- **Economique** (agricole puis industrielle)
- Puis **social** (cf. travaux sur la pauvreté finan entre 1880 et 1900 par les grands industriels Charles Booth et Benjamin Rowntree au RU à Manchester et à Londres, monographies de Frédéric le Play en France dans les années 1850)

3) Etatisation de la société et développement de la statistique

La statistique **contemporaine** :

- En faire au XIXeme une **activité scientifique avec un perfectionnement des méthodes**
- Amélioration est étroitement **liée** aux **formes d'intervention** de l'Etat

Depuis le début du **XIXème siècle**, il y a eu **4** étapes d'amélioration des méthodes quantitatives :

« Etat libéral »: le rôle de l'Etat est au **centre** = garantir le bon déroulement d'une économie de marche, état gendarme (adam smith) = donc il faut garantir une **transparence de marché** d'où une **centralisation de l'information économique**. Cette étape montre les **limites** de la méthode **anglaise** qui subit un discrédit et s'appuie sur l'école descriptive allemande. On voit aussi **l'apparition de bureau de statistique** pour rassembler l'info

Ex 1. Royaume-Uni, 1833 : la création du département statistique du Board of Trade (débat sur les Corn Laws)
Ex 2. Etats-Unis, fin XIXe s.: les lois anti-trusts (Sherman Antitrust Act, 1890) concurrence, baisse des prix

……

« Etat providence » : fin du XIX : le rôle de l'état change en raison de la **grande dépression** (1870-90) : lutter contre la pauvreté par une **protection législative et juridique du travail** (prémices retraite, protection sociale) = un renouvellement de la statistique grâce à la **création de bureaux de statistique du travail**. Il y a un renouvellement de la statistique publique : au niveau des **objets** (pas seulement les prix), nouvelles catégories d'intervention de l'état (chômage, salaire, budget ouvrier) + au niveau des **méthodes** (repose sur l'idée d'assurance des risques d'où l'idée de probabilité = création de : variance, écart type, sondage aléatoire).

Ex 1. Royaume-Uni, 1833 : la création du département statistique du Board of Trade (débat sur les Corn Laws)
Ex 2. Etats-Unis, fin XIXe s. : les lois anti-trusts (Sherman Antitrust Act, 1890)

……

« L'Etat keynésien » (1945-1980's) : répondre à la crise des années 30 et à la Guerre : il faut **réguler l'économie**

Pour cela, il faut mettre en place de nouveau instrument au niveau de la comptabilité nationale, tout unifié d'un point de vue macroéconomique. Il faut réorganiser la chaîne de production des statistiques. Cela oblige l'état à une collecte systématique des infos de l'économie et à une étude tout aussi systématique. C'est un processus indispensable pour l'élaboration des politiques publiques, sociales…

- **création d'instituts spécialisés** (INSEE en 1946, INED, etc.) (RU plutôt tardif en 1990)
- **« coordination statistique »** (loi de 1951 relative à l'obligation, à la coordination et au secret statistique = PdV de la France) = conception **de tout unifié** + prémices de l'informatique = identifier les lacunes, les mettre en relation pour **éviter contradictions et meilleures** analyses = organise transmission des données nationales et individuelles vers des instituts spécialisés (anonymat pour protection des données)
- **transformation des objets des enquêtes publiques** (budget ouvrier avant, ici flux économique élargie macroéconomique = ensemble de la population donc moins ciblés, restreint)

……

« L'Etat néo-libéral » (1980's - ?) : rôle de l'Etat : accompagner l'**adaptation** au marché international.

CSQ : Développement des indicateurs de performance, palmarès, big data (traces des comportements), etc…

CCL du préambule : Méthodes quantitatives pas un simple outil descriptif
Origines sociale et politique est sous-jacente à la mise en pratique de ces méthodes = contrainte à la rigueur pour les scientifiques

Partie 1 : Les méthodes quantitatives en sciences sociales

Chapitre 1 : La collecte des données quantitatives. Le questionnaire

Le questionnaire est une méthode utilisé par un très grand nombre d'acteurs pas seulement universitaires et scientifiques

I. L'enquête par questionnaire comme méthode scientifique

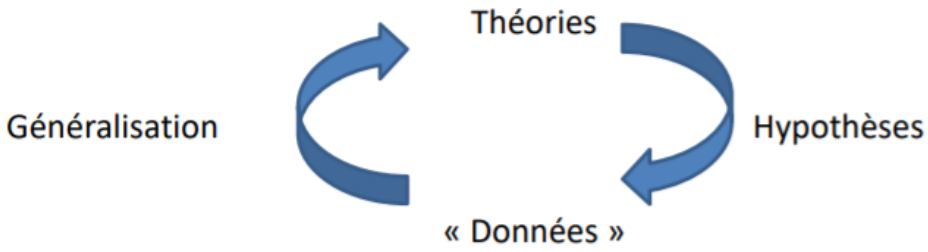
3 objectifs d'analyse quantitative :

- **Estimer une grandeur à partir d'un échantillon** (cf. institut de sondage) ex : pourcentage de vote, côte de popularité de tel candidat
- **Vérifier des hypothèses** : démarche causale

Dans les deux cas (travail sur échantillon): l'objectif s'inscrit dans le cadre d'un **raisonnement inférentiel** (significativité). On **déduit** le comportement de tous les français selon la **représentativité** de l'échantillon.

- **Décrire une population ou sous population** (quand inconnu, permet de se renseigner par questionnaire) → outil ancien de la statistique descriptive // On ne se préoccupe pas (nécessairement) de la représentativité statistique (Julien GROS, « Quantifier en ethnographe. Sur les enjeux d'une émancipation de la représentativité statistique », Genèses, n° 108, 2017) Le groupe étudié n'est pas un échantillon mais **un groupe social** : la base de donnée est nécessairement biaisée.

1 objectif explicatif : inséré dans le cadre d'une démarche **hypothéticodéductive** (enquêtes par questionnaire)



Le questionnaire est un instrument pour tester des hypothèses. Il permet de confronter une réalité empirique selon des données.

L'utilisation du questionnaire impose de se poser **5 types de questions** :

1) **Quel est l'objet de l'enquête ?** sur quelle population on cherche des informations ? questionnaire pas adaptés à toutes les populations, ciblées. **Pour que le questionnaire soient pertinent il faut que des institutions (intérêt économique, politique ou commercial) est travaillé l'objet pour lui donner des frontières (institutionnalisées, sociales). = objectivation du social** Homogénéisation rend l'objet plus pertinent, plus représentatif (vote alors que SDF plus informel, lui cibler une activité de ces derniers)

EX : Ministre de la culture, années 1980 : artiste peintre. Interroger qui ? Reconnus, phase de reconnaissance, étudiants (auto déclaration = filet très large) ou artiste peintre qui en vivent (filet très réduit, car ce sont les plus établis). Perdre toutes les situations intermédiaires. Comment les saisir ? Tirer parti de la technique de

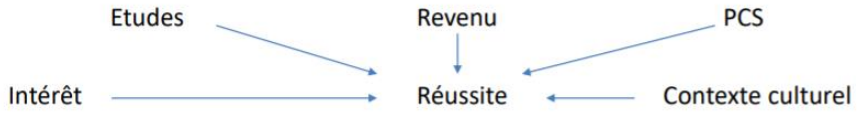
l'**objectivation** : tous les intermédiaires qui aident à leur reconnaissance (galeries, annuaires, critiques, musées…). Avec tous ces moyens, ils parviennent à établir une liste d'artistes.

2) **Quelle problématique adopter ?** quelles sont les questions que l'on se pose ? Pourquoi mettre en évidence certains facteurs ? Les faits ne parlent pas d'eux-mêmes, il faut **les faire parler en établissant un système d'hypothèse, un modèle d'analyse qui est à confronter avec une réalité empirique** (celle du questionnaire).

EX : les facteurs de la réussite à l'école primaire ?

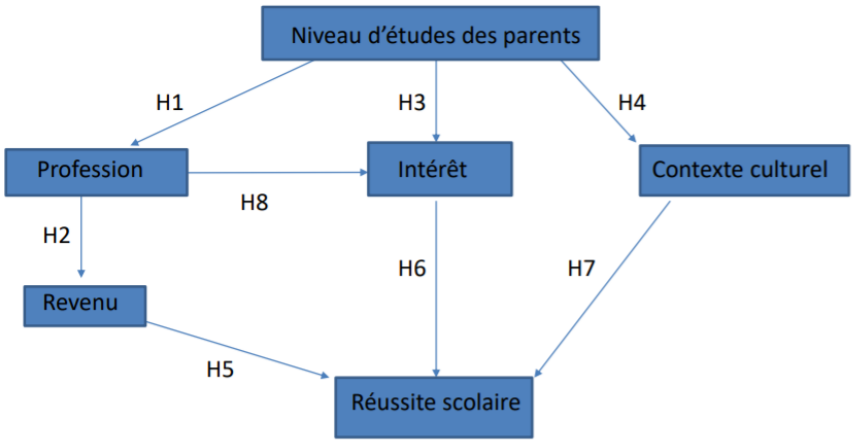
Hypothèses :

- Réussite plus fréquente dans milieux favorisés : **Revenu, PCS**
- Réussite plus fréquente quand parents sont disponibles : **intérêt**
- Réussite plus fréquente quand le capital culturel des parents s'élève **: études parents, contexte culturel**



Hypothèses isolées qui sont à hiérarchiser, à mettre en relation.

Il est nécessaire de construire un système de relations :



Pourquoi définir les choses de cette manière (système d'hypothèses) ?

- **Confirmer la relation** entre la variable la plus importante et la variable la plus élevée. Si les résultats infirment cela, il y a un pb.
- **Quelles sont les relations dans le système à enrichir**, à compléter ?

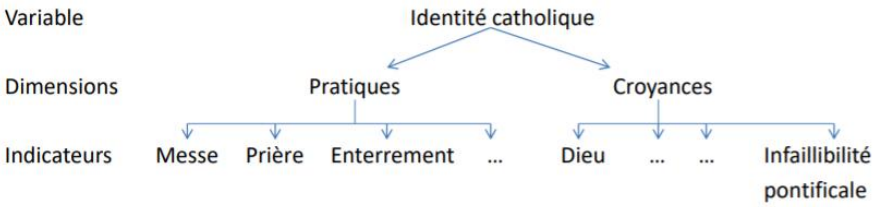
Comment rendre ce système de relation opératoire ?

Dès le début du XXème siècle, solution classique proposée (partiellement valable) : plusieurs étapes formalisées par le sociologue **Paul Lazarsfeld** (« Des concepts aux indices empiriques », 1958). Pour rendre opératoire le système, il va falloir trouver des **indicateurs (réalités empiriques) fortement liées au concept**. Propose de réfléchir de **manière analytique (décomposer le concept en élément simple).**

EX : prudence (variable) = assurance, conduite, pari (indicateurs) = as – t- elle eu des PV ? (questions) = réponses

Les variables sociologiques sont plus complexes donc il va falloir décomposer la variable en dimension intermédiaire (niveau intermédiaire). Il va falloir faire un lien ensuite entre le niveau théorique et le niveau de la « réalité ».

Ex : l'identité catholique (Guy Michelat, « L'identité catholique des Français », *Revue française de sociologie*, 1990)



Questions sur la fréquence….
Problème qui se pose : nombre d'indicateurs très élevé = quels indicateurs retenir ? Combien d'indicateurs retenir ?

Paul Lazarsfeld définit deux principes qui doivent guider le chercheur. Il les tire de ses **propres études**.

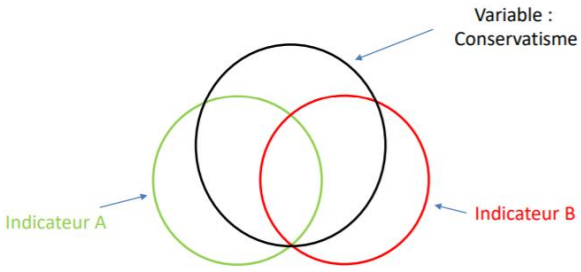
a) L'imperfection des indicateurs

Ex : l'étude du conservatisme des universitaires US pendant le maccarthysme.

A: « Seriez-vous favorable à autoriser les étudiants à inviter Owen Lattimore (spécialiste des questions d'Extrême-Orient, inculpé devant la commission d'enquête) ? »
B: « Pensez-vous qu'on doive permettre la formation d'un groupe de jeunesse socialiste dans cette université si certains en expriment le désir ? »

Analyse des résultats : 15% de conservateur et presque 80 de personnes tolérantes. Problème quand on croise les deux.

Tout indicateur est donc imparfait : la relation entre chaque indicateur et le concept (ou ses dimensions) est défini en termes de probabilité.



Quel indicateur est le meilleur ?

b) L'interchangeabilité des indicateurs

Mettre en relation les variables entre elles.

Ex : Lazarsfeld utilise les deux indicateurs précédents pour expliquer le choix entre individualisme et corporatisme, à travers la question : « Imaginons un enseignant accusé d'activités subversives ; pensez-vous qu'il soit plus important pour l'administration de l'université de protéger sa réputation en renvoyant l'enseignant ou les droits des membres du corps enseignant?

Quelques soit l'indicateur que l'on choisit, les proportions sont à peu près les mêmes. Des indicateurs pas identiques mais tout de même associés.

Si les indicateurs sont liés entre eux, les liens entre les variables restent globalement stables, quel que soit l' (les) indicateur(s) retenu(s). Bien le lien des variables qui nous intéressent et pas l'indicateur en tant que tel.

3) **Sur quoi et comment interroger?** = La formulation et construction du questionnaire
4) **Qui interroger?** = l'échantillon et le mode de passation (comment ils arrivent jusqu'à l'enquêté)
5) **Comment interpréter les résultats?** = l'analyse quantitative.

Peut aussi faire une analyse de seconde de main, avec des enquêtes déjà faites.

II. La formulation des questions et de la construction du questionnaire

2.1 Les type de questions

Contenu et forme

2.1.1. Questions de fait vs questions d'opinion

4 types de questions en fonction du contenu :

• **Questions qui décrivent les personnes** (identité sociale : revenu, origine sociale) : qui êtes-vous? → Phénomène d'explication, ces questions vont permettre d'analyser les réponses aux autres questions.

• **Questions sur les pratiques** : que faites-vous? → Déclaration sur ce type de pratique

• **Questions d'opinion,** d'évaluation, de jugement : que pensez-vous de…?

• **Questions cognitives** : que savez-vous de…? → Éviter de les multiplier, peut être dissuasif

Pratique et opinion : Dans les faits, la frontière fait / opinion peut être floue. Pq ? car pas d'observation des pratiques mais des **déclarations sur ses pratiques**, réponses qui s'apparent à des réponses d'opinions. Ex : des pratiques très **stigmatisantes ou très valorisantes** = sur le vote de l'extrême droite ou pratiques culturelles (musées…) ou alors des réponses approximatives qui peuvent être basé sur **l'image que l'on souhaite donnée de soi** (ex du nombre de livres).

2.1.2 Questions fermées VS questions ouvertes

Questions fermées : les personnes interrogées doivent choisir entre **des réponses déjà formulée à l'avance**. Réduire l'option de choix de l'enquêté.

Deux avantages, deux raisons de choisir ce type de questions :

- Le **coût** (passation plus direct et moins de personnel, standardisé)
- Un effet de **libération de la parole** (relation avec l'enquêté : thème abordée complexe ou difficile à traiter = cadre qui rassure la personne, réponse légitimée (acceptée))

Il faut bien choisir la liste des réponses, **bien connaître le terrain aux préalables**. Aussi non faible proportion de réponse, ne s'y retrouve pas ou une obligation de réponse et donc moins représentative **= une distorsion.**

***Questions ouvertes* :** on ne propose pas d'items de réponse.

Deux avantages :

- elles permettent une **expression plus libre**, voire même reformulation de la question (voir comment la question est comprise)
- les réponses permettent des **possibilités de codage plus importante** (ex : lecture des collégiens : quel est le dernier livre que tu as lu ? coder le niveau de l'ouvrage, scolaire ou non, thème, type, description donnée est-elle riche ou non ?)

Cependant de **nombreux inconvénients dans les questions ouvertes** :

- Elles accroissent les **effets liés à l'enquêteur** (comparabilité réduite)
- Elles font courir le risque d'un effet **d'éviction** important (=**exclusion**, elle présuppose une verbalisation des pratiques, qqc de non reparti socialement donc un risque par rapport à la représentativité de l'échantillon)
- Les réponses peuvent **être trop dispersées** pour permettre un recodage (pas de regroupement possible, recodage difficile)

Ex : L'expérience d'Howard Schuman et Stanley Presser (« Open and closed questions, American Sociological Review, 1979) = interroge des parents sur ce qui est le plus important à apprendre aux enfants pour les préparer dans la vie d'adulte. **40% des réponses ouvertes ne sont pas exploitables**.

	While we're talking about children, would you please look at this card. If you had to choose, which thing on this list would you pick as ***the most important*** thing for children to learn to prepare them for life?		While we're talking about children, would you please say what you think is ***the most important*** thing for children to learn to prepare them for life?	
1	To Obey	19%	To Obey	2,4%
2	To Be Well-Liked or Popular	0,2	To Be Well-Liked or Popular	0
3	To Think For Themselves	61,5	To Think For Themselves	4,6
4	To Work Hard	4,8	To Work Hard	1,3
5	To Help Others When They Need help	12,6	To Help Others When They Need help	0,9
6			To Be Self-Reliant	6,1
7			To Be Responsible	5,2
8			To Have Self-Respect	4,1
9			To Have Respect For Others	6,7
10			To Have Self-Discipline	3,5
11			To Be Honest	7,4
12			To Have Other Moral Qualities	3,0
13			To Be Religious	5,4
14			To Love Others	2,0
15			To Get an Education	12,8
16			To Learn a Trade or Job Skill	0,9
17			To Get Along With Others	5,0
96			Multiple Responses	16,1
97	Other	0	Other	9,3
98	D.K.	0	D.K.	1,3
99	N.A.	1,8	N.A.	2

→ Quelles solutions?

- **Limiter le nombre** de questions ouvertes dans le questionnaire (thème central, catégorie, pose peu de recodage)
- Utiliser des **questions semi-ouvertes (rubrique "autres")**

2.1.3 Les types de questions fermées

- ✓ les questions alternatives
- ✓ les questions préformées (à choix multiple)
    - Sur le **type de réponse**

Une seule réponse possible = Questions préformées à **choix forcé**
Plusieurs réponses possibles = Questions préformées à **choix libre**

- 3 types de questions préformées selon **le type d'item de réponse** :

Questions préformées à **échelles numériques** (nombre de mois, de fois…)
Questions préformées à **échelles ordinales** (échelle gauche-droite, échelle d'accord)
Questions préformées à **échelles nominales**

2.2 Choisir la formulation des questions

2.2.1 La pertinence des questions posées

Deux règles à respecter :

- **La question que se pose le chercheur n'est pas la question à poser à l'enquêté** (pourquoi : question complexe à décomposer + risque de biaiser les réponses)
- Toute personne interrogée **doit comprendre** la question, de manière similaire **(univocité)**. **Doit se rapprocher de ses préoccupations réelles, de son vécu quotidien, de son expérience pratique.**

Manière de jongler entre la théorie et la manière dont on parle aux enquêtés. Sinon :

- Risque d'un **décalage entre le sens donné par l'enquêteur et par les enquêtés**
- Risque que la **formulation** (question, items) soit totalement **étrangère** à l'univers des enquêtés (**non réponse**)

Ce problème concerne les questions de pratiques et d'opinions.

Questions **pratiques** exemple : **une étude sur les choix d'orientation en Brevet de technicien supérieur (Sophie Orange, « Interroger le choix des études supérieures », 2012)**
Grands nombres de non-réponses car séparation entre BTS ou IUT = change pour juste demander la liste des choix sans distinction. Pourquoi même avec reformulation, un grand nombre de non-réponse ou alors de « mauvaises » réponses = graves défauts de sa réflexion à elle = n'a pas tenu compte de la spécificité de la population qu'elle étudie
Le taux de non-réponses révèle des présupposés non identifiés dans la construction de l'objet :
- **Le choix des études supérieures** est un « choix » : or décisions des étudiants des quartiers populaires sont souvent **collectives**
- La **mémoire des « choix » est également répartie** (or « cadres sociaux de la mémoire », M. Halbwachs) = temporalité spécifique, retient mieux quand des actions individuelles (rencontre CO, recherches, discussions familiales).

Questions **d'opinion** : Négliger ces deux règles fait courir un risque d'**« imposition de problématique »** (terme de Pierre Bourdieu) :
« Effet exercé par toute enquête d'opinion qui résulte du fait que les **questions posées** ne sont **pas des questions qui se posent réellement** à toutes les personnes interrogées » (Pierre Bourdieu, « L'opinion publique n'existe pas », 1972)
→ obliger de répondre alors que pas réellement d'opinion sur le sujet = mais leur réponse va être analysé comme une opinion …

➔ **3 effets de cette « imposition de problématique »** (Daniel Gaxie, « Au-delà des apparences », Actes de la recherche en sciences sociales, 1990) :

1) ***La retraduction de problématique*** **:** l'enquêté interprète mal la question.

L'individu **retraduit le sens de la question posée donc pas une réponse directement voulue**. Il y a donc une double évaluation par la suite.

Ex : les questions d'opinion : Le libéralisme dans la société française
Question A : « Certains disent que pour s'en sortir, il faut introduire plus de libéralisme dans la société française. Êtes-vous tout à fait d'accord, plutôt d'accord, plutôt pas d'accord, pas du tout d'accord ? »

Question cognitive lors d'une enquête sur enquête :
Question B: « Pouvez-vous me dire en quelques mots ce que représente le libéralisme pour vous?

Question d'opinion : montre un effet d'imposition par les enquêteurs (2/3 contexte politique)

Réponse à la question B				
NR	Confusion *	Libéralisme = liberté	Libéralisme sens éthique**	Libéralisme sens politique
10,1%	8,1%	33%	13,2%	35,6%

Montre l'intérêt, la position sociale, l'âge, le sexe, la PCS, capital culturel = classe populaire, peu diplômé

2) *Les contradictions dans les réponses*

Répondent a priori de façon incohérente aux questions de l'enquête

Réponse à la question A en contradiction avec celle de la question C :
Question A : « Certains disent que pour s'en sortir, il faut introduire plus de libéralisme dans la société française. Êtes-vous tout à fait d'accord, plutôt d'accord, plutôt pas d'accord, pas du tout d'accord ? »
Question C : « L'actuelle majorité UDF-RPR [enquête en 1988] parle de renforcer le libéralisme. Êtes-vous tout à fait d'accord, plutôt d'accord, plutôt pas d'accord, pas du tout d'accord ? »

= montre l'intérêt politique = classe moyenne

3) *La « prudence tactique »*

C'est-à-dire des **réponses qui engagent peu.**
Une influence de la situation d'enquête : les enquêtés cherchent **à ne pas perdre la face** (Gaxie, 1990)
Ils cherchent à **identifier des bonnes réponses, plausibles, de manière scolaire.**

→ Conséquence : **réponse n'équivaut pas à opinion :**
Cet effet est connu depuis l'origine de la méthode, particulièrement aux USA

Ex 1. **Herbert Hyman** et al. **Interviewing in Social Resarch, 1954**: enquête à Memphis, Tennessee, 1942 → interrogé en fonction des groupes des enquêteurs : caucasien ou les noirs-américains
Cherche les réponses qui vont plaire, pas forcément d'opinion sur les sujets.

Ex 2 : enquête de sociologie cognitive : Enquête Démocratie 2000 : Cevipof / chercheurs US (N = 2148) (G. Grunberg, **N. Mayer, P. Sniderman,** La démocratie à l'épreuve, 2002) **Expérience « de la pommade »** --> **différence de réponse liée à la formulation de la question** : soit de manière neutre, soit en valorisant la contribution de l'enquêté

Avec "pommade"	Sans "pommade"
"Nous arrivons à la fin. Je voulais vous dire que j'ai beaucoup apprécié cet entretien et que vos réponses nous serons très utiles. Finalement ne pensez-vous pas, contrairement à ce que l'on entend souvent dire, que le vrai problème c'est qu'il y a trop de liberté [de démocratie] en France?"	"Je voulais vous poser une dernière question. Pensez-vous qu'il y a trop de liberté [de démocratie] en France?

	"Trop de liberté"		"Trop de démocratie"	
	Avec "pommade"	Sans "pommade"	Avec "pommade"	Sans "pommade"
TROP	**40%**	**28%**	**31%**	**16%**
PAS TROP	57%	70%	65%	81%
Sans "opinion"	3%	2%	4%	3%
Total	100%	100%	100%	100%

2ème exemple : **contre argumentation**
Pose la question, on contre-argument et on repose la question.

Conclusion : ne pas dire que les gens réponde n'importe comment. Ceux qui vont changer le plus d'avis sont ceux qui sont le plus éloignés de ces thématiques.
Effet n'est pas nul sur ceux qui connaissent le sujet. Etre vigilant sur les réponses des enquêtes d'opinions à thématique politique.

Expériences de contre-argumentation

1e question posée à tous			
Certains disent qu'il faudrait augmenter le montant des indemnités de chômage. Vous-même, êtes-vous plutôt d'accord ou plutôt pas d'accord avec cette opinion?			
D'accord : 54%		**Pas d'accord : 46%**	
Expérience de contre-argumentation de l'enquêteur			
Contre-argument "plein"	Contre-argument "vide"	Contre-argument "plein"	Contre-argument "vide"
Si les indemnités chômage sont trop confortables, les chômeurs ne sont pas encouragés à chercher du travail	*Si on pense à tous les problèmes que ça pose*	*Avec les indemnités actuelles, il y a trop de familles qui n'arrivent pas à s'en sortir*	*Si on pense à tous les problèmes que ça pose*
2e test			
Certains disent qu'il faudrait augmenter le montant des indemnités de chômage. Vous-même, êtes-vous plutôt d'accord ou plutôt pas d'accord avec cette opinion?			
D'accord : 60% Pas d'accord 37% SR : 3%	D'accord : 81% Pas d'accord 15% SR : 4%	D'accord : 29% Pas d'accord 66% SR : 5%	D'accord : 13% Pas d'accord 84% SR : 3%
Changement d'avis Diplômés de gauche : 19% Non-diplômés de droite : 55%		**Changement d'avis** Diplômés de droite : 10% Non-diplômés de gauche : 43%	

- **Une intériorisation de la domination sociale :**

Les **enquêtés qui ont un avis peuvent être incités à ne pas répondre s'ils perçoivent la connotation politique** de la question **(= sentiment d'illégitimité à répondre)**

Ex : enquête sur le libéralisme (Gaxie, 1990)
Question A : « Certains disent que pour s'en sortir, il faut introduire plus de libéralisme dans la société française. Êtes vous tout à fait d'accord, plutôt d'accord, plutôt pas d'accord, pas du tout d'accord ? »
Question C : « L'actuelle majorité UDF-RPR [enquête en 1988] parle de renforcer le libéralisme. Êtes vous tout à fait d'accord, plutôt d'accord, plutôt pas d'accord, pas du tout d'accord ? »

Le **sentiment d'illégitimité se réduit si les questions sont proches de l'expérience des enquêtés**. Donc poser des questions proches des valeurs des enquêtés, il faut qui est un lien plus direct, des questions plus concrètes.

= Absence de réponse n'équivaut pas à absence d'opinion

Donc, un **danger constat lié des présupposés des chercheurs** ou de celui qui formule questionnaire. Donc, **un risque permanent d'artefact :** *les variations de la « mesure » sont davantage imputables aux caractéristiques de l'instrument qu'à celles de l'objet mesuré.*
Méthode de recueil explique plus le taux de non réponse que les questions posées.

Quelles solutions ?
- Soigner la **construction** de l'objet d'étude
- Jouer sur le registre de **langage** (langage plus familier et concret)
- **Définir** les termes employés
- Recours aux **questions cognitives** pour « repérer ce que les enquêtés ont en tête » (Gaxie, 1990)
- Tester la fiabilité des réponses avec des **questions-leurre**

2.2.2 Le poids des mots

Mécanisme d'adhésion à la formulation qui est utilisée. Elle donne des attendus, des sens légitimes (effet de légitimité). **La formulation donne des indications sur ce qu'il est légitime / attendu de penser et répondre sur le sujet.**
Il y a un mécanisme d'adhésion qui influe sur l'enquêté (ex. enquête sur le clivage gauche-droite). Ils portent principalement sur les questions d'opinion et les questions de faits qui se comportent comme des questions d'opinion.

Ex 1 : le clivage des réponses

« On classe habituellement les Français sur une échelle de ce genre où, comme vous le voyez, il y a deux grands groupes : la gauche et la droite. Vous, personnellement, où vous situez-vous sur cette échelle ? »

|___| |___| |___| |___| |___| |___| |___|

Extrême gauche Gauche Droite Extrême droite

Source : Cevipof, *L'électeur français en question*, 1990

	1981				1988			
Se classent :	Gauche/ Extrême-gauche	Droite/ Centre-droit	Centre	Non Réponse	Gauche/ Extrême-gauche	Droite/ Centre-droit	Centre	Non Réponse
Ens. Français	30+14 = 44%	24%	27%	5%	31+12 = 43%	26%	25%	6%

Classement présenté de manière naturelle (habituellement), aucune autre possibilité à l'enquêté que de se classer. Il montre une idée d'un comportement anormal à cette enquête, d'où des non-réponses.

Cette enquête permet de montrer différente formulation.

« On range parfois les Français en deux catégories : ceux qui sont de gauche et ceux qui sont de droite. Dans laquelle de ces deux catégories vous rangeriez-vous : ceux qui sont de gauche ? ceux qui sont de droite ? »

	1981				1991			
Se classent :	A gauche	A droite	Refus de se classer	NR	A gauche	A droite	Refus de se classer	NR
Ens. Français	42	31	20	7	36	28	30	6
Commerç., artisans et chefs d'entreprise	29	44	21	6	18	39	36	7
Cadres et professions intellect. supérieures	44	34	20	2	27	36	35	2
Ouvriers	48	26	17	9	37	15	37	11
UDF	4	74	19	3	3	76	19	2
Mouvement écologiste	40	17	33	10	43	11	41	5
FN	-	-	-	-	0	54	46	0

Source : Sofres, *L'état de l'opinion 1992*

Erosion car de nouveaux mouvements politiques.
La formulation peut orienter même si aucune opinion préconstituée.

Ex 2 : La substitution de termes « synonymes » (contexte des années 40)
Les mots ne sont pas neutres, influence les réponses.

Ex 3 : L'effet d'ajout d'information
Ex 4 : le déséquilibre des items de réponse
Ex 5 : l'effet d'acquiescement : les « yeasayers » → question positive
Ex 6 : l'effet des mots introductifs (modalisateurs)
Ex : l'effet de liste

} Confere le diapo

Quelles solutions?

- **Pour les questions d'opinions, suggérer la légitimité de l'absence d'opinion** (oui ou non. si oui ? risque : voir le nombre d'absence d'opinion) ou **la pluralité des opinions** (cf. ex)

Ex : en pensant à l'école, pouvez-vous me dire avec laquelle de ces opinions êtes-vous d'accord?
L'école devrait avant tout donner le sens de la discipline ou de l'effort
L'école devrait avant tout former des gens à l'esprit éveillé et critique

- Pour les questions sur les **pratiques** : donner **la possibilité de proposer plus d'une réponse** (choix libre : ex : dernier roman lu cet été ?)

- Pour les questions à **échelle nominale** : établir un équilibre entre modalités positives et négatives de réponse
  Alterner aléatoirement l'ordre des items posés (passation au téléphone) ou procédure du « carton » (passation en face-à-face)

2.3 Les effets de halo

Les questions précédentes peuvent amener l'enquêté à donner une réponse différente.

2.3.1 Les effets de focalisation sur le thème : les questions qui **précèdent influencent la perception du sujet de la question par l'enquêté** (par les informations qu'elles contiennent).

Ex : « L'attitude des Français à l'égard de l'immigration étrangère, particulièrement le droit de vote des étrangers» (INED, 1974)

Forme A. Q. 1. *Est-ce que vous seriez d'accord ou pas d'accord pour qu'on donne aux étrangers résidant en France le droit de vote aux élections municipales ?*

Q. 2. *... le droit de vote aux élections législatives (élection des députés) ?*

Q. 3. *... le droit de vote aux élections présidentielles (élection du président de la République) ?*

Q. 4. *On a donné récemment aux travailleurs étrangers le droit de voter aux élections pour désigner les représentants du personnel et le droit d'être élus. Estimez-vous qu'il était souhaitable de prendre cette mesure ?*

Q. 5. *Dans une ville de Belgique, des élections ont été organisées parmi les travailleurs étrangers pour qu'ils désignent une sorte de Comité, ayant seulement voix consultative sans pouvoir de décision, et qui serait en rapport avec la municipalité. Est-ce que cela vous paraît une bonne chose, une mauvaise chose, ou une chose sans importance ?*

Forme B : ordre inverse.

2.3.2 L'effet de cohérence : les enquêtés ont une propension à éviter de se contredire avec leurs réponses antérieures.

Ex : la liberté de la presse aux Etats-Unis (1948 et 1983)
• Question 1 : « Do you think the United States should let Communist newspaper reporters from other countries come in here and send back to their papers the news as they see it?
• Question 2: Do you think a Communist country like Russia should let American newspaper reporters come in and send back to America the news as they see it?

2.4 La construction finale du questionnaire

La passation du questionnaire est une situation de communication : il faut donc soigner la progression et l'équilibre du questionnaire :

- ✓ Gérer le **temps** (= longueur du questionnaire) d'où un mode de transmettre de question (téléphone, en personne)
- ✓ Gérer la **difficulté** des questions, d'abord créer une relation de confiance
- ✓ **Alterner** les types de formulations des questions (faciles / difficiles, affirmatives / négatives; fermées et ouvertes) : aussi non des réponses stéréotypés
- ✓ **Eviter les ruptures thématiques** : éviter de passer du coq à l'âne
- ✓ Il faut **gérer l'hétérogénéité** de l'échantillon à l'aide de **questions filtre** (pour ceux qui ont répondu à certaines questions, plus ciblées certaines réponses)
- ✓ Il faut tester la version provisoire du questionnaire **(phase de « pré-test »)**

A ce stade, on a enfin le questionnaire, il va donc falloir savoir qui est l'objet ? Comment le rendre opératoire sur le terrain ?

III. Qui et comment interroger ?

Une fois que le questionnaire est prêt, il faut encore répondre à deux questions :

- **Choisir un échantillon représentatif** (= tel que les observations pourront être généralisées à l'ensemble de la population)

- **Choisir une méthode de passation**

Notions de vocabulaire

- **Population de référence** : population sur laquelle on est à la recherche d'informations.
- **Unité d'échantillonnage** : unités élémentaires de l'échantillon (individus).
- **Unité d'analyse :** niveau sur lequel porte l'analyse.
- **Unité déclarante** : la personne qui va répondre aux questionnaires
- **Taux de sondage** : t = n / N (n : effectif de l'échantillon; N : effectif de la population de référence). C'est la part de la population de référence qu'on interroge dans l'enquête
- **Base de sondage :** liste des unités d'échantillonnage

Notion de base de sondage pose une condition préalable à l'enquête par questionnaire :

➔ **la population de référence doit être « objectivée » dans la réalité :**

➔ Cet intérêt préalable permet de **définir la base de sondage** que l'on va pouvoir utilisé dans l'enquête

Cas 1 : Pour la population totale

Base dont on dispose :

- La liste des logements en France (INSEE) = la plus fiable
- Les listes électorales
- Les annuaires téléphoniques

Donc des **bases imparfaites :**

- **Liste électorale** : que les majeurs ayant le droit de vote (pas les étrangers, les mineurs, les condamnés, les non-inscrits…) et donc **comprend 85%** de la population majeure
- **Les annuaires téléphoniques :**
  - problème des **listes rouges**
  - problème des téléphones portables :
    1992 : annuaire France Telecom = 95% population adulte métropol.
    2010 : **13% ne sont pas joignables par tel fixe**, 82% sont joignables par tel mobile
    Solution : les **<u>procédures GANT</u>** : génération aléatoires de numéros de téléphone

Les bases de sondages **doivent être soumises à la critique** :

- Quel **point de vue les « auteurs** » de la base ont-ils adopté pour sélectionner les individus ?
- Quel <u>lien</u> la base entretient <u>avec la population de référence</u> ? attention aux distorsions éventuelles

Cas 2 : Lorsque la population est plus réduite

Des questions qu'il faut d'autant plus se poser.

Plusieurs **bases** possibles : le choix dépend de la **définition** de votre **objet** d'étude.

Ex: **Une étude par questionnaire sur les écrivains (cf. Bernard Lahire**, La condition littéraire – La double vie des écrivains, 2006)
Va essayer de décrire quantitativement les écrivains en France, à qui adresser le questionnaire ? LAHIRE va se tourner vers divers bases :

<u>Trois bases possibles</u> :

- Le **recensement** : seulement ceux **dont c'est la profession principale** → pas retenu par Lahire car on peut être écrivain avec une profession principale qui est autre.

- **La base de la sécurité sociale (AGESSA) : c**aractérisée par **une définition juridique de la population**, limitation à la fraction la plus légitime du groupe → pas retenu car ceux sont qui vont vivre de leur droit d'auteur, aucun critère de reconnaissance littéraire.

Se limite à la fraction la plus légitime des écrivains (reconnus, gagent leurs vies de métiers). **Le financement de la culture publique entraîne la création d'annuaire**, qui découle de leur lien avec le conseil régional. Si pas de base régionale, une réponse biaisée pour les résultats de l'enquête ;

- Des **bases régionales** + travail complémentaire d'enquête auprès des petits salons littéraires

Il faut donc comparer ceux qui sont dans la base et ceux de l'enquête.

Deux grandes méthodes pour définir l'échantillon dans les bases de sondage : deux grandes méthodes (3.1 ou 3.2)

3.1 Le sondage aléatoire

Principe: l'échantillon est construit à partir d'une procédure **de tirage aléatoire**. Tous les individus ont la **même probabilité d'être tirés au sort.**

1) Une fois la base de sondage définie, comment procéder?

Le **sondage aléatoire simple** (SAS) est une stratégie **contraignante** : il n'y a pas de « **sosie** »

Deux inconvénients :
- La **dispersion géographique** (coût) = surtout sur le cas d'une population totale
- Le **déséquilibre dans la composition interne** de l'échantillon = des groupes absents ou sous-représenté.

Face à ces difficultés, on définit un plan de sondage.

2) Pour répondre à ces problèmes, on définit des plans de sondage.

Le but est de définir l'échantillon par étape.

Il existe trois stratégies :

Stratégie 1: Le sondage par grappe

Pour **contourner le problème de la dispersion géograph**ique. Le principe est de **tirer au sort des « grappes »** (= unités géographiques (communes, agglomérations, …) ou institutionnelles (=classe..)). On peut tirer des grappes différentes : **Si tailles différentes**, il s'agit d'un sondage **« à probabilité inégales ».**

Il y a **deux types** de plan de sondage :
- Soit on interroge **tous** les individus de la grappe **(sondage à un degré)**
- Soit **on tire au sort** les individus à l'intérieur de la grappe **(sondage à deux degrés)**

Ex : une enquête conséquente d'un sondage à un degré : Enquête emploi (INSEE, trim.): aires de 20 logements, 1/6 est modifié chaque trimestre

Pourquoi est-ce pertinent ?

Les **avantages** :

- Permet de **résoudre les problèmes d'absence ou de défauts de la base de sondage** = conserver le principe de **représentativité**
- Sur la propension de réponse : permet de **réduire les refus de répondre**

Un **inconvénient** majeur : **les « effets de grappe »** = personnes ont une propension à se **ressembler**, il faut donc **veiller à l'hétérogénéité** des grappes (ex : sélectionner des grappes de **petites tailles comme le fait l'Insee)**.

Stratégie 2 : L'échantillon stratifié

Répond au problème de composition de l'échantillon, on souhaite que certain sous-groupes soient présents.

Principe :

- On procède à **une stratification de la population** : *besoin de la liste des individus et de leurs caractéristiques respectives pour constituer des groupes.*
- On procède **ensuite au tirage au sort des individus dans chaque strate**.

Ex : enquête sur les pratiques culturelles des Français en activité : N = 20M, n= 1000, taux de sondage t = 1/20000

Implique un taux de sondage plus importants pour les catégories les plus petites. Attention : dernier échantillon n'est pas représentatif, appliquer un CM.

Exemple fictif		Objectif 1		Objectif 2	
	Population totale	Taux de sondage	Effectif dans l'échantillon	Taux de sondage	Effectif dans l'échantillon
Professeurs et professions scientifiques	550.000	$\frac{1}{20000}$	27	$\frac{1}{5000}$	108
Cadres et professions intermédiaires de la fonction publique	700.000	$\frac{1}{20000}$	35	$\frac{1}{5000}$	140
Professions non libérales de santé	700.000	$\frac{1}{20000}$	35	$\frac{1}{5000}$	140
…		$\frac{1}{20000}$		$\frac{1}{20000}$	
Ouvriers	7.400.000	$\frac{1}{20000}$	370	$\frac{1}{20000}$	370
…		$\frac{1}{20000}$		$\frac{1}{20000}$	

Stratégie 3 : Le sondage en plusieurs phases

Quand on ne peut définir de base de sondage des individus recherchés

Principe :

- On **procède à une première enquête** (plus simple) auprès d'un **très large échantillon pour récolter l'information minimale recherchée que l'on recherche** (ex. état de santé)
- On procède **ensuite à cibler un 2e échantillon** plus réduit **en fonction de l'information** de la phase précédente

Ex 1 : enquête de l'INSEE : Vie quotidienne et santé (VQS) et Handicap et santé (HS) :

- Phase 1 100000 personnes, par voie postale
- Phase 2 : 20000 sélectionnées dans l'échantillon précédent, en face-à-face au domicile

Exemple lié aux imperfections des bases disponibles

Ex. 2: l'enquête Trajectoire et origines (INSEE-INED 2008) : enquête sur la diversité de la population en France. On n'a pas la possibilité d'effectuer des bases de sondage sur des personnes de deuxième génération. Comment faire donc pour étudier cette population ? Utilise une stratégie plus complexe :

- Une stratification en 5 sous-échantillons aux taux de sondages différents

	Echantillon final	Base de sondage
Immigrés	8 300	Recensement 2007
Descendants d'immigré(s)	8 200	Abs.
Originaires des Dom	700	Recensement 2007
Descendants d'originaire(s) des Dom	700	Abs
Population « majoritaire »	3 903	Recensement 2007
Total	21 803	

- Pour les descendants, utilisation du fichier nominatif du recensement et des bulletins de naissance anonymes

Principe illustré dans la dernière diapositive du diaporama MMS – 2020-2021 – Chapitre 1§ 3.1

3.2 Le sondage empirique ou méthode « à choix raisonnés »

C'est une méthode de construction qui est utilisée quand on n'a pas de base de sondage directe ou dérivée. Deux cas de figures principaux :

➢ Dans le premier on cherche à garder une sorte de tirage au sort : **l'échantillonnage sur place**

Principe : certaines personnes se **concentrent dans des endroits particuliers** (fumeurs aux bureaux de tabac, etc.). On procède à un **tirage au sort parmi les personnes présentes.**

Plusieurs risques de biais : donc *correction avec des tirages au sort dans les listes de lieux de passation, des moments donnés*

- **Problème du choix du lieu** : échantillonnage **spatial**
- Problème du **choix du moment** : échantillonnage **temporel**
- Problème **des gros utilisateurs** : la **probabilité** d'être tiré au sort n'est **pas égale**. *(difficile à corriger, pourquoi ne pas corriger en posant une question)*

Ex. **L'enquête « Sans domicile 2012 »** de l'INSEE (janv.-mars 2012, cf. Economie et statistique, sept. 2016, n° 488-489)

- Un tirage à trois degrés:
  - tirage des **agglomérations** : 80 agglomérations de plus de 20 000 habitants
  - tirage des **visites** : services x jours :

Constitution d'une base de sondage des services existants dans les 80 agglomérations

+ enquête téléphonique pour obtenir les informations nécessaire au tirage des visites

- tirage des **usagers** sur place le jour de la visite : 4 en face-à-face (francophones: n = 4419) et 12 au max. en auto-administré (non-francophones, 14 langues, n = 1500)

Echantillon final: 4419 en face-à-face et 1500 en auto-administré

- Une enquête qui ne couvre que la **fraction la plus intégrée des SDF**

➢ **La méthode des quotas** : on sort du tirage au sort, principe de modèle réduit

Principe : la **représentativité** de l'échantillon ne découle **pas du hasard** = modèle réduit de la population de référence.

L'idée est que les **variables attachées à l'individu ne sont pas indépendantes entre elles.**

Si l'échantillon est représentatif du point de vue des grandes variables, on le sera aussi du point de vue des grandes variables que l'on cherche à déterminer. **Il n'est pas représentatif en soi mais représentatif selon les critères que l'on a choisi.**

Deux étapes :

1) Le choix des quotas

Ex : sondage pré-électoral lors d'une élection municipale

Sexe		Age		PCS	
Homme	**48 172**	18-24 ans	**16 658**	Artisans, commerçants, ch. d'entreprise	**2 026**
Femme	**51 874**	25-44 ans	**32 519**	Cadres, professions intellectuelles supérieures	**5 025**
		45-64 ans	**29 745**	Professions intermédiaires, employés	**19 412**
		65 ans et plus	**21 124**	Ouvriers	**18 583**
				Retraités	**18 458**
				Elèves, étudiants	**11 410**
				Autres personnes sans activité professionnelle	**20 132**
Total	**100 046**		**100 046**		**100 046**

Dans certains cas, on peut procéder à un échantillonnage par degrés (plan de sondage) :

- Une « **stratification » géographique** (par catégorie d'agglomérations et « région »)
- Puis pour chaque strate**, la définition d'un quota en fonction de variables individuelles** (âge, sexe, PCS) confié à des enquêteurs

2) L'établissement de la feuille de route

Elle est établit par des enquêteurs qui vont sur le terrain.

Feuille de route

Nom de l'enquêteur...*J. Martin*..... Semaine du.....*3 au 7 mars*.....
Ville..... *Montpellier*..... Nombre d'interviews .....*10*.....

1 *Sexe*							
Féminin	1	2	3	4	5	**6**	7
Masculin	1	2	3	**4**	5	6	7
2 *Classe d'âge*							
18 à 24 ans	1	**2**	3	4	5	6	7
25 à 39 ans	1	2	**3**	4	5	6	7
40 à 54 ans	1	2	**3**	4	5	6	7
55 ans et plus	1	**2**	3	4	5	6	7
3 *PCS de la personne interrogée*							
Artisans, com., chefs d'entr.	**1**	2	3	4	5	6	7
Cadres, prof.intel.sup.	**1**	2	3	4	5	6	7
Prof. interm., employés	1	**2**	3	4	5	6	7
Ouvriers	1	**2**	3	4	5	6	7
Retraités	**1**	2	3	4	5	6	7
Élèves, étudiants	**1**	2	3	4	5	6	7
Autres inactifs	1	**2**	3	4	5	6	7

Contrairement à la méthode aléatoire, **on a le droit de prendre des sosies. Si la personne ne souhaite pas répondre, on peut trouver une autre personne qui répond aux critères de la feuille de route**.

ATTENTION : La méthode des quotas fait courir un risque important de biais:

- **Risque de manipulation** de l'échantillon par l'enquêteur lié à la définition des quotas **(« queue d'enquête »)** = change un certain nombre de caractéristiques **pour répondre aux délais de l'entreprise de sondage** (→ cf. texte du TD 1) surtout pour les enquêteurs des instituts de sondage privé qui sont soumis à moins de contrôle. (à la différence de l'Insee).
  Le bidonnage des enquêtes dans les instituts de sondage → plus on **multiplie les catégories**, plus on a de bidonnage.

Vous arrive-t-il de	Oui	Non, jamais	Ens.
Reformuler une question	86	14	100
Faire répondre des membres d'un panel personnel ou de votre entourage	74	26	100
Modifier l'âge d'une personne	52	48	100
Escamoter une liste d'items ou une question	51	49	100
Interpréter une réponse	45	55	100
Modifier les habitudes de consommation	41	59	100
Remplir entièrement un questionnaire sans répondant	21	79	100

Source : Rémy Caveng, « La production des enquêtes quantitatives », *Revue d'anthropologie des connaissances*, 2012

- **« Biais de sélection »** en raison du principe de substitution des individus refusant de répondre (« **sosie** » car pas contraint à liste nominal de personne à interroger, caractéristique principale de la méthode par quotas).

    - La méthode des quotas est **aveugle aux distorsions internes** (clivages internes ne vont pas être pris en compte par la méthode) :

        - La catégorie socio-professionnelle (exemple : pas de distinction ouvrier qualifiée ou non qualifiée)
        - Le diplôme
        - L'âge

    - Dans les passations **auto-administrées** (internet) depuis les années 2000 : opinion way, elab…: problème de l'access panel : **distorsion en fonction de l'équipement dont on dispose et des professionnel du sondage. L'échantillon est fait à partir d'une base de l'intermédiaire access panel qui est biaisée.**

Conclusion : quelle est la meilleure méthode d'échantillonnage ?

Le sondage aléatoire ? Oui mais il connaît également des limites liés à ces caractéristiques :

- ne peut pas **interroger de sosies donc des relances : problème en termes d'homogénéité temporelle** de l'enquête (relances). On va **prévoir une liste d'échantillon plus large** pour anticiper ses refus.
- Surtout, la méthode est très sensible au refus de répondre : **les refus s'accroissent**
  Ex : l'enquête Trajectoires et Origines (INSEE-INED, France, 2008) : Plus d'un tiers de refus de participé par rapport à 15% au début des sondages). 1 individu sur 2 refuse donc seulement 85% de la cible est atteinte malgré une anticipation des refus

Les études qui comparent échantillons aléatoires et par quotas montrent que **les quotas peuvent donner des résultats moins biaisés.**

Ex : échantillon français de l'enquête European Value Survey, 2008 : 3000 questionnaires, dont 1500 par quota et 1500 en aléatoire. Une distorsion plus forte sur la part de quotas aléatoires.

Des solutions ? Combiner les deux méthodes (aléatoire et quotas renforcés):

- **Au téléphone** : **génération aléatoire** des numéros de téléphone
- **En face-à-face** : **méthode des itinéraires ou méthode de Politz** (tirage au sort de quartiers et de routes aléatoires, et la feuille de route se fait sur quotas)
  Ex : eurobaromètres, European Social Survey, etc…

- **Passation en ligne :** en auto-administré (internet) : la constitution d'un **access panel à partir d'une enquête en face-à-face ou par sélection aléatoire** (base de sondage téléphonique) + équipement internet postérieur

3.3 L'administration des questionnaires = passation

Le choix des modalités d'administration : 3 stratégies :

- **Le face-à-face**
- **Le téléphone**
- **Le questionnaire auto-administré (papier ou internet**)

Ce choix des méthodes s'opère en fonction **de deux critères :**

1) **l'influence sur la qualité de l'échantillon** (sa représentativité, qu'elle n'est peu de biais) : arbitrer entre le coût de l'enquête et l'amélioration de la représentativité

Sur population totale : la méthode de **face à face** est la meilleure.
Pourquoi ?
Téléphone : limiter des questionnaires peu complexe, **problème de la durée** et sur l'effet de l'accessibilité de l'enquête
Auto-administré : taux de non-retour très élevé (rarement plus de 40% des personnes qui répondent), **incertitude sur l'identité** du répondant. Surtout, **non-respect des principes de représentativité** pour les passations sur internet.

Le **face à face est beaucoup plus fiable mais beaucoup plus coûteux, peut réduire le taux de non-réponse, jouer sur l'officialité de la procédure de l'INSEE. On peut aussi mener une enquête sur l'enquête.**

Principe illustré dans l'avant dernière diapositive du diaporama MMS – 2020-2021 – Chapitre 1§ 3.2 3.3

Sur population restreinte : l'auto-administré obtient de bons retours, face à face n'est plus la seule méthode.
Ex : Enquête sur les écrivains (Lahire, 2006) : 66% de répondants

2) **l'influence sur la qualité des réponses**

Le face-à-face est **le plus sensible aux « effets d'enquêteur», est donc la moins efficace**. Donner la meilleure image de soi-même donc **un effet de légitimité.**

Effet moindre sur les enquêtes par internet : recours au questionnaire auto-administré surtout pour les enquêtes sur **données sensibles** (enquête de « délinquance autorévélée », etc.) → **anonymat** préservé au maximum donc plus de réponse que si un enquêteur en face de l'individu.

Plan du chapitre 2 : Analyse des données quantitatives (partie 1)

1. La critique scientifique des données quantitatives 22
   1.1. La « critique d'exactitude » : Méthodes quantitatives et risque d'artefact 22
      1.1.1. Analyser les effets du protocole d'enquête : le questionnaire 22
      1.1.2. L'analyse secondaire des statistiques publiques 22
      1.1.3. Que faire avec les risques d'artefact ? 25
   1.2. La construction des variables 26
      1.2.1. Les types de variables 26
      1.2.2. La mise en forme quantitative des données 26
      1.2.3. Passer des indicateurs aux variables : les variables synthétiques 27
2. Les instruments de l'analyse quantitative 30
   2.1. Décrire : L'analyse univariée → polycopiée !!! 30
      2.1.1. Les paramètres de tendance centrale 30
      2.1.2. Les paramètres de dispersion 30
   2.2. Analyser les relations entre variables 30
      2.2.1. Le lien entre deux variables qualitatives 31
      2.2.2. L'analyse statistique des corrélations (variables quantitatives) 32
      2.2.3. De la (cor-)rélation à la causalité : introduction à l'analyse multivariée 34
   2.3. Modéliser la relation entre variables : l'analyse par régression 35
   2.4. Les tests statistiques : le test Khi-A d'indépendance

Partie 1 : Les méthodes quantitatives en sciences sociales

Chapitre 2 : L'analyse des données quantitatives

Principale difficulté : construire les données pour en faire qqch de pertinent par rapport à l'objet choisi
Il faut **donc procéder à une critique scientifique des résultats que l'on a obtenu**. Il y a donc une étape d'analyse précise des résultats que l'on avait anticipés.

On peut aussi analyser des données produites par d'autres acteurs : des données de seconde main (plus en plus de ce type de données avec le **développement des BIG DATA** mais produite avec une logique spécifique qui n'est pas celle de la recherche) :

Les analyses de secondes mains sont ….

… **issues** de **fichiers administratifs**
… issues des **recensements de la population**
… **recueillies par le moyen de questionnaires sur échantillons**
… données **prosopographiques** (= des méthodes utilisées en histoire, des archives permettent de rechercher des informations biographiques)
… données **textuelles** (ex = comparer les programmes des partis politiques, repérer quels termes sont souvent utilisés = corpus de textes et étude sur l'occurrence des mots…)

Toutes ces données vont **reposer sur un processus d'enregistrement que l'on ne contrôle pas donc elles ne sont pas immédiatement rigoureuses du point de vue de notre recherche**. Ces données brutes vont devoir passer à une **analyse de leurs conditions de productions pour évaluer les limites de leurs validités en particulier**. C'est ici qu'intervient **l'analyse critique des données** :

- **Double critique (François Simiand**, élève de Durkheim, « Méthode historique et sciences sociales », 1901) :
  - ➔ Critique **d'exactitude** : tournée vers **le rapport entre les données et la réalité**, y-a-t-il un effet du protocole d'enquête sur les résultats ?
  - ➔ Critique **d'aptitude** : **pertinence** de ses données par rapport aux **buts du chercheur :** cherche à répondre à une question qui est définit dans la problématique

1. La critique scientifique des données quantitatives

1.1. La « critique d'exactitude » : Méthodes quantitatives et risque d'artefact

Il faut réfléchir sur la production, de quelle manière elles ont été construites ? Risque important d'artefact donc il faut se prémunir : variations sont dû aux instruments de mesures et non aux variations.

1.1.1. Analyser les effets du protocole d'enquête : le questionnaire

D'autant plus important quand ce sont des enquêtes de seconde main.

- **quels effets du protocole d'enquête sur les résultats obtenus**?

Ex : **les enquêtes sur les conditions de travail** de l'INSEE / Min. travail : 1978, 1984, 1991 (M. Gollac, « Des chiffres insensés ? Pourquoi et comment on donne un sens aux données statistiques », 1997)

Problème : résultats inattendus par rapport aux années 1978 et 1984. En **1991**, on constate une **forte dégradation inattendue des conditions de travail** (fait de porter des charges lourdes : 21% à 31%)

Quelles **causes possibles** ? Se demander s'il n'y pas un **risque d'artefact** :

- Une **erreur informatique**? Pas le cas
- Les **modifications du questionnaire**? Formulation, une modification : travaux peu autonomes, attention flottante dans les deux premières questions. Nouvelle question en 1991 : concerne tous les salariés donc les enquêtés qu'ils n'avaient pas répondu, donnent un avis sur les charges lourdes.
- **Un changement des conditions de passation** ? relation qui change entre enquêtés et enquêteurs. Les responsables des enquêtes ont remarqué que les enquêteurs reformulent ou ne pose pas certaines questions (cadres, on ne lui pose pas les questions du W à la chaîne). Sociologie des enquêteurs de terrain a changé entre les trois enquêtes : les hommes fonctionnaires et retraités → enquêtrices, employé dans les bureaux : on peut s'attendre à ce qu'elles ne décident pas de la même manière de ce qui est à poser ou non dans l'échange.

Comment **départager**? Il faut pouvoir mesurer si ces effets sont effectivement produits

- Introduire dans l'enquête une **analyse des effets de formulation**
- Recourir aux **méthodes qualitatives** (observations par ex)
- Faire une **analyse quantitative détaillée des réponses** : solution *systématique* : grâce aux anticipations faîtes. Enquêtés par concernés = comparer les réponses, qui a déclaré une dégradation.

1.1.2. L'analyse secondaire des statistiques publiques

Risque d'artefact encore plus important en particulier car l'enregistrement statistique utilisé implique un grand nombre d'acteurs.

- les modalités de l'enregistrement statistique ont-elles des effets propres? L'enregistrement statistique = une chaîne statistique

  Deux acteurs statistiques :

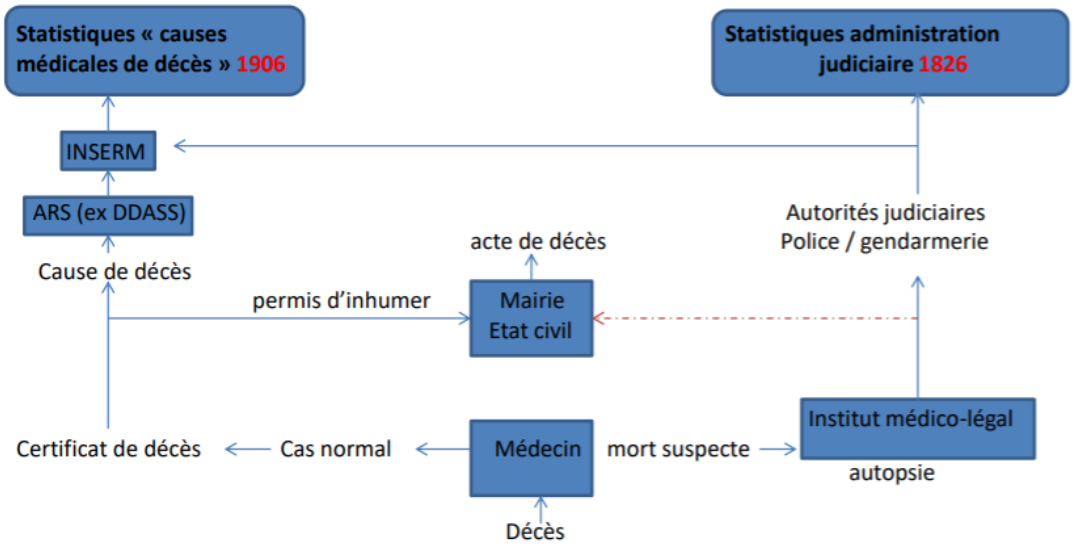
A) **les administrations :** chiffres produits sont des activités secondaires de l'administration donc un fort risque que les données produites n'ont pas pour finalité la connaissance
B) **instituts spécialisés dans la production d'enquête :** grande transparence des écritures donc normalement prémunis des effets d'artefact. Leur fonctionnement (de Insse, ined) peut avoir des effets.

Deux exemples pour illustrer des critiques des aptitudes de la statistique publique

➔ **Les dynamiques institutionnelles de l'enregistrement statistique**

1ère précaution : il faut tenir compte des **structures administratives de la chaîne statistique**

EX1. : Baudelot et Establet, Durkheïm et le suicide (1984)



Que montre cette étude : il compare les deux sources fin des 1970's : dans certaines villes, stat. judiciaires > « causes médicale de décès » Cause? Des différences bureaucratiques entre police et gendarmerie. En toute logique, les statistiques de l'administration judiciaire devraient être contenues dans les statistiques des « causes médicales de décès » (géré par l'INSERM qui fait des tables sur les causes de décès donc aussi sur le suicide). Or ce n'est pas le cas : on a plus de suicidés « judiciaires » que « médicalement approuvés ». C'est **donc un problème de classement**

Individu est classé mais pas au bon endroit, extérieur à l'administration ??

Ex 2 : les statistiques sur les « violences urbaines » de la section « Villes et banlieues » des Renseignements généraux dans les années 1990

Une **nouvelle catégorie** pour détecter les **lieux** pouvant donner naissance à une **émeute** (émeutes urbaines 1990-91; attentats du GIA en 1995-96). Un travail de **communication** pour la diffuser dans le grand **public**.

Sur **5 ans** à peine, une très forte croissance des phénomènes. En réalité, **elle s'explique par une forte croissance du nombre de quartier classés « sensibles » et un effet d'apprentissage** (les communes apprennent à « se servir de la

catégorie »). Il faut donc faire attention **aux effets de mise en place d'un dispositif sur le dénombrement d'un phénomène.**

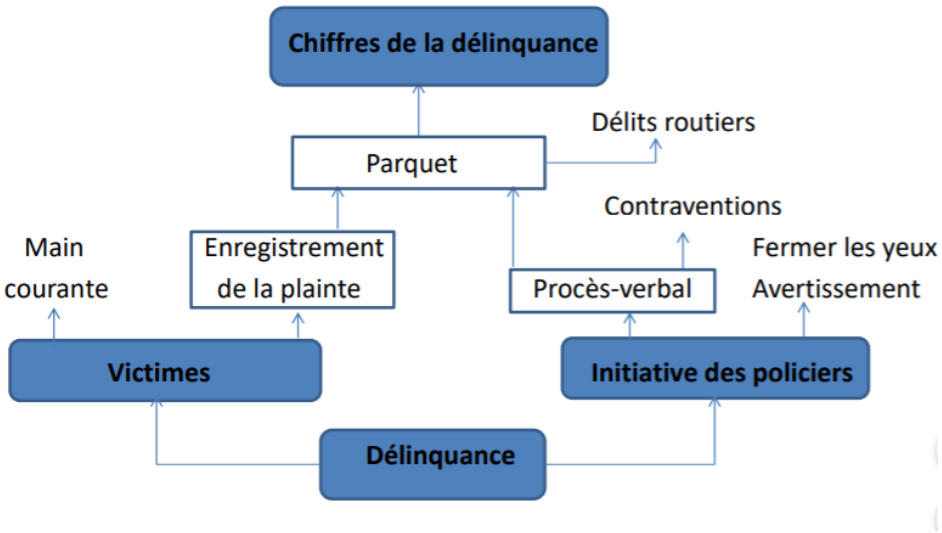
➔ Le comportement des acteurs de la chaîne statistique

2ème précaution : il faut tenir compte **des conditions de déclaration et d'identification des phénomènes** :

- **comment les faits recensés sont-ils portés à la connaissance de l'institution** ? Qui lui apporte l'information
- Comment sont-ils **enregistrés pour être intégrés aux statistiques**? Comment les informations vont être intégrées à la statistique ?

Exemple : l'enregistrement des statistiques de la délinquance (système « 4001 ») = recense les procès-verbaux transmis au ministère public (« faits constatés », « faits élucidés » = ce pour qui on a trouvé le suspect et « personnes mises en causes »). On exclut les contraventions (violences interpersonnelles de niveau réduit), la délinquance en col blanc (fisc), les affaires de circulation.

Il y a ainsi deux voies d'accès aux chiffres de la délinquance : soit ils passent par l'initiative des victimes (dépôt de plainte...) soit par l'initiative de la police (procès-verbal...). Néanmoins, il peut y avoir une partie qui peut ne pas être déclarée à la police. C'est donc le comportement des victimes qui explique que tous les faits ne soient pas enregistrés : ce comportement varie dans le temps et varie selon le type de délit (taux de plainte variant et différence de qualification pénale). Il y a donc un risque de biais dans la composition interne des types d'infraction.
Biais : effet statistique qui conduit à une distorsion dans la composition de la population



Enquête par questionnaires pour avoir une troisième voie d'entrée.

Les données qui sont issues du seul constat policier font courir un risque important d'artefact (logiques administratives : moyens, priorités, modes d'évaluation...). Il y a également un risque important de biais dans la composition des « personnes mises en cause » (lié au taux d'élucidation et lié à l'attitude des agents à l'égard des infractions et délits)

➔ La définition administrative des catégories

Le problème posé par les statistiques ne se limite pas à la question du dénombrement des phénomènes (les chiffres sont-ils « vrais » ?).

Il va falloir définir des catégories pour classer les objets. L'opération statistique consiste également à classer : une fois constatés**, les faits sont « labellisés** ».

EX. : l'évolution de la violence interpersonnelle en France (Lagrange, 1995). On constate a priori une hausse de plus de 50% des homicides et tentatives d'homicides. Si on y ajoute les infanticides et les coupes et blessures

volontaires ayant entraîné la mort sans intention de la donner, l'augmentation est de seulement 5,6%. Si on ramène les catégories 1 et 3 (sans infanticides donc) à la population totale, on observe une baisse d'environ 9%.

La **troisième précaution** à prendre dans l'analyse des statistiques publiques : déconstruire les catégories administratives en considérant l'acte de classement des agents de terrain eux-mêmes : OBJETCIF : **déconstruire la définition des catégories** et les **reconstruire pour les adapter à l'objet** qui est celui du travail.

- **Quel est le sens des catégories employées**? Est-il stable?
- Peut-on supposer une **homogénéité dans l'utilisation des critères** de classement?

1.1.3. Que faire avec les risques d'artefact ?

Un **premier type de réponse est de redresser ou corriger l'erreur** : l'idée est qu'existe un phénomène réel, un **effectif « réel », dont il est possible de s'approcher**.

Pour ce faire, on dispose de **2 stratégies** :

1°) **Le redressement statistique** : on peut par exemple **calculer et appliquer des coefficients correcteurs** pour redresser les données des échantillons.

Dans les cas des **violences** ou des **suicides** :

- **Dépouillement** des sources judiciaires et des causes médicales de décès (doubles comptages)
- **Calcul** de coefficients de correction qui dépendent des grandes caractéristiques (sexe, âge….)
- **Vérification** dans des localités test (comparaison prédiction à un dépouillement des sources réelles) : coûteuse et rarement faite.

2°) **le recours à des enquêtes complémentaires**

- Pour la délinquance : **il s'agit des enquêtes de « victimation » et de « délinquance autoreportée** » (questionnaires). Par exemple les victimization studies au Royaume-Uni qui ont supplanté les données administratives. Cela **sert de référence pour les politiques** britanniques, une tradition beaucoup plus importante.
- Les enquêtes **d'ethnographie quantitative** (observation). Par exemple : enquête à Paris sur les contrôles d'identité

L'erreur de comptage peut aussi être **une source de comptage**. L'analyse de l'activité d'enregistrement statistique permet aussi de montrer comment les institutions contribuent à la construction sociale de l'objet de l'analyse (ex. : Jean-Claude Chamboredon, « La délinquance juvénile », 1971 Point de départ : une sociologie de l'enregistrement statistique de la délinquance : Chercher les acteurs qui prédispose à enfreindre la règle). La mesure statistique n'est pas une mesure neutre, enregistrer c'est donner sens et en quelques sortes le produire. Question différente de recherche sur la délinquance. On s'intéresse au processus qui font que les individus sont déviants, extérieurs au groupe. En France, problématique importée dans les années 1970.

On **décentre alors l'analyse vers le rôle joué par les institutions dans la constitution même du délinquant**

- **labellisation** comme délinquant : réinterpréter la biographie de l'individu en fonction de la déviance
- **effet** de **prophétie auto-réalisatrice** : effet sur la manière dont il se définit lui-même

En fonction de son origine sociale, il aura plus ou moins d'opposer un discours cohérents aux institutions. Et le faire sortir de sa position de délinquant.

1.2. La construction des variables

1.2.1. Les types de variables

Les variables ne **s'imposent pas naturellement** et **sont à cheval entre l'empirique et le théorique**. Il va falloir **construire** et définir les variables. Elles **permettent de mesurer un phénomène ou une partie du phénomène en le replaçant dans un cadre théorique**. Cela peut être fait en amont ou en aval de l'enquête. **Le travail du chercheur n'a pas toujours la même influence sur l'objet :** certaines variables sont construites par le social quand d'autres sont construites par les sciences sociales.

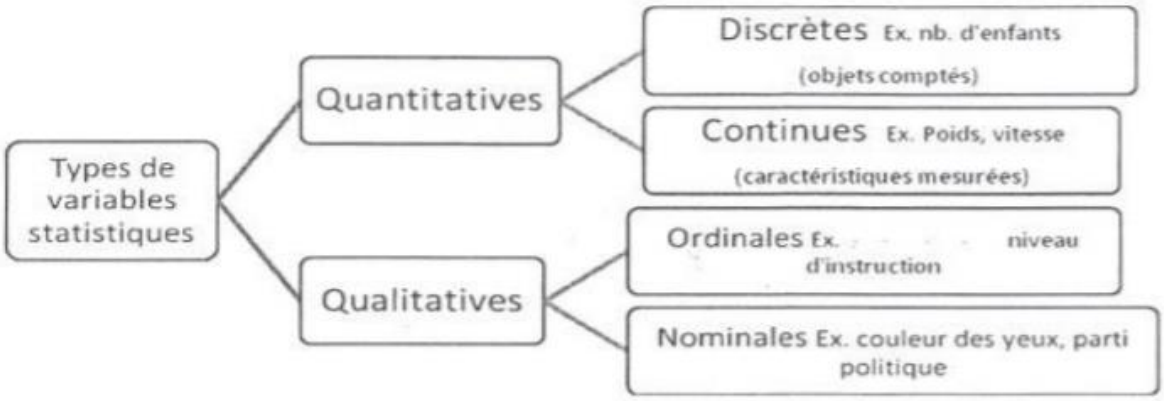
- **Variables « construites par le social » / variables « construites par les sciences sociales » (Olivier Giraud** in Apprendre la sociologie par l'exemple)

Dans le cas de variables construites par le **social**, elle **exprime des catégories ou des caractéristiques qui sont socialement et juridiquement déjà définies et déjà codées.** Il est « **impossible** » pour le chercheur de les **réinventer**. Elles peuvent être **recodées** pour **affiner leur perception.**

Dans le cas des variables construites par les **sciences sociales**, leur construction **passe par une décomposition des indicateurs et une reconstruction des variables pour les reprendre dans leur complexité**. La définition ne s'impose pas aux chercheurs.

- **Variables quantitatives / variables qualitatives**

On distingue aussi les variables quantitatives et les variables qualitatives. Les **quantitatives** peuvent être discrètes (âge) ou continues (poids, vitesses) = prennent **des valeurs numérique**. Les **qualitatives** peuvent être ordinales (niveau d'instruction) ou nominales (partis politique d'appartenance) = prennent **des valeurs littéraires.**



1.2.2. La mise en forme quantitative des données

Etape 1: La construction de la base de données.

Pour le questionnaire, il s'agit de **mettre les réponses sous une forme homogène** (codage).

Pour les questions fermées : certaines réponses **obligent à faire de nouveaux choix théoriques**.

Pour les questions ouvertes, il faut procéder à un **inventaire des réponses pour définir les modalités de la variable**.

Recoder car chacun répond à sa sauce, il va falloir faire l'inventaire de faire des réponses d'un échantillon et à partir de cela, on **fait un travail de catégorisation, on établit des regroupements**.

Ex : enquête par **questionnaire auto-administré auprès des instituteurs en formation** : à partir de leur mot, reconstruire les qualités qu'il est nécessaire d'avoir chez un futur enseignant. Rapprocher les descriptions qui se ressemblent le plus. Difficulté : des personnes répondent avec une qualité et d'autres plusieurs.

- Qualités **personnelles** de l'instituteur: Adaptation à la situation ; Capacité d'autocritique ; Écoute et contact des élèves ; Motivation ; Autres qualités personnelles
- Qualités **professionnelles** : Pédagogue, qui instruit, qui donne des connaissances ; Éducateur, qui contribue à l'éducation de l'enfant ; Autres fonctions actives

Enjeux : créer une catégorisation

Deux remarques :

1) Le codage est une **opération de catégorisation.**

Elle a une **double logique** (bien dans une traduction de la réalité avec des outils quantitatifs) :

- **Regrouper ce qui se ressemble**
- **Séparer ce qui est différent**

Etablir une frontière entre les groupes, c'est la condition pour avoir **une variable pertinente**.

2) **Le codage nécessairement réduit** la réalité : simplification (voire un appauvrissement)

Une fois que l'on a codé, que l'on a traduit nos réponses, il va falloir décrire la base.

Etape 2 : Décrire la base de données (tri à plat)

Pour chaque question, il faut **voir comment se répartisse les réponses**.

Important car première prise de connaissance après la passation. Il va **falloir réajuster certaines des anticipations que l'on a pu avoir.**

Distinguer deux types de variables :

- **Variables primaires** : **chaque question correspond à une variable**
  Ex : pour les variables ordinales (échelles d'accord) ou quantitative :
  - Prudence tactique
  - Réponse polarisée
  - Réponse unanimiste (accord ou désaccord)

Pas une question uniquement technique : **idée de la traduction du réel dans un langage mathématique** (agrégat statistique)

Alors que les **variables secondaires : ce que l'on va faire à partir de ces variables**.

1.2.3. Passer des indicateurs aux variables : les variables synthétiques

Moment important pour **comprendre les risques d'intérêt de l'analyse quantitative**

Comment est-ce que l'on regroupe ? Comment on associe ses différents indicateurs ?

Deux grandes stratégies (cf. deux -)

a) Les indices énumératifs :

On traduit un phénomène qualitatif en une mesure quantifié. On définit **une variable unidimensionnelle** (du moins au plus) en termes d'intensité.

Principe : on calcule **des scores pour les indicateurs** retenu. **L'indice est construit à partir de la somme** des scores. Mais **comment on définit les scores** ? Comment on les ajoute ?

Méthode 1 : **Construction par dénombrement** (indicateurs **dichotomisés**)

On va donner **deux valeurs à l'indicateur**, on simplifie : oui-non // 1-0

Ex : La variable « échelle de pratiques religieuses » (Michelat, 1990) en 9 modalités, réduites à 5 par regroupement

Document Michelat - indicateur énumératif : cf. MOODLE

On compte le nombre de fois où l'on obtient 1, indice de 0 au maximum de 8. On voit bien la difficulté : score de deux : pourquoi ? Messe une fois par mois, prier de temps en temps / baptisé et souhaite être enterré religieusement. Du pdv de cette variable, score faible par rapport à la pratique. **Indicateurs ont la même valeur par rapport à la pratique religieuse.**

➔ **Simplifie bcp l'information mais on ne garde pas la nuance**

Méthode 2 : **construction par combinaison**

Les indicateurs peuvent **prendre plus de deux valeurs, mais elles sont ordonnées (ordre d'intensité entre les valeurs)** (=condition). Ajouter les scores pour construire un indicateur.

Ex : mesure de l'intensité de l'intérêt pour la politique : dépend de l'information et de la participation électorale. 3 manœuvres que l'on ajoute. **Page 9 du diaporama**

Selon les hypothèses sur le poids de chaque indicateur (pondération), vous ne c**onstruirez pas la variable de la même manière**. Il faut donc **s'assurer de la robustesse de l'indice** (quand on fait varier de manière très faible la variable, il ne faut pas qu'il y est un changement sur le cpt de la personne).

Plus il y a d'indicateurs, plus il y a de valeurs (mais **moins il est possible de mener l'analyse** en croisant plusieurs variables entre elles), **d'où la nécessité de réduire le nombre de modalités de l'indice** (pour plus de praticité et de lisibilité)

Comment définir les pondérations ? Attention de ne pas voir les présupposés des enquêteurs.

b) Les indices typologiques

Comme les **indices énumératifs ne sont pas adaptés à toutes les variables**, il faut trouver une autre stratégie.

Exemple : statut social en fonction du revenu et prestige. Suivre une stratégie énumérative : méthode de combinaison. Diagonale : voit une difficulté : on mêle des groupes différents au même niveau → **pas dans le principe de catégorisation**

Donc recourir aux **indices typologiques : permettre de construire les groupes qui se différencient selon différentes dimensions.** Chaque indicateur est **dichotomisé**, mais les types ne sont pas ordonnés

- Croiser des **dimensions ordinales** :

Ici : savoir si le revenu est faible ou élevé, de même pour le prestige.

		Revenu	
		Faible	Elevé
Prestige	Faible	- -	- +
	Elevé	+ -	+ +

			Revenu			
			-		+	
			0	1	2	3
Prestige	-	0	0	1	2	3
		1	1	2	3	4
	+	2	2	3	4	5
		3	3	4	5	6

Groupe qui se distingue par son statut social faible, un groupe pour un prestige social élevé ; deux groupes intermédiaires : un pour le prestige, un pour le revenu.

Rapproche ce qui se ressemble

- Croiser des **dimensions nominales**

Ex 1 : les questions « quelle est votre religion » et « êtes-vous pratiquant » peuvent être regroupées.

Ex 2: variable âge X sexe

Ex 3: **L'intégration professionnelle (Serge Paugam)** : la transformation des formes d'intégration liée à la transformation du marché du travail (précarisation, dualisation).

Enjeux 1 : Serge Paugam **s'interroge sur la manière de construire sa variable.** Réflexion théorique qui l'amène à distinguer deux grandes dimensions : **reconnaissance dans le travail qui passe par la satisfaction / accès à des droits qui passe par l'emploi stable.**

Enjeux 2 : **établir la dichotomie** : à partir de quel niveau on est satisfait ou non ? à partir de quel niveau on est intégré ou non ?

Ce qui est important : **type qu'il distingue = on voit une cohérence entre eux**

1. Satisfaction + CDI (« intégration assurée »)
2. Satisfaction + CDD (« intégration incertaine »)
3. CDI + Insatisfaction (« intégration laborieuse »)
4. CDD + Insatisfaction (« intégration disqualifiant ») → Précariat depuis les années 90

Sert à illustrer le raisonnement typologique : réflexion théorique que l'on traduit par un questionnaire

Prend une batterie d'indicateurs pour mesurer la satisfaction dans le travail (=variable 1)
Variables directes (promotion, salaire, relations avec les collègues, …) • Variables indirectes portant sur l'entreprise (bonne organisation du travail, ambiance…) • Variables sur la trajectoire de l'entreprise (en constante évolution, bonnes possibilités d'améliorer sa formation), • Variables sur les gênes quotidiennes au travail (bruit, rythme, efforts physiques, isolement…)

Stabilité dans l'emploi (=variable 2) : réponse à la question : « pensez-vous que dans les deux prochaines années, il y a un risque que vous soyez licenciés ? » perception très corrélé avec des perceptions subjectives de l'entreprise.

	Adhésion syndicale	Confiance dans la représentation du personnel	Sentiment d'être inutile	Attitude « révolutionnaire »*
Intégration assurée	24,60%	64,50%	3,30%	7,50%
Intégration incertaine	21%	62,20%	4,40%	9,80%
Intégration laborieuse	36,20%	60%	8,20%	12,70%
Intégration disqualifiante	30,40%	49,50%	11,40%	14,40%

* Accord avec l'opinion : « il faut changer radicalement toute l'organisation de la société par une action révolutionnaire »

→ Analyse causale qui regarde les attitudes des salariés **cf diapositive 15 du MOODLE**

Paugam a conservé le maximum de nuance de la réalité tout en adaptant le nombre de variables pour le traitement qui vient par la suite

Pour **les indices typologiques, plusieurs stratégies de regroupement** (P. Lazarsfeld, « Sur les techniques de construction des typologies en sciences sociales », 1970)

Méthode 1 : La **réduction par simplification de dimensions**

Avant une distinction vue comme non nécessaire donc des réductions

Méthode 2 : La **réduction pragmatique: regrouper les types autour de types « purs »**

Idée **est de construire des types qui se ressemblent** (regrouper et séparer en gardant la nuance tout en facilitant le traitement). On réduit les types en suivant **une réduction pragmatique**.

Regroupe des types autour de types pures : logique de ressemblance/différenciation

Ex : **la mesure des goûts culturels des retraités** Trois dimensions : Lecture, TV, Cinéma (intensité de la lecture, fréquentation du cinéma, relation à la télévision.
Analyse typologique avec trois variables qui ont les mêmes trois modalités : 27 types au total

		Lecture								
		Haut			Moyen			Bas		
		TV			TV			TV		
		Haut	Moyen	Bas	Haut	Moyen	Bas	Haut	Moyen	Bas
Cinéma	Haut	HHH	HMH	HBH	MHH	MMH	MBH	BHH	BMH	BBH
	Moyen	HHM	HMM	HBM	MHM	MMM	MBM	BHM	BMM	BBM
	Bas	HHB	HMB	HBB	MHB	MMB	MBB	BHB	BMB	BBB

Qui sont les types purs ? Les consommateurs fréquents (autour de HHH), moyens (autour de MMM) et occasionnels (BBB)
Puis regrouper selon une logique sociologique : Les consommateurs curieux (BBH et assimilés), les éclectiques (HBM et assimilés), les « égarés » (HHB et assimilés)

2. Les instruments de l'analyse quantitative

2.1. Décrire : L'analyse univariée → polycopiée !!!

Possibilité d'avoir à la colle des exercices de réflexion sur ces notions

Pour les variables qualitatives : tri à plat et distributions de fréquences
Pour les variables quantitatives : statistiques descriptives (cf. polycopié)

2.1.1. Les paramètres de tendance centrale

- La moyenne arithmétique : X
- La médiane Mé et les quantiles : Q1, Q2, Q3; D1, D2…. D9; C1, C2… C99
- Le mode : Mo

2.1.2. Les paramètres de dispersion

- Les intervalles de dispersion : étendue, intervalle interquartile Q3-Q1
- La variance V(X) et l'écart-type $\sigma_X = \sqrt{V(X)}$
- Le coefficient de variation : $CV_X = \frac{\sigma_X}{\bar{X}}$

2.2. Analyser les relations entre variables

On parle analyse bivariée (liaison entre deux variables : une explicative, une expliquée)

On parle aussi **d'analyse mutlivariée** :

1) Les **deux variables sont quantitatives** ; outil de base : le tableau croisé

2) Elles **sont qualitatives** ; régression linéaire

3) L'une **est quantitative, l'autre est qualitative** : ANOVA (Analysis Of Variance) → 4ème année

2.2.1. Le lien entre deux variables qualitatives

Tableau croisé (ou « tri croisé ») est l'outil de base. Il permet de **tester l'hypothèse qu'une variable « indépendante » (ou « explicative ») influence une variable « dépendante » (ou « expliquée »)**

Ex : Travailler comme salarié pendant ses études selon l'origine sociale des étudiants

Origine sociale	Travaille	Ne travaille pas	Total
Agriculteurs			100%
Artisans, comm., ch. d'entrep.			100%
Cadres			100%
Professions intermédiaires			100%
Employés			100%
Ouvriers			100%
Ensemble			100%

Variable expliquée en colonne (travailler ou non en tant que salarié)

Variable explicative en ligne (les différentes CSP des parents)

Le pourcentage : variation de la variable expliquée en fonction de la variable explicative.

Un cas particulier : les « tables de mobilité » / table d'homogamie → Les pourcentages sont calculés sur l'ensemble du tableau. 100% est la somme de tous les pourcentages. Un **tableau de présence conjoint = nom mathématique**. Pourquoi ? **Diagonale sont les situations d'homogamie ou de mobilité sociale.**

Lire un tableau croisé :

Sexe	Philosophie	Economie	Sociologie	Total
Garçons	278	336	173	787
Filles	118	407	684	1209
Ensemble	396	743	857	1996

Sexe	Philosophie	Economie	Sociologie	Total
Garçons	35,3	42,7	22	100%
Filles	9,8	33,6	56,6	100%
Ensemble	19,9	37,2	42,9	100%

Exemple : Le choix de la filière universitaire est-il influencé par le sexe ? Echantillon : 1996 étudiants de 3 filières de l'Université de Strasbourg

- **Lecture majoritaire : on s'intéresse à la répartition de chaque valeur explicative (ligne)**. On ressort la modalité la plus fréquente. On peut ensuite analyser les écarts. DESCRIPTIF
- **Lecture différentielle : on s'intéresse à chaque modalité de la variable expliquée (colonne).** On compare les pourcentages colonnes par colonnes.

Ecarts à la moyenne	Philosophie	Economie	Sociologie
Garçons	(35,3-19,9)=+15,4	(42,7-37,2)=+5,5	(22,0-42,9)=-20,9
Filles	(9,8-19,9)=-10,1	(33,6-37,2)=-3,6	(56,6-42,9)=+13,7
Ensemble	19,9	37,2	42,9

Un cas particulier de lecture différentielle : **les écarts à la moyenne.** Analyse importante. **La moyenne** (= la distribution marginale du tableau) **correspond à la situation d'indépendance** (**pas de relation entre les 2 variables).**

Définir l'intensité moyenne de l'influence d'une des deux variables. A partir de quel niveau l'écart est-il important ? On **détermine un seuil qui dépend de la taille des sous-groupes.**

Remonter la lecture du tableau à la variable : du tri croisé à l'analyse

Deux exemples :

Ex : enquête sur l'usage du téléphone portable par les lycéens

Plus le diplôme du père s'élève, plus cela semble corrélé avec le nombre d'amis que possède le lycéen. Diplôme du père : capital K, amis : liens forts que se constitue le lycéen, sa capacité. Il faut monter en généralité, on ne peut pas se contenter à une simple lecture du tableau

Ex 2 : L'enquête sur les conditions de travail (Insee, 1991) : tri croisé

Que fait l'équipe ?

- **Analyse des processus aboutissant à la formation des réponses**

Enquêté réponde de manière subjective alors pourquoi faire un enquête quantitative

Utilisation de schème

- **Reconstituer les schèmes mis en œuvre pour répondre** : mise en évidence du **rôle de l'objectivation des conditions de travail sur leur perception / déclaration** (M. Gollac, 1997) :

- Rôle des collectifs de travail

- Rôle des syndicats et mouvements sociaux

- Rôle de l'Etat

2.2.2. L'analyse statistique des corrélations (variables quantitatives)

Comment on procède ? **Va recourir à la modélisation en ayant recours à des régressions**

1) Nuages de points et types de liaisons

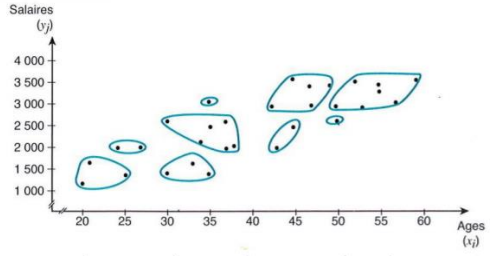
Exemple : PME de 30 salariés. On essaye d'étudier lien entre âge et salaire en euros de ces salariés. On obtient 30 couples de coordonnées.

Représentation sous forme d'un nuage de point

Idée : **résumer ce nuage de point par une droite. Modélisation en la simplifiant en établissant une relation linéaire** (entre âge et salaire).

a) **La liaison « nulle »:** les **deux variables sont « indépendantes »** : **nuage de point n'a pas de forme** (ex : salaire/taille)
b) **La liaison « fonctionnelle » ou « totale »** : la **relation correspond à une fonction connue** (linéaire, exponentielle, polynomiale…). Il **faut sortir des sciences sociales** : en physique, le lien entre la pression d'un gaz et son volume.
c) **La liaison « relative »** : **direction avec des exceptions : cas intermédiaire**

Reprendre exemple des PME : regrouper par classe



Il semble y avoir un lien entre âge et salaire, mais il y a des exceptions. (3 trentenaires qui sont encore au niveau de salaire des plus jeunes / un trentenaire qui gagne come le plus âgés) Quel est l'intensité du lien? Plus le nuage aura une forme allongée, plus la liaison sera forte.

Quel est l'intensité du lien entre les deux variables ? plus le nuage a une forme allongé plus il se rapproche d'une liaison totale, donc forte.

On présente une méthode différente du tri croise : méthode par régression

2) Régression, ajustement linéaire et corrélation

Dans le tri croisé on décrit les données**. <u>Dans le cas de la modélisation</u>, de la régression** on cherche à **définir la courbe qui résume le mieux le nuage de points.** Quand la courbe est **une droite = « régression linéaire simple » (= une variable)**

En définissant cette droite, on **commet une erreur dû au raisonnement par approximation** (surévaluation ou sous-évaluation du salaire comme dans l'exemple). **Epsilon désigne l'erreur ou le résidu que l'on commet** en résumant le nuage par une droite.

<u>Quel est la meilleure droite ?</u> Celle pour laquelle **l'erreur est la plus faible**.

Elle est déterminée par un **calcul de la distance de tous les points à l'égard de la courbe**, et on va essayer de la minimiser (calcul du minimum d'une fonction) = **« Méthode des moindres carrés ».**

- On calcule les écarts,
- on va les élever au carré ses écarts (positifs ou négatifs),
- les ajouter et trouver la droite qui va minimiser cette distance.

Minimiser cette distance totale revient donc à **calculer le minimum d'une fonction à deux variables** :

$\sum (Y_i - aX_i - b)^2$ (Dérivées partielles)

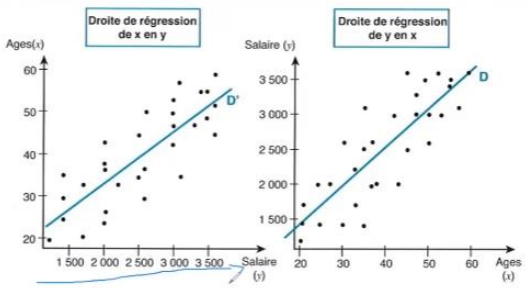
On résume le nuage par une droite : D : y = a x + b

Cette « meilleure droite » ne signifie pas qu'elle est bonne, peut toujours être très éloigné de chacun des points.

Donc … <u>Peut-on déterminer un seuil à partir duquel la relation est bonne ou mauvaise ?</u>

➔ **créer une nouvelle droite**

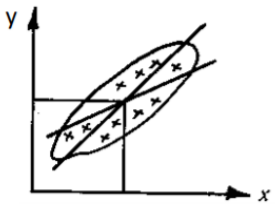
On **résume donc le nuage par deux droites :** **D' : x = a' y + b'** et **D : y = a x + b**



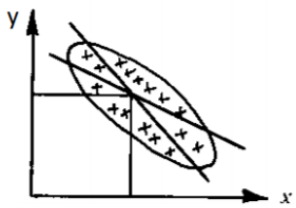
On peut faire un nouvel ajustement en projetant parallèlement à l'axe des abscisses = expliquer la variable explicative ne fonction de la valeur expliquée.

La méthode est la même : on cherche à minimiser l'écart par les **méthodes des points carrés. = chercher le minimum de la fonction**

Remarque : Deux types de **corrélation :**



Corrélation positive



Corrélation négative

- **Une corrélation positive** (les deux variables varient dans le même sens).
- **Une corrélation négative** (les deux variables varient en sens inverse)

Définir ces deux droites **permet de comprendre la qualité des ajustements**. Si la droite meilleure est la bonne corrélation.

➔ **Mais à partir de quel seuil la corrélation sera bonne ou au contraire mauvaise?**

Intuitivement, on peut le mesurer par l'angle α entre la droite D et D'.

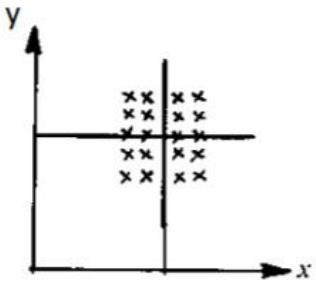
- Plus l'angle est proche de zéro, plus le nuage de points sera allongé, liaison totale
- Plus l'angle est grand (proche de 90°), plus le nuage sera volumineux et arrondi

Mathématiquement, la qualité de la corrélation se mesure par le **« coefficient de corrélation linéaire »** (noté R = coefficient de détermination), tel que : **R^2 = pente de la droite D x pente de la droite D' = (cosinus α)²**

- **s'il y a indépendance** entre les variables :

D est parallèle à Ox et D' est parallèle à Oy.

Donc **$R^2 = 0$** : la corrélation (linéaire) est nulle [cos 90° = 0]



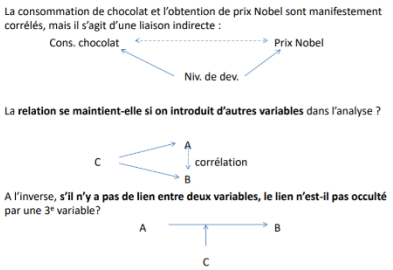
- Si D et D' sont confondues, $R^2 = 1$: il y a au contraire **Liaison totale**

- R^2 est donc compris ente 0 et 1 et R est compris entre -1 (**corrélation totale négative** : cos 180° = -1) et + 1 (**corrélation totale positive**: cos 0° = 1)

Dans les faits : la **corrélation** est considéré comme **bonne, si $R^2 > 0,76$**

NB : **le coefficient de corrélation (linéaire) R ne mesure que l'existence ou l'absence d'une corrélation LINEAIRE**. R **peut être nul** (= droites de régression perpendiculaires), alors que les deux variables sont **manifestement liées.** Dans ces cas, il **faudra procéder à un ajustement du nuage, non par une droite, mais par une courbe particulière** (exponentielle, parabolique, logistique, …)

2.2.3. De la (cor-)rélation à la causalité : introduction à l'analyse multivariée



Cette intuition se traduit dans le **principe de l'analyse multivariée**, systématisé en sociologie des sciences politique par **Lazarsfeld.** On développe ici son intuition à partir de l'analyse tabulaire multivariée qui est une extension du tableau tri croisé.

Il a présenté cette analyse dans « L'interprétation des relations statistiques comme procédure de recherche », 1946

Exemple : Enquête par questionnaire sur les préférences radiophoniques de 2300 auditeurs US. Est-ce que l'âge est lié au programme d'écoute des programmes radiophoniques ?

	PR	T. Pol	M. Cl.	Autres	Total
Jeunes	170	340	300	190	1000
	17%	34%	30%		0,81
Vieux	330	580	384	6	1300
	26%	45%	29%		
Ensemble	500	920	684	196	2300

PR = programme religieux; T. Pol. : tribunes politiques; M. Cl : Musique classique

Relation ente l'âge et l'écoute de programme religieux, même chose pour les programmes politiques. En revanche pour la musique classique pas une différence (1 point de %)

On va introduire une **variable de contrôle** pour limiter l'effet propre de la variable âge. --> Niveau de diplôme, d'instruction.

On va ensuite décomposer chacune des variables initiales à partir de la variable de contrôle.

- Etude la 1ère relation : âge / écoute de programme religieux

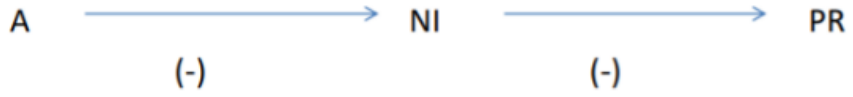
Jeunes			Vieux		
17%		<	26%		
Niv. instruction supérieur			Niv. instruction inférieur		
Jeunes		Vieux	Jeunes		Vieux
9%	≈	11%	29%	≈	32%
10%			30,80%		

Le lien entre Age et écoute des programmes religieux disparaît. Il y a en revanche un lien entre NI et PR

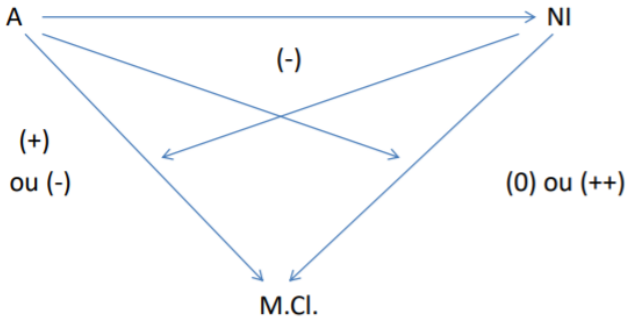
Pourquoi y avait-il relation? Il y a un lien entre âge et NI.

De même pour les deux autres variables : fin du diapo

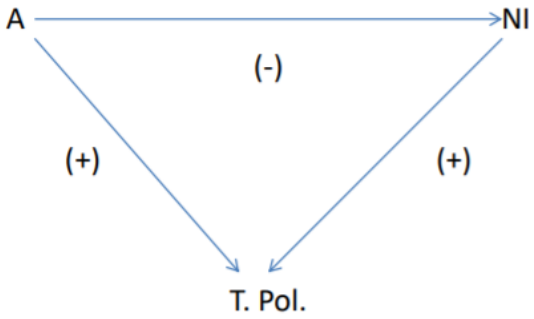
Structure causale simple



Structure causale à double niveau d'interaction



Structure causale complexe



BREF / On cherche à isoler l'effet propre de la variable en choisissant des variables de contrôle **= raisonnement « toutes choses égales par ailleurs »**. Cela revient **à constituer des sous-populations en fonction des modalités des variables de contrôle, et on vérifie si la relation se maintient ou disparaît.**

Plusieurs cas de figure à l'issue de l'analyse:

- La corrélation initiale **disparaît**
- Elle se **maintient**
- Une **nouvelle corrélation** apparaît entre deux des variables de l'analyse

Des cas intermédiaires : où la corrélation initiale s'affaiblie sans disparaitre ou au contraire s'accentuer ou même totalement s'inverser !

2.3. Modéliser la relation entre variables : l'analyse par régression

➔ ***Les limites de l'analyse tabulaire multivariée***

- **impraticable** quand on a recours à **un grand nombre de variables de contrôle : pb de significativité.** Question du nombre est d'autant plus importante quand on reste dans une analyse tabulaire : si l'on multiplie les modalités des variables, on multiplie les sous populations à analyser. Très rapidement, la comparaison de fréquence devient impraticable.

- Pas applicable aux liens entre variables quantitatives

➔ ***Résultat : recours aux méthodes de modélisation par régression :*** Plusieurs méthodes possibles selon les variables

- si toutes les variables sont quantitatives = **régression linéaire multiple :** on va essayer d'expliquer une variable en fonction de deux ou plusieurs variables explicatifs. Dans le cas d'un logiciel, on va donner un programme en fournissant les donnant et il calcule les coefficients de régression.
  Ex : Y (F (X1; X2) = a X1 + b X2 + c : permet d'expliquer le lien qui unissent les variables.
- si la variable dépendante est qualitative = **régression logistique :** il faut trouver autre chose que le CM pour simplifier les données.
  La difficulté qui se pose : il va falloir passer d'un raisonnement qualitatif à un raisonnement quantitatif pour pouvoir procéder à un raisonnement par régression.
  Difficulté, fréquence (utilisé précédemment) est comprise entre 0 et 100 donc on va devoir passer par d'autres méthodes :

1) les odds, les chances : construite à partir des fréquences :

Reprise de l'exemple de Lazarsfeld sur les programmes radiaux.

Chance consiste à proportion de ceux qui écoutent programmes de musique classique divisée par proportion de ceux qui n'écoutent pas.

Chance = **Probabilité de réussite** (de trouver quelqu'un qui est caractérisé par la variable expliquée) **divisée par probabilité d'échec** (ne pas trouver quelqu'un qui est caractérisé par la variable expliquée) → **rapport de probabilité**

Elle permet de contourner les limites des fréquences !

½ = 0,5

1/1 = 1

1/0,5 = 2

1/0,25 = 4

1/0,1 = 10

1/0,01 = 100

1/0,001 = 1000

Plus la probabilité de trouver quelqu'un qui caractérise la variable expliquée est grand, plus on s'approche de l'infini

NEANMOINS, les chances peuvent venir complexifier cette transition entre qualitatif et quantitatif donc….

2) Les rapports de chance ou odds ratios

On prend une **référence** et calcule **les rapports de chance des sous populations par rapport à la population de référence.**

$\frac{A}{B} > 1$ => A > B : la sous-population a plus de chances d'adopter la caractéristique ou la pratique étudiée que la référence

$\frac{A}{B} = 1$ => A = B

$\frac{A}{B} < 1$ => A < B

Odds ratio = $\dfrac{p_i / (1 - p_i)}{p_{ref} / (1 - p_{ref})}$

Mais cela s'applique dans la cadre des analyses tabulaires multivariées donc on ne contourne pas le problème que si l'on multiplie nos variables, il y aura trop de sous population et donc des incohérences. DONC …

3) Intuition de la régression logistique : on va **chercher à décomposer en effets simples pour chaque variable (modèle)**

Exemple de Lazarsfeld : écoute musique classique et diplôme

Si l'une des variables explicatives a plus de deux modalités (n) : l'effet de la variable sera résumé par n-1 coefficients.

Comment lire les résultats des logiciels ?

Le logiciel donne les odds ratios aux différentes modalités des variables.

Ex : choix d'un « nouveau » style éducatif par les parents (enquête des 1980's)

Rég. log. sur nouveau style éducatif		Chances
Constante = Sit. Réf.		0,2469
		Odds ratio
Sexe	Masc.	Réf.
	Fem.	1,31
Politique	Droite	0,45
	Non marqué	Réf.
	Gauche	1
	Ecologiste	1,59
Retard scolaire	Oui	2,47
	Non	Réf.
Religion	Oui	0,6
	Non	Réf.

Référence : Homme, sans orientation politique marquée, pas croyant pratiquant, enfant sans retard scolaire.

Les femmes ont 1,31 fois plus de chances que les hommes d'adopter un « nouveau » style éducatif, toutes choses égales par ailleurs.

Les femmes écologistes ont 1,31x1,59=2,08 fois plus de chances que la référence d'adopter un « nouveau » style éducatif. On fait le produit pour ajouter les deux variables.

Les personnes de Droite ont 55% moins de chances que ceux qui n'ont pas d'orientation politique marquée d'adopter un « nouveau » style éducatif

La modalité par régression a permis de définir un ensemble d'odds ratios**. Pour qu'ils soient pertinents, il faut bien définir les variables explicatives. Un des problèmes de l'analyse par régression ….**

Biais des variables omises : En oubliant une variable à la régression, on a pour conséquence qu'une des variables qui va lui être liée va avoir une importance supérieure à ce qu'est son importance réelle.

Pour essayer de contrôler cet effet, on **peut comparer plusieurs modèles pour faire apparaître le périmètre des variables choisies et on compare l'évolution des odds rations** en fonction du nombre et du type de variable introduit dans la régression.

Comparaison de trois modèles : Ex 2 : La perception d'une injustice dans les décisions d'orientation scolaire (enquête Trajectoires et origines, INED-INSEE, 2008)

Modèle 1 : régression simple : **deux variables seulement** : la perception d'une injustice dans l'orientation (var. expliquée) et l'origine géographique de la famille (var. explicative)

Modèle 2 : on introduit les propriétés socio-démographiques de l'enquêté et de sa famille comme **variables de contrôle.**

- **La relation initiale s'atténue, mais reste forte et significative** à l'exception du Portugal et de l'Asie du Sud-Est
- **Les variables de contrôle jouent un rôle négligeable** (PCS, diplôme) ou secondaire, pour la plupart
- On peut définir l'effet combiné des variables en multipliant les « odds ratio »

L'expression **« toutes choses égales par ailleurs » est abusive** : seules les variables de contrôle représentent ce qui est égal par ailleurs

Modèle 3 : on ajoute les propriétés scolaires de l'individu aux variables de contrôle du modèle 2

A propriétés sociales et scolaires équivalentes, l'origine géographique exerce une forte influence sur le ressenti d'une injustice au cours de l'orientation scolaire

2.4. Les tests statistiques : le test Khi-A d'indépendance

Rappel : depuis le début, on travaille sur échantillon.

La question qui se pose **: peut-on généraliser cette conclusion à l'ensemble de la population ? C'est le principe d'un raisonnement inférentiel.**

Le problème qui se pose : **l'échantillon est peut être atypique** (seulement des filles si l'on prend une répartition fille, garçon). Ainsi **on doit s'interroger sur le risque que la liaison statistique due « au hasard »** : on cherche à **déterminer la probabilité que ce soit produit une erreur liée au hasard**, au tirage au sort, à la manière de construire cet échantillon.

Pour répondre à cette question, on va avoir **recours au test statistique.**

Exemple entre deux variables qualitatives : le test du Khi-2 :

Pour le présenter, on reprend l'exemple entre le choix de la filière universitaire et le sexe.

- ⇨ **Raisonnement par l'absurde** : on raisonne sur **une population de référence où la relation entre les deux variables n'existe pas** : c'est une **hypothèse fictive dite « nulle »** (H0)
- ⇨ A partir de cette hypothèse, on se pose une autre question : **quelle est la probabilité que l'échantillon observé soit issu de cette population ?**

Comment obtenir cette probabilité ?

Pour comprendre : exemple : taille (en cm) des personnes composant la promotion de 1ère année. Il faut définir quelques choses de référence qui est le sol.

Reprise générale : il faut caractériser les distributions d'échantillons possibles issus d'une population. On regarde la distance du Khi-2 (mesurée par rapport à un échantillon de référence). Plus la probabilité est élevé plus il sera risqué de la généraliser à la population réelle.

Plusieurs questions : comment déterminer la référence ? Comment on calcule la distance ? Comment interpréter le résultat ? Comment conclure sur le risque de généralisation d'un résultat sur la population réelle ?

1) 1ère question : échantillon de référence = échantillon d'indépendance

C'est un échantillon où les deux variables auxquelles on s'intéresse sont indépendantes. Pour que l'on puisse comparer, on va prendre **un échantillon qui a la même structure que l'échantillon réel.**

Dans le cas de l'exemple sur le lien entre le choix de la filière universitaire et le sexe, l'échantillon d'indépendance signifie que l'on a la même propension à choisir une filière que l'on soit un garçon ou une fille.

Comme **on cherche à caractériser l'échantillon**, on va déterminer combien de garçon serait inscrit en philosophie **s'il y avait indépendance.**

Effectif modalité variable dépendante / taille échantillon x effectif modalité variable indépendante

Effectifs théoriques	Choix de la filière universitaire			
Sexe	Philosophie	Economie	Sociologie	Total
Garçons	(396/1996) x 787 = 156			787
Filles				1209
Ensemble	396	743	857	1996

On calcule les effectifs théoriques pour les filles selon le même principe

	Choix de la filière universitaire			
Sexe	Philosophie	Economie	Sociologie	Total
Garçons	(396/1996) x 787 = 156	(743/1996) x 787 = 293	(857/1996) x 787 = 338	787
Filles	(396/1996) x 1209 = 240	(743/1996) x 1209 = 450	(857/1996) x 1209 = 519	1209
Ensemble	396	743	857	1996

2) 2ème question : comment mesurer cette distance entre l'échantillon de référence et les autres

Ex : pour un échantillon similaire l'échantillon réel

Solution 1 : les «écarts « absolus » : effectifs observées – effectifs théoriques (celui qui correspond à l'échantillon d'indépendance)

- Permet de voir des personnes en plus ou en moins

Solution 2 : les écarts relatifs : $\frac{effectif\ observé - effectif\ théorique}{effectif\ théorique}$

Cependant, ces écarts mettent sur le même plan des écarts calculés sur des effectifs très différents.

Solution 3 : on combine les deux premières méthodes **: le produit des deux écarts: (Écart absolu) X (écart relatif)**

$$\frac{(effectif\ observée - effectif\ théorique)^2}{effet\ théorique}$$

On pondère l'écart absolu par l'écart relatif.
On obtient donc une mesure conventionnelle qui va pour chaque **situation définir une distorsion.**
On obtient donc une mesure synthétique de la distance : **la distance du khi 2 =**

$$\sum_{ensemble\ des\ cellules\ du\ tableau} \frac{(effectif\ observée - effectif\ théorique)^2}{effet\ théorique}$$

3) Comment interpréter le résultat ?

On peut calculer une distance du Khi-2. S**i on est dans une situation d'hypothèse nulle, la distribution des distances va prendre un taux très particulier.**

Plus la distance va être élevé, plus la probabilité de trouver un échantillon de ce type va être faible. Cela signifie que **l'on peut déterminer une valeur seuil (= valeur critique) tel que la probabilité de trouver un échantillon qui a une valeur supérieure à ce seuil prenne une valeur déterminé.**

Reformulation de la question que l'on se posait au début : Quelle est la probabilité qu'un échantillon tiré au sort de cette population fictive ait une distance du khi-2 égale à celle calculé pour l'échantillon réel ?

Conclusion : **si la population réelle était identique à la population fictive** (= si les variables sexe et choix de filière étaient indépendantes), **la probabilité d'obtenir un échantillon ayant la même distance du Khi-2 que notre échantillon réel est très inférieure à 1%.**

On peut conclure avec **un infime risque de se tromper (= la probabilité calculée) que la population réelle n'est pas la population fictive:** il n'y a donc pas indépendance entre la variable de genre et de choix de filière dans la population totale.

On peut donc généraliser le résultat de l'échantillon à la population totale.

On voit un petit risque de se tromper : on va donc accepter un seuil alpha d'erreur qui va affecter la généralisation. On retient comme pertinent un seuil de 1 à 5%. Cela signifie que dans 5% des cas, on se trompe en rejetant le cas.

Avant de rejeter l'hypothèse nulle, il **faut prendre en compte certaines limites, précautions du test du Khi-2 au moment de l'interprétation :**

- ***Taille de l'échantillon*** **: le test du Khi-2 est peut concluant lorsque l'on a un petit échantillon**

- **Distinguer entre significativité (possibilité de généraliser le résultat) et importance de l'effet (existence ou non d'une relation entre les deux variables : significance fallacy.**
  Le fait de conclure de cette non significativité, une absence de relation entre les deux variables. C'est une erreur car **le test n'est pas symétrique**. Un résultat qui est significatif (= la probabilité de se tromper est faible en le généralisant) autorise à affirmer l'existence de l'effet, on peut rejeter l'hypothèse nulle. En revanche, un résultat non significatif ne permet pas de conclure qu'il n'y a pas d'effet.

- Il **faut s'interroger sur la nature de la relation entre les deux variables quand le tableau est « creux »,** c'est-à-dire que **la distance du Khi-2 va dépendre de quelques situations seulement et non pas d'une relation globale entre les variables.**
  Exemple : tableau creux : lien entre pays d'implantation et la forme juridique des entreprises. La moitié de la distance du Khi-2 est liée au statut des entreprises allemandes (SARL). Seulement 4 situations influences la distance.
  Pour déterminer si le tableau est creux est de calculer une contribution moyenne : quelle valeur on devrait avoir dans chacune des cases si toutes situations participaient également ?
  ⇨ **distance / nb de cases du tableau sans les marges.**

Pourquoi on s'intéresse à ce test ?

- avoir la **relation entre deux variables**
- la **question de la significativité va être présente dans la plupart des régressions logistiques :** les résultats sont jugés significatifs ou non en fonction d'un indicateur qui suit une distribution Khi2.

Seuil **** = une erreur < à 0.001 en généralisant
Seuil *** = une erreur < 0.001
Seuil ** = 1% (0.01) < erreur < 5% (0.05)
Seuil * = 10% d'erreur (0.1)
Ns = non significatif

Partie 2 : Les méthodes qualitatives en sciences sociales

Préambule : Méthodes quantitatives vs méthodes qualitatives?

On les distingue souvent par leurs techniques d'enquêtes et les méthodes d'analyse. Mais on peut aussi les combiner.

Méthodes quantitatives : ont pour objectif de **recueillir des données mesurables et comparables entre elles**: l'idée est de quantifier le réel pour pouvoir l'étudier. On peut les mesurer avec des stats ou des questionnaires pour produire ces données. Mais se heurtent à **des limites** au niveau de leur analyse. 2 grandes limites :

- *Une conception analytique (variables),* on peut fractionner les individus selon les grandes variables (isoler des situations) pour les comparer. On a **une connaissance extensive MAIS qui risque d'appauvrir les informations obtenue**s, on perd ce que l'on peut connaître par familiarité avec les personnes.
- *Une généralisation fondée sur la représentativité* (càd l'échantillon va être réduit, tiré de la population à laquelle on s'intéresse. Il ne fait sens qu'en tant que protocole de recherches : c'est un instrument) et **l'exhaustivité** (on cherche à rendre compte de la population à partir d'un échantillon), MAIS avec le risque de créer des « agrégats statistiques» qui n'ont pas de réalité (compter ensemble des choses #)

Méthodes qualitatives : permettent d'échapper à ces limites, difficultés d'appauvrissement des données ou risques d'étudier des groupes qui ne sont pas réel. Diverses méthodes qualitatives :

- **Entretien** : on interroge un nb de personnes, de manière longue ou non, sur sujets particuliers
- **Observation** : comment vive les populations ? Ethnologue
- **Traces de l'histoire** : archives

Points communs de ces méthodes :

- Centrées sur **un petit nombre de cas** (connaissance « intensive », bcp d'infos sur un petit nombre de cas, opposé du quantitatif) : démarche de compréhension, on cherche à découvre le sens que les personnes donnent au monde social qui les entoure.
- Une généralisation fondée **sur l'idée de totalité et d'exemplarité** (raisonnement par cas) = le but n'est pas de rendre compte de l'exhaustivité de la situation mais de la totalité (comment s'articulent les représentations entre-elles ?) // fonctionnement plus synthétique.

Ne pas surestimer l'opposition entre les deux : L'opposition n'est cependant pas absolue :
Ex : Guy Michelat, Michel Simon, Classe, religion et comportement politique, 1977: deux temps : 77 entretiens non-directifs puis 1780 questionnaires : orientations du vote (droite conservatrice, catholique VS gauche, relation de classe)

Chapitre 3 : L'observation en sciences sociales

Les précurseurs : William Foote Whyte, Street Corner Society

- Recours à l'observation au sein de l'ethnologie naissante (fin XIXe s.) : société coloniale, non occidentale
- Réadapter : Premières études sur les sociétés occidentales reposant sur une utilisation systématique de l'observation : Etats-Unis, années 1920-1930 (« Ecole de Chicago »)
- Un exemple : Street Corner Society, 1943 : sur le quartier italo-américain du North End (Boston) ; post face à l'ouvrage qui décrit méthodes et difficultés sur le terrain : qu'est-ce que l'on fait sur le terrain ? quelle démarche scientifique à suivre ?

La méthode de l'observation participante

A l'origine de la recherche WFW, il y a **une inspiration à la fois scientifique et sociale**.

Scientifique : il vient de la classe moyenne, parents profs, blancs, anglais. Au cours de sa visite dans le quartier pauvre de Philadelphie (études), il va en faire la base de son questionnement. Il obtient une bourse pour Harvard et s'installe dans le quartier (d'abord famille d'italiens puis avec sa femme). Il observe de près et doit après obtenir des alliées. La qualité du matériau dépend d'une chance inouïe : la rencontre de Doc, chef de bande dans le quartier. Pour le trouver, il suit les réseaux sociaux du terrain : foyer sociaux culturel. Le rôle du Doc va l'intégrer dans le quartier, il s'assure une protection et un moyen d'acceptabilité – donc rôle d'introducteur. Il va aussi avoir un rôle d'instructeur ET d'informateur, il est coproducteur de la recherche. Ce rôle va évoluer pendant ces quatre années : il soumet des versions # de sa thèse pour avoir sa réaction.

Entre les deux : Le Doc va finir par être un Ami

Sociale : WFW occupe de nombreux rôles dans la communauté :

- corner boys, bande de Docs
- college boys, étudiants
- secrétaire du Club de la communauté italienne, structurer les relations
- agent de campagne électorale, structuration politique, qui intervient et comment les mobiliser ?
- « entrepreneur de mobilisation » : marche pour revendiquer des mesures en faveur du quartier – qque sorte un membre du quartier

Par cette enquête, il **expérimente les limites de la méthode de l'observation (les choses à ne pas faire d'où l'intérêt de la post face)**. Il cherche à rester neutre mais il est pris dans les activités du quartier. En restant trop longtemps dans le quartier, **il y a un risque de devenir « indigène » : « going native ».** (On le voit, il organise une mobilisation). Il va commettre plusieurs erreurs :

- **Reprendre expressions / attitudes des quartiers** => sort de son rôle, il faut conserver ses limites donc : « pour être accepté par les gens, vous n'êtes pas obligé de faire tout ce qu'ils font »
- **Fraude de la campagne électorale**

Quels résultats ? Comment va élaborer son d'objet ? Très # du questionnaire.

. L'élaboration inductive de l'objet d'étude

- Au début, il n'a pas d'idée claire de son objet.

Deux inspirations théoriques :

- **Positives :** on importe les méthodes des ethnologues sur des villes moyennes (répartition du pouvoir, la famille…) => les « études de communauté » (Robert & Helen Lynd, Middletown, 1929)
- **Négatives :** un rejet des travaux sur les quartiers pauvres en termes de désorganisation sociale (William Thomas, F. Znaniecki, Le paysan polonais, 1919)

- A mi-parcours, un mémo pour renouveler sa bourse :

- Il se rend compte qu'il ne faisait pas ce qu'il pensait initialement : il ne fait pas une « étude de communauté » (manque famille, religion) mais **un travail sur les bandes** (Corner boys, collège boys) **et les institutions sociales du quartier** (politique, raquette, les clubs de la communauté italienne…). Pas nécessaire d'étudier tous les groupes mais d'établir **les ramifications qui intègrent les bandes par rapport à l'intégration du quartier dans la ville et dans l'état** (totalité, possibilité de généraliser). Avec ce qu'il a, il peut généraliser

- Une **découverte au hasard de son hypothèse centrale** : le lien entre performance individuelle et structure du groupe (bowling) : donc un besoin de resituer.

- Le résultat va **lui permettre de contredire les travaux sur les quartiers pauvres** en termes de désorganisation sociale (William Thomas, F. Znaniecki, Le paysan polonais, 1919)

Illustre bien les difficultés : dépendance de Doc et subjective car implication ! Cependant, quelle scientificité de la démarche?

L'ambivalence épistémologique de la méthode de l'observation

Une méthode adaptée à des objets spécifiques

Il faut avoir certaines conditions :

- ✓ **qu'une situation limitée**
- ✓ **facile d'accès au regard**
- ✓ **autorisant une présence prolongée** (le + important) : difficile pour socialisation par lien parents/enfants

Historiquement, 3 objets privilégiés :

- **Les petites communautés rurales ou urbaines** (ex Street Corner Society, « Ecole de Chicago »)
- **Le monde du travail industriel** : fast food, uberisation
- **Les services et leurs interactions** : hôpitaux (patients/médecins)

Points communs de ces objets :

- ✓ Pratiques qui ont un caractère occulté ou non officiel : **question du secret**
- ✓ Pratiques relatives à la **vie quotidienne**
- ✓ Pratiques professionnelles perçues comme trop peu « légitimes » pour être évoqués spontanément ou qui **échappent à la conscience des enquêtés**

Les risques du travail de terrain

➔ Au centre de la **méthode de l'observation : le « terrain ».**

Cf. Claude Levi-Strauss, *Le regard éloigné*, 1983: le terrain a un effet particulier sur le chercheur : aller sur le terrain, c'est produire une **« révolution intérieure », faire du chercheur « un homme nouveau »**. En allant sur le terrain, on fait l'expérience de l'altérité, de l'autre, de l'étonnement. On va se poser des questions sur ce que l'on voit et soit même en train d'observer. **Cela a une influence sur la méthode**. Le premier objet de l'enquête n'est pas de poser les questions mais de découvertes des questions avec une innovation de concept. Avec du temps, on se transforme, on va modifier son regard ;

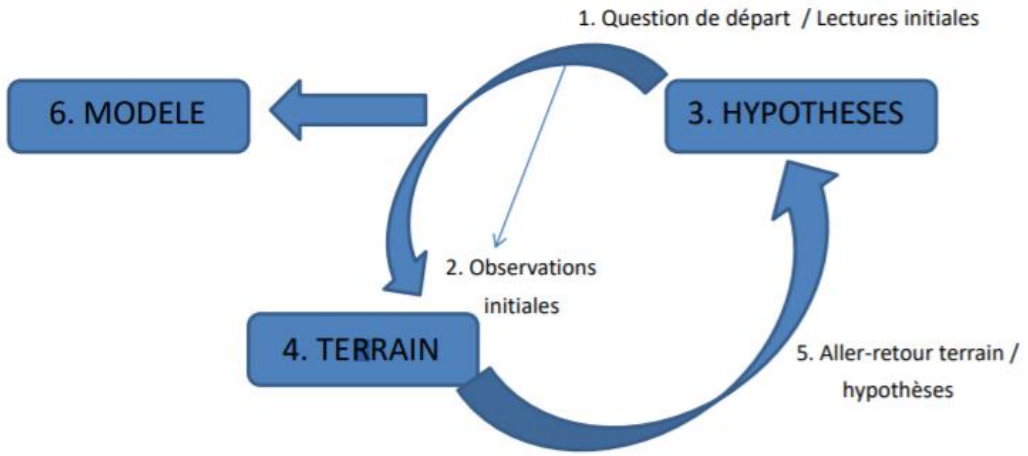
➔ Csq : **le protocole de recherche s'inscrit dans le cadre d'une démarche inductive**, on remonte du terrain pour produire de l'intelligibilité.

Norbert Elias*, Engagement et distanciation*, 1983 : le savoir commun n'est pas un non savoir, c'est une richesse qu'il faut comprendre et analyser. **La posture du chercheur repose donc sur un aller-retour entre :**

- ✓ **« engagement » sur le terrain**
- ✓ **et « distanciation » pour se démarquer des idées préconçues.**

La construction de l'objet ne se fait pas au départ MAIS la construction de l'objet n'est achevée qu'au terme du travail (# approche hypothéticodéductive). Oui il y a des hypothèses mais elles sont très larges, pour recueillir un max d'observations sur le terrain.

Le terrain est la base mais il y a des **risques importants**



- Le risque d'**ethnocentrisme** : Ne pas s'ouvrir au terrain et plaquer des grilles préconçues selon ce que le chercheur est dans l'espace social. Plus on reste longtemps, plus on échappe à ce risque.
- **Le « paradoxe de l'observateur » (William Labov,** Le parler ordinaire, 1978) : au cœur de la démarche, il y a la nécessité du terrain. Mais si on l'observe, il y a un risque qu'il se comporte autrement, perturbable. Est-ce que le chercheur observe existe indépendamment de sa présence ?

> Ex : L'enquête d'Elton Mayo dans les usines Hawthorne de la Western Electric (Cicero, années 1930): Enquête de psychologie expérimentale sur le lien entre conditions de travail et productivité. Dans un premier temps, on améliore conditions de travail, productivité augmente. A l'inverse, la productivité augmente aussi. Pourquoi ? Donc dans un 2e temps, on isole 6 ouvrières : étude par observation et entretien. Ici, vraie plaisir à participer à l'enquête donc collaboration, attention de la direction et ne veulent pas décevoir. Donc pas les conditions de travail mais plus due à l'observation.

⇨ On peut en tirer deux conclusions sur ce paradoxe :

- ✓ **Ce qui est observé peut être qu'un artefact** : un résultat qui découle de la démarche elle-même et non pas de la réalité que l'on observe : les effets de la méthode. Attention, résultats analysables, effet de la collaboration
- ✓ **L'observateur ne peut rester neutre** : on est forcément intégrer dans les activités

Transition : Donc quel rôle avoir ?

Rôles de l'observateur et « auto-analyse »

Le chercheur n'est pas désarmé : deux instruments :

- **Minimum de choix de sa « position d'observateur »** (= formes d'intégration dans le groupe)
- **Possibilité d'une réflexivité = « auto-analyse »** (Florence Weber, Le travail à côté, 1989)

2.3.1. L'observation incognito

= **se faire passer pour un membre du groupe pour ne pas être perçu comme observateur**

Exemples :

- Günter Walraff, Tête de Turc, 1985 et Parmi les perdants du meilleur des mondes, 2010.

Il teint sa peau en plomb. Il veut se faire passer pour un immigré turc. Comment les allemands vont réagir ?

- Florence Aubenas, Quai de Ouistreham, 2010

Cherche un travail mais ment sur son diplôme, dit avoir qu'un bac

Mais plus d'inconvénients que d'avantages :

- *Méthodologiques*
  - **Limitation des positions d'observations** : réduit les éventails d'observation auquel on est confronté (à la # WFW) + poser les questions, prendre les notes
  - **Incertitude du terrain** : durée, risque d'être découvert, rendre un travail postérieur complexe pour collègues
- *Déontologiques* : les enquêtés n'ont-ils pas le droit d'être informés sur les objectifs de l'enquête? On **dévoile une intimité des personnes sans autorisation.**
  Ex : Laud Humphreys, *Tearoom Trade* (1970), Trad. *Le commerce des pissotières*, 2007 : homosexualité masculine : parmi aussi les personnes respectables, pas seulement chez les déviants. Plaque d'immatriculation, va chez les gens sous un autre rôle (fonctionnaire municipale, assurance maladie…) pour obtenir infos

2.3.2. L'observation à découvert

On est observé et on observe !

Cette position d'observation **n'est pas toujours contraignante:**

- **Possibilité de phases d'observation incognito** (ex : Vincent Dubois, La vie au guichet, 1999 : au sein de la CAF, connu par certains de la CAF comme chercheur mais stagiaire pour les autres)
- La participation sur longue **durée réduit le « paradoxe de l'observateur »** : pourquoi ? lus difficile de conserver une façade à l'égard du chercheur, relation de confiance d'où une nécessité d'avoir une analyse réflexive selon comment le chercheur est identifié dans le temps.

Dans les deux cas (observation incognito et à découvert), l' « auto-analyse » est centrale : il faut :

- **distinguer** pour chaque situation de qui on est connu comme observateur, et qui vous prend pour quelqu'un d'autre
- **repérer les inflexions** dans la manière dont on vous identifie dans le temps

D'où une possibilité d'avoir **plusieurs positions du chercheur** : (Patricia Adler, Peter Adler, *Membership Roles in Field Research*, 1987):

- **L'observation participante « périphérique »** : les caractéristiques en tant qu'observateur ou du propre rapport au terrain oblige à **rester en retrait par rapport au groupe. en quelques sortes : exclu.** Ou alors des cas où le chercheur prend volontairement de la distance (ex : Adler, travaillent sur le monde du trafic de drogue, héberge les personnes qui sortent de prison MAIS il faut prendre le distance pour limite les attentes du groupes, pour qu'il ne cherche pas à les recruter)
- **L'observation participante « active » :** (ex. Street Corner society ; WFW) : **chercher à jouer un rôle dans le groupe, acquérir un statut.**

➔ Ces deux positions sont différentes mais une **ressemblance : l'observateur est d'abord extérieur au groupe puis prend part dans ses activités selon les cas.**

- **L'observation participante « à part entière » ou « complète »** : le chercheur est **déjà membre du groupe** (ex : Howard Becker, *Outsiders*, 1963 : pianiste de jazz). Il est **sociologue de son propre groupe**. Ici, il faut s'en distancier pour faire du groupe un objet. **Risque d'ethnocentrisme, et de présenter comme scientifique le sens commun du groupe**. Il faut une **vigilance épistémologique**.

Le rôle à jouer dans le terrain est central dans l'observation de terrain. Être un bon observateur c'est savoir choisir les bons rôles, OU, si on ne peut choisir, être conscient des rôles qu'on vous fait jouer (vigilance épistémologique).

Conclusion : **une « ambivalence épistémologique »** (**Olivier Schwartz**, spécialiste du monde ouvrier en France, «L'empirisme irréductible», 1993) : il y a ambivalence car 2 caractéristiques :

- Les **matériaux réunis par l'observation ne respectent pas pleinement les contraintes de la validation scientifique**. On dépend des passeurs qui font rentrer l'observateur dans le terrain, il y a aussi les qualités de l'observateur ; il faut donc critiquer les résultats et redouble d'une exigence de réflexivité.
- **L'exigence de rigueur a des limites : pour que la méthode soit féconde, il faut tolérer une part de «bricolage»,** de souplesse. Il faut faire attention à ne pas mutiler ce qui est observé !

Le travail d'observation

Le journal de terrain

Le problème de l'observation : **l'observateur est le seul instrument pour créer le matériel** contrairement au questionnaire, à l'entretien. L'enjeu est de **prendre des notes** : il faut noter à chaud le maximum de ce qu'on a vu. Pourquoi écrire ?

- **Objectiver des impressions** : elles deviennent matérielles une fois l'intersection ou processus terminé. Elles sont des preuves empiriques, des choses que l'on va montrer aux lecteurs.
- On voit, on ressent les choses mais **il nécessaire de prendre de la distance**. Elle est un moyen de comprendre ce que l'on a fait. Il faut situer le moment (contexte, date…)
- Elle est **déjà une forme d'analyse** : on se force à relever ce qui est important, à hiérarchiser ce qui est vu et à revenir sur ce qui est vécu. On se met sur la voix de ce qui va être nos résultats = découvrir les questions à se poser.

Il faut donc **tenir un journal de terrain :** = un **document recensant toutes les données de terrain.** On peut distinguer trois types de dimensions, de données que l'on consigne dans ce journal :

- ***Notes descriptives*** (= « journal d'enquête ») : **relate l'observation au sens strict**. Noter c'est enregistrer le plus possible, surtout **au début de l'enquête** ; pourquoi ? car on ne sait pas encore ce qui est le plus important. (cf. le bowling chez W. Foote Whyte : se rend compte avec recul que les performances évolues, cet élément était plus tard intéressant)

  Que noter? **Le cadre ; Le moment ; Les individus rencontrés et leurs caractéristiques ; Leurs attitudes, comportements ; Leurs interactions (**verbales, et physiques).

  Quelles techniques de recueil? **La prise de note ; Les schémas des lieux réguliers ; Les photos ; La récupération de traces.**

  Ex : Loïc Wacquant, Corps et âme. Carnets ethnographiques d'un apprenti boxeur, 2001 : apprentissage de la boxe dans un gymnase du ghetto noir de Chicago (1988-1992)

- ***Notes d'analyse*** (= « journal de recherche ») : **interprétations**, références théoriques, rapprochements avec d'autres phénomènes, etc. Déjà une **étape de distanciation**. Cette partie est très importante car le matériau brut (enquête) ne peut pas être utilisé telle quelle.

  Problème principal de l'observation : **comment voir davantage que ce qu'on est prêt à voir?** (S. Beaud, F. Weber, Guide de l'enquête de terrain). Plusieurs facteurs permettent de voir davantage :

  - **Rôle des qualités personnelles** (cas des parcours de personnes qui connaissent une trajectoire sociale, confronté à différents milieux sociaux)
  - **Rôle du savoir-faire que l'on a accumulée** : se nourrir des réflexions des autres (revue *Genèses, Terrains et Travaux* : cairn.info). Faire une évaluation réflexive en revenant sur le terrain de qn comme WFW qui revient critiquer son propre analyse.
  - **Prise de distance par la construction théorique : « l'observation sans armes est vide, l'observation trop armée n'apprend rien » (S. Beaud**, F. Weber, *Guide de l'enquête de terrain*) : ici que l'on retrouve l'ambivalence de la méthode : il faut structurer de plus en plus l'observation (ex : élaborer un guide d'observation) tout en prenant en compte ce qu'on nous donne à voir. Il ne faut pas trop systématiser pour éviter d'être prisonnier d'un schéma prédéfini et donc inattentifs à l'inattendu.

Ex : Yasmine Siblot, *Faire valoir ses droits au quotidien*. *Les services publics dans les quartiers populaires*, 2006

- Guide d'observation : fait une méthode assez proche de celle quanti ☺ puis après va vers du quali

  - ***<u>Réflexions personnelles</u>*** (= « journal intime »): le rapport aux observés, les émotions ressenties. Il n'y a **pas de neutralité pendant l'observation**. L'évolution de ces émotions met sur la voie de l'évolution de votre position sur le terrain. Son but est **d'aider l'analyse réflexive.**

Partie 2 : Les méthodes qualitatives en sciences sociales

Chapitre 4 : L'entretien en sciences sociales

1. L'approche classique de l'entretien de recherche : la « neutralité bienveillante »

Double inspiration historique de l'entretien :

- La **tradition de l'interrogatoire social** (Le Play, Rowntree au XIXe siècle) : observateur est accompagné d'un représentant de l'ordre. C'est un cadre d'interrogatoire contraint.
- **L'entretien thérapeutique** : psychologie, psychanalyse. La patient est actif dans les troubles dont il souffre.

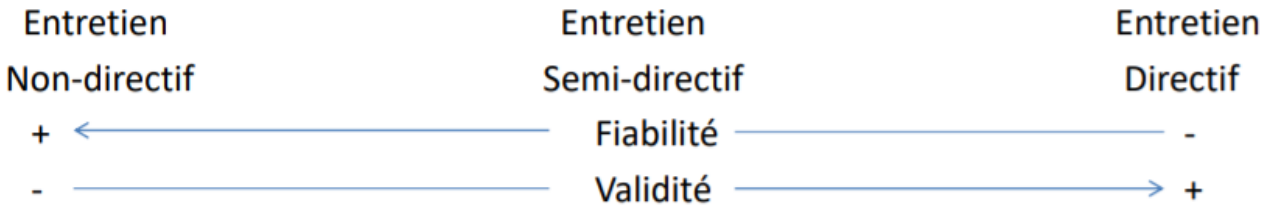
Exemple qui utilise cette double inspiration : 1 e formulation systématique de **l'entretien en sciences sociales : l'étude sur la productivité industrielle à la Western Electric dans les 1930's** (Elton Mayo) F. Roethlisberger, W. Dickson, *Management and the Worker*, 1939

- Très grande échelle : 20000 entretiens pendant 2 ans dans une perspective de « recherche-action ». ils vont porter sur les sentiments que les salariés ont à l'égard de l'encadrement du travail dans l'entreprise etc. Le but est ensuite de chercher des formations pour améliorer le moral des salariés.
- Ils le font par tatônnement :
  - **« approche directe »** (= questionnaire, série des questions ouvertes posées de façon standardisé = même ordre). Des réponses stéréotypées, les employés sont contraints et veulent aborder des choses qui sortent du cadre.
  - **approche « indirecte »** (entretien avec prise de notes) => assouplissement de la démarche
  - puis **enregistrement** = > utilisation des moyens techniques

Cela traduit une évolution progressive de l'objet et de la démarche. Initialement, l'équipe **cherche des réponses à des questions**, on a des hypothèses et des idées. Progressivement, **une évolution vers l'entretien** : la recherche d'informations sur l'évaluation des enquêtés sur ce qui les entoure. Il faut partir des questions telles que les enquêtés les formule, il va falloir écouter leur vécus. On passe de **l'hypothéticodéductive à l'inductive**, c'est-à-dire que l'on cherche les « réponses » à des « questions » que se posent les enquêtés.

- Néanmoins il y a de **grandes différences entre entretien thérapeutique et entretien de recherches :**
  - ⇨ **Relation entretenu par le patient alors qu'ici c'est la sociologue qui permet la rencontre.** Il faut donc avoir des stratégies d'intéressement pour que l'enquêté accepte.
  - ⇨ **Le patient maitrise la durée alors que là c'est le sociologue d'où la formulation d'un ensemble de règles qui inspire les recommandations classiques sur l'entretien:** techniques de **« neutralité bienveillante »** (Carl R. Rogers, « The Non Directive Method as A Technique for Social Research », 1945) :
    - ❖ *écoute patiente et bienveillante :* chaque personne est intéressante et ce qui prend du temps à être dit est le vécu le plus difficile
    - ❖ *pas de rapport d'autorité :* car déjà des rapports de subordination
    - ❖ *intervention minimale (relances) :* que pour aider ou aiguiller vers des thématiques qu'il a peu abordées

      Le degré d'intervention dépend du type d'entretien : dilemme entre « fiabilité » (= matériau correspondant à ce que pense l'enquêté) / « validité » (matériau conforme aux objectifs de la recherche)



Un instrument de compromis pour l'entretien semi-directif : le guide d'entretien (cf. moodle) qui permet de vérifier la validité (Robert Merton et al, *The Focused Interview*, 1956), mais son usage doit rester souple

⇨ **La gestion classique du déroulement de l'entretien :**

1) **Démarrer l'entretien par une consigne de départ** : Eviter une question factuelle. Cerner les thèmes centraux de l'enquête, pas plus de deux phrases.
2) **Utiliser des relances** : nourrit un nouveau discours

- non verbales : hochet la tête par exemple
- verbales : reformulation, demandes d'informations complémentaires
- Intervention-miroir : signe non verbaux de l'enquêté que l'on utilise
- Contradiction : à utiliser avec précaution : idée de rester neutre (la contradiction dans l'entretien des ouvriers peut rapidement réintroduire le rapport de force vu que l'enquête est commanditée par le patron)

3) **Enregistrer l'entretien, mais prendre des notes** (écoute active) pour mieux relancer l'enquêté
4) **Utiliser le guide d'entretien de manière souple**

➔ Ces techniques ne suffisent pas : la situation d'entretien est une situation sociale « artificielle » qui produit des effets sur les échanges entre enquêtés et enquêteurs !!

2. Relation d'enquête et limites de l'entretien

2.1. L'enquêteur comme producteur du discours

La conception classique de l'entretien néglige la violence symbolique inhérente à la situation d'entretien. Cf. Pierre Bourdieu, « Comprendre », in La Misère du Monde, 1992 : lutte entre deux définitions de la situation :

- Une définition « officielle » par l'enquêteur
- Une **définition implicite comme situation de jugement réciproque entre enquêteur et enquêté :**
  - Conduit à mettre en place une **façade de représentation** (Erving Goffman, *La mise en scène de la vie quotidienne*, 1959) : plus continuer de contourner comme dans le cas de l'observation
  - **Exerce des effets intimidants sur celui qui paraît socialement dominé dans l'interaction** (Pierre Bourdieu) : exemple : ignore des questions, donne des conseils sur comment mener l'entretien

Il faut voir les **effets de cette violence symbolique** sur les enquêtés mais peut aussi s'exercer sur le chercheur :

➔ Ex : Michel Pinçon, Monique Pinçon-Charlot, Voyage en grande bourgeoisie : journal d'enquête, 2005. Il y plusieurs risques :

- **Dépossession relative** quant à la définition officielle de la situation : on doit informer sur la définition de l'enquête
- **Ressources dont disposes les enquêtes pour résister à l'objectivation** : càd échapper, contourner l'effort du chercheur à essayer de saisir la trajectoire de l'enquêté : la réserve : plus discret sur sa biographie alors que pour sciences sociales se sont des infos importantes : l'enquêteur est moins au contrôle dans cette situation. Même quand aucune esquivassions, qqch de liée à la formation de l'enquêté : les sciences sociales font parties de la science sociale des élites, donc très souvent enquêtés connaissent la démarche et savent où l'on veut en venir → tendance à esthétiser sa trajectoire pour rencontrer la trajectoire du chercheur (ex : rester très générale et donner une interprétation de sa formation).

Quelles sont les solutions à ces formes de violence symbolique : Pierre Bourdieu, « Comprendre », in La Misère du monde :

1) **Réduire la distance sociale** entre enquêteurs et enquêtés. Dans ce travail collectif de Bourdieu, ils vont faire une expérience de soit recruter des enquêteurs dans le milieu et de les former soit que les enquêteurs choisissent des personnes du même milieu que lui. Mais l'analyse n'est pas faite par l'enquêteur, il va y avoir des non-dits (du fait du même milieu) et donc des infos pas exploitables.

Attention à **la présentation de soi de l'enquêteur** pour observer des versions différentes de la présentation de soi de l'enquêté : permet d'ouvrir l'éventail de choses que l'enquêté va dire
Ex : questions du jugement des sans-abris : la verticalité opposé à être avachi. Si présentation de soi #, les choses auraient pu être différente. Au lieu d'être en costard, être plus détente.

2) **Renoncer à la « neutralité bienveillante »** au profit de l'engagement dans l'entretien. Cette solution a pour intérêt **d'amener une présentation de soi différente de l'enquêté.** Néanmoins, elle est fortement critiqué : Gérard Grunberg, Etienne Schweisguth, « Bourdieu et la misère: une approche réductionniste », 1996. Lorsqu'on manifeste son avis dans l'entretien, et que d'un coup on arrête son jugement, cela peut être vu comme négatif surtout quand l'enquêté est dominé.

Toutefois, **quand enquêteur dominé, Recours à des stratégies de contournement :**

Ex : Sylvain Laurens, « Comment et pourquoi poser les questions qui fâchent? », Genèses, 2007 : Les hauts fonctionnaires et la politique de l'immigration française des années 1970.

- **recours aux matériaux d'archives** (nationales ou privés) : il va pouvoir confronté les enquêtés aux documents qu'ils ont eux-mêmes produits. → sortir de la réserve.
- **« détour statistique »** : vers la fin de l'entretien, il est possible de dire à l'enquêté que l'on cherche à faire une base statistique des caractéristiques du groupe : amener à parler à la biographie personnelle. Imparfait dû à esthétisation de leur trajectoire. Dans d'autres cas, amène à un deuxième entretien.

ATTENTION : il faut éviter le risque du méthodologisme : penser que l'on peut trouver des éléments de méthode pour réduire ou effacer l'effet de l'enquêteur propre à la relation de l'enquête. Même s'il y a des effets, cela n'empêche pas d'avoir une démarche rigoureuse et scientifique à partir des résultats obtenus grâce à une analyse réflexive

3) **Recours à l'analyse réflexive** = réflexion sur les **conditions de possibilité de la recherche et de ses résultats**. Cette conception ethnographique de l'entretien va demander à tenir un journal de terrain pour noter la manière dont on a mené l'entretien ou les refus d'accès au terrain. Cela va permettre de :

- Contrôler la qualité du matériau au moment de l'analyse
- Instrument de connaissance sur limite de l'entretien.

Croire à l'existence d'une « vérité » des pratiques est une illusion positiviste. Réduire et se rapprocher de la réalité des pratiques est une illusion car l'enquêté est tout aussi vrai dans l'entretien qu'il est dans d'autres situations sociales. L'enquêté quand il adapte son comportement il le fait à partir de ce qu'il est, à partir de schèmes de classement des enquêtés.

→ **Analyser la situation d'entretien peut permettre de dégager les schèmes de classement des enquêtés**.

Ex : Murielle Darmon, « Le psychiatre, la sociologue et la boulangère : analyse d'un refus de terrain », *Genèses*, 2005 : étude ethnographique sur le traitement et la construction médicale de l'anorexie. Elle va contacter des médecins pour avoir accès au patient. Elle ne veut pas se cantonner à l'analyse sur les régimes et les discours de l'anorexie. Ce qu'il l'intéresse c'est le traitement de l'anorexie. Elle cherche à rendre social l'anorexie. (cf. cours de Marine DELASALLE).

2.2. La constitution du corpus d'entretien

Qui va –t-on interroger ? L'objectif n'est pas de **constituer un corpus d'entretiens** « représentatifs » (#quanti) mais **« significatifs », « contrastés ».** On cherche à repérer des logiques d'action, des représentations. On cherche à les rapprocher à des processus sociaux plus généraux. On cherche à faire une typologie la plus proche de l'exhaustivité possible.

Ce qui préside aux choix des entretiens, ce n'est pas être représentatif mais c'est une **logique de diversification maximale.** On cherche des personnes # sur plusieurs critères : un corpus contrasté.

Ex : Anselm Strauss, *Awareness of Dying*, 1965 : traitement des mourants par les infirmières dans les services de soins intensifs. Relations infirmières, patients en fin de vie (soins intensifs, palliatifs, grands prématurés). Ils vont interroger et comparer ses interactions puis d'autres types d'hôpitaux (public, clinique privé…). Le travail se fait à SF et vont même aller au Japon → aucune représentativité mais # types d'interaction possible. Si l'objt n'est pas la représentativité, il n'y a pas de base de sondage, l'enquêteur va se laisser guider par le milieu d'enquête. Il dispose de deux stratégies :

- Définir le corpus au fur et à mesure des rencontres sur le terrain (entretiens ethnographiques) : il va falloir tenir un journal de terrain qui va permettre de recenser les démarches et de repérer les groupes qui vont rester en retrait de relation pour corriger par la suite cette absence
- Partir de l'organigramme d'une institution puis procéder par « boule de neige » (capital social et réseaux de relations des enquêtés) : enquêté recommande des personnes qui sont intéressantes pour le travail : on bénéficie capital social et permet de suivre des réseaux.
  Là aussi danger : certains parties de l'espace social ne rentre pas dans la relation. On suit des réseaux particuliers, on risque de parcourir une partie de l'espace social et laisser dans un clair-obscur d'autres groupes intéressants pour l'enquête.

Différence entre les deux stratégies : la façon dont se déroule l'enquête (rencontres de terrain, prise de rdv par la suite, plus lier à l'observation VS prise de contact par tel / mail / courrier pour obtenir rdv sur recommandation de qn)

→ Donc nécessité de corriger le corpus et créer un « corpus raisonné ». On analyse de manière réflexive le corpus et ajouter des éléments pour corriger ces biais.

3. L'analyse de l'entretien

3.1. Les objectifs de l'analyse de l'entretien

1er objectif : la « description simple » : La « description simple » : le chercheur utilise **une théorie existante dans la discipline pour forger un schéma d'analyse a priori qui lui permette de classer son matériau** (= grille d'analyse)

2 cas de figure :

- **Entretien inscrit dans démarche hypothético-déductive** (questionnaire) = approche classique de l'entretien : on cherche à répondre à des hypothèses : directif avec questions ouvertes standardisés
- **Approche inductive** : on fait remonter l'info à partir du terrain des entretiens (ex : Michel Crozier, Erhard Fiedberg, *L'acteur et le système*)
  Pourquoi description simple : on ne constitue pas la grille à partir du matériau

2ème objectif : la « description analytique » : on peut **vouloir créer sa grille d'entretien** : c'est le schéma général d'analyse lui-même qui est construit de manière inductive à partir des matériaux. Ses travaux **sont couplés à l'observation, forte part d'ethnographie.** Le but est d'éviter d'entamer son raisonnement avec du préconçu. On veut éviter au terrain de d'être forcé à rentrer dans la grille. L'idée est de découvrir des choses nouvelles, des concepts nouveaux.

Un cas particulier, un prolongement : **la « théorie fondée »** (Barney Glaser, Anselm Strauss, The Discovery of Grounded Theory, 1967) : pourquoi ? **L'objectif de la démarche inductive est de fondées de nouvelles théories à partir du terrain de recherches.** On cherche à construire des modèles.
2 grands types de théorie possible :

- Des théories locales **(subsantive theory**) ajustées à un terrain de recherche particulier
- Des théories formelles ou générales (**formal theory**) qui renouvellent les concepts de la discipline.

Comment construire cette grille d'analyse de manière inductive ?

3 grandes étapes :

3.2. Les étapes de l'analyse qualitative d'entretiens

Comment générer la grille d'analyse à partir des matériaux ?

3.2.1. Le travail de découverte : analyse des premiers entretiens

Deux ppes :

1) **L'analyse interne** : **analyse du contenu de l'entretien, assimilé à un texte**

➔ Les **mots ne sont pas neutres** : ils sont l'expression de transformation plus globale, on cherche à les faire apparaitre en analysant l'entretien.

➔ Il y a parfois **même un inconscient du texte qu'il faut dégager**. Pourquoi un mot est employé plutôt qu'un autre
Ex : G. Michelat, M. Simon, Classe, religion et comportement politique, 1977 : « Pour être bien accueilli chez les Russes, il suffit d'être soviétique »:
- Ironie : les Russes ne sont pas accueillants pour les gens de l'extérieur (anticommunisme)
- Il s'agit d'une ouvrière CGT : lapsus : « soviétique » = « communiste », les communistes sont des étrangers, pas des français comme les autres

Pour voir cela, il faut être attentif à la forme du discours :

- **Les mots et expressions « indigènes »**
- **Le silence et les non-dits**
- **Les malentendus**

L'interprétation ne se limite pas au contenu, elle est aussi guidée par le contenu !!!

2) L'Analyse externe

Quel est le contexte dans l'entretien de recherche ? Il y en a 2

- Le **contexte de la relation d'entretien** (question de violence symbolique….) : le sens des mots en dépend. Il faut aussi voir le déroulement de l'entretien (passer du nous au jeu…)
- Le **contexte social dans lequel est pris chaque enquêtés** : l'enquêté livre des données factuelles « objectives », des situations vécues, des jugements.

Ces informations sont-elles fiables ? à condition d'analyser leur statut

- **Les omissions / mensonges** délibérés ne sont pas plus importants pour l'entretien que pour la méthode du questionnaire
- **Un récit n'est jamais utilisé seul dans l'analyse** : la confrontation des entretiens peut favoriser une analyse plus précise
- A travers leurs récits, les **enquêtés livrent des indices sur des processus sociaux plus larges dont ils ne sont pas (entièrement) conscients**
- Le **récit est lié à ce que la personne est socialement aujourd'hui :** le récit de vie est médiatisé par la « totalisation subjective » de l'expérience (« illusion biographique »)

3.2.2. Le travail de réduction des matériaux

Exemple : La **construction d'une biographie** = reconstituer les faits, mais aussi les mettre en relation par des interprétations.
Suppose un travail de sélection des éléments pertinents

Dans l'analyse d'entretiens, la sélection / réduction passe par :

- La comparaison des entretiens (récurrences)
- La mise au point de concepts pour nommer ces récurrences

Exemple : **Anselm Strauss,** *Awareness of Dying*, 1965 :

- **Termes « indigènes »**

La « maîtrise de soi » des infirmières : **« j'étais effrayée de ne plus me maîtriser quand la famille a commencé à pleurer au-dessus de l'enfant » maîtrise de soi.**

- **Concepts descriptifs** englobant les premières observations

Notion de « perte sociale » : **« il était si jeune »**, **« il allait être docteur »**, **« il avait une vie remplie »**, **« que vont faire ses enfants et son mari sans elle ? »**.

- Tri des nouveaux matériaux d'après ces premiers concepts pour formuler des hypothèses

Exemple : **Anselm Strauss**, ***Awareness of Dying*, 1965** :

- Attributs utilisés sont visibles, d'autres appris progressivement au contact du patient
- La perception de la perte sociale peut changer dans le temps
- Variation de la nature des soins en fonction de la « perte sociale »
- Faire des synthèses théoriques régulières pour faire évoluer vos hypothèses et élaborer le fil conducteur de l'analyse empirique (= question centrale) :

Pas une question empirique, mais une « **mise en énigme de la réalité sociale** »

- Vos hypothèses sont-elles valables ?
- **Principe de la saturation** de l'information (A. Strauss) : pas de nouvelles infos, concepts, chaque nouvel entretien simplement un cas particulier.
- Il faut mettre en danger votre modèle, chercher les **cas négatifs** qui ne rentrerais pas dans les hypothèses.

En quoi la nôtre est acceptable et rigoureuse ?

Exemple : A. Lindesmith, *Opiate Addiction*, 1949 : étude sur les héroïnomanes. Mis en évidence de l'expérience du « flash » dans l'addiction. **Cas négatif** : les patients ayant reçu de la morphine sans le savoir (pas d'addiction)

Ceci explique la pratique de l'échantillonnage raisonné pour constituer le corpus d'entretiens

1) Le travail de comparaison systématique et de construction du plan

Travail d'application de la grille d'analyse finale à l'ensemble du matériau.

Deux opérations :

- Une **comparaison « horizontale »** systématique des entretiens
- Une **comparaison « verticale »** (cohérence de la trajectoire)