
PONENCIA IV

ANÁLISIS DE DATOS CUALITATIVOS EN LA INVESTIGACIÓN SOBRE LA DIFERENCIACIÓN EDUCATIVA

por

Eduardo García Jiménez, Javier Gil Flores, Gregorio Rodríguez Gómez
Universidad de Sevilla

EL DATO CUALITATIVO EN EL ESTUDIO DE LA DIFERENCIACIÓN EDUCATIVA

El estudio del hecho diferencial en educación, y por consiguiente la formulación de propuestas educativas que respeten las diferencias, se enfrenta en la actualidad con problemas que no pueden abordarse desde el mero análisis de las diferencias individuales. La comprensión de cuestiones como el rendimiento académico del estudiante no puede hacerse sin incluir —junto a los elementos tradicionalmente considerados— al centro escolar como unidad de análisis y, por ende, a aspectos tales como el clima y la cultura escolar o el intercambio con el entorno y el medio social. Del mismo modo, la intervención diferenciada requiere hoy propuestas que consideren la acción educativa en su conjunto, incluyendo planteamientos organizativos y pedagógicos: programas formales con estrategias de enseñanza adaptadas; actividades educativas no formales e informales; y esquemas organizativos que mejoren el clima de relaciones en el aula y el centro, y faciliten la implicación de la comunidad.

La explicación de estos nuevos y viejos problemas o la mejora de esas situaciones educativas, posiblemente esté reclamando un nuevo marco conceptual y metodológico desde el que entender la problemática de la diferenciación educativa. Así, algunos investigadores están proponiendo aproximaciones conceptuales que se fundamentan en planteamientos sociológicos o antropológicos y aproximaciones metodológicas que se apoyan en planteamientos de investigación que podríamos identificar como cualitativos. Bartolomé Pina (1992), al analizar el problema de las diferencias culturales, reclama un tipo de investigación cualitativa que, sobre todo,

asegure que las categorías culturales a introducir en la investigación sean comprendidas por los sujetos, que tenga en cuenta el pluralismo cultural y lingüístico, que triangule datos, observaciones y métodos en diferentes contextos; insistiendo en la adecuación de la observación persistente y prolongada como medio de acercamiento a una comprensión más profunda del fenómeno multicultural. En el campo de la integración de los sujetos con necesidades educativas especiales, Bodgan (1986) reclama una aproximación sociológica y el uso de la investigación cualitativa como medios para abordar el análisis del contexto donde se desarrollan los procesos educativos; y Hegarty (1988) considera que procesos como la organización de los servicios de educación especial o el desarrollo de los proyectos de integración son excesivamente complejos como para ser controlados, existiendo variables claves que no pueden ser cuantificadas y dificultades para formular hipótesis en los términos adecuados para la utilización de diseños experimentales.

La idoneidad de las estrategias de investigación cualitativa para dar respuesta a algunos de los problemas que tiene hoy planteados la diferenciación educativa, está también refrendada por estudios empíricos realizados desde campos tan diversos como el bilingüismo, las diferencias debidas al sexo, la educación multicultural, la atención de los déficits socioculturales o el estudio de las diferencias individuales en razón de los procesos cognitivos.

Desde la perspectiva de la educación multicultural, en una revisión realizada sobre campos como la integración del contenido, los procesos de construcción del conocimiento, la reducción del prejuicio, la pedagogía de la equidad y el desarrollo de la cultura escolar, se recogen diferentes estudios que utilizan enfoques etnográficos al aplicar técnicas tales como entrevistas, documentos, observaciones...etc. (Banks, 1993):

Dentro de la educación infantil y en el estudio de problemas relativos a la intervención temprana, la educación de los padres, la continuidad y coherencia entre ciclos y niveles, la evaluación del contexto en que tiene lugar la interacción propia de todo proceso de enseñanza-aprendizaje, las áreas de contenido y las políticas educativas en general, se están planteando formas de indagación intrapersonales y centradas en el contexto, tales como estudios de caso, observaciones estructuradas, entrevistas... etc. (Bowman, 1993): En el campo de la educación bilingüe, autores como Fillmore y Valadez (1986) consideran que los métodos de investigación etnográficos resultan adecuados para estudiar los modelos de interacción social y lingüística entre los alumnos y los profesores en clases bilingües. Asimismo, en los estudios sobre compensación educativa se destaca la importancia de la investigación-acción como un componente necesario en la planificación, ejecución y evaluación de los programas de educación compensatoria (Espín, 1991).

Las investigaciones sobre las diferencias individuales en función del sexo, realizadas por autores como Alberdi (1985, 1986), Alberdi y Escario (1987), utilizan técnicas cualitativas como los grupos de discusión para analizar las actitudes y comportamientos de los profesores; o autores como Brullet y Subirats (1990) abogan por la necesidad de contar con la «investigación-acción» como medio funda-

mental en el proceso de transformación de la realidad educativa sexista; a lo que habría que añadir la necesidad de contar con medios y estrategias que nos acerquen a un mayor conocimiento de lo que sucede en las aulas (Caselles, 1991).

La investigación sobre la eficacia escolar o sobre la mejora de las escuelas, ya desde su inicio, ha utilizado estrategias metodológicas de corte cualitativo, recordemos en este sentido que uno de los primeros trabajos realizados fue el estudio de casos llevado a cabo por parte de Weber (1971): En los últimos años se viene reclamando la necesidad de utilizar estrategias metodológicas que saquen a la luz la especificidad de las características idiosincráticas de cada centro educativo, para adaptar su educación a las características contextuales donde se inserta (Rodríguez, 1992).

Esta progresiva inclusión de dimensiones cualitativas en la investigación sobre la diferenciación educativa, está propiciando nuevas formas de evidencia (datos) más contextualizadas y holísticas, próximas a la realidad y a la práctica, y que ofrecen una información más rica, específica y matizada. Este tipo de evidencias que, en adelante, identificaremos como datos cualitativos constituyen la base sobre la que se sitúa el desarrollo de esta ponencia.

El dato cualitativo puede definirse como una elaboración primaria que nos informa acerca de la existencia de una realidad, sus propiedades o el grado en que éstas se manifiestan. Dada su naturaleza de aproximación inicial a la realidad, el dato cualitativo ofrece una información básica con un relativamente bajo nivel de estructuración teórica. De hecho, los datos cualitativos se nos presentan como simples registros o asientos que dan fe de una realidad o informan de sus cualidades¹.

Mientras el número indica la cuantía o grado de las diferencias educativas, el dato cualitativo permite explicar la diferenciación desde un contexto holístico, que aborda la realidad en su conjunto sin separar los elementos que la explican. El dato cualitativo supone partir de una realidad educativa inicial sin apenas modificaciones; es, por tanto, una elaboración de bajo nivel, pues al investigador puede no interesarle modificar la realidad hasta no conocerla tal como era inicialmente. El dato cualitativo, además, realiza una explicación contextual de las diferencias permitiendo profundizar en los matices, contradicciones, anécdotas, peculiaridades, particularidades de personas, grupos, instituciones, etc.

El dato cuantitativo establece una diferenciación parcial, en función de un número limitado de características medidas en un momento preciso de la vida del sujeto; el dato cualitativo podría permitir una diferenciación global, en función de un elevado número de rasgos conocidos del sujeto, el grupo, la institución, su historia y trayectoria a lo largo del tiempo, significado y consecuencias de las características o de su contexto. El dato cuantitativo conduce a propuestas de adaptación educativa

¹ El dato cuantitativo por contra es altamente deductivo, una vez codificado no hay manera de recuperar la manifestación conductual original. Es como si el dato hubiese perdido la memoria de la conducta, situación y contexto de los que procede.

basadas en determinados tipos y rasgos previamente identificados, sobre la base de un individuo o grupo normativos (adaptación de carácter nomotético); el dato cualitativo lleva a propuestas de adaptación educativa acordes a las necesidades de cada sujeto concreto, teniendo en cuenta todas sus características individuales y su contexto socioeducativo (adaptación de carácter ideográfico).

Precisemos la definición de dato cualitativo a partir de dos de sus componentes, el nivel de elaboración y la forma de registro, y analicemos cómo se interpreta la naturaleza de este tipo de datos desde diferentes posiciones de investigación.

Desde determinados programas de investigación o «colegios invisibles», que cabría identificar con planteamientos positivistas o neopositivistas, el bajo nivel de elaboración teórica presente en los datos cualitativos responde a la ausencia misma de un cuerpo teórico suficientemente contrastado, de suerte que el dato cualitativo es entendido como una manifestación del nivel protocientífico en que se encuentra determinado conocimiento de la realidad. Desde esta perspectiva, si existen datos cualitativos es porque nuestra capacidad explicativa no nos permite una formulación de problemas más fundamentales, un acercamiento más selectivo a la realidad o mediciones más precisas de un objeto de conocimiento.

No conocemos los datos físicos más directa u objetivamente que los datos sociales, excepto en la medida en que hemos desarrollado instrumentos y símbolos más adecuados para registrar los primeros (Lundberg, 1968. Citado en Tejedor, 1988b:15):

Desde otros programas de investigación, que podrían identificarse con esquemas de trabajo antropológicos o sociológicos, el bajo nivel de elaboración teórica presente en los datos cualitativos es una consecuencia de la posición de partida adoptada por el propio investigador: la construcción del conocimiento es una tarea social o una consecuencia de la interacción entre un sujeto y la unidad social de la que forma parte, no el resultado de una elaboración científica previa. El conocimiento que se genera desde este tipo de programas se presenta a menudo bajo la forma de esquemas explicativos o interpretativos, generalmente de carácter operativo o práctico, que pueden llegar a tomar la forma de teorías de la acción o teorías operativas, es decir, aquellas cuya intencionalidad es la de servir de guías para el diseño de la acción educativa.

El registro de los datos cualitativos permite que la información contenida en los mismos pueda ser efectivamente manipulada y comunicada para responder a un problema de investigación. Para ello pueden emplearse distintos soportes físicos o formas de lenguaje que permitan registrar y expresar la información elaborada. Los datos cualitativos pueden quedar registrados en forma textual: transcripciones de entrevistas, notas de campo, diarios, y otros documentos. Incluso cuando los datos aparecen elaborados inicialmente en forma de imágenes o sonidos, suelen ser traducidos a una forma textual para proceder a su análisis. No obstante, es habitual que los datos cualitativos recogidos inicialmente en forma de textos puedan ser expresa-

dos con posterioridad mediante otra forma de lenguaje. El recuento de la aparición de determinados términos o tópicos dentro de un conjunto de datos textuales, llevaría a construir datos cualitativos expresados numéricamente.

Los datos cualitativos de naturaleza textual son los que pueden realizar un aporte más novedoso y conceptualmente importante en el estudio de la diferenciación educativa. Nos referimos a los datos cualitativos registrados inicialmente en forma textual, o transformados en textos, bien se mantenga su naturaleza textual a lo largo del proceso de análisis o bien se transformen posteriormente en valores numéricos. Son precisamente este tipo de datos los que favorecen la explicación contextualizada de las diferencias, los que permiten profundizar en los elementos singulares de la diferenciación educativa, o los que aportan una visión más global y/o integrada de este problema. Los datos cualitativos registrados directamente mediante símbolos numéricos no realizan una aportación diferente a la que representan los datos cuantitativos recogidos mediante tests, inventarios, escalas, cuestionarios, etc.

ANÁLISIS DE DATOS CUALITATIVOS

Los datos que obtiene el investigador no suelen ofrecer suficiente información para resolver los complejos problemas planteados en la investigación sobre la diferenciación educativa, si no son organizados y manipulados en alguna forma. La información que aportan los datos *per se* no permite al investigador establecer relaciones, interpretar y extraer significados relevantes de cara al problema abordado, si no se realizan determinadas actividades y operaciones. En este sentido amplio solemos hablar de análisis de datos.

Las ideas básicas presentes en el concepto de análisis sostenido por una mayoría de autores hacen referencia al examen de las partes de algo por separado y de las relaciones existentes entre ellas, para reconstruir un todo significativo (Spradley, 1980): El análisis de datos es una actividad que implica un *conjunto de manipulaciones, transformaciones, reflexiones y comprobaciones realizadas a partir de los datos con el fin de extraer significado relevante para un problema de investigación*. Según Miles y Huberman (1984b), todas estas acciones presentes en un proceso de análisis de datos pueden resumirse en tres tareas básicas estrechamente relacionadas: reducción de la información, disposición o presentación de la misma, obtención de resultados y verificación de conclusiones (ver Figura 1):

Tradicionalmente se suele caracterizar el proceso de investigación por su linealidad temporal. Las etapas que lo componen tienen un sentido secuencial en el que el análisis de datos constituye una fase que sigue a la recogida de los datos y precede a su interpretación. Sin embargo, a menudo, en las notas de campo y junto a la descripción de situaciones educativas suelen aparecer juicios, opiniones, sospechas, dudas, reflexiones e interpretaciones del investigador, haciendo que el análisis y la elaboración de los datos se alternen o superpongan en el proceso de investigación. Lofland y Lofland (1984:132) presentan de un modo gráfico el desarrollo simultáneo de análisis y recogida de datos (ver Figura 2): Para estos autores, la

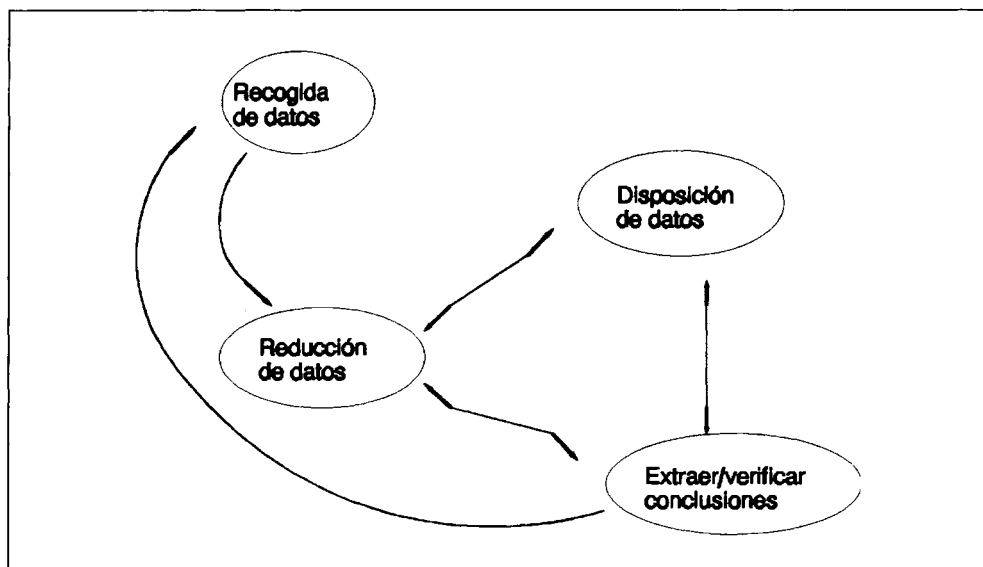


Figura 1
Componentes del análisis de datos (Miles y Huberman, 1984b: 23).

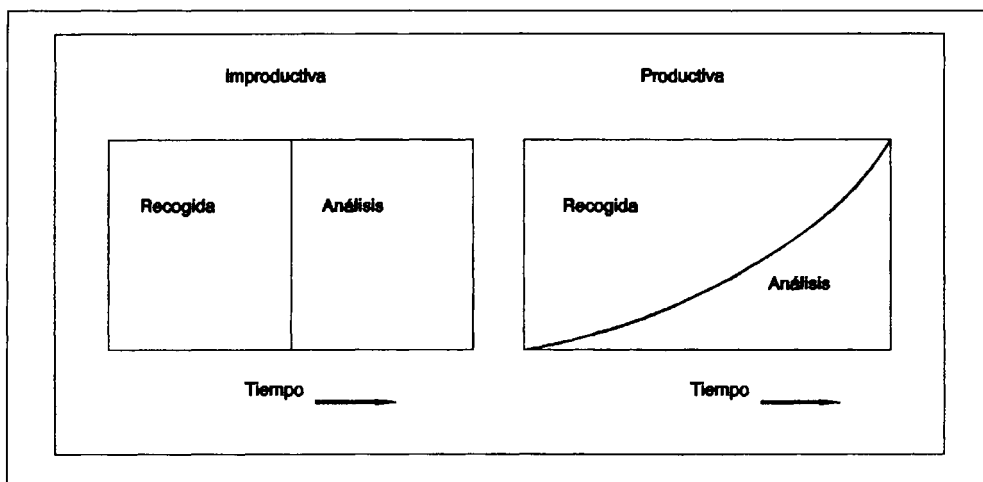


Figura 2
Relaciones productiva e improductiva entre recogida y análisis de datos (Lofland y Lofland, 1984: 132).

consideración de ambos procesos como fases separadas del trabajo del investigador resulta poco productiva, inclinándose por la concomitancia de tales tareas con un progresivo incremento del análisis, que prosigue una vez finalizada la recogida de

datos. Durante este último período se desarrollan y ponen en orden ideas surgidas previamente.

El bajo nivel de elaboración y la riqueza de significados intrínsecos a los datos cualitativos de naturaleza textual ha hecho de su análisis una tarea que ofrece no pocas dificultades. Enfrentarse al análisis de este tipo de datos ha representado para muchos investigadores una aproximación intuitiva hasta lograr estrategias particulares adecuadas a sus necesidades de análisis. La escasa definición de los métodos de análisis así generados hace que muchos investigadores vean a aquéllos simplemente como una «molestia atractiva» (Miles, 1985): La indefinición existente en relación a los procedimientos de análisis se ve agravada por el carácter plural de la investigación cualitativa, en la que coexisten un amplio número de tradiciones de investigación (Jacob, 1988).

Salvo quizás en el caso de los datos cualitativos textuales transformados en valores numéricos, no siempre el investigador cuenta con vías definidas o convenciones claras que orienten el modo de analizar los datos cualitativos, por lo que los procedimientos utilizados son de tal variedad y singularidad que a veces resultan particulares de un solo investigador o grupo de investigadores. Tanto es así que podría afirmarse que el único punto de acuerdo entre los investigadores es la idea de que el análisis es el proceso de extraer sentido de los datos (Tesch, 1990): Aún en el caso de que los investigadores lleguen a explicitar sus métodos, la falta de un lenguaje común hace que resulten poco claros. No sólo existe, por tanto, desacuerdo en el modo en que los investigadores realizan el análisis de sus datos cualitativos, sino también en la terminología que usan para referirse a las actividades y operaciones implicadas en él. Ello supone una dificultad añadida para el empleo de las técnicas de análisis por otros investigadores.

En síntesis, frente al carácter sistemático y formalizado de las técnicas de análisis aplicadas a los datos cuantitativos, los métodos usados con datos cualitativos presentan una gran diversidad y se basan con frecuencia en *la intuición y la experiencia del investigador*. A veces se ha afirmado que la tarea del analista es más un arte que una técnica, lo que supone sumar a las habilidades y tareas propias de todo proceso analítico otras de carácter especial. Para Woods (1989), el analista tendría que poseer actitudes y cualidades de *creatividad, imaginación* y un «*olfato de detective*» para extraer temas de interés a partir de indicios observados y llegar a producciones teóricas. Glaser y Strauss (1967) añaden a la imaginación la *ingenuidad*, y hablan de la *sensibilidad teórica* como característica que ha de poseer el investigador para conceptualizar y formular una teoría que emerja de los datos cualitativos. Otros aluden a la habilidad del investigador para *trascender su propia perspectiva* y llegar a conocer las perspectivas de aquellos a quienes está estudiando (Wilson, 1987): Autoras como Goetz y Lecompte (1988) hablan de la necesidad de contar con un *pensamiento divergente* y un *talento artístico*, aunque no refiriéndose a todo el proceso de análisis sino a algunos de los elementos instrumentales presentes en la interpretación (metáforas, analogías, etc.).

La singularidad y privacidad de los métodos de análisis cualitativos y la falta de

detalle con que son descritos en los informes de investigación, concentran parte de las críticas relativas a la credibilidad en la investigación cualitativa. Diversos autores han intentado dar rigor y exhaustividad a las técnicas cualitativas de análisis, tratando de que se aproximen en sistematización y fiabilidad a los procedimientos cuantitativos, si bien, como afirman Goetz y Lecompte (1988), hay autores que rechazan ese rigor y sistematización procesual por considerar que esclerotizan todo el proceso en perjuicio de su naturaleza intuitiva y creativa. Estos intentos de sistematización, en el análisis de datos cualitativos, se incluyen en lo que Firestone y Dawson (1982) denominan enfoques «de procedimiento», que incluyen una serie de reglas y pasos a seguir, frente a enfoques «intuitivos» o «intersubjetivos», basados respectivamente en la intuición y en la interacción entre investigadores o entre investigadores y participantes en los contextos estudiados. Los enfoques de procedimiento hacen esfuerzos por describir, con mayor o menor detalle, una serie de operaciones y pasos concretos, ofrecen consejos, recomendaciones y advertencias para el manejo, la disposición o la presentación de los datos. El avance que representan tales aportaciones en el análisis de datos cualitativos no ha evitado sin embargo que aún sean muchos los investigadores que recurren en algún momento de su proceso analítico a las técnicas estadísticas, con la intención de complementar o contrastar las conclusiones obtenidas por otras vías.

Efectivamente, la razón que apoya el uso de técnicas estadísticas en el análisis de datos cualitativos, de suerte que se sustituyan las palabras por números, puede encontrarse en el hecho de que aunque las palabras aportan mayor significado, los números son procesados con mayor economía. Por esta razón, la mayor parte de los investigadores prefiere trabajar con números, trasladando los textos a números lo más rápidamente posible. Para los investigadores que toman este último camino, el terreno está bien marcado por teorías de medición, reglas de decisión, niveles de confianza, algoritmos para el cálculo, etc.

De lo dicho anteriormente se deduce la existencia de dos tipos de datos cualitativos: los datos que conservan su naturaleza textual y los datos textuales que son transformados en valores numéricos. También han de considerarse como datos cualitativos, aquellos que se expresan directamente en forma numérica y están medidos en una escala nominal. No obstante, como ya se ha apuntado con anterioridad, este último tipo de datos cualitativos no representa una aportación novedosa o informativamente diferente a la que desde hace tiempo vienen ofreciendo los datos cuantitativos en el estudio de la diferenciación educativa.

La utilización que el investigador educativo hace de los datos cualitativos, manteniendo en éstos su naturaleza textual o bien transformándolos en valores numéricos, ha generado en la actualidad dos enfoques de análisis diferentes. Así, el análisis de datos cualitativos de naturaleza textual supone un proceso por el que se reduce y dispone la información recogida, se obtienen resultados y verifican conclusiones acudiendo a procedimientos preferentemente no estadísticos. Por su parte, el análisis de datos cualitativos transformados en valores numéricos se realiza mediante procedimientos estadísticos. A pesar de esta diferenciación, para el análisis de datos

cualitativos, es posible reconocer la existencia de un proceso común, en muchos de sus elementos, tanto a los datos de naturaleza textual como a los transformados en valores numéricos.

PROCESO GENERAL DE ANÁLISIS DE DATOS CUALITATIVOS

El análisis de datos cualitativos va a ser considerado aquí como un proceso realizado con un cierto grado de sistematización que, a veces, permanece implícita en las actuaciones emprendidas por el investigador. En este sentido, resulta difícil

CUADRO I
 PROCESO GENERAL DE ANÁLISIS DE DATOS CUALITATIVOS

TAREAS	ACTIVIDADES	OPERACIONES
	Separación de unidades	Determinación de criterios de separación físicos, temáticos, gramaticales, conversacionales y sociales.
Reducción de datos	Identificación y clasificación de elementos	Categorización y codificación
	Síntesis y agrupamiento	Agrupamiento físico, creación de metacategorías, obtención de estadísticos, métodos estadísticos de agrupamiento y síntesis.
Disposición y transformación de datos	Disposición	Elaboración de tablas numéricas, gráficos, modelos, matrices y sistemas de redes
	Transformación	Expresión de los datos en otro lenguaje (numérico, gráfico)
Obtención de resultados y verificación de conclusiones	Proceso para obtener resultados	<i>Datos textuales:</i> descripción e interpretación; recuento y coocurrencia de códigos; comparación y contextualización. <i>Datos transformados en valores numéricos:</i> técnicas estadísticas; comparación y contextualización
	Proceso para alcanzar conclusiones	<i>Datos textuales:</i> consolidación teórica, aplicación de otras teorías, uso de metáforas y analogías, síntesis con resultados de otros investigadores. <i>Datos transformados en valores numéricos:</i> uso de reglas de decisión (comparación de los resultados con modelos teóricos; recurso a la perspicacia y experiencia del analista).
	Verificación de conclusiones	Comprobación o incremento de la validez mediante presencia prolongada en el campo, intercambio de opiniones con otros investigadores, triangulación, comprobación con los participantes, establecimiento de adecuación referencial, ponderación de la evidencia, comprobación de la coherencia estructural.

hablar de una estrategia o procedimiento general de análisis de datos cualitativos, fuera de lo que pueda inferirse a partir de las acciones identificadas en un análisis ya realizado. No obstante, tomando como base estas inferencias, es posible establecer una serie de tareas u operaciones que constituyen el proceso analítico básico, común a la mayoría de los estudios en que se trabaja con datos cualitativos. Estas tareas serían: reducción, disposición y transformación, y obtención de resultados y verificación de conclusiones.

En cada una de estas tareas es posible distinguir, asimismo, una serie de actividades y operaciones concretas que son realizadas durante el análisis de datos, aunque no necesariamente todas ellas estén presentes en el trabajo de cada analista. En ocasiones, determinadas actividades pueden extenderse hasta constituir por sí mismas el proceso de análisis o, por el contrario, pueden no ser tenidas en cuenta en el tratamiento de los datos, de acuerdo con los objetivos del trabajo, el enfoque de investigación, las características del investigador... etc. Entre ellas no siempre se establece una sucesión en el tiempo, y pueden ocurrir de forma simultánea, o incluso estar presentes varias de ellas dentro de un mismo tipo de tarea. El marco general desde el que suelen analizarse los datos cualitativos podría esquematizarse en las siguientes tareas, actividades y operaciones recogidas en el cuadro I.

A) Reducción de datos

La limitada capacidad humana para procesar grandes cantidades de información hace necesaria su reducción a unidades elementales, fácilmente analizables, pero éstas a su vez, han de ser comprensivas, relevantes y significativamente densas. Para ello se suele recurrir a tres actividades básicas: separación, clasificación y síntesis, que incluyen operaciones tales como la categorización y codificación. Estas tres actividades son complementarias y, a veces, indisociables en la práctica analítica del investigador.

La reducción de datos no es una tarea asociada necesariamente a una fase inicial del análisis. En momentos avanzados del análisis, la reducción de datos puede hacerse sobre los resultados de las operaciones analíticas realizadas con anterioridad tomando como base los datos originales. Es posible advertir la función reductora incluso antes de contar con los datos. Miles y Huberman (1984b) consideran que puede darse una reducción de datos anticipada cuando focalizamos y delimitamos la recogida de datos, aspectos constituyentes de una forma de preanálisis que comporta descartar determinadas variables y relaciones y atender a otras. De alguna manera, los datos recogidos suponen un conjunto conscientemente reducido del universo de datos que podrían haberse obtenido. Finalmente, la reducción está presente cuando el investigador resume o esquematiza sus notas de campo o cuando selecciona entre los datos obtenidos aquellos que considera relevantes para sus cuestiones de investigación.

A.1. Separación de unidades

De acuerdo con el significado básico del término «análisis», una de sus primeras actividades es la separación de las unidades que conforman el cuerpo de datos objeto de estudio. Tesch (1990:95) afirma, como uno de los principios y prácticas más características del análisis cualitativo, que los datos pueden ser segmentados, es decir, «divididos en unidades relevantes y significativas».

Los investigadores que trabajan con datos cualitativos, en ocasiones, no suelen explicitar en sus estudios los criterios seguidos para la división del corpus de datos en unidades, e incluso se llega a obviar este importante aspecto. Lofland y Lofland (1984), por ejemplo, describen minuciosamente la operación de cortar los textos en fragmentos (unidades) pero omiten el modo en que llegan a ser delimitados tales fragmentos. No obstante, esta separación puede realizarse sobre la base de criterios físicos (espaciales y temporales), temáticos, gramaticales, conversacionales y sociales.

Las consideraciones *espaciales o temporales* permiten definir unidades de carácter artificial. De acuerdo con tales criterios, constituirían unidades los bloques de un determinado número de líneas, las páginas, los bloques de minutos, etc. Es más frecuente, y probablemente más adecuado, considerar unidades en función del *tema abordado*. Erickson (1982), al analizar datos audiovisuales, propone un modelo de análisis en el que se desciende desde el todo a las partes: partiendo del suceso como un todo, diferencia elementos constituyentes cada vez más pequeños, considerando conversaciones, sucesos, actividades que ocurren en la situación observada; a partir de ellos, es posible encontrar segmentos que hablan de un mismo tema. Sin embargo, el criterio temático conlleva una pequeña dificultad cuando las unidades son objeto de cómputo para posteriores análisis. Podríamos encontrarnos tanto fragmentos de gran extensión como fragmentos reducidos que hablen de un mismo tema, resultando unidades que llegan a ser consideradas equivalentes en análisis ulteriores, a pesar de que no lo son. Este problema ha sido soslayado en el análisis de contenido mediante la distinción entre unidades de registro y unidades de enumeración (Cartwright, 1978): Mientras las primeras determinan fragmentos referidos a un mismo tópico, las segundas aluden a la unidad sobre la base de la cual se realiza el recuento.

Las unidades pueden establecerse a partir de *criterios gramaticales* (el párrafo, la oración o la palabra podrían ser considerados las unidades básicas de un texto) o *conversacionales* (declaraciones, turnos de palabra, etc.): En ellas no es necesario un juicio sobre el significado para establecer los límites de la unidad. En algunas investigaciones, las unidades también se han elaborado a partir de *criterios sociales*. Cada unidad se haría corresponder con el segmento de texto generado por la intervención de un sujeto con determinado status o rol social (p.e. profesor, director, alumno, inspector, etc.):

A.2. Identificación y clasificación

La identificación y clasificación de elementos consiste en examinar las unidades de datos para encontrar en ellas determinados componentes generalmente temáticos que nos permitan clasificarlas en una u otra categoría de contenido. Las operaciones conocidas como categorización y codificación son las más representativas dentro de las actividades de identificación y clasificación.

La *categorización* hace posible agrupar conceptualmente las unidades que son cubiertas por un mismo tópico. Mucchielli (1988:34) define la categoría como «una noción general que representa a un conjunto o una clase de significados». Para Bogdan y Biklen (1982), las categorías pueden referirse a situaciones y contextos, actividades y acontecimientos, relaciones entre personas, comportamientos, perspectivas sobre un problema, métodos y estrategias, procesos. La categorización se realiza conjuntamente a la división en unidades cuando ésta se hace atendiendo a criterios temáticos. Si una unidad es separada por referirse a determinado tópico, automáticamente puede ser incluida en la categoría correspondiente a ese tópico.

El proceso de *codificación* no es más que la operación concreta por la que se asigna a cada unidad un indicativo (código) propio de la categoría en la que se incluye. Es el proceso físico, manipulativo, mediante el cual se realiza la categorización. Por esta razón, no es extraño que el estudio de ambas operaciones se aborde a menudo de forma conjunta, y que incluso en algún momento lleguen a identificarse como un único proceso, aspecto que queda reflejado en afirmaciones tales como «categorizar es asignar etiquetas a algo» (Bliss, Monk y Ogborn, 1983:8), o «los códigos son categorías» (Miles y Huberman, 1984b:56).

Los códigos, que representan a las categorías, pueden ser números con los que se marca cada unidad de datos (Bogdan y Biklen, 1982), aunque es más frecuente utilizar abreviaturas de palabras como categorías (Tesch, 1990): Por tanto, físicamente el código es «una abreviatura o símbolo aplicado a un segmento de texto — muchas veces una frase o un párrafo de notas de campo transcritas— con el fin de clasificarlo» (Miles y Huberman, 1984b:56).

Las categorías utilizadas en un estudio pueden establecerse a priori, de acuerdo con un marco teórico y conceptual previo, las cuestiones o hipótesis que guían la investigación, las categorías ya usadas en estudios de otros investigadores o los instrumentos de investigación empleados (el uso de cuestionarios o guiones de entrevista, por ejemplo, permite utilizar las cuestiones planteadas como fuente de categorías o modalidades útiles para organizar los datos): En ocasiones, las categorías que utiliza un investigador se relacionan entre sí dando lugar a estructuras de cierta complejidad conceptual conocidas como *sistemas de categorías*. Estos sistemas suponen un nivel explicativo superior al de la simple enumeración de temas o categorías. La realidad educativa analizada a través de esos sistemas de categorías se explica desde la localización física, el orden, la covariación, la causalidad, etc.

El partir de un sistema de categorías a priori no es una condición suficiente, sino un requisito previo para conseguir una misma codificación en distintos codificado-

res. Es preciso además una definición operativa de categorías, ofreciendo reglas que especifiquen los aspectos del contenido que deben tomarse como criterio de pertenencia a una categoría (Cartwright, 1978): Ha habido intentos de proporcionar sistemas de categorías válidos para múltiples estudios, como el ampliamente difundido sistema de categorías de Bales para el análisis de conductas lingüísticas en un grupo (ver Anguera, 1978): Apoyadas en este planteamiento a priori se han diseñado aplicaciones informáticas para el análisis de datos textuales capaces de realizar el proceso de codificación de modo automático, aunque con la limitación de que las categorías deben estar definidas por la aparición de un término o frase en los datos que señalen la relevancia del fragmento para una categoría (Tesch, 1990): La codificación automática eliminaría los efectos asociados a las características personales y circunstanciales del codificador, las cuales afectan al proceso de reducción de datos (Molly, 1984).

Los sistemas de categorías también pueden elaborarse inductivamente a partir de los propios datos. Al examinar los datos nos preguntaríamos por el tópico capaz de cubrir cada unidad, generando así una categoría. Posteriores comparaciones entre los datos agrupados bajo un mismo tópico o entre tópicos permitirían refinar las categorías emergentes (Glaser y Strauss, 1967): En el proceso denominado «codificación abierta» (Strauss, 1987), se parte de la búsqueda de conceptos que traten de cubrir los datos. En muchos casos, el sistema de categorías deja de ser únicamente una herramienta que utiliza el investigador para organizar sus datos y se constituye por sí mismo en un producto del análisis, un resultado de investigación. Para Spradley (1979, 1980), uno de los objetivos es llegar a elicitar las categorías (dominios, en la terminología de este autor) en que se clasifican los términos de una cultura y reducir de ese modo la complejidad de la experiencia humana. Su interés está en descubrir cómo las personas organizan su conocimiento, por lo que las categorías son un fin en sí mismas; no tiene sentido imponer categorías analíticas a los datos obtenidos durante el estudio.

Por lo general, procedimientos deductivos e inductivos se combinan en el trabajo del analista, de modo que a partir de unas categorías a priori se producen modificaciones para adaptarlas al conjunto de datos a que están siendo aplicadas. Tanto si las categorías son definidas a partir de los datos como si son preelaboradas (en función del estudio o tomados de sistemas generales), codificación y categorización no se desarrollan linealmente, sino que son procesos recurrentes, durante los cuales el sistema de categorías puede ser alterado. A medida que el investigador posee un mayor conocimiento del material que analiza, puede alterarse de forma no deliberada el modo en que se asignan los códigos. Cuando surjan incongruencias derivadas de la aplicación inconsistente de códigos, habrá que revisar lo realizado hasta el momento. De hecho, durante la codificación es posible advertir que determinados códigos podrían haber sido agrupados en uno solo o, por el contrario, que cabría dividir uno de ellos en otros diversos, ciertas categorías podrían «necesitar ser renombradas, modificadas en su contenido o sustituidas por otras nuevas» (Tesch, 1990:91):

Miles y Huberman (1984b) señalan que la codificación puede llevarse a cabo en diferentes momentos de la investigación y distinguen entre códigos *descriptivos* que atribuyen una unidad a una clase de fenómenos y son utilizados en un primer momento de la codificación, y códigos con un mayor contenido inferencial (*interpretativos y explicativos*) utilizados posteriormente. Por otra parte, los sistemas de categorías no son únicos, sino que mediante las categorías no buscamos sino extraer distinciones y nombrarlas, y esas distinciones pueden hacerse en varias dimensiones independientes (Bliss, Monk y Ogborn, 1983): Según el aspecto en que nos fijemos, un texto puede ser categorizado de distinto modo.

Autores que ven en la categorización un modo de transformar los datos textuales en datos susceptibles de medición y tratamiento estadístico (Ander-Egg, 1990; Cartwright, 1978; Kerlinger, 1975; Sánchez Carrión, 1985, etc.) demandan de los sistemas de categorías una corrección lógica que se alcanza con los requerimientos clásicos de *exhaustividad* de las categorías (cualquier unidad puede ser ubicada en alguna de las categorías), *exclusión mutua* (cada unidad se incluye en una sola categoría) o el estar elaboradas desde un *único principio clasificadorio*. Mucchielli (1988) habla también de la *objetividad* (inteligibles para distintos codificadores), y la *pertinencia* en relación a los objetivos del estudio y al propio contenido analizado. Sin embargo, concepciones más recientes en la investigación en ciencias humanas abogan por categorías que no están claramente delimitadas unas de otras, con superposición entre las áreas temáticas que cubren. Se habla en estos casos de *membreía* en una categoría, para expresar en qué grado determinada unidad puede formar parte o ser miembro de ella, y de la noción de *prototipo*, que representa el tipo ideal con un cien por ciento de pertenencia a una sola categoría. Tal elemento no existe realmente; es un constructo mental que sirve como punto de referencia al que comparar las unidades de datos reales, para determinar su grado de membreía (Tesch, 1990).

A.3. Síntesis y agrupamiento

La identificación y clasificación de elementos están estrechamente unidas a la síntesis. Cuando categorizamos estamos sintetizando diferentes unidades de datos en un mismo tópico o concepto teórico. La categorización supone en sí misma una operación conceptual de síntesis, en la que se reducen diversas unidades a un concepto que las representa. En el caso de los datos cualitativos textuales, la síntesis permite realizar un agrupamiento físico en el que se reúnen todas las unidades que forman parte de la misma categoría. En el caso de los datos cualitativos transformados en valores numéricos, la unión de modalidades de variables supone una operación de agrupamiento, mientras la obtención de estadísticos de tendencia central supone procesos de síntesis.

En el caso del análisis de contenido automático se ha recurrido a dos procedimientos para llegar a agrupar unidades (generalmente palabras): Los *sistemas a priori* buscan palabras identificadas con conceptos teóricos relevantes para el pro-

blema investigado (Mochmann, 1985): Es necesario definir previamente la relación entre conceptos y palabras, creando lo que se denomina un diccionario de análisis de contenido. Mediante este diccionario el investigador sintetiza previamente el significado de las palabras elaborando un concepto que las agrupa. El ordenador se encargará de localizar todas las palabras que se agrupan bajo cada concepto y ofrecer sus frecuencias de aparición en las unidades consideradas para permitir cálculos estadísticos. Existen diccionarios generales, apoyados en teorías de gran alcance, y también diccionarios especializados o incluso específicos de cada estudio (Mohler, 1985):

Otros enfoques, denominados *empíricos*, agrupan automáticamente las palabras (excluyendo las palabras funcionales: artículos, conjunciones, etc.) sin utilizar categorías preestablecidas, y evitan así toda influencia del investigador en el agrupamiento resultante. Las síntesis (categorías) surgen mediante procedimientos estadísticos como los análisis factoriales o los análisis de conglomerados realizados sobre matrices de correlaciones obtenidas a partir de la frecuencia que cada palabra alcanza en una unidad diferenciada en el texto (Iker y Harway, 1969):

Los procesos de agrupamiento y síntesis también están presentes en el proceso de análisis cuando, por ejemplo, agrupamos varios códigos en un metacódigo o cuando sintetizamos la información contenida en varias categorías que tienen algo en común en una metacategoría, o elaboramos un sistema de indización.

Las actividades y operaciones revisadas hasta aquí constituyen modos de contribuir a la reducción de datos que en el fondo presupone todo análisis: se parte de un conjunto amplio y complejo de datos para llegar a unidades elementales más manejables que permitan establecer relaciones y sacar conclusiones.

B) Disposición y transformación de los datos

Como afirma Levine (1985), al aportar razones para apoyar la conveniencia de un análisis sistemático de la información verbal, los datos de este tipo no son recogidos siguiendo, en una secuencia lineal, a los problemas de la investigación, sino en función de la interacción del investigador y el informante o el contexto de estudio. Además, los textos narrativos resultan dispersos, ofrecen una visión secuencial de hechos simultáneos, dificultan el examen de varias variables a la vez y están vagamente ordenados (Miles y Huberman, 1984b): Cuando los datos cualitativos se transforman numéricamente el analista también se encuentra ante dificultades para estudiar directamente dichos datos, de modo que sea posible describir la tendencia central de cada variable o para identificar directamente la intensidad de las posibles relaciones entre las mismas.

Dadas estas dificultades, es necesaria una *transformación y disposición* de los datos que permita presentarlos de manera abarcable y operativa de cara a resolver las cuestiones de investigación. Disposición y transformación son dos actividades que facilitan el examen y la comprensión de los datos, permitiendo extraer conclusiones, y condicionando posteriores decisiones en el proceso de análisis. La dispo-

sición supone organizar los datos, presentándolos en alguna forma espacial ordenada, de tal modo que se simplifique la información y se posibilite su procesamiento posterior. Existen diferentes procedimientos aplicables a la disposición de los datos cualitativos.

Un primer procedimiento lo constituyen las *tablas numéricas*, entre las que podemos diferenciar los siguientes tipos (Lebart, Morineau y Fenelon, 1985):

- a) Tablas de contingencia, en las que una misma población puede ser objeto de dos particiones, que se sitúan en filas y columnas respectivamente. En la intersección de éstas aparecerá el número de sujetos o casos que pertenece simultáneamente a dos partes consideradas. Estas tablas tienen por finalidad estudiar la relación entre dos o más variables de tipo nominal u ordinal.
- b) Tablas de incidencia, o también denominadas tablas de presencia-ausencia. Sujetos y variables son trasladados a filas y columnas. Según un individuo u objeto presente o no determinada característica o atributo, se consignará 1 ó 0 en la celda de intersección correspondiente.
- c) Tablas de valores numéricos, recogiendo en las intersecciones el valor de las variables ubicadas en las columnas para los individuos colocados en las filas.

La presentación de datos en forma numérica puede también darse una vez que se ha llevado a cabo un cierto procesamiento estadístico de la información, recurriendo en estos casos a tablas que recogen distribuciones de frecuencias, proporciones, porcentajes, etc. Otras tablas o matrices numéricas presentan ordenadamente datos tales como correlaciones, pesos y puntuaciones factoriales, distancias entre sujetos, etc. obtenidos tras diversos análisis. En cualquier caso, las tablas cumplen una función analítica. Su finalidad no es almacenar datos, sino comunicarlos en algún modo ordenado que permita extraer conclusiones (Wainer, 1992):

Un segundo procedimiento de disposición de los datos lo constituyen los *gráficos*, medios para presentar información que ha podido ser objeto de un mayor o menor grado de elaboración. Los histogramas, curvas de nivel, diagramas de sectores circulares son utilizados para describir gráficamente distribuciones de frecuencias. Gráficos de caja, postes telegráficos, tallo y hojas, matrices ordenables, matrices ponderadas, etc. son procedimientos que se utilizan en la disposición de los datos cualitativos. Asimismo, técnicas como el análisis factorial de correspondencias, el escalamiento multidimensional o el análisis de conglomerados conllevan representaciones gráficas que muestran la situación de los elementos respecto a ejes factoriales, su posicionamiento en un espacio geométrico, o el modo en que se agrupan (dendograma): Los gráficos permiten no sólo presentar los datos, sino advertir relaciones y descubrir su estructura profunda (Wainer, 1992).

Como un tercer grupo de procedimientos, pueden destacarse las técnicas de *modelado gráfico* muy extendidas entre los investigadores que trabajan con datos cualitativos de naturaleza verbal. Spradley (1979,1980) utiliza representaciones taxonómicas en forma de perfil general («outline»), diagrama de caja («box diagram»)

o diagrama en árbol («lines and nodes»): También se apoya en disposiciones denominadas paradigmas, que se acercan más a un formato de matriz, en las que se señala la presencia o no de ciertos atributos (columnas) en un grupo de términos considerados (filas): Otra técnica de modelado gráfico la constituye los diagramas esquemáticos, «representaciones gráficas o imágenes visuales de las relaciones entre conceptos» (Strauss y Corbin, 1990:198).

Las *matrices* constituyen otra forma de disposición gráfica de los datos. Consisten en tablas de doble entrada en cuyas celdas se aloja una breve información verbal (fragmentos de texto, citas, frases, abreviaciones o figuras simbólicas), de acuerdo con los aspectos especificados por filas y columnas. Es frecuente construirlas para sintetizar los fragmentos codificados en una misma categoría (columna) para diferentes sujetos, situaciones, casos, etc. (filas): Algunas de las dimensiones sobre la base de las cuales se elaboran pueden implicar la ordenación de fenómenos o componentes de fenómenos a lo largo del tiempo, en función de los roles desempeñados, etc. Determinadas matrices presentan procesos y productos, relaciones de causa-efecto, agrupamientos conceptuales o empíricos de datos.

Bliss, Monk y Ogborn (1983) proponen un *sistema de redes* («networks») mediante las cuales tratan de clarificar y presentar las distinciones que delimitan a las diversas categorías consideradas en el análisis del material cualitativo, las cuales a su vez pueden estar subdivididas. Para ello usan una notación precisa, con la que indican nombres de categorías, subdivisiones de éstas, diferencias, circunstancias, restricciones, etc.

La presentación de un mismo corpus de datos puede hacerse en múltiples formas, según el objetivo que el analista persiga con ello. A veces, el interés es exclusivamente descriptivo, tratando de ofrecer la información disponible y permitir una visión de conjunto que lleve a análisis o presentaciones más detalladas. En otras ocasiones, ese interés se centra en la interpretación y el analista trata de presentar los datos de forma que permitan explicar y crear generalizaciones, desarrollar teorías, etc. También la disposición de los datos puede favorecer la verificación de supuestos, teorías, etc. o el estudio de prácticas e intervenciones educativas.

Cuando la disposición de los datos no sólo implica un cambio en la forma de presentarlos espacialmente sino también una modificación del lenguaje, hablamos de transformación de los datos. Las transformaciones más frecuentes son las que permiten pasar de una expresión verbal de los datos a una expresión numérica, o bien de cualquiera de estas formas de expresión a una de tipo gráfico. Se realizan transformaciones cuando contamos códigos, palabras, segmentos de palabras,... y proseguimos el análisis a partir del recuento efectuado y no de la forma verbal en que aparecían los datos. También son ejemplos de transformación aquellos procedimientos de disposición que implican una expresión gráfica de los datos (histogramas, modelos gráficos, representaciones taxonómicas): Como también ocurre con los procedimientos empleados para la disposición, la transformación de los datos facilita la descripción, interpretación, verificación o valoración de las diferencias educativas.

C) Obtención de resultados y verificación de conclusiones

C.1. *Proceso para obtener resultados*

Un importante grupo de actividades del proceso de análisis se dirige a la obtención de resultados, en relación a los objetivos particulares de cada investigación. Obtener resultados implicaría, de acuerdo con las connotaciones semánticas del término análisis, ensamblar de nuevo los elementos ya diferenciados en el proceso analítico para reconstruir un todo estructurado y significativo. Van Maanen (1985) denomina a los resultados «conceptos de segundo orden», pues se construyen a partir de los datos, o conceptos de primer orden. La obtención de resultados no tiene lugar necesariamente al término del proceso de análisis, sino que éstos son extraídos por el investigador durante todo el proceso de recogida de datos y durante otros momentos del análisis (Miles y Huberman, 1984b): Las tareas de reducción y presentación de datos le permiten hacer afirmaciones que progresivamente avanzan desde lo descriptivo a lo explicativo y desde lo concreto a lo abstracto. Las categorías obtenidas en el proceso de reducción pueden considerarse, asimismo, resultados del estudio y, como se señaló más arriba, la ordenación de los datos en una matriz, un cuadro, una figura, etc. presupone también algún tipo de conocimiento respecto a su estructura.

Cuando se analizan los datos cualitativos conservando su naturaleza textual, los resultados se obtienen en función del modo en que se dispone y organiza la información, a partir del recuento o de la coocurrencia de determinados códigos. Los resultados también surgen a partir de una interpretación de los datos una vez reducidos y dispuestos en alguna forma: el investigador alcanza una visión global de los datos que no necesariamente coincidirá con la de otros investigadores que se enfrentaran a los mismos datos. La lectura de una gran parte de los trabajos que se publican sugeriría que los resultados emergen directamente a partir del corpus de datos, sobre todo una vez que éste ha sido adecuadamente reducido y presentado.

Operaciones que lleva a cabo el investigador para obtener resultados a partir de datos de naturaleza textual son la comparación y la contextualización. La primera de ellas, permite destacar las semejanzas y diferencias entre las unidades incluidas en una categoría, o entre distintas categorías, y hace posible la formulación de sus propiedades fundamentales, a partir de las cuales puede llegarse a una definición, ilustración y verificación de esa categoría (Fielding y Fielding, 1986): La comparación facilita la obtención de resultados a partir de las matrices y otras figuras usadas para la presentación de los datos. La comparación de filas o columnas de una matriz, por ejemplo, constituye la principal vía para extraer resultados (Miles y Huberman, 1984b): Los resultados pueden emanar de la comparación con otros escenarios, casos, situaciones, etc. similares al estudiado. También la comparación puede realizarse respecto a un criterio particular (una meta u objetivo) o a una norma, que sirven de base para el establecimiento de un juicio de valor. Los resultados se obtienen además merced a la contextualización de los datos. Así, los

procesos, las relaciones, los sistemas, las opiniones presentes en los datos pueden ser referidos al marco cultural en el que se han generado, de suerte que lo que se describe, interpreta, verifica o evalúa forma parte de un todo más comprensivo que da unidad al conjunto de elementos considerados en el análisis.

Cuando se trata de datos cualitativos transformados en valores numéricos el proceso de obtención de resultados es más inequívoco y transparente. El investigador cuenta con un amplio repertorio de algoritmos de cálculo, técnicas estadísticas y operaciones matemáticas que traducen un razonamiento lógico de naturaleza formal, todos ellos le permiten obtener elaboraciones de segundo orden fácilmente replicables por otros investigadores. También con este tipo de datos se utiliza la comparación (inter o intra) como herramienta para obtener resultados. El análisis de tablas de contingencia, por ejemplo, se basa en la comparación del modo en que se distribuyen los datos por filas o por columnas. Por su parte, la contextualización está presente cuando disponemos espacialmente los datos poniendo de manifiesto la posición relativa de cada elemento respecto al conjunto.

C.2. *Proceso para alcanzar conclusiones*

Otro momento en el análisis de los datos supone la interpretación de los resultados obtenidos hasta *alcanzar conclusiones*. La interpretación nos lleva a conclusiones que representan una lectura más comprensiva de los resultados y que pueden conducir a la creación y explicación de generalizaciones, el planteamiento de nuevas cuestiones o problemas, el desarrollo de nuevos conceptos, la redefinición de conceptos ya existentes, la clarificación y comprensión de la complejidad de algo o la generación de una nueva teoría o explicación.

En el caso de los procedimientos de análisis de datos cualitativos que conservan la forma textual, pueden ser utilizadas estrategias como las propuestas por Goetz y Lecompte (1988) para alcanzar conclusiones a partir de la interpretación de los resultados. Estas estrategias, que suponen integrar los resultados en marcos teóricos y de investigación más amplios, serían las siguientes:

- Consolidación teórica. Consiste en confrontar los datos obtenidos con los marcos teóricos desarrollados originalmente, modificando éstos para intentar que se ajusten más a los datos. Este enfoque ha sido ampliamente desarrollado por Glaser y Strauss (1967), quienes han defendido la utilidad de generar teoría fundamentada en los datos.
- Aplicación de otras teorías. La búsqueda de otros marcos más generales en los que integrar los datos permite la generalización de los resultados.
- Uso de metáforas y analogías. Son medios útiles para establecer conexiones entre temas aparentemente no relacionados, o conectar los resultados a la teoría (Miles y Huberman, 1984b): Preguntarse «¿a qué me recuerda esto?», acerca de los diferentes aspectos estudiados puede ser una estrategia adecuada para encontrar analogías y metáforas (Bogdan y Biklen, 1982:153).
- Síntesis de los resultados con los obtenidos por otros investigadores, para lo

que es necesario comparar, contrastar, agregar, ordenar, establecer relaciones y especular. Es decir, se repite el proceso analítico pero esta vez con los resultados alcanzados y, por tanto, a niveles más elevados de abstracción y generalización.

En el análisis de datos cualitativos numéricos, el proceso que nos permite llegar a conclusiones a partir de los resultados obtenidos se ve facilitado por reglas de decisión previamente establecidas, basadas en la comparación de los resultados con modelos teóricos de distribución del estadístico de contraste. Los procedimientos estadísticos que no se basan en la inferencia también utilizan reglas de decisión, si bien ésta se apoya en parte en la experiencia y perspicacia del analista. Por ejemplo, ante la representación gráfica resultante de un análisis factorial de correspondencias, el analista tendría que ser capaz de identificar un rasgo o característica que es poseída con distinto signo y peso por los elementos situados a uno u otro lado de un determinado eje; el factor representado por ese eje puede quedar identificado con la característica en cuestión.

C.3. Verificación de conclusiones

Una vez alcanzadas las conclusiones de un estudio, es necesario verificarlas. Esta tarea supone comprobar el valor de verdad de los descubrimientos realizados, es decir, contrastar su validez. Entendida como un estado ideal que se pretende alcanzar y no como un logro, la validez ha sido definida desde distintas posiciones epistemológicas que han elaborado diferentes patrones para verificar las conclusiones de un estudio. Desde determinada posición epistemológica, identificable con programas de investigación positivistas o neopositivistas, se entiende que las conclusiones son válidas en la medida en que existe una realidad educativa que las sostiene o que al menos no las contradice. Para favorecer o incrementar la obtención de conclusiones válidas, desde esta posición, se hacen esfuerzos por conseguir que el proceso de investigación no se vea afectado por diferentes fuentes de error y que los resultados puedan ser generalizados.

Desde otra posición epistemológica, identificable con programas de investigación interpretativos, se considera que la validez es un ideal que se verifica en función de la concordancia entre las conclusiones extraídas durante el análisis y las perspectivas mantenidas por los sujetos o grupos estudiados. Podemos hablar de validez cuando las personas comparten —quizás tras un proceso de diálogo y negociación— una descripción o interpretación (Smith y Heshusius, 1986:9): Autores como Guba (1983) plantean además la necesidad de que se dé una adecuación entre datos e interpretaciones y entre proposiciones y referentes.

Cuando los datos cualitativos se analizan desde esta posición, la validez de las conclusiones puede incrementarse, o simplemente ser calibrada, con estrategias desarrolladas en las fases de diseño, recogida de datos, análisis de datos o redacción del informe (Dawson, 1982): Por el hecho de estar presente, el investigador provoca

en los participantes conductas que ordinariamente no habrían ocurrido y es, al mismo tiempo, influenciado por ellos. La *presencia prolongada en el campo* hace que el investigador forme parte de la situación estudiada permitiéndole soslayar los efectos de su presencia, le permite comprobar los prejuicios y percepciones propios o de los participantes, y comprender lo que es esencial o característico (Guba, 1983), haciendo posible el ajuste entre las categorías de análisis extraídas y la realidad. El *intercambio de opiniones* con otros investigadores, exponiéndose a sus preguntas y críticas puede ser útil para llegar a conclusiones válidas. La publicación de resultados constituye un modo de presentarlos a la revisión por parte de otros expertos (Goetz y Lecompte, 1988): La *triangulación* es un procedimiento ampliamente usado que consiste en el uso de diferentes fuentes de datos, investigadores, perspectivas o metodologías para contrastar los datos e interpretaciones (Denzin, 1978): Si diversas clases de datos conducen a la misma conclusión podemos estar algo más seguros de tal conclusión (Hammersley y Atkinson, 1983): Sin embargo, se han hecho objeciones a la triangulación en el caso en que métodos de diferentes perspectivas metodológicas son usados conjuntamente a pesar de responder a asunciones ontológicas y epistemológicas diferentes (Blaikie, 1991): La *comprobación con los participantes* supone contrastar los datos e interpretaciones de los mismos con los sujetos que constituyen la fuente de esos datos. Como profundos conocedores de la realidad que se investiga, los participantes podrían actuar como jueces que evalúen los principales descubrimientos de un estudio (Denzin, 1978): Pueden realizarse esas comprobaciones de los resultados frente a datos nuevos procedentes de otras fuentes (documentos, grabaciones, etc.) recogidos con tal fin durante el estudio, estrategia denominada por Guba (1983) *establecimiento de adecuación referencial*.

Miles y Huberman (1984b) proponen la *ponderación de la evidencia* como técnica que incrementaría la validez de los resultados. Su idea es la de considerar ciertos datos mejores que otros a la hora de extraer conclusiones, ya sea por su procedencia de determinados informantes, por las circunstancias en que fueron recogidos (al final o al principio de la estancia en el campo, revelados en privado o en presencia de otros informantes, etc.) o por haber sido validados mediante estrategias como la triangulación o la comprobación con los participantes. Debe ser comprobada la *coherencia estructural* para asegurarse de que entre los datos e interpretaciones no se dan contradicciones o incoherencias (Guba, 1983): Un modo de aproximarse a ella es la *búsqueda de evidencia negativa* (Miles y Huberman, 1984b), consistente en preguntarse, una vez alcanzadas las conclusiones, si hay datos que se oponen o son inconsistentes con esas conclusiones.

MÉTODOS DE ANÁLISIS DE DATOS CUALITATIVOS

La investigación sobre la diferenciación educativa puede apoyarse en diferentes métodos para el análisis de datos cualitativos. La elección de uno u otro procedimiento analítico va a estar condicionada por el tipo de problemas abordados en la

investigación, por el diseño utilizado para comprender y resolver dichos problemas y, entre otros aspectos, por el tipo de datos cualitativos generados. Teniendo en cuenta esta última variante, vamos a recoger aquí algunos de los métodos más conocidos y utilizados en el análisis de datos cualitativos.

Datos cualitativos textuales

Existe una gran diversidad de procedimientos que podríamos identificar como *interpretativos* (generalmente asociados a enfoques etnográficos, basados en el estudio de casos, sociológicos, etc.) para el análisis de datos cualitativos textuales, aunque la mayoría de ellos no representan sino variantes del proceso general antes descrito que se utilizan en un número reducido de investigaciones. No obstante, en algunos de estos procedimientos interpretativos se han realizado intentos de sistematización por parte de algunos autores que han contado con mayor aceptación y están ampliamente referenciadas en la literatura. Este es el caso de procedimientos como la inducción analítica, creado por Znaniecki (1934), o el método de la comparación constante de Glaser y Strauss (1967).

Uno de los métodos analíticos más característicos y sistematizados entre los investigadores interpretativos posiblemente sea la *inducción analítica*. Desde que este método fuera utilizado por Znaniecki (1934), ha sido empleado y modificado configurándose en un método riguroso, adecuado en investigaciones que se focalizan en un tema concreto. Se emplea para la generación de teoría y requiere de múltiples lugares o sujetos para su desarrollo. Consta de diversas etapas, a lo largo de las cuales se trata de ir progresivamente modificando una formulación inicial para que llegue a explicar la totalidad de los casos estudiados:

- a) Definición aproximada del fenómeno a estudiar.
- b) Formulación teórica que explique el fenómeno.
- c) Comprobación del ajuste de la hipótesis a los datos a medida que éstos van siendo recogidos.
- d) Reformulación de la hipótesis o redefinición del fenómeno en caso de no darse tal ajuste.
- e) Búsqueda deliberada de casos negativos que puedan no adaptarse a la formulación.
- f) Reformulación o redefinición en caso de hallarlos.
- g) Continuación del proceso hasta haber examinado una amplia gama de casos y haber probado adecuadamente la hipótesis.

Otro método para el análisis de datos cualitativos, también de carácter inductivo, es el desarrollado por Glaser y Strauss (1967): Estos autores, que ponen el énfasis en la generación de teoría fundamentada (basada en los datos), proponen el *método de la comparación constante* que es desarrollado mediante un procedimiento continuo en el que podemos diferenciar cuatro momentos característicos:

- a) Codificación de los datos en categorías, comparando entre sí los incluidos en

- un mismo grupo para generar propiedades teóricas de cada categoría. Simultáneamente se elaboran resúmenes recogiendo las ideas que surgen (hipótesis, relaciones, etc.).
- b) Integración de categorías y sus propiedades. La comparación de nuevos datos procedentes de los nuevos casos seleccionados (muestreo teórico) con las propiedades de la categoría, permite describir relaciones y nuevos dominios dentro de las categorías, que se van integrando en una teoría coherente.
 - c) Delimitación de la teoría y de las categorías. Cuando nuevas comparaciones producen modificaciones en la teoría, se va clarificando la lógica relacional, descartando propiedades irrelevantes, incorporando detalles, reduciendo categorías, hasta que un caso nuevo no consiga añadir nada (saturación teórica): Las teorías se reducen a conceptos de mayor nivel.
 - d) Explicitación de la teoría. Para escribirla se recurre a los resúmenes de cada categoría, se vuelve a los datos codificados para validar afirmaciones, hipótesis y proveer ilustraciones.

Basado en este enfoque se ha diseñado una aplicación informática específicamente concebida para el análisis de datos cualitativos expresados en forma de textos. Esta aplicación, conocida como NUDIST (Non-numerical Unstructured Data Indexing, Search and Theorizing), ha sido desarrollada en la Universidad de La Trobe (Australia), y será tratada con más detalle en los trabajos «NUDIST: Una herramienta informática para el análisis de datos cualitativos» (Corrales et al.), y «NUDIST versus AQUAD: Dos posibilidades informáticas para el análisis de datos cualitativos» (Rodríguez et al.) presentados ambos a este Seminario.

Otros procedimientos analíticos, no siempre suficientemente caracterizados, menos difundidos pero también utilizados con datos cualitativos de naturaleza textual, son los que proponen autores como Spradley, Goetz y Lecompte, Bogdan y Biklen, Taylor y Bogdan, Lofland y Lofland, Erickson, Miles y Huberman, y otros. Spradley (1979, 1980), interesado en describir los sistemas de significado que grupos culturales han creado y utilizan en el modo de actuar e interpretar sus experiencias, entiende el análisis como una herramienta para el descubrimiento de significado cultural a partir de los datos obtenidos mediante la entrevista en profundidad y la observación participante. El procedimiento de análisis supone la identificación de dominios (categorías), el desarrollo de taxonomías, la búsqueda de contrastes y la generación de temas (ideas síntesis), en relación con otras tantas estrategias presentes en la recogida de información.

Goetz y Lecompte (1988), quienes están interesadas por la descripción o reconstrucción analítica de los escenarios y grupos culturales a partir de las concepciones de los participantes, ponen en práctica estrategias como el análisis tipológico (clasificación en categorías), la enumeración y la aplicación de protocolos observacionales estandarizados, y procedimientos analíticos generales como los ya mencionados: método de la comparación constante e inducción analítica. Consideran, asimismo, que tanto estos procedimientos como las estrategias de selección secuencial se

apoyan en determinados modos genéricos de pensamiento y manipulación sobre los datos, a los que denominan actividades de teorización y que comprenden: a) la percepción; b) la comparación, contrastación, agregación y ordenación; c) la determinación de vínculos y relaciones; y, d) la especulación. Para estas autoras, los métodos señalados y las tareas de teorización son los instrumentos conceptuales de que dispone el investigador para desarrollar su proceso de análisis, combinándolos en cada investigación de acuerdo con sus fines y marco teórico.

Miles y Huberman (1984b) son autores de una obra dedicada específicamente al análisis de datos cualitativos. Con ella intentan aportar procedimientos sistemáticos, replicables, que permitan extraer y comprobar conclusiones. Mencionan tres grupos de técnicas constituyentes del proceso de análisis: métodos de reducción de datos (agrupamiento de códigos, sumarización, partición,...), presentación de datos (matrices, gráficos, cuadros, redes,...) y extracción y verificación de conclusiones (recuento, construcción de metáforas, detección de relaciones,... y triangulación, comprobación de la representatividad,...): Basado en el procedimiento propuesto por estos autores, Huber (1988) ha elaborado el programa AQUAD (Analysis of Qualitative Data), que responde en su concepción a los procedimientos de reducción y presentación de datos, extracción y verificación de conclusiones. Un análisis más detallado de este programa puede verse en los trabajos «Formulación y comprobación de hipótesis cualitativas con ayuda del programa AQUAD» (Gil et al.) y «NUDIST versus AQUAD: Dos posibilidades informáticas para el análisis de datos cualitativos» (Rodríguez et al.) presentados ambos a este Seminario.

Dado el coste en esfuerzo y tiempo que requiere el desarrollo de los procedimientos de análisis de datos cualitativos, se han diseñado en los últimos años un gran número de programas específicamente creados para facilitar estas tareas. Tales programas (TAP, ETHNOGRAPH, QUALPRO, QUALOG, etc.) se enmarcan en una perspectiva interpretativa y han sido ampliamente descritos en obras como las de Tesch (1988, 1989, 1990, 1991).

Datos cualitativos expresados mediante valores numéricos²

La diferenciación educativa genera datos cualitativos textuales susceptibles de transformación en valores numéricos, bien tras las operaciones de categorización y codificación, bien como resultado del recuento de unidades diferenciadas en los textos. En ambos tipos de datos (transformados mediante la categorización y codificación o procedentes del recuento de unidades en los textos) y en función del problema estudiado el investigador puede utilizar diferentes métodos de análisis.

2 Para este tipo de datos existe una serie de paquetes de programas de análisis estadístico, en los que se abordan diferentes métodos de análisis, tales como SYSTAT, SPSS, BMDP, GENSTAT, SAS... etc.

1. Datos transformados tras las operaciones de categorización y codificación

Distinguiremos en este apartado entre métodos de interdependencia entre variables y métodos de dependencia de variables. En el primer grupo de métodos de análisis, **basados en relaciones de interdependencia**, el interés se centra en la asociación mutua entre variables sin establecer distinciones (dependientes e independientes) entre ellas. Se persigue la organización y reducción de los datos en una forma más fácil de manejar por el investigador y que posibilite una mejor comprensión global de los mismos. Dentro de este grupo de métodos para el estudio de la interdependencia vamos a considerar algunas modalidades:

a) *Desarrollados a partir de tablas de contingencia*

El uso de tablas de contingencia en el análisis de datos cualitativos ha respondido a finalidades diferentes. En ocasiones, se persigue únicamente la descripción empírica de la distribución conjunta de los valores de dos variables, sin recurrir a ningún tipo de técnica estadística que permita comprender las relaciones presentes en los datos. Otras veces, no se va más allá del cálculo de alguna medida de ajuste o asociación tales como la diferencia de razones, chi-cuadrado, coeficiente de contingencia, V de Cramer, etc., medidas que sin embargo ofrecen indicaciones claras sobre la intensidad y/o el sentido de la relación entre las variables. No obstante, quizás la aplicación más atractiva de este tipo de tablas sea la que las vincule al desarrollado procedimientos analíticos de mayor complejidad, para datos de naturaleza secuencial y no secuencial. Este es el caso del análisis secuencial, el sistema de diferencia de proporciones, el análisis lineal-logarítmico o el análisis factorial confirmatorio.

Las *técnicas secuenciales* se utilizan para solucionar problemas de ordenación de los datos en el tiempo, lo que a su vez permite comprender cómo esos datos se producen momento a momento. En este sentido, el análisis secuencial recoge datos relativamente momentáneos y no centrados en un desarrollo que se extiende a lo largo de meses o años: los datos secuenciales recogen la continuidad entre unidades sucesivas codificadas (Bakeman y Gottman, 1989:80): El análisis secuencial parte de la obtención de una serie de estadísticos descriptivos simples (duración media y probabilidad simple del evento, probabilidad de transición), calculados a partir de una tabla de contingencia en la que cada celda indica el número de veces que ocurre un determinado evento dentro de una secuencia. Estos valores se comparan con modelos teóricos, equiprobables o no, a partir de estadísticos como chi-cuadrado de bondad de ajuste o la prueba binomial de puntuaciones z. Aunque el procedimiento es relativamente simple y puede ser desarrollado de forma manual, existen programas informáticos como ELAG que nos facilitan el desarrollo de este tipo de análisis (Bakeman, 1983).

El *sistema de diferencia de proporciones* se aplica al estudio de tablas de contingencia multidimensionales y se desarrolla de acuerdo con tres ideas básicas:

la medida base de asociación es la diferencia de proporciones o porcentajes; parte del «paradigma de la elaboración», desarrollado por Lazarsfeld (1955) y Rosenberg (1968), asignando además valores a las relaciones conceptuales del paradigma; sigue los principios del gráfico de flujo, para representar los sistemas de ecuaciones lineales en forma de grafos. El análisis, siguiendo el sistema de diferencia de proporciones, o sistema de la D, descompone una tabla de contingencia multidimensional en una serie de ecuaciones para más tarde representar éstas en forma de grafos. El procedimiento que se sigue para estimar los coeficientes de relación «d» en el grafo supone calcular proporciones y estandarizarlas. Estos coeficientes «d» son susceptibles de ser sometidos a prueba estadística calculando su intervalo de confianza (Sánchez Carrión, 1984): Un programa informático que permite el análisis de la diferencia de proporciones es el programa CHIP (Bogart y Conner, 1986).

Las técnicas denominadas *modelos lineales logarítmicos* (Kennedy, 1983) también permiten al investigador estudiar tablas de contingencia multidimensionales, midiendo y contrastando las asociaciones e interacciones que aparecen en ellas. Se basan en un modelo aditivo, de acuerdo con el cual es posible determinar el logaritmo en base e de la frecuencia esperada en cada casilla a partir de los efectos principales y debidos a las interacciones entre las variables que forman la tabla. Constituyen una respuesta satisfactoria al análisis de los datos relativos a variables categóricas obtenidas tras los procesos de reducción de datos cualitativos en forma de registros narrativos, citas textuales de sujetos o documentos (Anguera, 1985).

También es posible desarrollar procedimientos de análisis factorial booleano y confirmatorio. En el primer caso, el *análisis factorial booleano* es una alternativa al análisis factorial clásico para variables dicotómicas basadas en el álgebra de Boole. Su objetivo es expresar un conjunto de variables observadas por medio de un grupo menor de factores o variables latentes. La obtención de los factores se realiza a partir del cálculo de matrices de discrepancia entre el valor observado y el estimado. Las puntuaciones factoriales de un individuo tienen valor uno si éste presenta una discrepancia para cualquiera de las variables dominantes en el factor (las que no tienen una saturación cero) y cero en el caso contrario.

Por su parte, el *análisis factorial confirmatorio* es una técnica que nos permite descubrir una serie de variables latentes estudiando la covariación entre un grupo de variables observadas, y atendiendo a una serie de restricciones: los factores comunes no tienen por qué estar necesariamente correlacionados; las variables observadas no precisan tener un factor único asociado a ellas; es posible encontrar correlaciones entre los factores únicos. Cuando se aplica a datos cualitativos, el análisis se inicia con la construcción de una tabla de contingencia a partir de la cual se calcula la correlación tetracórica entre las variables. Este tipo de correlación implica realizar una estimación de la correlación entre variables latentes que presentan una distribución conjunta normal bivalente. La estimación puede desarrollarse por diferentes métodos. Los programas PRELIS y LISREL7 (Jöreskog y Sörbom, 1989) facilitan los cálculos necesarios para la contrastación del modelo diseñado en cada caso.

b) *Orientados a la descripción mediante relaciones topológicas*

La dificultad de asociar valores numéricos que representen isomórficamente a las realidades educativas ha llevado a considerar métodos estadísticos basados en medidas ordinales o nominales, y generalmente interesados por modelos de representación gráficos más que por la expresión algebraica numérica. Como alternativa, se ha desarrollado una línea de análisis estadísticos que se apoya en las relaciones ordinales entre los objetos; lo esencial en el análisis no es la precisión numérica sino la estructura de un sistema de posiciones y distancias relativas entre objetos que es reflejado mediante un lenguaje geométrico-espacial (Conde, 1987): Entre ellos, pueden citarse los métodos factoriales desarrollados a partir de las aportaciones de Benzécri (1973) y los métodos de escalamiento no métrico.

Los métodos factoriales operan a través de cálculos basados en el álgebra lineal, y proporcionan representaciones gráficas en las que se sintetiza un amplio número de valores numéricos, situando los objetos a describir sobre ejes factoriales. El antecedente de estas técnicas se encuentra en el análisis factorial clásico, pero a diferencia de éste, en el que se considera que cada sujeto puede ser resultado de la combinación de unos factores comunes, unos factores específicos y una perturbación aleatoria, en los procedimientos de la escuela francesa no se parte de un modelo a priori. Entre las técnicas más empleadas se encuentra el *análisis de correspondencias*, el cual proporciona al analista una representación geométrica de la información contenida en una matriz de datos que adopta la forma de una tabla de contingencia. Los resultados de un análisis de correspondencias se presentan en gráficos que indican las configuraciones de puntos en proyecciones planas formadas por los ejes principales tomados de dos en dos (Lebart, Morineau y Warwick, 1984:44): La inspección de estos gráficos permitiría identificar los rasgos que se corresponden con cada eje y los puntos-fila y/o puntos-columna agrupados en una misma región del espacio. Existen un conjunto de paquetes de programas que facilitan este método de análisis desde diferentes perspectivas, tales como el SPAD-N (Lebart, Morineau y Lambert, 1988), ANCORSIM (Cornejo, 1988), la serie Dual3 (Nishisato y Nishisato, 1984), Mapwise (Goodnow, 1988), los programas CORAN y CORRESP del paquete PC-MDS (Carroll et al., 1986) o el programa SimCA (Greenacre, 1986).

El *escalamiento multidimensional* incluye una variedad de técnicas cuyo propósito común es hacer explícito el modelo o estructura que se encuentra oculto en una matriz de datos, y representar esa estructura en algún modo que resulte mucho más accesible para el ojo humano (Shepard, Rommey y Nerlove, 1972): Estas técnicas operan sobre una matriz cuadrada de distancias entre objetos o individuos para obtener una representación espacial consistente en una configuración geométrica de puntos, a modo de mapa, que refleja la estructura oculta de los datos y hace que las relaciones entre éstos sean más fáciles de comprender (Kruskal y Wish, 1986): En el escalamiento multidimensional *no métrico*, las distancias calculadas entre cada par de objetos son números que expresan una similitud o diferencia entre ambos. El

objetivo de este procedimiento es encontrar un conjunto de puntos en un espacio de dimensión mínima tales que los datos de disimilaridad sean una función monótona (ordenada) de las distancias en ese espacio. Es decir, para todo conjunto de datos que estén ordenados del mismo modo siempre obtendremos idéntica solución (Coxon y Jones, 1984): El programa MDSCAL de Kruskal fue el primero de los creados para llevar a cabo el escalamiento multidimensional no métrico. Entre los programas más recientes se encuentra, con el mismo nombre, el elaborado por Orloci y Kenkell (1987).

c) *Basados en procedimientos de clasificación*

Se trata de técnicas que persiguen la asignación de un nuevo objeto u observación a su propio lugar en un conjunto bien establecido de categorías, conglomerados o clases. Cuando trabajamos con datos cualitativos, podemos realizar clasificaciones de naturaleza estadística mediante el empleo de métodos multivariantes tales como los análisis de conglomerados o las técnicas de clasificación de la escuela francesa. Los primeros, los métodos denominados *análisis de conglomerados*, tienen por finalidad agrupar objetos o individuos que son similares o tienen algunas características comunes. Se parte de una matriz de n sujetos sobre los que se han medido p variables. Esta matriz se transforma en otra de dimensiones $n \times n$ donde se recogen las semejanzas (o desemejanzas) entre cada pareja de objetos, de acuerdo con un criterio de proximidad (o lejanía): A partir de ella se forman los grupos. Dunn-Rankin (1983:122) distingue cuatro pasos en el proceso del análisis de conglomerados: recogida de la información sobre los objetos, obtención de las distancias entre todos los objetos, agrupamiento de los objetos en base a una medida de proximidad (o lejanía) y representación gráfica de los grupos que han sido formados.

Los *métodos de clasificación* desarrollados por la escuela francesa de análisis de datos son similares a los análisis de conglomerados mencionados más arriba. Permiten agrupar en clases o jerarquías de clases objetos o individuos sobre los cuales poseemos cierta información. Utilizan procedimientos basados en cálculos algorítmicos y producen representaciones gráficas generalmente en forma de árbol invertido. Lebart, Morineau y Fenelon (1985) presentan diversas técnicas de clasificación que emplean *algoritmos ascendentes o aglomerativos*, en los cuales se procede a la construcción de las clases por aglomeraciones sucesivas de elementos dos a dos, dando lugar a una jerarquía de la partición de los elementos; *algoritmos descendentes o divisivos*, que proceden mediante dicotomías sucesivas del conjunto de los elementos, propiciando igualmente una jerarquía de la partición; y algoritmos que conducen directamente a las *particiones*. Como programa informático para el desarrollo de los procedimientos de clasificación podemos utilizar SPAD-N (Lebart, Morineau y Lambert, 1988):

Las técnicas **basadas en relaciones de dependencia** consideran previamente variables explicativas y variables a explicar. Por tanto, resultan adecuadas cuando

interesa predecir determinadas variables en función de otras. Básicamente, los dos métodos para el estudio de la dependencia entre variables utilizados con datos cualitativos son los modelos logit y el análisis de correlación canónica. Los *modelos logit* se interesan por el estudio del efecto de un conjunto de factores (variables independientes) sobre una variable dependiente considerada como respuesta. En este análisis está presente la idea de la regresión múltiple. A diferencia de los modelos lineales logarítmicos, cuando se ajusta un modelo logit se consideran todas las asociaciones e interacciones significativas que incluyen la variable dependiente, junto con la interacción entre las variables independientes. Se trata de encontrar qué efectos de los que incluyen la variable dependiente son significativos (Sánchez Carrión, 1984:287-8).

El *análisis de correlación canónica*, por su parte, es adecuado cuando pretendemos predecir los valores de un conjunto de variables (criterios) en función de los obtenidos para otro conjunto de ellas (predictores): Se trata de obtener dos combinaciones lineales, una de las variables criterio y otra de las variables predictores, tales que la correlación simple entre la variable independiente compuesta y la variable dependiente compuesta sea máxima (Warwick, 1975): El uso del análisis de correlación canónica con propósitos descriptivos no requiere asunciones sobre la distribución de las variables; en tales casos, las variables predictoras y criterios pueden ser medidas en el nivel nominal u ordinal (Dillon y Goldstein, 1984:339).

2. Datos transformados procedentes del recuento de unidades diferenciadas en los textos

Como hemos expresado anteriormente, una de las vías habituales en el análisis de datos cualitativos textuales consiste en la aplicación de técnicas estadísticas a los valores numéricos resultantes del recuento de unidades en los textos. El análisis interpretativo de los datos suele verse complementado con un recuento de las unidades incluidas en cada categoría considerada durante el proceso analítico. A veces, la cuantificación se utiliza únicamente para valorar la importancia de determinados tópicos o conceptos dentro del corpus de datos, pero es posible también construir tablas de contingencia cruzando las categorías con distintos sujetos, instituciones, contextos, etc. estudiados. En este caso, los datos resultantes de la cuantificación de los textos pueden ser objeto de las técnicas que se han presentado ya al hablar de los datos cualitativos transformados en valores numéricos tras las operaciones de categorización y codificación.

Sin embargo, en los últimos años han cobrado un impulso considerable los enfoques analíticos que se basan en la cuantificación directa de los textos, sin que medie una reducción o transformación previa de los datos por parte del analista. El procedimiento consiste en contar las unidades elementales presentes en los textos (palabras o secuencias breves de palabras) y realizar análisis estadísticos a partir de los recuentos. El análisis estadístico de textos tiene su origen en los análisis cuantitativos realizados sobre obras literarias, que iban dirigidos al recuento de palabras,

el estudio de la distribución del vocabulario, la comparación del léxico empleado por distintos autores o por un mismo autor en diferentes períodos creativos. Las investigaciones de autores como Yule (1944), Zipf (1946), Guiraud (1960), Muller (1968), entre otros, y el posterior desarrollo y popularización de la informática se encuentran en la base de los métodos de la denominada estadística textual, que han acabado aplicándose al estudio de los datos textuales en muy diversos ámbitos: historia, literatura, sociología, etc.

La aparición del programa GENERAL INQUIRER (Stone y otros, 1966) marcó el inicio del análisis automático de textos. El programa se basaba en la búsqueda y recuento de palabras y frases previamente identificadas por el analista mediante la definición de un diccionario confeccionado para el análisis. Mochmann (1983) describe programas más recientes basados en estos mismos principios: EVA, SPENCE, COFTA, COTAG, TEXPACK. Otros programas, en lugar de partir de un diccionario previo, han tratado de extraer los temas presentes en un texto sometiendo a tratamientos estadísticos las frecuencias de cada palabra en cada una de las unidades consideradas. Es el caso del programa WORDS (Iker, 1975): La idea base sigue siendo el despiece automático de un texto en unidades elementales, para operar cálculos estadísticos a partir de su recuento.

Los métodos de la escuela francesa de análisis de datos, resultan especialmente adecuados para el análisis de grandes matrices de datos como las originadas al examinar la distribución de unidades elementales dentro de un texto. Enmarcados en esta línea, se han difundido programas específicamente diseñados para el análisis de datos textuales, tales como LEXICLOUD (Salem, 1987), ALCESTE (Reinert, 1986) o SPAD.T (Lebart, Morineau y Bécue, 1989): Las técnicas que permiten desarrollar se basan, en todos ellos, en el análisis de correspondencias, las clasificaciones jerárquicas o el cálculo de medidas de asociación (chi-cuadrado, método de las especificidades):

El procedimiento general de análisis de datos cualitativos así como los métodos específicos en que éste puede concretarse, representan propuestas útiles para reducir, disponer y obtener conclusiones significativas a partir de la evidencia recogida en relación con los diferentes problemas a los que se enfrenta la investigación sobre la diferenciación educativa. Las técnicas presentadas para el análisis de datos numéricos, procedentes de la transformación de datos textuales, han sido empleadas para abordar algunas de las cuestiones que, desde hace años, vienen preocupando a los investigadores que trabajan en este campo. Partiendo de datos cualitativos numéricos, para la identificación de rasgos se recomienda el análisis factorial o para la construcción de tipologías el análisis de conglomerados (Pérez Juste, 1985): Por otra parte, a un nivel más general, Tejedor (1988a) plantea la utilización del análisis factorial de correspondencias o del escalamiento multidimensional para describir o reducir datos; el uso del análisis canónico con una intención predictiva; o la utilidad del análisis de conglomerados para agrupar la información recogida.

No obstante, queremos insistir también en el interés que presentan las técnicas

de análisis de datos textuales, no basadas en procedimientos estadísticos. Entendemos que las principales aportaciones del análisis de datos cualitativos al campo de la diferenciación educativa cabe situarlas en la resolución de problemas en los que el pluralismo cultural y lingüístico intervengan como elemento fundamental. Del mismo modo, en aquellos problemas en los que las categorías culturales o significados propios de cada grupo social resulten variables fundamentales para explicar o intervenir en determinadas situaciones educativas. En estos casos, resulta más apropiado un tipo de análisis que mantenga en esencia la estructura y los términos que utiliza un individuo o una comunidad para expresar y comunicar su enfoque de la realidad.

El análisis de datos cualitativos puede permitir también aproximarse con éxito a aquellas investigaciones o intervenciones educativas preocupadas por reflejar las interacciones propias del proceso de enseñanza-aprendizaje para los diferentes niveles educativos o en función de las diferencias debidas al sexo. Este tipo de análisis resulta adecuado para estudiar modelos de interacción social y lingüística, como los que se desarrollan entre alumnos y profesores en las clases o en situaciones educativas informales.

Este tipo de análisis nos permite asimismo profundizar en las características idiosincráticas de cada centro educativo en su lucha por adaptar la educación a las características contextuales donde se inserta. Recoger la cultura y el clima de relaciones propios del centro, explicar las múltiples interacciones que tienen lugar dentro de las instituciones educativas, y las que éstas mantienen con los diferentes elementos del contexto en que se insertan, valorar los cambios curriculares introducidos en un centro y/ o en una clase, son algunos de los problemas para los que el análisis de datos cualitativos puede ofrecer respuestas satisfactorias. De igual modo, este tipo de análisis nos ayuda a profundizar en los matices, contradicciones, anécdotas, peculiaridades o particularidades del discurso de personas, grupos o instituciones implicadas en acciones educativas, y que constituyen puntos de referencia fundamentales para explicar ciertas diferencias.

Una aplicación más del análisis de datos cualitativos está vinculada con la explicación global de las diferencias educativas. Este tipo de análisis nos permite alcanzar una comprensión ecológica de los fenómenos, de modo que sea posible entender las relaciones entre un sujeto (grupo, institución, etc.) y su medio, y las relaciones entre los diferentes medios en los que vive ese sujeto. Elementos tales como su historia personal (y la del propio medio en que vive), su enfoque de la realidad (y el mantenido por el grupo social del que es parte), su conductas escolares y no escolares, etc. pueden interpretarse holísticamente bajo formas de análisis de datos cualitativos.

REFERENCIAS BIBLIOGRÁFICAS

- ALBERDI, I. (Coord.) (1985): *Mujer y educación*. Madrid, Instituto de la Mujer. Serie Documentos 3.
 ALBERDI, I. (1986): La educación de la mujer en España, en C. BORREGUERO: *La mujer española: de la tradición a la modernidad*, Madrid, Tecnos.

- ALBERDI, I. y ESCARIO, P. (1987): Sexismo en la enseñanza media, en INSTITUTO DE LA MUJER. *La investigación en España sobre mujer y educación*. Madrid, Instituto de la Mujer.
- ANDER-EGG, E. (1990): *Técnicas de Investigación Social*. Buenos Aires, Editorial Humanitas.
- ANGUERA, M. T. (1985): Posibilidades de la metodología cualitativa vs. cuantitativa. *Revista de Investigación Educativa*, 3 (6), 127-144.
- BAKEMAN, R. (1983): Computing Lag Sequential Statistics: The ELAG Program. *Behavior Research Methods and Instrumentation*, 15, 530-535.
- BAKEMAN, R. y GOTTMAN, J. M. (1989): *Observación de la interacción: Introducción al análisis secuencial*. Madrid, Morata.
- BANKS, J. (1993): Multicultural education: Historical development, dimensions and practice, *Review of Research in Education*, 19, 3-49.
- BARTOLOMÉ PINA, M. (1992): Diseños y metodologías de investigación desde la perspectiva de la educación intercultural. En X Congreso Nacional de Pedagogía. *Educación Intercultural en la Perspectiva de la Europa Unida*. Salamanca, (pp. 647-674).
- BENZECRI, J. P. (1973): *L'Analyse des Données*. París, Dunod.
- BLAIKIE, N. W. (1991): A critique of the use of triangulation in social research. *Quality & Quantity*, 25, 115-136.
- BLISS, J., MONK, M. y OGBORN, J. (1983): *Qualitative Data Analysis for Educational Research*. Londres, Croom Helm.
- BOGDAN, R. C. (1986): A sociology of special education, en R. J. MORRIS y B. BLATT (Eds.): *Special education: Research and trends*. Oxford, Pergamon, (pp. 334-359).
- BOGDAN, R. C. y BIKLEN, S. K. (1982): *Qualitative Research for Education. An Introduction to Theory and Methods*. Boston, Allyn and Bacon, Inc.
- BOWMAN, B. (1993): Early Childhood Education, *Review of Research in Education*, 19, 101-134.
- BRULLET, C. y SUBIRATS, M. (1990): *La coeducación*. Madrid, MEC
- CARTWRIGHT, D. P. (1978): Análisis del Material Cualitativo, en L. FESTINGER y D. KATZ (Eds.): *Los Métodos de Investigación en las Ciencias Sociales*. Buenos Aires, Paidós.
- CARROLL, J. D. et al. (1986): Interpoint Distance Comparisons in Correspondence Analysis. *Journal of Marketing Research*, (23), 271-280.
- CASELLES, J. F. (1991): Pedagogía diferencial, sexo y educación: El sexismo en el sistema educativo mixto y la coeducación como alternativa. En C. JIMÉNEZ FERNÁNDEZ, *Lecturas de Pedagogía Diferencial*. Madrid: Dykinson, (pp. 105-134).
- CORRALES, A. et al.: «NUDIST»: Una herramienta informática para el análisis de datos cualitativos. Comunicación presentada al VI Seminario de Métodos de Investigación, Madrid.
- CONDE, F. (1987): Una propuesta de uso conjunto de las técnicas cuantitativas y cualitativas en la investigación social. El isomorfismo de las dimensiones topológicas de ambas técnicas. *Revista Española de Investigaciones Sociológicas*, (39), 213-224.
- COXON, A. P. y JONES, C. L. (1984): Escalas multidimensionales, en J. J. SÁNCHEZ CARRIÓN *Introducción a las técnicas de análisis multivariable aplicadas a las ciencias sociales*. Madrid, Centro de Investigaciones Científicas, pp. 107-129.
- DAWSON, J. A. (1982): *Qualitative research findings: what do we do to improve and estimate their validity?* Comunicación presentada al Annual Meeting of the American Educational Research Association, Nueva York. ERIC Document Reproduction Service N° ED218330.
- DENZIN, N. K. (1978): *The research act*. Nueva York, McGraw Hill.
- DILLON, W. R. y GOLDSTEIN, M. (1984): *Multivariate analysis*. Nueva York, John Wiley & Sons.
- DUNN-RANKIN, P. (1985): *Scaling methods*. Hillsdale, New Jersey, Lawrence Erlbaum Associates, Publishers.

- ERICKSON, F. (1989): Métodos Cualitativos de Investigación sobre la Enseñanza, en M. C. WITTRICK (Ed.): *La Investigación de la Enseñanza, II*. Madrid, Paidós-MEC, pp. 125-301.
- ERICKSON, F. (1982): Audiovisual Records as a Primary Data Source, en A. GRIMSHAW (Ed.): *Sociological Methods and Research (Special Issue on Sound-Image Records in Social Interaction Research)*, 11, 2, 213-232.
- ESPÍN, J. V. (1991): Los programas de educación compensatoria. ¿Una respuesta a las diferencias socioculturales desde la educación?, en C. JIMÉNEZ FERNÁNDEZ: *Lecturas de Pedagogía Diferencial*. Madrid, Dykinson, pp. 105-134.
- FIELDING, N. G. y FIELDING, J. L. (1986): *Linking data*. Sage University Paper serie on Qualitative Research Methods, 4. Beverly Hills, Sage Publications.
- FILLMORE, L. W. y VALADEZ, C. (1986): Teaching Bilingual Learners. En WITTRICK, M. C. (Ed.): *Handbook of Research on Teaching*. Nueva York: MacMillan.
- FIRESTONE, W. A. y DAWSON, J. A. (1982): *Approaches to qualitative data analysis: intuitive, procedural and intersubjective*. Comunicación presentada al Annual Meeting of the American Educational Research Association. Nueva York.
- FRANZOSI, R. (1990): Strategies for the prevention, detection, and correction of measurement error in data collected from textual sources. *Sociological Methods and Research*, 18 (4), 442-472.
- GAVIRIA SOTO, J. L. (1986): El enfoque diferencial en la investigación pedagógica. *Bordón*, (262), 199-221.
- GIL, J. et al.: *Formulación y comprobación de hipótesis cualitativas con ayuda del programa AQUAD*. Comunicación presentada al VI Seminario de Métodos de Investigación, Madrid.
- GLASER, B. G. y STRAUSS, A. L. (1967): *The Discovery of Grounded Theory*. Nueva York, Aldine Publishing Company.
- GOETZ, J. P. y LECOMPTE, M. D. (1988): *Etnografía y Diseño Cualitativo en Investigación Educativa*. Madrid, Morata.
- GOODNOW, W. E. (1988): *Brief User's Manual*. Chicago, IL: Research Software Inc.
- GREENACRE, M. J. (1986): SimCA: A Program to Perform Simple Correspondence Analysis. *The American Statistician*, (51), 230-231.
- GUBA, E. G. (1983): Criterios de credibilidad en la investigación naturalista, en A. PÉREZ y J. GIMENO, (Eds.): *La enseñanza: su teoría y su práctica*. Madrid, Akal, pp. 148-165.
- GUIRAUD, P. (1960): *Problèmes et méthodes de la statistique linguistique*. París, PUF.
- HAMMERSLEY, M. y ATKINSON, P. (1983): *Ethnography: principles and practice*. Londres, Tavistock.
- HEGARTY, S. (1988): Qualitative research: Introduction, en S. HEGARTY y P. EVANS (Eds.): *Research methods in special education*, Oxford, NFER-Nelson, (pp. 109-113).
- IKER, H. P. (1975): *Words, system manual*. Rochester, Nueva York, Computer Print.
- IKER, H. P. y HARWAY, N. J. (1969): A computer systems approach toward the recognition and analysis of content, en G. A. GERBNER y otros (Eds.): *The analysis of communication content*. Nueva York, 381-405.
- JACOB, E. (1988): Clarifying qualitative research: a focus on traditions. *Educational Researcher*, 17 (1), 16-24.
- JÖRESKOG, K. G. y SÖRBOM, D. (1989): *Lisrel 7. User's Reference Guide*. Mooresville, IN, Scientific Software, Inc.
- KENNEDY, J. J. (1983): *Analyzing Qualitative Data. Introductory Log-Linear Analysis for Behavioral Research*. Praeger.
- KERLINGER, F. N. (1975): *Investigación del Comportamiento*. México, Nueva Editorial Interamericana.
- KRUSKAL, J. B. y WISH, M. (1986): *Multidimensional scaling*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 11. Beverly Hills, Sage Publications.

- LAZARSELD, P. F. y ROSENBERG, M. (1955): *The language of social research: A reader in the methodology of social research*. Nueva York: Free Press.
- LEBART, L., MORINEAU, A. y FENELON, J. P. (1985): *Tratamiento Estadístico de Datos*. Barcelona, Marcombo.
- LEBART, L., MORINEAU, A. y LAMBERT, T. (1988): *SPAD.N: Système Portable pour l'Analyse des Données*. París: CISIA
- LEBART, L., MORINEAU, A. y WARWICK, K. M. (1984): *Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices*. Nueva York: John Wiley.
- LEVINE, H. G. (1985): Principles of data storage and retrieval for use in qualitative evaluations. *Educational Evaluation and Policy Analysis*, 7 (2), 169-186.
- LOFLAND, J. y LOFLAND, L. H. (1984): *Analyzing social settings*. Belmont, Wodsworth, Publishing Company.
- MCINTYRE, D. y MACLEON, G. (1986): The characteristics and uses of sistematic classroom observation, en M. HAMMERSLEY (Ed.): *Controversies in classroom research*, Milton Keynes, Open University Press, 10-24.
- MILES, M. B. (1985): Qualitative data as an attractive nuisance: the problem of analysis, en J. VAN MAANEN (Ed.): *Qualitative Methodology*. Beverly Hills, Sage Publications, pp. 117-134.
- MILES, M. B. y HUBERMAN, A. M. (1984a): Drawing valid meaning from qualitative data: toward a shared craft. *Educational Researcher*, 13 (4), 20-30.
- MILES, M. B. y HUBERMAN, A. M. (1984b): *Qualitative Data Analysis. A Sourcebook of New Methods*. Beverly Hills, Sage Publications.
- MOCHMANN, E. (1985): Análisis de Contenido mediante Ordenador Aplicado a las Ciencias Sociales. *Revista Internacional de Sociología*, 43 (1), 11-44.
- MOHLER, P. Ph. (1985): Algunas Observaciones Prácticas sobre la Utilización del Análisis de Contenido en Ordenador. *Revista Internacional de Sociología*, 43 (1), 45-57.
- MOLLY, E. (1984): *Factors affecting the reduction of narrative data*. Comunicación presentada al Annual Meeting of the American Educational Research Association, New Orleans. ERIC Document Reproduction Service N° ED247300.
- MUCCHIELLI, R. (1988): *L'analyse de contenu des documents et des communications*. París, Les Editions ESF-Entreprise Moderne d'Édition.
- MULLER, C. (1968): *Initiation à la statistique linguistique*. París, Larousse.
- MULLER, C. (1977): *Principes et methodes de statistique lexicale*. París, Hachette.
- ORLOCI, L. y KENKEL, N. C. (1987): *Data analysis in population and community ecology*. Draft Copy, Universidad de Hawaii, Honolulu & New Mexico State University, Las Cruces.
- PÉREZ JUSTE, R. (1985): *Elementos de Pedagogía Diferencial*, Madrid, UNED.
- REINERT, M. (1986): Un logiciel d'analyse lexical (ALCESTE): *Les Cahiers de l'Analyse des Données*, XI (4), 471-484.
- RODRÍGUEZ, G. et al.: *NUDIST versus AQUAD: Dos posibilidades informáticas para el análisis de datos cualitativos*. Comunicación presentada al VI Seminario de Métodos de Investigación, Madrid.
- RODRÍGUEZ, G. (1992): La investigación sobre la eficacia escolar. En C. GARCÍA PASTOR: *La investigación sobre la integración: Tópicos, aproximaciones y procedimientos*. Salamanca, Amarú.
- ROSENBERG, M. (1968): *The logic of survey analysis*. Nueva York: Basic Books.
- SALEM, A. (1987): *Le Lexicloud. Programmes pour le traitement lexicométrique des textes*. Ecole Normale de Fontenay-Saint Cloud.
- SÁNCHEZ CARRIÓN, J. J. (1984): *Introducción a las técnicas de análisis multivariados aplicados a las ciencias sociales*. Madrid, Centro de Investigaciones Sociológicas.

- SÁNCHEZ CARRIÓN, J. J. (1985): Técnicas de Análisis de los Textos mediante Codificación Manual. *Revista Internacional de Sociología*, 43 (1), 89-118.
- SÁNCHEZ CARRIÓN, J. J. (1989): *Análisis de tablas de contingencia. El uso de los porcentajes en las ciencias sociales*. Madrid, Centro de Investigaciones Sociológicas - Siglo XXI de España Editores, S.A.
- SHEPARD, R. N.; ROMNEY, A. K. y NERLOVE, J. B. (1972): *Multidimensional scaling. Theory and applications in the behavioral sciences*. Nueva York, Seminar Press.
- SMITH, J. K. y HESHUSIUS, L. (1986): Closing down the conversation: the end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher*, 15 (1), 4-12.
- SPRADLEY, J. P. (1979): *The ethnographic interview*. Nueva York, Holt, Rinehart and Winston.
- SPRADLEY, J. P. (1980): *Participant observation*. Nueva York, Holt, Rinehart and Winston.
- STONE, P. J. y otros, (1966): *The General Inquirer: a computer approach to content analysis*. Cambridge, Massachusetts, the M.I.T. Press.
- STRAUSS, A. L. (1987): *Qualitative analysis for social scientists*. MA, Cambridge University Press.
- STRAUSS, A. y CORBIN, J. (1990): *Basics of qualitative research. Grounded theory procedures and techniques*. Newbury Park, Sage Publications.
- TEJEDOR, F. J. (1988a): La Pedagogía Diferencial en su conceptualización metodológica. *Bordón*, 40 (4), 589-598.
- TEJEDOR, F. J. (1988b): *Perspectivas de la investigación cuantitativa en educación*. Cádiz. Documento policopiado.
- TESCH, R. (1988): *The impact of the computer on qualitative data analysis*. Paper presented at the Annual Meeting of the American Educational Research Association.
- TESCH, R. (1989): Basic Qualitative Analysis with QUALPRO. *Qualitative Studies in Education*, 2 (4), 367-371.
- TESCH, R. (1990): *Qualitative Research: Analysis Types and Software Tools*. Bristol, The Falmer Press.
- TESCH, R. (1991): Software for qualitative researchers: analysis needs and program capabilities, en N. G. FIELDING y R. M. LEE (Eds.): *Using computers in qualitative research*. Londres, Sage Publications, pp. 16-37.
- VAN MAANEN, J. (1985): The fact of fiction in organizational ethnography, en J. VAN MAANEN (Ed.): *Qualitative Methodology*, Beverly Hills, Sage Publications, pp. 37-55.
- WAINER, H. (1992): Understanding graphs and tables. *Educational Researcher*, 21 (1), 14-23.
- WARWICK, P. V. (1975): Canonical correlation analysis, en N. H. NIE et al.: *SPSS: Statistical package for the social sciences*, Nueva York, McGraw Hill, pp. 515-527.
- WEBER, G. (1971): *Inner-city children can be taught to read: Four succesful schools*. Washington DC: Council for Basic Education.
- WILSON, S. (1977): The use of ethnographic techniques in educational research. *Review of Educational Research*, 47 (1), 245-265.
- WOLCOTT, H. F. (1986): *Problem finding in qualitative research*. Universidad de Oregon, Documento de trabajo.
- WOODS, P. (1989): *La escuela por dentro*. Barcelona, Paidós-MEC.
- YULE, G. U. (1944): *A statistical study of vocabulary*. Cambridge, University Press.
- ZIPF, G. K. (1946): *The psychobiology of language, an introduction to dynamic philology*. Boston, Houghton-Mifflin.
- ZNANIECKI, F. (1934): *The method of sociology*. Nueva York, Farrar & Rinehart.