

Ésta es exactamente la forma en que aparecen los datos si hay una tarjeta perforada para cada unidad, una columna para cada variable y una perforación para cada valor, y se pasan las tarjetas por una máquina que registra para cada tarjeta lo que está perforado en ella. Además, ésta es la forma que reviste el tipo de datos que tiene un director de escuela en su registro, con los nombres de los alumnos puestos verticalmente, las materias enseñadas en la escuela puestas horizontalmente y la calificación obtenida por el estudiante  $n_i$  en la materia  $n_j$  puesta en la celda apropiada.

En términos técnicos, la matriz de datos nos da para cada par en el producto cartesiano  $O$  por  $S$  el valor correspondiente en  $R$ .  $O_i$ , en otras palabras, la idea es, simplemente, que debe haber un valor  $R$  para cada combinación  $O_i S_j$ . La forma misma de la matriz de datos y esta definición nos conducen inmediatamente a los primeros 3 principios de la recolección de datos:

- 1) Principio de comparabilidad: la proposición '( $O_i, S_j$ ) da el valor correspondiente en  $R_k$ ' debe ser verdadera o falsa para cada  $i, j$  y  $k$ .
- 2) Principio de clasificación: para cada estímulo  $S_i$  la serie de las categorías de respuestas  $R_{ik}$  debe producir una clasificación de todos los pares ( $O_i S_j$ ) ( $i = 1, \dots, m$ ).
- 3) Principio de integridad: para cada par ( $O_i, S_j$ ) debe hallarse empíricamente un valor  $R_k$ .

Estos tres principios exigen algunos comentarios:

Si se observa una columna cualquiera de la matriz  $M$  en la tabla 1.1.3, se ve que la idea es, simplemente, que el mismo estímulo debe ser presentado a todas las unidades de análisis, por ejemplo, se debe formular la misma pregunta a todos los interpelados. Lo importante es que sea la misma pregunta; el estímulo no debe ser cambiado de un objeto a otro. Esto, sin embargo, es solo lo que ya estaba en la definición de la matriz. El principio de comparabilidad agrega ahora que cada una de tales combinaciones de un estímulo determinado con una unidad de análisis debe tener sentido, es decir, debe ser verdadero o falso que una combinación ( $O_i S_j$ ) da un valor determinado  $R_k$ . Por medio de este principio se hacen comparables los estímulos, las unidades y las respuestas. Si  $O_i$  es una nación,  $S_j$  una pregunta directa acerca del estado civil y  $R_k$  la lista de los elementos "casados, solteros, viudos y divorciados", entonces cualquier combinación no es ni cierta ni falsa, sino que simplemente no tiene sentido, como no lo tendría preguntar acerca del sexo de una nación.

Por lo tanto, se obtiene la comparabilidad cuando las tres series se ajustan las unas a las otras de tal manera que se satisfaga la condición mencionada en el principio 1). Esto quiere decir, por ejemplo, que las preguntas deben ser adecuadamente traducidas en una encuesta comparativa, pero no quiere decir que una pregunta en una encuesta 'debe tener el mismo significado' para todos los interpelados, siendo suficiente que tenga algún significado. La comparabilidad en nuestra acepción es un requisito menos exigente; sin embargo, significa que una línea puede ser comparada con otra porque las  $R$  en las celdas no carecen de sentido.

Si se descubre que dos unidades son diferentes en cuanto a una dimensión  $S_j$ , se puede descomponer la diferencia en dos componentes, uno 'real' y otro 'semántico', y atribuir la diferencia observada a alguna función de estos dos componentes. Se puede, incluso, sostener que toda la diferencia se debe a un componente semántico, como en el caso en que existe un mero desacuerdo verbal y no un desacuerdo en los hechos.

Puede también darse el caso de que los componentes semántico y real se anulen recíprocamente, de tal manera que el resultado neto sea una semejanza observada que puede calificarse de espuria. Pero todo esto es un problema de interpretación y no pertenece al campo de la recolección de datos, salvo si se ha decidido reducir al máximo el componente semántico. Debe notarse, sin embargo, que no es evidente por sí mismo que el factor semántico constituya un 'error' en un estudio tipo encuesta o que las tasas de criminalidad de las naciones no puedan ser comparadas porque se basan en códigos penales diferentes. Una frase ambigua es, con todo, una frase, y una investigación acerca de cómo reaccionan las personas o una frase de este tipo puede producir información importante acerca de cómo reaccionarían a tal frase en la vida diaria, y aunque las tasas de criminalidad variarían si se basaran en un criterio universal, se pueden llevar a cabo muchas investigaciones basadas en la comparación de tasas fundadas en definiciones diferentes, por la sencilla razón de que las tasas pueden decirnos algo acerca de cómo cambia de un país a otro 'eso que se llama crimen'.

El principio de clasificación agrega tres limitaciones más para la selección de las series  $S_j$ ,  $O_i$  y  $R_k$ . Lo que dice este principio es que para cada estímulo  $S_j$  la serie de respuestas  $R_{ik}$  debe formar una clasificación; por ejemplo, para cada par ( $O_i S_j$ ) debe haber una (exhaustividad) y solo una (exclusión mutua)  $R_{ik}$  en que el par pueda ser ubicado. (Idealmente, debe haber también un fundamentum divisionis.) Cuando se satisface este principio en la práctica, estamos seguros de que hay una, y solo una respuesta para cada posible combinación de unidades y estímulos. En la construcción de cuestionarios esto mismo se conoce como el principio según el cual cada interpelado debe marcar una, y solo una respuesta o categoría. Cuando se concede al interpelado la llamada 'selección múltiple'

ple', se le está dando en los hechos una serie de estímulos, uno para cada categoría, de tal manera que hoy dos respuestas para cada estímulo, aceptación o rechazo. Esto es una clasificación tan buena como cualquier otra, es una dicotomía.

De los tres principios mencionados, los dos primeros se refieren a la lógica de la matriz y a las tres series sobre las que se basa. El tercero, principio de integridad, se refiere al trabajo empírico de llenar la matriz, y dice simplemente esto: no dejar ninguna celda vacía. Este desideratum se anuncia fácilmente, pero no se alcanza tan fácilmente en la práctica, de tal manera que todo lo que contiene el principio es una norma según la cual la cantidad de celdas que contengan 'ninguna información', 'ninguna respuesta', etc., debe ser mantenida tan bajo como sea posible. Mientras se pueda cumplir con los dos primeros principios antes de la recolección de los datos, cuidando que los estímulos o las variables se apliquen a todas las unidades y que todos las series de categorías de respuestas formen clasificaciones, solamente se puede cumplir con el tercer principio en el proceso de la recolección de los datos y con posterioridad. Aquí entran en juego cosas tales como las técnicas para mantener bajas las tasas de rechazo a través de una legitimación adecuada de la investigación, o las maneras de rastrear datos casi imposibles de obtener. Pero existe también la posibilidad post hoc de examinar la distribución de las celdas vacías en M y de eliminar las O y S más adecuadas para liberarse del máximo de celdas vacías con el mínimo de suposiciones. Como regla general de tipo práctico, puede decirse que un 10% es el máximo absoluto de celdas vacías admisibles en cualquier columna o cualquier fila de M y que un 5% es un máximo más aconsejable.

En algunos casos, sin embargo, puede tener algún valor mantener una O o una S en la matriz aun infringiendo los dos principios relativos a las celdas vacías; una razón es que 'ninguna respuesta' también puede ser una respuesta y tratárela como tal.

Generalizando, la técnica consiste, por lo tanto, en contar el número de las 'DK's y de las 'NA's para cada O y para cada S y obtener su distribución. Muy a menudo, un pequeño porcentaje de las O y de las S absorberá un gran porcentaje de las 'DK's y de las 'NA's, y, naturalmente, éstas son nuestras candidatas obvias para el rechazo. Solo en raras oportunidades las 'DK's y las 'NA's estarán distribuidas al azar en la matriz de datos.

Si, como de costumbre, m es el número de unidades, n el número de las dimensiones y r el número promedio de las categorías de respuestas, la cifra  $W = m \times n \times r$ , o, mejor, escrita de este modo: (m, n, r), es la

\* 'DK', forma abreviada de don't know, no sé; 'NA', forma abreviada de no answer, falta de respuesta. (N. del T.)

característica de la matriz. W es una cifra que da una idea de cuánto trabajo hay en un proceso de recolección de los datos, aunque debe darse más importancia a m y a n que a r; r es una medida de cuánta medición discriminatoria se hace y, por lo tanto, debe ser igual, por lo menos, a 2. Si  $r = 1$ , no se puede hacer ninguna discriminación entre diferentes unidades y la dimensión no tiene valor descriptivo ni analítico, puesto que el resultado siempre es conocido y conocido como siendo siempre igual. Por cierto, r puede variar y en general varía de variable en variable, de tal manera que r debe ser concebida como algún tipo de valor promedio entre los diferentes 'r's.

La elección básica en cualquier plan de investigaciones tendrá que referirse a los números m y n. Un proyecto con las dimensiones (300, 20) es obviamente muy diferente de un proyecto del tipo (20, 300). El investigador tendrá que elegir. Llamaremos al par ordenado (m, n) la estrategia de investigación del proyecto; el investigador tiene que elegir su estrategia para atacar el problema que encara y la próxima sección se dedica a hacer algunas reflexiones acerca de diferentes tipos de estrategias.

Por último, diremos algunas palabras acerca de cómo r depende de m y de n. No hay ninguna razón por la que deba depender de n: agregar una variable no cambia el número de valores de las variables precedentes. Sin embargo, se puede considerar a r como dependiendo de m. Estrictamente hablando, siempre podemos tener  $r \leq m$ , puesto que no necesitamos más valores que elementos o unidades por medir. Si tenemos una unidad, necesitamos un solo valor; no hay que especificar el resto (por lo tanto, la variable consistirá en ese valor y en todos los otros valores posibles agrupados). No obstante, en la medida en que m aumenta, r aumentará cada vez menos. Pronto se hará evidente la amplitud (range) y con más unidades habrá más dificultades en discriminar entre ellas y mayor tendencia a agruparlas. A menudo la variable utilizada no será estrictamente unidimensional, de modo que la ubicación exacta de las unidades con respecto a otras unidades será discutible, conflicto que solo puede ser resuelto estableciendo categorías más crudas, esto es, bajos valores de r. El investigador procura capturar la ubicación genérica y deja a un lado los detalles y los problemas que le vienen a la mente precisamente cuando se encuentra con muchas unidades.

## 1.2. El número de unidades y de variables

Los números m y n son de importancia crucial para la evaluación de cualquier proceso de recolección de datos, tanto desde un punto de vista teórico como desde un punto de vista muy práctico. El valor más bajo de m es 1, como en el caso en que se investiga a una sola persona o a una

mitivo de que se dispone cuando cumple las dos condiciones siguientes: en primer lugar, es necesario que se encuentre en sociedades cuya organización no es superada en simplicidad por ninguna otra; además, es necesario que sea posible explicarla sin hacer intervenir ningún elemento tomado de alguna religión anterior".<sup>5</sup>

Más adelante dice:

"En la base de todos los sistemas de creencias y de todos los cultos, debe necesariamente existir un cierto número de representaciones fundamentales y de actitudes rituales, que, a pesar de la diversidad de formas que unas y otras han podido revestir, tienen en todas partes la misma significación objetiva y cumplen en todas partes las mismas funciones. Son estos elementos permanentes los que constituyen lo que hoy de eterno y humano en la religión; ellos son todo el contenido objetivo - de la idea que se expresa cuando se habla de la religión en general".<sup>6</sup>

Aquí tenemos dos ideas importantes. En primer lugar, la división de una institución social en dos partes que pueden llamarse lo esencial y lo accidental, y la sencilla ecuación:

$$\text{Institución social} = \text{Parte esencial} + \text{elementos accidentales} \\ \text{cosa social}$$

Lo esencial, evidentemente, representa la condición si ne qua non, "la idea que se expresa cuando se habla de la religión en general". Durkheim expresa esta ecuación casi en nuestras mismas palabras cuando dice:

"La imaginación popular o sacerdotal no ha tenido todavía ni el tiempo ni el medio de refinar y transformar la materia prima de las ideas y de las prácticas religiosas; ella se muestra, por lo tanto, al desnudo y se ofrece por sí misma a la observación, la que no tiene que hacer mayor esfuerzo para descubrirla. Lo accesorio, lo secundario, los desarrollos de lujo no han sobrevivido aún a ocultar lo principal. Todo está reducido a lo indispensable, aquello sin lo cual no podría haber religión. Pero lo indispensable es también lo esencial, es decir, lo que nos importa conocer ante todo".<sup>7</sup>

En segundo lugar, tenemos la idea de que la parte esencial se encuentra en su forma más pura en las sociedades primitivas y, en general, mientras más primitiva la sociedad, mejor. En otras palabras, la idea es que la sociedad de alguna manera parte con la "materia prima", y a ésta se agrega, por

procesos de evolución y difusión, "lo accesorio, lo secundario, los desarrollos de lujo".

Puesto que nadie parece haber refutado una de las proposiciones básicas de los evolucionistas, que el cambio social tiende a producirse en dirección a formas más complejas y no a formas menos complejas (con división del trabajo y su concomitante, la complejidad creciente de la red de status, como una de las dimensiones), no se impugnará que algunas instituciones sociales puedan ser más simples a niveles más bajos de la organización social en general. Pero esto, aun siendo más defendible de lo que, por lo común, estamos dispuestos a admitir, no es lo mismo que decir que estas instituciones son más puras. La simplicidad no es lo mismo que la pureza.

Dos ideas muy fundamentales pueden facilitar la aceptación de la idea de Durkheim de trabajar con el caso puro, que le condujo al estudio intensivo de la tribu de los Avunta con datos de segunda mano. En primer lugar, la clásica distinción entre los universales y los particulares, que corresponde a nuestra distinción entre lo esencial y lo accidental, una distinción entre la idea general y las características adicionales no esenciales (tal como el famoso "color" del caballo, exactamente "qué" objetos se definen como sagrados, etc.). Si ahora introducimos la igualmente clásica discusión entre realistas y nominalistas, podemos pensar en continuo de actitudes en este conflicto, en vez de los dos extremos tradicionales: a saber, el extremo realista ante rem (los universales tienen una existencia independiente anterior a las cosas, que son sus manifestaciones) y el extremo nominalista in rebus (los universales no tienen tal existencia, son solamente nombres de abstracciones obtenidas a partir de las cosas). Las siguientes serían algunas posiciones intermedias entre ambos extremos.

Se puede adaptar la posición de que "las cosas" (por ejemplo, las denotadas de la idea, de los universales) pueden ser ordenadas según cuantos particulares tengan, partiendo desde las aproximaciones más puras a la idea misma y agregándole particulares, "lo secundario", para utilizar la terminología empleada, hasta llegar a las formas más complejas que exhiben más elementos accidentales. Admitido que sea posible llegar a una ordenación lineal o parcial de las "cosas", ello solo tendría interés si esta variable (el grado de "pureza") estuviera altamente correlacionada con alguna otra variable. En este punto se introduce la segunda idea cultural básica: la idea de que la variable está correlacionada con las etapas de la evolución, cualquier cosa que ésta sea. Aunque Durkheim no tuvo esta idea, tuvo la idea de "sociedades en que la organización no es superada en simplicidad por ninguna otra", y la idea de que es aquí donde habría que buscar para encontrar "las formas elementales de la vida religiosa". La dificultad resi-

La investigación puede a continuación extenderse legítimamente por la horizontal haciendo a un cierto número de personas la misma pregunta, o por la vertical, formulando a la misma persona un número de preguntas diferentes. Tardé o temprano, las dos dimensiones de las unidades y de las variables probablemente tendrán que ser juntadas, por lo menos implícitamente, y esto conduce al proceso en dirección a la diagonal, a un nivel alto o bajo, dependiendo más de los recursos que de la ideología metodológica.

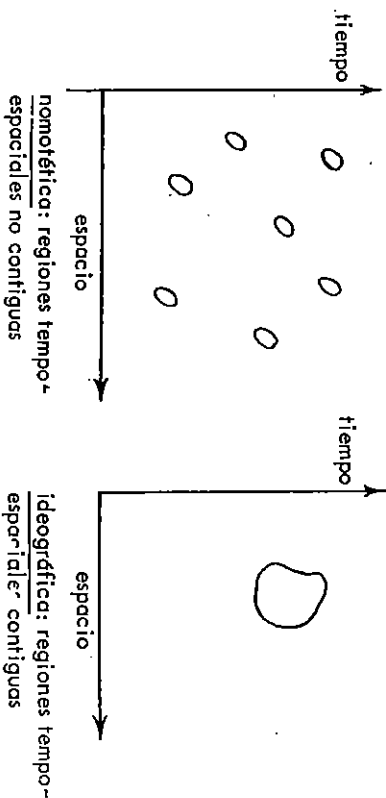
Puede ser que la comprobación última deba obtenerse por medio de proyectos ubicados en la diagonal o cerca de ella, pero también puede ser que las primeras ideas (y tal vez las perdurables) se obtengan cerca de los ejes. Por lo tanto, no confrontamos tanto las estrategias de investigación extensivas con las intensivas, como confrontamos las estrategias asimétricas con las simétricas. Una investigación muy intensiva tendrá probablemente un bajo rendimiento en valor comprobatorio en cuanto a proposiciones generales, y se supone que el propósito perseguido es una cierta generalidad, debido al bajo número de unidades. De la misma manera, una investigación muy extensiva no contendrá un número suficiente de variables para controlar condiciones relevantes y poner a prueba por lo menos algunos de las muchas hipótesis alternativas que tendrán que aparecer en la mente del investigador. Por lo tanto, la filosofía más fructífera en relación a la tabla parece ser una combinación de dos 'tanto como': tanto una investigación de tipo extensivo como una investigación de tipo intensivo, y, con tal fin, tanto una  $m$  grande como una  $n$  grande.

Si nuestro argumento de que las diferentes ideologías referentes a la elección de la estrategia de investigación representan diferentes fases en el proceso científico es correcto, debe esperarse que los sostenedores de estas diferentes ideologías tengan algún tipo de conexión con estas diferentes fases. No disponemos de datos empíricos al respecto, sino solamente la impresión más bien intuitiva de que el investigador 'orientado hacia la idea', en busca de perspectivas más que de comprobación, tiende a argumentar en favor de las estrategias que están lejos de la diagonal, y más a menudo en favor de las investigaciones intensivas, estudios de casos, etc. De la misma manera, el investigador 'orientado hacia la comprobación', menos interesado en las 'perspectivas' y más interesado en los 'dificiles hechos', se encuentra a menudo argumentando en favor de su especialidad, en la esquina superior derecha del cuadro, y a menudo también en favor de las estrategias extensivas. En este último caso, el investigador puede a menudo ser derrotado con sus propios argumentos, puesto que un número escaso de variables difícilmente proporcionará la firme base necesaria para explorar alternativas y, eventualmente, descartar  $n-1$  hipótesis en favor de la hipótesis  $n$ . Sea de ello lo que fue-

re, es de escasa importancia, puesto que lo que nos interesa es la metodología normativa y no la metodología descriptiva - nos interesa lo que los investigadores deben hacer y no lo que los investigadores hacen realmente. Señalamos solo al pasar que las ideologías simétricas en este campo parecen ser difíciles de acoger, por razones bien conocidas a los estudiosos de la tolerancia de la ambivalencia y, particularmente, porque el folklore de la ciencia contemporánea parece haber impuesto cierto tipo de oposición entre ambas: a menudo parece que hay que estar en favor de la investigación intensiva o en favor de la investigación extensiva, estar orientado hacia el 'caso' o hacia la 'encuesta'.

Esto mismo se aplica a la diferencia entre las llamadas ciencias nomotéticas ('generalizadoras') y las ciencias ideográficas ('singularizadas'): tal diferencia a menudo se exagera. Ambos tipos de ciencias desarrollan proposiciones y ambos tipos de ciencia conectan las proposiciones en teorías por medio de relaciones de implicación. En otros términos, ambos tipos tienen como su meta tanto la descripción como la explicación. No obstante, mientras las proposiciones de las ciencias ideográficas se refieren a regiones limitadas y antiguas tanto en el tiempo como en el espacio (Francia bajo Luis XIV, la ciudad-estado italiana durante el Renacimiento, etc.), las proposiciones de las ciencias nomotéticas se refieren a fenómenos que (por lo menos en principio) se pueden encontrar en una vasta área de regiones no contiguas en el espacio y en el tiempo ('La relación entre la industrialización rápida y la estabilidad política'). Por lo tanto, se puede representar la relación de la manera siguiente:

Diagrama 1.2.3. La relación entre las ciencias nomotéticas e ideográficas



Frases hechas tales como la diferencia entre 'decir muy poco acerca de - mucho' y 'decir mucho acerca de muy poco' no captan el punto esencial: Las proposiciones de las ciencias ideográficas programáticamente están limitadas a una región contigua (habitualmente un intervalo en la historia de una nación, o de algunos aspectos seleccionados de una nación), mientras que las proposiciones de las ciencias nomotéticas no conocen tales límites. Las regiones que son relevantes para su prueba a prueba habitualmente se encuentran dispersas por un amplio campo y no existe necesidad de llenar con conocimiento los huecos entre ellas.

Obviamente, el campo total de una monografía ideográfica puede ser más amplio que el campo total de una monografía nomotética. como cuando una persona escribe la historia de la revolución industrial en Gran Bretaña y otra persona pone a prueba hipótesis acerca de las relaciones entre el trabajador y el empresario en una cierta cantidad de establecimientos, en un período dado y en una nación dada. Obviamente también, los límites entre ambas de ninguna manera son precisos. Por el contrario, representan dos enfoques diferentes: la descripción concreta, detallada, realista y la explicación de algo que es significativamente interrelacionado, y la prueba a prueba de proposiciones generales, menos realista, más abstractas acerca de algo que se encuentra disperso y que no está interrelacionado en el mismo sentido.

Como sucede a menudo, las dicotomías tienen una tendencia a polarizar no solo el pensamiento sino también a la gente. Es tan obvio que ambas se complementan, que no necesita justificación: la ciencia generalizadora puede ser más o menos una ciencia singularizadora en cuanto a lo que hoy que buscar, que tenga interés teórico; la ciencia singularizadora puede dar a la ciencia generalizadora datos para poner a prueba sus hipótesis. Sin embargo, la diferencia no se encuentra simplemente a lo largo de un eje explicación-descripción. La ciencia ideográfica tiene la ventaja de dar explicaciones que son realistas en el caso singular, en el sentido de tomar en cuenta un número aproximadamente más suficiente de factores; la ciencia generalizadora tiene la ventaja descriptiva de examinar un campo más amplio, razón por la cual puede desarrollar categorías que tienen sentido en un contexto más amplio.

Esto es particularmente importante en relación con la muy a menudo discutida relación entre la ciencia social sincrónica (que trata fenómenos que tienen lugar en cualquier sitio en el espacio, pero dentro de un intervalo de tiempo relativamente estrecho) y la ciencia social diacrónica (que trata fenómenos que tienen lugar en cualquier punto del tiempo, pero dentro de una región limitada del espacio). Tal vez se puedan representar las relaciones entre ellas de la manera siguiente.

Tabla 1.2.4. La relación entre la ciencia social sincrónica y la ciencia social diacrónica.

	tiempo	
sincrónica		
sociología		
ciencia política		
economía		
antropología		
arqueología		
	diacrónica	espacio
	historia	

Sin embargo, la diferencia entre, por ejemplo, la sociología y la historia no se plantea meramente en términos de lo sincrónico versus lo diacrónico. La sociología tiende a ser más nomotética, la historia a ser más ideográfica. El conjunto total del trabajo sociológico puede cubrir relativamente bien el espacio con un conjunto de monografías interconectadas, una referente a las relaciones familiares en cualquier sitio, otra, a la estructura del poder en cualquier sitio, etc. Lo que falta es ciencia ideográfica, sincrónica: un conjunto realmente bueno de tratados acerca de la sociología de una región dada en el espacio, por ejemplo una nación. Esto se encuentra en antropología y en los estudios de comunidad, pero tal vez a costa de las ideas generales. De la misma manera, el conjunto total del trabajo histórico y de gran parte del trabajo antropológico puede llenar tanto el tiempo como el espacio, reuniendo trabajos acerca de regiones adyacentes y contiguas en el espacio y en el tiempo. Lo que falta aquí es la ciencia nomotética, diacrónica, que conecta todo este conocimiento dentro de marcos teóricos que no conocen limitaciones en el espacio y en el tiempo.

Es un deplorable estado de cosas el que mantiene separadas a la sociología y a la historia institucionalmente en las universidades, separando así no solo los enfoques sincrónico y diacrónico sino también, en una medida considerable, los enfoques nomotéticos e ideográficos. Se podría imaginar que estas paredes fueran derribadas para abrir el entero campo de las relaciones humanas en el tiempo y en el espacio, sin ninguna limitación, de manera concorde con el actual estado de la humanidad, con la emergencia rápida del hombre internacional y de una ciudadanía tiempo-espacial hasta ahora

desconocida que lleva a abrir incluso el cosmos. Es obvio que esto exige el trabajo de equipo y la cooperación institucional, y es probable que esto es haga más fácilmente en universidades nuevas, tal vez incluso en países en desarrollo, en que las desafortunadas prácticas académicas del pasado no han sedimentado en una costura demasiado gruesa, de modo que solo un terremoto intelectual pueda derribar las murallas. Teóricamente, sería posible adelantar una ciencia del hombre más general, explorando relaciones tanto en el espacio como en el tiempo, en regiones contiguas y no contiguas en los niveles explicatorio y descriptivo, todo esto dentro del mismo proyecto de investigación. Por cierto, esto requeriría una buena ciencia social sincrónica y diacrónica, descriptiva (sociografía, sociología descriptiva) y de ciencia social diacrónica, descriptiva (historiografía, historia descriptiva) de la que tenemos hoy en día. Sin embargo, no debería ser demasiado difícil hacerlo con los modernos métodos de almacenamiento de datos.

Puesto que al enfoque de este libro es principalmente nomotético, puede valer la pena dar más sustancia a estas ideas, destacando lo que significarían en términos de tarjetas IBM. La unidad clásica de la ciencia social sincrónica ha sido el individuo. Se da un gran paso adelante agregando unidades de nivel más alto, provincias, naciones, etc. Se da otro gran paso adelante por medio del análisis de pares, conjuntos de tres, etc. unidades a cualquier nivel (individual o colectivo), puesto que ello permite la inclusión de variables relacionales (véase 2.1). De este modo, el paso que sugerimos ahora es simplemente agregar la dimensión tiempo. Así, se puede considerar una nación como una unidad espacial, si no como "un trozo tiempo-espacial". Se pueden tomar los últimos 150 años de historia de la nación (siempre que haya existido durante ese período) y dividirlas en intervalos de, por ejemplo, 5 años, llegando a un total de 30 unidades espacio-temporales. Si se puede hacer lo mismo para, por ejemplo todas las naciones latinoamericanas, se obtendrán cerca de 600 unidades de análisis. Si se pudiera hacer lo mismo para todos los países de naciones latinoamericanas, incluso dentro de un período solamente de 50 años, se obtendrán  $10 \times (20) = 1.900$  unidades de análisis y se podría explorar la interacción de tiempo y espacio sobre las relaciones internacionales. Manteniendo constante la nación, se obtendría análisis diacrónico; manteniendo constante el tiempo, se obtendría análisis sincrónico; permitiendo la variación de ambas dimensiones, se obtendría el tipo de análisis (variación nomotética) hacia el que la ciencia social, sin duda, se está dirigiendo.

En consecuencia, los argumentos en favor de tener más de una unidad de análisis en ningún sentido son solamente argumentos aplicables a las encuestas y a los estudios de opinión pública aunque puede decirse que estos

La matriz de datos

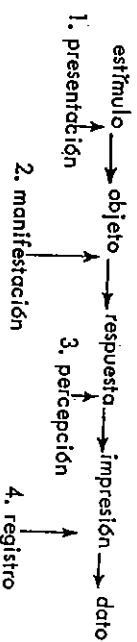
21

estudios han mostrado el camino, ya que eran más fáciles de llevar a cabo. Mas bien son argumentos en favor de una mayor imaginación en la elección de las unidades dentro de la región tiempo-espacial en que se pueden contener los datos y argumentos en favor de la destrucción de las barreras institucionales en la elección de tales unidades.

### 1.3. Observación e inferencia

Los datos en las ciencias sociales, como en otras ciencias, se fundan en observaciones de los sentidos y un criterio para ordenar los datos según su calidad se refiere a la proximidad del dato. Una marca en un cuadrado vacío puesto al lado de una respuesta es un dato y no lo es la inferencia de que el encuestado está en favor o en contra de algo. Un dato es un dato y no lo es la inferencia de que la persona sonriente está de buen humor, etc. Un dato es lo que es observado, manifiesto o fenotípico; no lo que es inferido, latente o genotípico. Lo que nos interesa en esta exposición de la recolección de los datos son siempre los datos manifiestos y nunca lo que está latente, lo que puede ser inferido. Aunque la palabra observación se refiere explícitamente a la percepción visual, se la utiliza aquí y en otras partes para referirse a todas las formas de percepción por medio de los sentidos utilizadas para registrar las respuestas tal como se presentan a nuestros sentidos. Sin embargo, hacemos una distinción entre una respuesta y un dato: una respuesta es cierto tipo manifiesto de acción, un dato es el producto del registro de la respuesta. El camino desde la respuesta al dato es torcido y presenta, entre otras cosas, dificultades tales como variaciones sensoriales intra e interpersonales, variaciones en el uso de símbolos para registrar las impresiones de las respuestas (el aspecto semántico), etc. Por lo tanto, es un desideratum evidente hacer tan corta como sea posible la última parte de la siguiente secuencia:

Tabla 1.3.1. La secuencia estímulo-dato



Una respuesta es observable, un dato es observado.

Las observaciones científicas, al revés de otras observaciones, deben cumplir dos importantes condiciones, y estas condiciones delimitan hasta cierto punto lo que puede ser considerado propiamente como datos manifiestos. Supóngase que el fenómeno observado es repetitivo y constante (por ejemplo, un hombre que da la misma respuesta a la misma pregunta experimental, etc.). Parece muy razonable exigir que las observaciones repetidas de tal objeto produzcan datos constantes, independientemente del hecho de que las observaciones sean realizadas por un observador o por observadores diferentes. Si repetidas observaciones de un fenómeno constante por el mismo observador producen datos constantes, se dice que la observación es intrasubjetiva o confiable. Si observaciones repetidas de un fenómeno constante por observadores diferentes producen datos constantes, se dice que la observación es intersubjetiva. (Debe notarse que en ambos casos estamos tratando con definiciones por reducción emplearse juntos como el definians del término 'objetividad', tan vagamente empleado. Tales requisitos permiten excluir de la esfera de la observación científica todas las observaciones que sean inestables, precipitadas y vagas, insuficientemente estandarizadas, y todas las observaciones que pertenezcan a la percepción subjetiva de una persona determinada solamente y que no pueden ser compartidas por otras personas. Obviamente, es necesario hablar de grados de confiabilidad y de intersubjetividad, y tal vez sea mejor decidir en cada caso particular por dónde pasa exactamente la línea divisoria.

La confiabilidad y la intersubjetividad se refieren a los dos componentes del proceso de la observación que hemos llamado percepción y registro, y no se refieren ni a la presentación del estímulo ni a la manifestación de las respuestas, que tienen lugar en el interpelado. Si un estímulo constante produce una respuesta constante o no, es un problema enteramente diferente: se refiere (por definición) a la constancia del objeto (la unidad de análisis) y no tiene nada que ver con el proceso de la observación. A menudo se considera deseable encontrar estímulos que produzcan fácilmente respuestas constantes, pero este desiderium tiene cierto cariz de irrealdad. En primer lugar, los estímulos pueden parecer constantes y no serlo, porque el contexto en que ellos aparecen releva para los objetos no es constante; por ejemplo, cuando se formula la 'misma' pregunta por segunda vez en un estudio de tipo 'panel'. En segundo lugar, si los estímulos dan lugar a respuestas constantes, esto puede ser una indicación de que los estímulos son demasiado crudos, puesto que las respuestas no reflejan las pequeñas oscilaciones que, podemos presumir sin riesgo, ocurren en el espíritu humano. Este último, además, es

La matriz de datos

un argumento para utilizar más de un estímulo para abarcar una misma dimensión, puesto que, por cierto, no existe un desiderium que exista constancia también para los objetos.

En resumen: además de los principios para la construcción de una matriz de datos, hemos agregado dos nuevos principios: 9

4) Principio de intrasubjetividad o confiabilidad: observaciones repetidas de las mismas respuestas por el mismo observador deben producir los mismos datos.

5) Principio de la intersubjetividad u objetividad: observaciones repetidas de las mismas respuestas por observadores diferentes deben producir los mismos datos.

En este punto a menudo se menciona un principio de validez, con la idea de que una observación es válida si se ha observado lo que se quería observar. En consecuencia, la validez, obviamente, se refiere a la relación entre lo manifiesto y lo latente. Puesto que, por definición, las observaciones se hacen al nivel manifiesto, podemos formular el principio de la manera siguiente:

6) Principio de validez: deben obtenerse datos de tal naturaleza y de tal manera que puedan hacerse inferencias legítimas acerca del nivel latente a partir del nivel manifiesto.

Puesto que, en verdad, los conceptos de manifiesto y latente no han sido definidos, dejaremos para más adelante otros argumentos y no haremos uso del principio de validez. En realidad, los datos siempre son válidos en algún contexto, con referencia a alguna dimensión latente, y esto debilita en algo el concepto de validez. La validez se transforma en una pregunta acerca de si los datos pueden ser utilizados exactamente en una pregunta acerca del propósito para el que fueron concebidos; una vez más, es un punto del que hoy que tener conciencia antes de la recolección de datos y que pueda ser de gran importancia, pero también desempeñar un rol menor, como en el caso de estudios más descriptivos y exploratorios.

#### 1.4. Clases de variables y de modelos de investigación

En los análisis de la ciencia social, las unidades de análisis más frecuentemente utilizadas son los individuos. El individuo puede ser caracterizado, y en el hecho ha sido caracterizado, por medio de un número de variables enomadamente grande. Tal número es alto por dos razones: el indi-

viduo es el nivel habitual de la interacción humana y, por consiguiente, es el nivel al que la necesidad de discriminar en la percepción y en la comunicación se siente más agudamente y está más desarrollada; en segundo lugar, la posibilidad de la interacción verbal con individuos hace posibles las respuestas verbales, con su inmensa variedad.

En medio de tal variedad, se requiere algún tipo de ordenación, alguna tipología. Sugierimos dos criterios para clasificar a las variables que caracterizan a individuos.

En primer lugar, el valor que un individuo tiene en una variable de terminada puede ser público o privado. Entendemos por una 'variable pública' aquella en que los valores individuales son conocidos por otras personas y se sabe que son conocidos. Algunos ejemplos: la edad, el sexo, la raza, la religión (a menudo), la ocupación, el ingreso (por lo menos en términos generales: véase la práctica establecida del acceso público a los registros de impuestos), la dirección, la familia de origen, los datos de recreación acerca de la familia, etc. Por 'variable privada' entendemos no variable en que los valores individuales, por cierto, pueden ser conocidos por otros, pero los demás no pueden tener pretensiones legítimas de conocer tales valores. Algunos ejemplos son el coeficiente de inteligencia, los proyectos para el futuro, si una persona es autoritaria o no, las actitudes y los patrones de conducta de diferentes tipos. Una persona puede hacer pública su opinión con respecto a la Comunidad Económica Europea o a la OTAN, pero esto no implica que su actitud sea una variable pública, porque también puede mantenerla fuera de la interacción humana.

En segundo lugar, el valor que una persona tiene en una variable puede ser permanente o temporal. Con esto último queremos significar que el conocimiento que se tenga del valor que una persona alcanza en tal variable en un momento determinado proporciona una base relativamente firme para formular una predicción acerca del valor que tal persona tendrá en la variable en un momento posterior; en cambio, si tal valor es permanentemente, la base para la predicción es muy firme. Obviamente, la línea divisoria entre lo que puede llamarse más propiamente un valor 'temporal' y lo que puede ser llamado un valor 'permanente' está lejos de ser clara, y lo mismo se aplica a la línea divisoria entre el valor 'público' y el 'privado': ¿cuánta gente debe tener una pretensión legítima de conocer un valor para que ese valor sea público? ¿El coeficiente de inteligencia, conocido por el psicólogo de la escuela, es una variable pública? Relativamente, ¿cuáles sistemas debe estimarse la 'pública'?, etc. Sin embargo, tales distinciones tienen un gran valor interpretativo.

Si las combinamos, obtenemos la siguiente tabla 2 x 2:

Tabla 1.4.1. Una tipología de variables para caracterizar a individuos

Valores públicos	Valores permanentes		Valores temporales	
	(variables de base)		imposible	
Valores privados	(variables de personalidad)		(variables de actitud y conductuales)	

Esta tipología presenta cuatro celdas, pero una de ellas está vacía debido a una proposición sociológica que puede ser formulada de la manera siguiente: la interacción institucionalizada está fundada en variables relevantes 'públicas' (por ejemplo, variables en que los valores de un individuo específico son conocidos públicamente). Y, por lo tanto, pueden ser legítimamente utilizados en la interacción; pero si las variables tienen que proporcionar una base para una interacción duradera, deben ser predecibles, lo que significa que los valores individuales tienen que mostrar un cierto grado de 'permanencia'. Los valores deben ser conocidos y reconocidos, lo que significa que las variables públicas no pueden ser temporales. Debe agregarse que un gran número de interacciones humanas no están institucionalizadas y tienen lugar sobre la base de indicios de un carácter más evanescente y efímero. Una característica de la interacción es que ella puede partir basándose en valores bastante privados e incluso temporales, pero en cuanto la interacción se pone en marcha y se vuelve institucionalizada, hoy un cambio en el carácter de los valores utilizados como base, en dirección hacia los valores públicos y permanentes. Un valor público jamás puede ser temporal, porque un valor temporal no proporciona una base suficiente para la interacción pública.

Se han puesto algunos términos en las restantes tres celdas. Ellos no deben ser considerados como completamente coextensivos con el significado que se ha dado a estas celdas por definición. Más bien, el significado de una 'variable de base' es el de una variable que es pública y permanente dentro de un sistema dado de interacción, por ejemplo, la sociedad noruega. Ser el presidente de una organización local no es un valor público cuando se toma como sistema de referencia a Noruega como un todo, pero 'la ocupación' lo es, debido a su relevancia en el contexto nacional. En el contexto de la organización local, sin embargo, puede ser irrelevante la ocupación, mientras que ser el presidente puede ser muy relevante. Variables de base típicas son la edad, el sexo, la ubica-



ción geográfica, la ocupación, etc. Cuando decimos que son permanentes, no negamos la posibilidad de que cambien; se trata solo de que son permanentes dentro de un período.

Las variables que son 'públicas y permanentes' colocan al individuo en la matriz social o, si se desea usar otro término, en la estructura social. El conjunto de valores en las variables de base da, para cada individuo, su personalidad social. Por tal razón, estas variables a menudo se llaman 'variables estructurales', término que preferimos no utilizar porque puede conducir a confundir los niveles de análisis. Las variables de base caracterizan a individuos, pero cuando reciben el nombre de 'estructurales' se piensa en la estructura de las sociedades.

Por 'variables de personalidad' entendemos aquellas variables que pueden ser desconocidas incluso para el individuo mismo y que exigen la habilidad de un psicólogo para ser descubiertas. Su característica esencial es la permanencia. Puesto que habitualmente se sostiene que las variables permanentes son de un carácter más bien latente, mientras que las variables manifiestas dan lugar a valores menos permanentes para los individuos, por lo común las variables de personalidad coincidirán con las que hemos llamado variables latentes. Debe destacarse, de nuevo, que 'permanencia' no significa que sea imposible todo cambio, sino que solo significa que hay intervalos suficientemente largos como para ser importantes, es decir, largos dentro de los cuales las fluctuaciones son insignificantes.

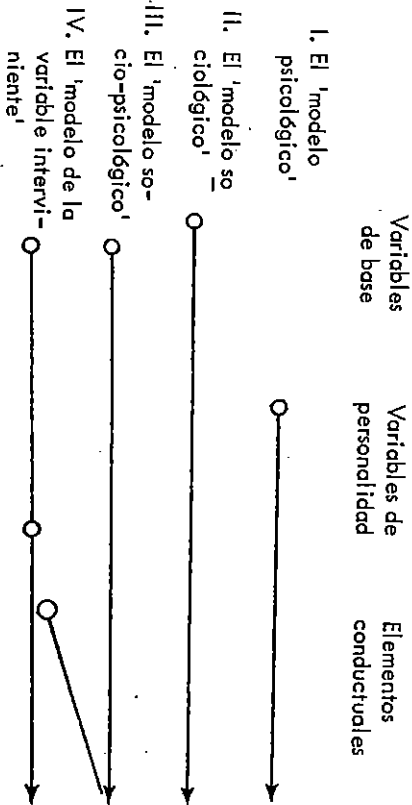
La tercera categoría contiene el resto, es decir, todos los tipos de variables acitudinales y conductuales u otras variables que puedan ser utilizadas para caracterizar a un individuo y que ni son públicas ni son permanentes. Muchos de los datos que se obtienen por medio de un cuestionario o de una entrevista caen dentro de esta categoría. Puesto que, de conformidad al esquema propuesto, el único factor que estas variables tienen en común es su carácter de variables temporales, no disponemos de un calificativo mejor que el de 'elementos conductuales'. Por lo tanto, los respuestas a entrevistas y a cuestionarios se clasifican como conducta (reacciones de palabra, conducta verbal).

En consecuencia, el valor de esta tipología reside en permitir una clasificación de las variables que caracterizan a individuos en variables de base, variables de personalidad y elementos conductuales. En realidad puede utilizarse un esquema muy similar, por ejemplo, respecto de las naciones; sin embargo, la aplicación más fructífera de este esquema reside en el campo del análisis de unidades que sean seres humanos.

Habitualmente, tales análisis se llevan a cabo con un cierto modelo en la mente, que puede haber sido explicitado por el investigador o no.

El razonamiento típico se desarrolla considerando a los elementos conductuales, que hemos calificado de 'temporales' como producidos o causados o siendo de alguna manera el resultado de las variables de base, de las variables de personalidad o de ambas. Esencialmente, los siguientes parecen ser los cuatro modelos más frecuentes en la literatura.

Tabla 1.4.2. Modelos para las relaciones entre los tipos de variables:



El investigador debe tener una imagen relativamente clara de lo que tiene en la mente. Los términos elegidos para calificar los diferentes modelos no son, una vez más, coextensivos con el contenido dado a los modelos por el diagrama, pero son más a menudo verdaderos que falsos. 10

La diferencia entre los modelos III y IV es la siguiente: en el modelo IV la idea es que una cierta condición social, definida por un conjunto de variables de base, produce (en la mayoría de los casos) cierta estructura de la personalidad, y esta estructura de la personalidad, a su vez, produce (en la mayoría de los casos) ciertos elementos conductuales. En cambio, en el tercer modelo las características de la personalidad no se consideran como peldaños en la escalera, por decirlo así, sino como causas independientes, que dan origen a los elementos conductuales juntamente con la característica del contexto social.

Supóngase que se han desarrollado índices muy poderosos de base social y de personalidad, de tal manera que ambas puedan ser analizadas extensamente por medio de una sola variable (por ejemplo, por medio de alguna medida de la dimensión 'centro-periferia' para contexto social y de

la dimensión 'mente abiertamente cerrada' para la característica de la personalidad; la selección de tales dimensiones, desde otros puntos de vista, incorpora gran parte del pensamiento contemporáneo. El siguiente sería un diseño poderoso para el estudio de los elementos conductuales, tanto manifiestos (acciones) como latentes (actitudes):

Tabla 1.4.3. Diseño para el estudio de los elementos conductuales

Variable de personalidad	Variable de base social					Total
	1	2	3	...	...	
1						
2						
3						
.						
.						
.						
b						
Total						

El diseño tendría que estudiar cómo varían los elementos conductuales al variar tanto el contexto social como la personalidad, y para hacerlo es necesario un diseño a x b, suponiendo que la variable social tiene a niveles y la variable de la personalidad b niveles. El diseño conduce por sí mismo tanto a interpretaciones del tipo III como del IV.

Para apreciar en qué consiste la idea del diseño, considérense algunos típicos diseños truncos que se usan a menudo. Modelos del tipo I y II utilizarían solamente los marginales horizontales y verticales, respectivamente, del diseño. Esto puede no ser objetable, aunque las relaciones serán probablemente más bajas de lo que podrían ser, debido a que en cada nivel social habrá una apreciable cantidad de variación en la personalidad que no se ha tomado en cuenta, y en cada nivel de la personalidad habrá una apreciable cantidad de variación social que no se ha tomado en cuenta. Sin embargo, se pueden obtener tendencias generales en términos de estructura social o de estructura de la personalidad. Lo que es realmente objetable es fundar el análisis en modelos del tipo I o II y utilizar muestras solamente de algunas de las columnas o

filas en el diseño. En la investigación sociológica esto significaría tomar muestras solamente entre gente con ciertas características de la personalidad; en la investigación psicológica significaría tomar muestras solamente entre gente con ciertas características de contexto social. El sociólogo se salva más a menudo de esta falacia debido a la dificultad de utilizar datos psicológicos como criterios de muestreo, pero el psicólogo sentirá a menudo la tentación de sacar las muestras más fácilmente asequibles o más extraordinarias, seleccionando en exceso las categorías sociales centrales y periféricas, respectivamente. Éste es uno de los puntos principales de la importante crítica metodológica a "La personalidad autoritaria" hecha por Hyman y Sheatsley.<sup>11</sup> Tal como ellos dicen en cuanto a "tomar una muestra de la heterogénea población nacional":

"..... tal población tiene una virtud, a menudo pasada por alto, que habría mejorado la calidad del análisis y que suscita algunas dudas en cuanto a ciertas conclusiones de "La personalidad autoritaria". La razón es que una muestra semejante revela la extensión y la ubicación social de las actitudes. Sin este conocimiento, la interpretación es difícil y a menudo peligrosa. Así, se puede perfectamente encontrar en una muestra nacional que una actitud específica es casi universal, y sin embargo esta misma actitud estudiada en un pequeño grupo homogéneo puede aparecer como desviada. Al no disponer de las normas nacionales, se podría intentar explicar el último hallazgo en términos de algún proceso de la personalidad idiosincrásico, cuando en realidad está muy correlacionado con un hecho social principal. .... pero si el investigador, debido al diseño de su muestra, no puede decir si una determinada actitud es normativa para la población en general, puede ir a buscar la explicación en un lugar equivocado".<sup>12</sup>

Se puede objetar al razonamiento presentado aquí que si la palabra 'grupo' en la frase "pequeño grupo homogéneo", se refiere a un grupo social y no a una categoría social, entonces lo raro es también lo desviado, incluso si es común en contextos mayores. Sin embargo, el argumento genérico es importante y se extiende no solo a las tasas y a los porcentajes sino también a las correlaciones, por ejemplo, a la organización de las actitudes:

"..... las investigaciones acerca de la organización de los sentimientos no pueden permitirse ignorar las consideraciones relativas al muestreo durante la preparación del diseño del estudio y el análisis de los resultados. Nos parece que el olvido de este asunto por parte de

los autores significa una metodología superficial y, desde un punto de vista práctico, tiende a perpetuar la implicación de que el nivel de organización de los sentimientos es una especie de universal, un proceso intrapsíquico que tiene escasa relación con las condiciones del contorno". 13

Demostraremos más adelante que el analista puede en muchos casos obtener casi cualquier correlación que desee entre las actitudes, dis-  
tensionando adecuadamente su muestra.

En la idea misma de tomar una muestra de una categoría social específica, por ejemplo, los famosos estudiantes que siguen un curso introductorio de psicología, se encuentra un caso, muy fácilmente reconocible, de pensamiento del tipo 'el caso puro' o del tipo 'una especie de universal, como lo expresan Hyman y Sheatsley. A continuación, los autores citan un gran número de investigaciones que demuestran que las correlaciones de actitud varían con factores tales como la edad ('los niños llegan a ser más consistentes en sus características morales según van creciendo') y la educación (una investigación que demuestra una correlación creciente entre el prejuicio y el autoritarismo con la educación). Obviamente, no son admisibles las generalizaciones desde una sola categoría social a toda la sociedad; de aquí se desprende la importancia del diseño completo.

En tal diseño no hay ningún supuesto implícito en el sentido de que el contexto social y la personalidad sean independientes; por el contrario, una tradición de investigación muy considerable en la psicología social demuestra que ellas son dependientes. Lo que importa es tratar de obtener datos que llenen todas las celdas, incluso cuando el diseño debe ser manejado de alguna de las maneras extremas (modelos I y II). Si se recogen datos sobre ambas variables básicas para todas las unidades, existe siempre la posibilidad de controlar el efecto de la segunda variable, de tal manera que se pueda ir en dirección a los modelos III y IV.

La relación entre las dos variables puede ser explorada de mejor manera por medio de un modelo de investigación de un V tipo:

Tabla 1.4.4. Un modelo para las relaciones entre los tipos de variables:

Variables de base	Variables de personalidad	Elementos conductuales
-------------------	---------------------------	------------------------



No hemos incluido este modelo entre los otros, puesto que no toca directamente los elementos conductuales, y obviamente tendré que estar basado en investigaciones realizadas en conformidad a los modelos III o IV, puesto que debemos presumir que las variables de personalidad son latentes; ya que los elementos de conducta son manifiestos, y que solo podemos obtener datos manifiestos. Este modelo debe dar por resultado una distribución de las características de la personalidad a lo largo de las categorías sociales y, por lo tanto, debe producir un tipo de conocimiento que es indispensable en las investigaciones que se llevan a cabo en conformidad al modelo IV.

Debe destacarse que la dirección de las flechas en los cinco modelos no debe ser tomada más que como una simple regla de carácter práctico. Así, se puede considerar a la personalidad como una variable independiente y a variables de base tales como la ocupación y la educación como variables dependientes -aunque se pueden conservar los modelos más fácilmente o, incluso, más fructuosamente, definiendo la búsqueda de la ocupación y de educación como un elemento conductual. En general, los modelos reflejan el estado actual del pensamiento en las ciencias sociales, pero ellos tendrán que ser refinados en los capítulos siguientes, de manera que sirvan como guías para la investigación.

Los modelos III y IV son interdisciplinarios y exigen poner en común las habilidades del sociólogo y del psicólogo, cuando las unidades son individuos. Por lo tanto, son modelos para equipos de investigación, a menos que se pueda presumir que un mismo investigador posee conocimientos adecuados, o que el sociólogo pueda facilitar al psicólogo y viceversa, instrumentos suficientemente buenos, y que puedan ser usados por individuos no especializados.

A menudo se hace una distinción entre variables independientes y dependientes. En matemáticas, estos términos se usan muy a menudo para indicar cierta asimetría en la manera como se presenta una relación funcional, pero significa muy poco en términos de significación matemática.

En las ciencias empíricas los términos solamente pueden ser utilizados si existe algún modelo, incluso del tipo más vago, en cuanto a las relaciones entre las variables estudiadas. Los modelos pueden tomar muchas formas; por ejemplo, los cuatro modelos básicos en la investigación sociológica mencionados más arriba sin embargo, ellos habitualmente tienen en común la idea de 'algo que de alguna manera produce otro algo'. En esta idea se ven claramente los dos elementos que conducen a las concepciones de las variables dependientes o independientes. Se pueden dar tres precisiones ulteriores:

- 1) Si el modelo es de alguno de los cuatro tipos discutidos más arriba (o de cualquier otro especie que abarque tipos diferentes de los tres tipos de variables), entonces la variable independiente es la variable más a la izquierda, con la dependiente a la derecha.
- 2) Si el modelo es causal, entonces la variable causal es la variable independiente, y el efecto es la variable dependiente.
- 3) Si existe un orden en el tiempo entre las variables, entonces la variable precedente es la variable independiente.

Por último, hay que mencionar un punto relativo a la elección de las variables en la sociología aplicada. La ciencia social aplicada debe dar una base para tomar resoluciones en favor del cambio social cuando algo 'anda mal' y en favor de la estabilidad social cuando algo 'anda bien', pero las resoluciones en sí mismas pueden ser formuladas por otras personas. La ciencia social aplicada se diferencia del discurso sociológico general solo en una cosa: debe formular sus conclusiones en términos de variables que sean manejables. Una relación causal en la que no existe posibilidad de manejar la variable independiente por razones técnicas, económicas, éticas o de otro tipo, es de escasa valor para el práctico. En los hechos, éste es un argumento contra la investigación social aplicada que se lleva a cabo en conformidad al modelo I señalado más arriba, puesto que los factores de la personalidad son tan difíciles de cambiar.<sup>14</sup>

Sin embargo, no es un argumento muy efectivo contra los modelos II y IV, particularmente en una cultura en que 'la ingeniería social' de alguna manera es considerada más practicable que la 'ingeniería humana'.<sup>15</sup> Cualquiera sea la filosofía que se tenga en cuanto a este punto, la elección del modelo básico debe ser considerada cuidadosamente antes de comenzar con un proyecto, y debe examinarse el propósito general del estudio antes de formular decisiones irreversibles en cuanto a la elección de las variables independientes.

## NOTAS DEL CAPÍTULO 1

- 1 Esto no significa que la solución sea trivial o que se alcance fácilmente. El químico tiene en su mente un caso ideal, 'puro' sufrido o cualquier otro sustancia que se define por tener solo un tipo de moléculas. Este hecho solo se puede indagar indirectamente, tomando una muestra del material y examinando el valor que alcanza una cierta cantidad de indicadores (características de fusión, refracción, dispersión, etc.). Sin embargo, los indicadores pueden no discriminar lo suficiente o la muestra puede ser demasiado exigente. La metodología no es demasiado diferente de la metodología de las ciencias sociales, pero el científico social a menudo no tiene ni siquiera un concepto de caso puro en su mente ni, mucho menos, un método para acercarse a él operacionalmente.
- 2 A menudo un número grande es necesario, incluso para 'sacar las diferencias'.
- 3 Sorokin, en las páginas 57 y siguientes de *Social and Cultural Dynamics*, Boston, Porter, Sagent, 1957, formula la siguiente observación contra la posición de que los eventos humanos son cíclicos: "No se puede narrar ningún proceso histórico sin admitir, explícito o implícitamente, que muchas características esenciales se repiten. Tomemos, por ejemplo, la religión romana. Si cualquier momento de cualquier evento histórico o proceso socio-cultural fuera único, inmediatamente surgiría una dificultad: ¿qué momento de la religión romana hoy que describir? ¿Debe ser la religión romana de los 20 del día 1 de julio del año 321, q.C. o la de las 7 del día 10 de mayo del año 322 a.C.? Hoy invariantes en la "unicidad" pues de otro modo no sería posible la descripción; pero también hay variaciones, y nuestro enfoque propondría cortar el proceso en trozos adecuados, unidades, y recoger cuanto información sea posible acerca de cada uno, incluyendo el tiempo. Cortes distintos producirán resultados diferentes y habrá que combinar tales resultados. Del mismo modo, en la ciencia social sincrónica: se harían cortes en el espacio (véase 2.2.) y se explorarían analíticamente las variaciones. Por lo tanto, para poner a prueba hipótesis se obtendría una muestra de las unidades de interés y se sometería sistemáticamente a prueba la hipótesis. McClelland, en *The Achieving Society*, contrasta hábilmente esto con la metodología tradicional de la cita-illustración, dentro de la cual se escogen los casos que están de acuerdo con la hipótesis y se rechazan hipótesis si se encuentra un solo caso desviado. Ésta es una concepción muy ingenua acerca de las proposiciones de la ciencia social; si solo se admitieran correlaciones perfectas, la ciencia social no iría muy lejos.
- 4 Paris, 1912.
- 5 *Ibidem*, pág. 1.
- 6 *Ibidem*, pág. 6.
- 7 *Ibidem*, pág. 8.
- 8 Una exposición de este punto, al nivel introductorio, se encuentra en *Methods in Social Research*, New York, McGraw-Hill, Goode y Hart, 1952; Cap. 5, y especialmente pág. 42.
- 9 Una exposición de este punto puede encontrarse en Johan Galtung, "An Inquiry into the Concepts of 'Reliability', 'Intersubjectivity' and 'Constancy', *Inquiry*, 1959, pp. 107-125.

**A tradução de conceitos complexos para o entrevistado**

No pesquisa realizado pelo Doria Uff para o Ceop, foi feita uma pergunta sobre o opinião dos pessoas — se favorável ou não — quanto à reserva de vogos para negros nas universidades. Primeiro, perguntou-se ao entrevistado se ele já tinha ouvido falar dessa política. Em caso afirmativo, perguntou-se se ele era contra ou a favor. Para aqueles que nunca tinham ouvido falar no assunto, explicou-se o que era o reserva de vogos e, então, perguntou-se o ele se era contra ou a favor.

A explicação funcionou. Isso foi comprovado pelos resultados. Nos dois casos — os que já conheciam e os que precisaram da explicação —, o nível de apoio à reserva de vogos foi idêntico, variando dentro do margem de erro do pesquisa.

**Conclusão**

É possível formular as perguntas de um questionário de modo a que produzam resultados que agradem aos clientes da pesquisa? A partir do que foi exposto, pode-se concluir que há formas suítes e não-suítes de se elaborar perguntas que produzam os resultados desejados por clientes ou pesquisadores. Um coordenador de pesquisa inescrupuloso que trabalhe para um candidato que tenha como plataforma a melhoria dos serviços de educação pode formular um questionário que apresente resultados sistematicamente enviesados em favor dessa plataforma. Basta que as perguntas sobre a importância relativa desse tema sejam colocadas após uma bateria de perguntas que trate de educação e, em particular, dos problemas dessa área. Mas não é necessário que um coordenador de pesquisa aja de má-fé para que surjam problemas de erro sistemático de medição que prejudiquem as conclusões de uma pesquisa, orientando de forma errada quem a usa. Basta que se cometam erros técnicos. O que o contratante e os usuários de pesquisa podem fazer para evitar esse problema? A sugestão é que o contratante tome as seguintes providências:

1. Tenha acesso à íntegra do questionário de pesquisa a ser utilizado no trabalho de campo.
2. Leia com atenção o questionário, procurando identificar — mesmo como leigo — possíveis fontes de erros sistemáticos.
3. Solicite ao coordenador da pesquisa uma justificativa técnica para a redação de cada pergunta e para o ordenamento das mesmas no questionário.

4. Solicite ao coordenador da pesquisa, caso julgue necessário, a realização de um pré-teste completo do questionário, desde a aplicação até a tabulação dos resultados.

Note-se que, se todas essas sugestões forem seguidas pelo contratante, o custo da pesquisa provavelmente aumentará. Isso é previsível, uma vez que o controle e a qualidade de uma pesquisa acarretam gastos. O espírito das sugestões é recomendar aos contratantes de pesquisas que criem incentivos para que institutos e coordenadores de pesquisas não trabalhem no "piloto automático", exigindo que constantemente avaliem seu próprio trabalho.

Já o cidadão comum que tem acesso às pesquisas por meio da mídia deve buscar pelo menos uma informação: a redação, na íntegra, das perguntas cujos resultados são divulgados. É comum os jornais e os demais meios de comunicação, por motivos de espaço e para manter a atenção de seu público, não apresentarem as perguntas tal como foram formuladas e, sim, apenas seu conteúdo de maneira sucinta e atraente. Trata-se de uma divulgação correta do ponto de vista jornalístico, mas errada do ponto de vista científico. O aperfeiçoamento da utilização das pesquisas pelo grande público depende em grande medida da melhor divulgação delas pelas empresas de jornalismo.

**O que ler sobre os assuntos tratados neste capítulo**

- O livro de Robinson, Phillip e Lawrence é especializado na medição de valores sociais e políticos. Os demais apresentam capítulos práticos e interessantes sobre elaboração de questionários.
- Asher, H. *Polling and the public*. Washington, D. C., CQ Press, 1998.
  - Babbie, Earl. *Métodos de pesquisas de survey*. Belo Horizonte, UFMG, 1999.
  - Fowler Jr., Floyd J. *Survey research methods*. London, Sage, 1993.
  - Robinson, J.; Phillip, R. S. & Lawrence, S. W. *Measures of political attitudes*. New York, Academic Press, 1999.
  - Traugott, M. & Lavrakas, P. *The voter's guide to election polls*. New York, Chatham House, 2000.
  - Zaller, J. R. *The nature and origins of mass opinion*. Cambridge, Cambridge University Press, 1998.