

ALGO MAS QUE RECUPERAR PALABRAS Y CONTAR FRECUENCIAS: LA AYUDA DEL ORDENADOR EN EL ANALISIS DE DATOS CUALITATIVOS

GÜNTER HUBER
CARLOS MARCELO GARCIA

Una de las razones por las cuales muchos investigadores sociales reaccionan al concepto de «investigación cualitativa» con –dicho eufemísticamente– discreción es la falta de procedimientos concretos para el análisis de datos cualitativos. En este sentido, Miles (1983) caracterizó a los datos cualitativos como una «molestia atractiva», y su descripción aún sigue siendo acertada. Los investigadores que utilizan métodos cuantitativos pueden confiar, para la mayoría de sus estudios, en instrumentos de medida estandarizados y posteriormente escoger de entre un surtido «stock» de procedimientos de análisis. Los investigadores cualitativos, de otra parte, pueden dirigirse a un amplio repertorio de métodos para recopilar datos, que en su caso significa revelar la unicidad de situaciones sociales específicas o la subjetividad de la visión personal del mundo. Pero en este momento, estos investigadores se quedan solos con una cantidad abrumadora de ricos y sugerentes datos (la mayor parte de ellos informes verbales), ¡y con la ingenuidad o la intuición de darles algún sentido!

TAREAS FUNDAMENTALES DEL ANALISIS CUALITATIVO

«Dar sentido» a los datos cualitativos, significa *reducir* notas de campo, descripciones, explicaciones, justificaciones, etc. más o menos prolijas, hasta llegar a una cantidad manejable de unidades significativas y manejables, significa también *estructurar* y exponer estos items, y por último, extraer y confirmar unas *conclusiones* más comprensivas. Miles y Huberman (1984) destacan lo que ellos consideran que es el «análisis cualitativo» mediante estas tres formas de actividad, pero al mismo tiempo los autores puntualizan que estas actividades de análisis, junto con la recopilación de datos forman un proceso interactivo y cíclico. En otras palabras, dependiendo de las necesidades del proceso de investigación, los aspectos de reducción de datos, estructuración y presentación de conclusiones se interrelacionan e influyen unos con otros. Una breve descripción de lo que puede ocurrir durante el análisis de entrevistas nos ayudará a tener una impresión más concreta, tanto de la naturaleza dinámica de la investigación cualitativa, como de las posibles contribuciones del ordenador en esta tarea.

Ya durante la recopilación de datos, el investigador a menudo empieza a buscar, de forma más o menos sistemática, las ideas generales que están detrás de las declaraciones de los sujetos, por ejemplo, buscar unidades centrales de significado en los textos de las entrevistas. Normalmente, estas ideas generales se representan por códigos. Los códigos, no sólo reducen declaraciones individuales llamativas y

destacadas en un sistema menos complejo y ambiguo de anotación que el lenguaje ordinario, también representan un esfuerzo interpretativo y explicativo del investigador. Por tanto, conociendo que la codificación es principalmente un proceso de interpretación, es preciso que los resultados deban ser controlados para que la investigación sea objetiva, fiable y válida, u otros criterios más adecuados para la investigación cualitativa (cf. Lincon y Guba, 1985). Para garantizar sobre todo la validez interna como un requisito previo para posteriores consideraciones, es necesario verificar si los códigos se utilizaron de forma consistente, si las áreas de significado que definen se observaron en todos los casos, si el significado simbolizado por códigos específicos realmente coincide con el contenido de aquellos pasajes de texto que reducen, etc. (Huber, 1989). Una vez que la tarea de interpretar o codificar un texto se ha llevado a cabo con éxito, al menos de momento, los datos originales se reducen considerablemente.

Tener una relación de códigos con validez interna, sin embargo, no es suficiente como producto de un análisis cualitativo, incluso si a través de la descripción de la frecuencia de los códigos y del modo/s de su distribución, llegan a hacerse visibles los puntos cruciales del texto original. En el mejor de los casos esto es una aproximación a la estructura del significado en el texto codificado. Lo que se requiere es mostrar configuraciones de códigos comunes y/o repetidas. Miles y Huberman (1984) sugieren y usan en su libro principalmente el formato bidimensional de matrices para responder a las demandas de la estructuración de los datos, por ejemplo: proporcionar una información bien ordenada de forma simultánea (en lugar de secuencial); permitir la comparación con los resultados de otros sujetos, situaciones, tiempos, estudios, etc.; así como dar directrices para realizar nuevos y más refinados análisis. Y si es verdad que una imagen vale más que mil palabras, entonces, las matrices reemplazan al menos a 100 palabras, sobre todo porque sitúan los datos y los procesos analíticos juntos en una situación, así como indican la necesidad de posteriores análisis.

Sin embargo, incluso una sofisticada exposición a partir de los datos reducidos, no es una conclusión suficiente sino un requisito previo, de considerable ayuda, para extraer conclusiones. Cuando se llevan a cabo análisis de alto orden, se deben de especificar las condiciones complejas, que normalmente incluyen diferentes códigos o unidades de significado y sus relaciones lógicas específicas, en algunos casos también se incluyen datos cuantitativos y socio-demográficos. Al llevar a cabo un formato bidimensional, éste puede resultar sobrecargado por las demandas del análisis, incluso cuando se llega al acuerdo de seleccionar las matrices más amplias y las más específicas, se corre el riesgo de perder claridad. Por el contrario, un instrumento que parece ser de gran ayuda es la combinación de condiciones relevantes para llegar a una conclusión en la forma de una simple pregunta. Podemos entender que al responder a tal pregunta, estamos contrastando hipótesis complejas. En el último capítulo de este artículo se presenta un ejemplo acerca del desarrollo de hipótesis en este tipo de análisis. El último paso en investigación cualitativa descrito por Miles y Huberman (1984) es «representar conclusiones» (drawing conclusions). Pueden ser conclusiones respecto configuraciones de unidades de significado típicas y repetidas, ya sean secuencias específicas, agrupaciones, estructuras jerárquicas y dimensionales, o relaciones causales.

La interacción entre la reducción de datos cualitativos y la fase de representar conclusiones es evidente. Si modificamos una interpretación, tanto al cambiar un código como un conjunto de códigos, las conclusiones que incluyen a estos códigos comienzan a ser cuestionadas. Si por otro lado, identificamos en un texto un patrón

de códigos que es incongruente con las reglas utilizadas para la interpretación hasta ese momento, entonces hemos de controlar los otros textos y probablemente debamos modificar las definiciones de algunos códigos. Ello implica volver a empezar con el proceso de reducción de datos desde el principio, cambiando, como consecuencia, la representación de los datos y las conclusiones. Procesos circulares como este son característicos de los estudios cualitativos. Después de un primer círculo a través de procesos elementales de análisis de datos, siempre tenemos más familiaridad con nuestros sujetos, con sus visiones del mundo, con los principios fundamentales que definen una situación social específica, etc. que al principio. Usando nuestra intuición podemos comenzar una segunda vuelta y corregir lo que parece ahora ser una incorrecta interpretación. Tal como establece la noción del círculo hermenéutico, esto es en principio un proceso que nunca acaba. Desgraciadamente estos procesos circulares no sólo definen el procedimiento general en investigación cualitativa, sino también sus fases más concretas. Con frecuencia, las reglas básicas de la codificación cambian durante el proceso de reducción de datos, y a veces no somos conscientes inmediatamente de este cambio. Por ejemplo: una entrevista se interpreta de una forma y la siguiente de una forma diferente; tan pronto como nos damos cuenta de la inconsistencia en nuestro fichero de códigos, hemos de volver al principio. Ello implica que hemos de leer de nuevo muchas páginas del texto, así como buscar unidades de significado específicas. Asimismo puede ocurrir que notemos al final del proceso de reducción de datos que hemos combinado algunos códigos, o que hemos diferenciado otras unidades de significado. De nuevo, hemos de luchar con los cientos de páginas y todas las dificultades que mientras tanto adjuntamos al texto. Los problemas se multiplican si dificultades como estas aparecen durante la fase del contraste de hipótesis. En este caso hemos de repetir la búsqueda simultánea para dos, tres o incluso más códigos que aparecen en determinadas configuraciones en todas nuestras transcripciones.

Sin exagerar, podemos resumir que un análisis complejo de datos cualitativos, pronto se convierte en intrincado, y sin un cuidadoso y sistemático procedimiento de control, siempre se está en peligro de perder la propia orientación. La probabilidad de perder la pista de los elementos de una hipótesis y sus relaciones es alta, sobre todo durante la representación de conclusiones, mientras se comprueban cientos de páginas de transcripciones de este patrón específico. Hace falta una herramienta en investigación cualitativa que ayude a los investigadores a mantener una visión sobre sus decisiones y a modificarlas sin dificultad.

LA CONTRIBUCION DE LOS ORDENADORES EN EL ANALISIS DE ALTO ORDEN

Siempre que exista un adecuado «software», los ordenadores pueden ser esa herramienta de ayuda. Después de algunos intentos poco satisfactorios con programas de bases de datos ya elaborados, la sugerencia realizada por Shelly (1986) en relación a la «vida después de la codificación» motivó al primer autor de este artículo para desarrollar sus propios programas «AQUAD» (Analysis of QUALitative Data), adaptados a las características peculiares de la investigación cualitativa. Shelly ha descrito «QUALOG», un paquete de programas para el manejo de datos cualitativos, el cual no sólo ayuda a sistematizar el proceso de reducción de datos, sino que también ofrece un apoyo eficaz en el contraste de hipótesis. Desgraciadamente, estos programas estaban escritos en una extensión de LISP, un lenguaje no utilizable por ordenadores personales. Por el mismo tiempo, sin embargo, se hizo accesible para ordenadores personales el lenguaje de programación basado en Inteligencia Artificial denomi-

nado TURBO-PROLOG. Y por encima de todo, Shelly y Sibert (1985) fueron tan amables de enviar el manual de usuario de QUALOG, a partir del cual se pudo derivar la arquitectura del programa. Existen algunas ventajas que hacen que la programación lógica sea especialmente apropiada para investigación cualitativa (cf. Sibert y Shelly, 1985, en LOGLISP):

- PROLOG es un lenguaje descriptivo. Al contrario que otros lenguajes procedimentales o «imperativos» más extendidos (como BASIC o PASCAL), que obligan al programador a definir cada paso en el proceso de datos, en PROLOG es suficiente –pero sólo en principio– describir el espacio problema (por ejemplo, listar todos los datos y reglas conocidas) y formular una declaración («goal») seleccionada para ser válida en ese espacio problema. PROLOG proporciona al ordenador un mecanismo de razonamiento deductivo retroactivo que se aplica a las hipótesis del usuario, es decir, la máquina intenta encontrar al menos una combinación de datos (y reglas) que prueben la hipótesis.

- Información simbólica, por ejemplo, se pueden manejar los datos verbales sin necesidad de trasladarlos a otro sistema de anotación, como por ejemplo números.

- Todos los datos pueden introducirse sin preocuparse por su número, tamaño, tipo, etc. de campos, directamente como ficheros ASCII a partir de cualquier procesador de textos.

- Todos los datos introducidos se pueden modificar con rapidez y flexibilidad si fuera necesario, por ejemplo durante uno de los procesos circulares de interpretación anteriormente descritos.

- Incluso las relaciones muy complejas entre datos se pueden analizar de forma poco complicada. Al contrario que ocurre con los programas usuales de bases de datos, PROLOG compara cada elemento de su banco de datos con cada uno de los otros elementos, es decir, la lógica interna se aplica también a la comparación a lo largo de todas las entradas, no sólo de «registros» simples. Por lo tanto, no es necesario combinar los componentes de un «registro» con los de otros «registros» mediante clasificación múltiple.

Estas características permitieron construir un instrumento que ayuda a la investigación cualitativa en las fases de reducción de datos, representación, y presentación de conclusiones.

CONTRIBUCIONES EN LA REDUCCION DE DATOS

Durante el proceso de codificación surgen constantemente problemas, que pueden resolverse fácilmente con la ayuda de un sistema de «software». En los estudios que se llevan a cabo sin este recurso, a menudo no se explicita la forma como los investigadores han abordado estos problemas, o si por el contrario, no los han tenido en cuenta. Estos problemas atacan aquellos discutidos bajo el epígrafe de objetividad, fiabilidad, y validez en investigación cualitativa (cf. LeCompte y Goetz, 1982):

- ¿Las reglas de codificación se utilizaron de forma consistente? O formulada la pregunta al contrario: ¿Cambiaron los significados de las categorías durante el proceso de codificación? Se puede responder a estas preguntas buscando y comparando todos los pasajes en cada texto que fue interpretado con el mismo código. Sin ayuda técnica, incluso un simple control de fiabilidad como es éste, supondría grandes esfuerzos, buscando cada código a lo largo de cientos de páginas, se podrían preparar tarjetas de códigos específicos, y no pensemos en el costo adicional de sólo uno o dos cambios en el sistema de codificación.

– ¿Permanecemos dentro de los límites de nuestros códigos o se produjeron solapamientos de significados durante el proceso de interpretación? Podría suceder que los códigos se utilizaran de forma consistente, pero siempre de forma incorrecta. Para comprobar este problema, se debe recuperar y comparar todas las unidades de significado que se supone se han definido con códigos solapados.

– ¿Podemos encontrar en los correspondientes textos los significados representados por códigos particulares? En términos de los criterios clásicos, preguntamos por la validez de contenido de los códigos.

– ¿Hemos introducido nuevos aspectos de interpretación durante el proceso de reducción de datos, por ejemplo «hipótesis valorativas» (Dreher, et al., 1985)? En la medida en que ello no nos dirige a los problemas descritos anteriormente, hemos de asegurarnos de en qué momento, en qué texto, en conexión con qué aspectos, etc., comenzó a aparecer esta nueva perspectiva. En el caso en que notemos tal cambio de estrategias, tendremos que repetir las anteriores interpretaciones, tomando en consideración estos nuevos cambios.

Incluso cuando se utilizan estrategias deductivas de reducción de datos –aplicando un sistema de categorías pre-elaborado– resulta bastante difícil responder a las preguntas sobre fiabilidad de los análisis cualitativos. Los problemas aumentan de forma considerable cuando utilizamos un esquema conceptual preliminar que debe extraerse de forma inductiva. En este caso, no podemos esperar obtener una codificación consistente a la primera vez, por el contrario, hemos de establecer la consistencia a la segunda, tercera ocasión mediante la modificación de las reglas de codificación. Incluso más complicado resulta si reducimos los datos de forma inductiva e intentamos atrapar los pensamientos de nuestros sujetos por medio de categorías que se desarrollan durante el proceso de codificación: ¿Los mismos/diferentes códigos significan las mismas/diferentes unidades de significado? Para poder responder a preguntas como éstas, es necesario elaborar una sinopsis de los momentos críticos a partir de los textos completos, sinopsis que se puede enriquecer con datos verbales contextualizados. Algunas veces sospechamos que estas cosas no preocupan mucho a los investigadores cualitativos que no utilizan el recurso del ordenador.

CONTRIBUCIONES EN LA REPRESENTACION DE LOS DATOS

Cada análisis realizado con AQUAD produce una salida «output» bien en el monitor, en la impresora o en disco (flexible o duro). Las Tablas 2-4 representan ejemplos de este tipo de representación. AQUAD incluye tres funciones que se programaron específicamente con el propósito de representar datos:

– En relación con el análisis convencional de textos, podemos recuperar «palabras claves en contexto» (cf. Weber, 1985). En algunos casos de interpretación de textos o de control para realizar análisis estandarizados, puede servir de ayuda recibir un resumen rápido sobre palabras críticas que puedan jugar un importante papel en la definición de algunos códigos. Esta función presenta tales palabras junto con datos acerca de su localización en el texto y su contexto. Esta representación, sirve de ayuda si queremos comprobar la significación de la frecuencia de palabras.

– Las matrices parecen ser el formato de representación más importante, tal como describimos anteriormente. Con la ayuda del ordenador, podemos definir una matriz bidimensional para representar pasajes de textos representativos, en los que hemos de indicar las cabeceras de filas y columnas. Para las columnas hemos de se-

leccionar códigos «simples», es decir, códigos que por definición sólo aparecen una vez en el fichero de códigos para un texto concreto, por ejemplo, sexo o edad de los sujetos, posición de la base de datos en la secuencia de entrevistas, etc. Las filas incluyen cualquier código de interés. En el estudio que describiremos a continuación, podríamos, por ejemplo, obtener una visión general pidiendo una representación matricial de pasajes de texto estructurados por el criterio de sexo (dos columnas) y el criterio contenido (tres columnas): CRE (creencias), PRE (preocupación), y SIM (sí mismo) (cf. Tabla 1). Como representación, podemos tener todos los pasajes de texto de todos los textos que cumplan la definición de nuestra matriz de 2 por 3 celdillas.

– Las matrices de texto a menudo consumen mucho papel, y puede que necesitemos de una pared vacía para componer el contenido de las celdas en el formato de dos dimensiones. Una tercera ayuda que posibilita la representación de los datos cualitativos consiste en sustituir el texto original por los códigos (utilizados para la reducción de los datos), junto con su localización (números de líneas) en el texto original. Especialmente si completamos los códigos con breves comentarios conseguimos de una parte listados de impresora más claros y espaciosos, mientras que de otra parte, en muchos casos son suficientes para las decisiones a tomar sobre posteriores análisis, resúmenes, etc.

CONTRIBUCIONES EN LA PRESENTACION DE CONCLUSIONES

Con la ayuda de un programa de ordenador, podemos formular hipótesis acerca de las relaciones entre unidades de significado, como unidades de significado de alto orden, y contrastarlas de forma rápida y fiable. Ya anteriormente establecimos que algunas conclusiones a partir de los datos, pueden ser de interés:

Secuencias de unidades de significado: Simplemente con buscar sucesiones específicas de códigos podemos identificar configuraciones secuenciales de especial importancia. En el estudio que a continuación describiremos, intentaremos encontrar si profesores principiantes, cuando hablan sobre sus primeros meses de enseñanza, a menudo describen los antecedentes socioeconómicos de los alumnos (código «PRO»; ver Tabla 1) después de responder sobre el rendimiento o sobre problemas de conducta (códigos «REN» o «CON»).

Agrupaciones de unidades de significado: Podríamos buscar también unidades de significado que estuvieran localizadas en el texto más cerca de unas unidades de significado que de otras. ¡Por supuesto que en este caso hemos de asegurarnos de que esa corta distancia sea significativa! Ofrecemos un ejemplo de ello en la Tabla 3. Guiados por consideraciones didácticas, hemos buscado todos los códigos con una distancia máxima de cinco líneas de texto a partir de segmentos de texto codificados como «MET» (declaraciones de los profesores principiantes sobre aspectos referidos a los métodos de enseñanza). Una estrategia menos específica consistiría en buscar, generalmente, todos los códigos que estén dentro de una determinada distancia unos de otros.

Estructuras jerárquicas: Las unidades de significado que se entrecruzan a menudo, revelan un significado adicional. Podría ser interesante, por ejemplo, identificar en el estudio ya mencionado, profesores principiantes que hablan acerca del sistema educativo, y comprobar si dentro de dichos párrafos aparecen incluido el código «ECE» (experiencias como estudiante).

Estructuras dimensionales: Podemos también tener acceso a estructuras complejas de significado a través de la combinación de unidades de significado o sus códigos, que estén correlacionadas, para llegar a estructuras cuasi-factoriales o «meta»-códigos. En el estudio sobre profesores principiantes, podríamos encontrar combinaciones específicas de códigos a partir de la categoría 2.2. («alumnos», ver Tabla 1). Este hallazgo podría suponer que los profesores principiantes usan «teorías implícitas acerca de los estudiantes».

Relaciones causales: Comparado con las posibilidades de representación de conclusiones que anteriormente hemos mostrado, podemos entender las relaciones causales como un contraste de hipótesis en un sentido bien distinto. Por lo tanto, este tipo de conclusiones parece adecuarse sólo para los análisis cualitativos que se refieren a fundamentaciones teóricas o a datos empíricos válidos y ya disponibles.

PRINCIPALES CARACTERISTICAS DEL PROGRAMA AQUAD

Las características más importantes del programa para análisis de datos cualitativos, en nuestra opinión son la flexibilidad en la codificación y las facilidades deductivas que posee para comprobar relaciones complejas entre códigos, no sólo entre simples datos, sino a través de combinaciones opcionales de todos los elementos del banco de datos (ficheros de códigos). Dado que Turbo-PROLOG ofrece estas características, éstas se utilizaron para desarrollar el paquete de programas AQUAD 2.2., que se compone de cinco elementos:

El Componente AQD-TM (Textos y Apuntes), permite:

- numerar las líneas de los ficheros de textos importados a partir de un procesador de textos en formato ASCII;
- añadir códigos al texto al final de cada línea (codificación múltiple, solapamiento, entrecruce de códigos);
- producir un fichero que incluya sólo los códigos a partir de un fichero de texto codificado; estos ficheros de códigos incluyen solamente el número de la base de datos, número de la primera y última línea codificada, y un código;
- editar textos e interpretaciones (los códigos añadidos a las líneas de textos);
- imprimir ficheros de texto original, textos numerados o ficheros de textos codificados;
- recuperar todos los pasajes del texto que se codifican con un código determinado;
- escribir, grabar, y editar apuntes, así como recuperarlos.

El Componente AQD-C (Códigos), permite:

- producir ficheros de códigos sin el texto original;
- añadir códigos a ficheros de códigos ya existentes;
- editar ficheros de códigos con las funciones principales de un procesador de texto (recuperar, sustituir, copiar, mover, borrar);
- imprimir ficheros de códigos;
- clasificar códigos alfabéticamente y/o por números de líneas;
- recuperar códigos o patrones específicos de codificación (solapamiento, orden jerárquico, codificación múltiple de la misma unidad de significado, códigos dentro de una determinada distancia en términos de números de líneas);
- contar la frecuencia de todos o algunos códigos.

El Componente AQD-H (Hipótesis)

Es la parte central del paquete. En él se utiliza la lógica deductiva de TurboPROLOG para el contraste de hipótesis sobre relaciones entre unidades de significado en los textos. El siguiente capítulo describe el proceso a partir de los primeros análisis más simples, hasta contrastar una hipótesis más compleja en un estudio en desarrollo. En este estudio, desarrollamos una hipótesis muy especializada. Con el propósito de hacer más fácil a los usuarios que no están familiarizados con añadir declaraciones al programa principal, el componente AQD-H cuenta con nueve estructuras de hipótesis, es decir, hipótesis que ya están formuladas y que incluyen variables en lugar de códigos o definiciones de distancias y que permiten el análisis de alto orden de las relaciones entre los datos verbales. Ayudado por el menú del programa, el usuario puede añadir códigos y distancias que respondan a sus datos. Puesto que estas estructuras se desarrollaron en interacción con usuarios de AQD durante varios cursos de doctorado en España y Alemania, incluyen las más frecuentes hipótesis o problemas de investigación.

El Componente AQD-A (Apéndice), ofrece la posibilidad de:

- introducir, imprimir y editar la información adicional que puede ser necesaria para posteriores análisis. Esta información consiste en inventarios de palabras que se pueden recuperar o contar en el texto original, o la definición de matrices para la representación de pasajes de texto concretos, así como las reglas para construir metacódigos;
- contar palabras o una parte de ellas en el fichero de textos;
- contar sufijos útil al menos en idioma alemán para la identificación del estilo cognitivo del hablante (concreto vs. abstracto);
- recuperar e imprimir palabras concretas en el contexto de una línea del texto;
- recuperar segmentos de texto codificado de acuerdo con la definición de celdillas incluidas en una matriz de códigos;
- crear y recuperar metacódigos.

UNA APLICACION PRACTICA DEL AQUAD

Vamos a describir, en esta parte del artículo, una de las aplicaciones del programa AQUAD para el análisis de datos cualitativos. Los resultados que describimos son una parte de un proyecto de investigación más amplio en el que se combinan metodologías cualitativas y cuantitativas de análisis de datos.

Fundamentación teórica

La investigación a la que nos referimos se centra en un estudio descriptivo del período de iniciación a la enseñanza para los profesores principiantes. Esta investigación se refiere al análisis del primer año de docencia con la intención de describir concepciones, creencias, preocupaciones, problemas, etc. de profesores principiantes de enseñanza elemental y secundaria.

El proceso de iniciación profesional, se ha entendido por algunos autores como un momento clave para la socialización del profesor. Desde esta perspectiva, se han realizado algunas investigaciones en las que se han pretendido identificar las diversas fuentes de influencia tanto estructurales como personales que influyen y determinan la adopción por parte del profesor principiante de la «cultura de la enseñanza».

Jordell (1987) ha elaborado en este sentido un modelo en el que de forma clara se diferencian las influencias que a distintos niveles se producen sobre los profe-

sores principiantes. Un primer nivel de influencia (NIVEL PERSONAL) viene presentado por las experiencias previas (biografía), así como la experiencia en la institución de formación de profesorado. Este nivel influye de forma interiorizada (teorías implícitas) sobre los profesores principiantes, tal como mostrara Lortie en su investigación.

En segundo lugar, destacan las influencias a *NIVEL DE CLASE*. Las investigaciones muestran que los alumnos así como los elementos estructurales que caracterizan a la enseñanza: multidimensionalidad, simultaneidad, inmediatez, impredecibilidad, publicidad, historia, socializan a los profesores en este nivel. De esta forma, los estudiantes, el ambiente de clase, la interacción en el aula, son elementos que influyen en la socialización de los profesores principiantes. Un tercer nivel de influencias es el NIVEL INSTITUCIONAL. En este nivel se han de considerar influencias tales como las de los propios colegas, los directores y padres. También se incluyen aquí las influencias del curriculum y la administración.

Un último nivel de influencia, más lejano y oculto lo representa la estructura económica, social y política en que la escuela se inserte.

Problemas a explorar

1. Describir las preocupaciones y problemas de profesores principiantes en su primer año de docencia.
2. Estudiar los diferentes niveles de influencia socializadora: personal, didáctico (clase) e institucional durante el primer año de docencia.
3. Identificar necesidades formativas de profesores principiantes.

Sujetos y métodos

Los sujetos de esta investigación fueron 105 profesores principiantes de Enseñanza Elemental y Secundaria; 30 de Enseñanza Elemental; 40 de Bachillerato y 32 de Formación Profesional. Del total de sujetos, 46 eran varones y 59 mujeres. Los 105 profesores fueron completando los siguientes instrumentos: «Inventario de Creencias del Profesor», e «Inventario de Problemas Docentes». Los resultados de estos inventarios no serán analizados en este artículo. Además los sujetos fueron *entrevistados* a través de entrevista semiestructurada, grabada en audio y posteriormente transcrita. Las preguntas formuladas a todos los profesores fueron las siguientes:

1. ¿Podrías describir brevemente qué han supuesto para ti estos primeros meses como profesor? Comenta tus impresiones, aprendizajes y cambios durante este período.
2. ¿Podrías decirme cuáles han sido las preocupaciones más importantes que has tenido durante este período de tiempo?
3. ¿Qué impresión tienes del centro en el que enseñas?: Instalaciones, colaboración de los padres.
4. ¿Estás satisfecho de las relaciones con tus compañeros de centro, de departamento?
5. ¿Podrías contar como transcurre una clase normal de un día normal?
6. ¿Qué aspectos de tu enseñanza consideras o encuentras más problemáticos según tu experiencia?
7. ¿Podrías definirte como profesor?
8. ¿Podrías comentar algún aspecto que consideres de interés referido a tus relaciones con los alumnos? ¿Cómo las percibes?
9. ¿Cómo valoras la formación pedagógica y didáctica que has recibido? ¿En qué medida te ha sido útil?

Análisis de datos

Una vez que realizamos las entrevistas, procedimos a su transcripción para su posterior análisis. Siguiendo a Miles y Huberman (1984), procedimos a identificar códigos significativos que nos permitieran analizar la información contenida en las entrevistas. Comenzamos revisando investigaciones y modelos teóricos sobre el período de iniciación a la enseñanza. Encontramos que la propuesta de Jordell (1987), descrita anteriormente en la cual identificaba diferentes niveles de influencia en la socialización de los profesores principiantes: personal, educativo e institucional, era un enfoque adecuado para nuestras intenciones, por lo que redactamos una relación de códigos en función de esas tres dimensiones.

El equipo de investigación analizó esta primera propuesta de códigos, y en función de la experiencia adquirida a través de la realización de las entrevistas con los sujetos investigados, se fueron ampliando o eliminando los códigos incluidos en esta primera aproximación. Como consecuencia de ello, el equipo de investigación identificó los códigos que aparecen en la Tabla 1.

1. DIMENSION PERSONAL
 - EDP. Experiencias Docentes Previas
 - EFP. Experiencias en Formación de Profesorado
 - ECE. Experiencias Como Estudiante
 - SIM. SI-Mismo
 - PRE. PREocupaciones
 - APR. APReendizaje
 - CRE. CREencias
 - NEF. NEcesidades Formativas
 - CDO. Carga DOcente
2. DIMENSION CLASE
 - LA CLASE
 - RPA. Ratio Profesor/Alumno
 - EFC. Espacio Físico de la Clase
 - EQU. EQUipamientos
 - AMB. AMBiente de Clase
 - LOS ALUMNOS
 - CON. CONducta
 - REN. RENdimiento
 - COM. COMprensión
 - CNP. CoNocimientos Previos
 - REL. RELaciones profesor-alumno
 - PAR. PARTicipación
 - EXP. EXPepectativas
 - PRO. PROcedencia
 - PLA. PLANIFICACION
 - CON. CONtenidos
 - ENSEÑANZA INTERACTIVA
 - MET. METodología
 - ACT. ACTividades
 - DIS. DISciplina
 - MOT. MOTivación
 - GES. GEStión de clase
 - EVA. EVAluación
3. DIMENSION INSTITUCIONAL
 - 3.1. EL CENTRO
 - COL. COLEGAS
 - MAT. Dotación de MATerial
 - AMC. AMbiente de Centro
 - CUR. CURiculum
 - IDE. IDEario de centro
 - ORG. ORGanos de gestión

- 3.2. EL ENTORNO
 - PAD. PADRES
 - ENT. Relaciones con el ENTorno
- 3.3. EL SISTEMA
 - ADM. ADMInistración educativa
 - LIM. Normas y LIMItaciones del Sistema

Tabla 1. Inventario de Códigos

Las siguientes descripciones se refieren a una muestra de 10 entrevistas que hemos tomado de dicho estudio. No presentamos en detalle los resultados porque el análisis aún no está concluido, pero queremos demostrar los pasos que hemos seguido, ayudados por el programa AQUAD. En primer lugar procedimos a obtener información sobre los tópicos más y menos destacados en las entrevistas realizadas. A través del programa AQD-C procedimos a realizar el cálculo de la frecuencia con que cada código aparecía en las 10 entrevistas seleccionadas. Como se puede comprobar en la Tabla 2, el código que con mayor frecuencia aparece es MET (Metodología), seguido de SIM (Sí mismo), y en tercer lugar, la categoría PRE (Preocupación).

EDP: 14	EFP: 10	ECE: 12	SIM: 25	PRE: 23
APR: 0	CRE: 4	NEF: 8	CDO: 4	RPA: 5
EFC: 2	EQU: 3	AMB: 2	CND: 2	REN: 4
COM: 2	CNP: 9	REL: 17	PAR: 4	EXP: 3
PRO: 1	PLA: 8	CON: 7	MET: 29	ACT: 7
DIS: 6	MOT: 12	EVA: 9	COL: 18	MAT: 15
AMC: 6	CUR: 2	IDE: 0	ORG: 2	PAD: 14
ENT: 0	ADM: 1	LIM: 4		

Tabla 2. Frecuencias de códigos.

Una vez analizada la frecuencia de los códigos, y utilizando otra de las opciones del programa AQD-C, hemos pretendido identificar qué códigos van asociados en cuanto a cercanía en las entrevistas con respecto a los que han aparecido con mayor frecuencia. De esta forma, podemos conocer algo acerca de patrones potenciales de significado en las producciones verbales de nuestros sujetos. En esta ocasión decidimos utilizar una distancia máxima de cinco líneas como criterio para la búsqueda. En la Tabla 3, aparecen los códigos que están en relación próxima, es decir, dentro de cinco líneas con respecto al más código utilizado con mayor frecuencia. La Tabla contiene los resultados en el formato original de AQD-C de algunas entrevistas seleccionadas.

Con los resultados de este análisis, y guiados por las líneas numeradas en el listado de ordenador, podemos dar el próximo paso que consiste en volver a los textos de las entrevistas. Hemos de asegurarnos de que existe realmente una relación significativa entre códigos críticos como MET o PRE y aquellos códigos que hemos recuperado dentro de una determinada distancia. Si no es así, podríamos alterar la distancia tomada como criterio y/o excluir principalmente los códigos que no tengan relación a partir del siguiente análisis más concreto: vamos a formular y probar hipótesis específicas.

```
***** ini001.cod
001:          98          120          MET / 122          130          REN
001:          132          134          MET / 134          136          CRE
001:          132          134          MET / 137          140          EVA
***** ini002.cod
002:          278          285          MET / 287          291          DIS
```

```

***** ini003.cod
003:      85      94      MET / 98      111      EQU
003:     241     246      MET / 249     259      ACT
003:     454     456      MET / 458     460      MOT
***** ini004.cod
004:     172     181      MET / 182     186      ACT
***** ini022.cod
022:     103     108      MET / 110     113      CON
022:     507     512      MET / 516     528      EDP
***** ini025.cod
025:     191     214      MET / 214     228      LIM
***** ini039.cod
039:     187     200      MET / 201     204      PAR
039:     187     200      MET / 201     204      PRE

```

Códigos que aparecen posteriormente al código MET con una distancia máxima de 5 líneas.

En nuestro caso, una hipótesis podría establecer que algunos profesores principiantes que se refieren a los métodos de enseñanza, lo hacen desde una estrecha perspectiva: las unidades de significado en sus entrevistas se centran alrededor de los métodos de enseñanza relacionando principalmente los métodos con otros aspectos de la enseñanza interactiva o de la enseñanza en general, pero sólo en pocas ocasiones se refieren a la dimensión personal o institucional. Otra hipótesis se podría centrar en la relación entre el código PRE (preocupación) y el código EVA (evaluación), partiendo del caso de que algunos profesores principiantes están preocupados sobre todo con cuestiones referidas a cómo evaluar a sus alumnos. Si por ejemplo, la búsqueda de profesores principiantes con puntos de vista limitados respecto de aspectos metodológicos tiene éxito, entonces podríamos diferenciar entre dos tipos de profesores. Su perspectiva limitada vs. amplia se podría relacionar con los problemas experimentados en clase (a través del instrumento cuantitativo utilizado), y puede que después de reinterpretaciones más específicas de partes relevantes de sus entrevistas, se encuentren relaciones respecto de los alumnos y de los colegas. Si esta hipótesis no pudiera ser confirmada, recibiríamos los resultados negativos en unos pocos minutos, de forma que, podremos ahorrar mucho tiempo y aburrimiento comprobando en 105 entrevistas página por página la co-ocurrencia de dos unidades de significado, en donde la segunda mitad se define por uno de entre 19 códigos, los restantes códigos se refieren a «dimensión enseñanza».

Como ejemplo del contraste de hipótesis, así como de la ayuda que proporciona AQUAD, vamos a demostrar la traducción de esta hipótesis al lenguaje (TurboProlog) de AQUAD. En español podríamos establecer que:

Existen algunos profesores del total de la muestra cuyas reflexiones sobre los métodos de enseñanza se relacionan con consideraciones acerca de otros aspectos de la enseñanza, pero no con aspectos personales o institucionales.

Por supuesto, podría haber sido mejor incluir la declaración alternativa de que existen profesores que relacionan los métodos de enseñanza con aspectos institucionales o personales. Debido al espacio disponible y a la claridad del ejemplo, en este caso demostramos sólo una versión «frugal» de la hipótesis. De cara a reducir diferenciaciones innecesarias entre los códigos y reducir el tiempo de procesamiento, en primer lugar, hacemos uso de la función «meta-códigos» del componente AQD-A y transformamos todos

- los códigos incluidos en «dimensión personal» a DIP,
- los códigos incluidos en «dimensión institucional» a DII,
- los códigos incluidos en «dimensión enseñanza» (excepto MET, por supuesto) a DIT.

De esta forma, creamos unos ficheros adicionales de códigos que en este caso sólo contienen cuatro diferentes códigos. El paso siguiente, consistirá en reformular la hipótesis original de forma que tenga en cuenta las especificidades del proceso deductivo de contraste de hipótesis con PROLOG:

La hipótesis 1 es verdadera si:

existe para un profesor en particular (número de la entrevista) un código en el fichero de códigos referido a MET,
y otro elemento en el banco de datos referidos a DIT,
y no existen elementos referidos a DIP o DII o a ambas dimensiones,
y sólo son relevantes los códigos que aparezcan dentro de cinco líneas a partir del código MET,
y queremos recibir para todos los profesores los resultados o un informe negativo.

Para poder realizar el paso final de la traslación, hemos de tener en cuenta que todas las variables han de escribirse en PROLOG, al final del programa AQD-H, y que hay que saber utilizar cinco elementos de PROLOG:

1. *Estructura de los elementos en el banco de datos*: Normalmente, cada elemento del banco de datos de PROLOG toma la forma de un predicado, que consiste en una relación y un determinado número de objetos. En el caso de AQUAD, se ha definido la estructura para la introducción de códigos mediante la relación *K* (del alemán Kode) y para los objetos *número del fichero de texto*, *primera línea*, *última línea*, *código*. La primera parte de la segunda línea en la tabla 3 por ejemplo, nos proporciona la información contenida en el fichero de códigos para el texto (entrevista) número 001:

K (001, 98, 120 MET)

Esto significa: En la entrevista 001, existe una unidad de significado que comienza en la línea 98, que termina en la línea 120, y que se refiere a MET (métodos de enseñanza).

2. *Predicado para el análisis de las cadenas de códigos*: Deseamos comparar los códigos de nuestro banco de datos con los códigos criterios. Si para esta codificación entendemos que sólo son relevantes partes de códigos (por ejemplo las primeras dos letras), es necesario conocer la función del sistema de predicados. En nuestro caso, dado que estamos buscando MET, DIP, DIT y DII, y no partes de códigos, no necesitamos de tal predicado.

3. *Transformación de «string» en «integers»*: Por diversas razones, todos los componentes de los códigos deben definirse como «string». En este momento tenemos que realizar un simple cálculo aritmético: tenemos que computar la diferencia entre los códigos críticos, debido a que hemos introducido la condición «dentro de cinco líneas a partir de MET». Por lo tanto, hay que transformar los «strings» que contienen la información respecto de los números de líneas (primera y última línea de unidad de significado) en «integers» mediante el sistema de predicados:

str_int (string, integer)

4. *Formato de impresión*: PROLOG imprime todos los resultados de forma secuencial, carácter a carácter. Para que incluya espacios entre unidades de información es preciso utilizar el sistema de predicado «new lines»:

nl

de cara a recibir unos listados presentados con claridad.

5. *Análisis completo de los bancos de datos*: Normalmente, el programa podría detenerse después de descubrir la primera confirmación de nuestra declaración hipotética, aun cuando existieran más elementos en el banco de datos que cumplan las condiciones de la hipótesis. Nos interesa conocer todas las soluciones, y por lo tanto debemos forzar al programa a que pruebe todas las posibles combinaciones de los datos que poseemos. Ello se consigue con el sistema predicado:

fail

como último elemento de la hipótesis que hemos trasladado. Este predicado propicia una interrupción. Después de una confirmación de la hipótesis, la búsqueda se detiene y a continuación, el programa vuelve a correr hasta que no queden datos en la base de datos que no hayan sido comprobados.

En este momento estamos dispuestos para trasladar nuestra hipótesis a PROLOG:

```

hy (1, 1): -
K (N, F1, L1, C1), C1 = «MET» (1)
str_int (F1, IF1), str_int (L1, IL1), (2)
write (N, «:», F1, L1, C1, «→»), nl, (3)
k (N, F2, L2, C2), C2 = «TDI», (4)
str_int (F2, IF2), IF2-IF1 >= 0, IF2-IL1 <= 5, (5)
write (« », F2, L2, C2), nl, (6)
k (N, F3, L3, C3), C3 <> «MET», (7)
k (N, F3, L3, C3), CE <> «TDI», (8)
str_int (F3, IF3), IF3-IF1 = 0, IF3-IL1 <= 5, (9)
write (« », F3, L3, C3, «negativo»), nl, fail (10)

```

En el programa AQD-H cualquier secuencia de comandos se repite para cualquiera de los ficheros de códigos de nuestra base de datos, imprimiendo el número del texto al que estos códigos pertenecen. La línea (1) de la hipótesis formulada hace que el programa busque en el primer fichero de códigos el código «MET». Para un análisis más específico se transforman los números de la primera y última línea de la unidad de significado en la línea (2) a través del cambio de «string» en «integer». La línea (3) hace que la información sobre el éxito de la búsqueda quede impresa, seguido de una línea en blanco. En la línea (4) comienza la búsqueda en el fichero de códigos N de un código que contenga el criterio «DIT». Si tal búsqueda tiene éxito, en la línea (5) la notación de F2 (número de la primera línea de la unidad de significado codificada como «DIT») se transforma en un «integer». En este momento se contrasta la condición en la que estamos interesados, es decir «DIT» muy cerca de «MET», ni antes de «MET», ni más de cinco líneas después de «MET». La línea (6) produce una salida por impresora si en la línea 5 se tiene éxito. De lo contrario, el programa seguiría buscando el siguiente código «MET» partiendo de nuevo de la línea (1). Las líneas (7) a (10) hacen que el programa compruebe si existen algunas referencias a otras dimensiones además de «DIT» («DIP» o «DII») también en estrecha relación con «MET». Sólo bajo la estricta condición de que no exista «DIP» o «DII» atribuiremos a los profesores principiantes «una perspectiva limitada en relación a los métodos de enseñanza». Por lo tanto, si el programa encuentra otras referencias además de «DIT» las imprimirá en la línea (10) junto con el comentario «negativo» (en relación a nuestra hipótesis) y una línea en blanco. El predicado final «fail» hace que el programa busque otras soluciones. Con la muestra de entrevistas y ficheros codificados a partir de la Tabla 3, podríamos obtener un listado con los resultados que se muestran en la Tabla 4.

***** ini001.cod				
001:	98	120	MET	-
	122	130	TDI	
	132	134	MET	-
001:	137	140	TDI	
	134	136	PDI negative	
***** ini002.cod				
002:	278	285	MET	-
	287	291	TDI	
***** ini003.cod				
003:	85	94	MET	-
	98	111	TDI	
003:	241	246	MET	-
	249	259	TDI	
003:	454	456	MET	
	458	460	TDI	
***** ini004.cod				
004:	172	181	MET	-
	182	186	TDI	
***** ini022.cod				
022:	103	108	MET	-
	110	113	TDI	
022:	507	512	MET	-
***** ini025.cod				
025:	191	214	MET	-
***** ini039.cod				
039:	187	200	MET	-
	201	204	TDI	
	201	204	PDI negative	

Tabla 4. Contraste de hipótesis

Como se puede observar, los sujetos 002, 003 y 004 cumplen completamente nuestras expectativas; siempre que hablan acerca de los métodos de enseñanza, relacionan este aspecto con dimensiones docentes pero nunca con dimensiones personales o institucionales. El sujeto 025 no relaciona las reflexiones sobre los métodos de enseñanza con otros tópicos. Por otra parte, los resultados de otros sujetos (por ejemplos el sujeto 001: dos casos positivos y uno negativo; el sujeto 022 con MET en las líneas 507-512 sin relación con DIT) necesitarían posteriores consideraciones. Y por supuesto, podríamos haber hecho que el programa contara e imprimiera los resultados positivos y negativos para cada sujeto/entrevista adicionalmente.

Este ha sido sólo un ejemplo de cómo el programa AQUAD puede resultar de mucha utilidad en la investigación cualitativa, ya que facilita el trabajo de los investigadores. Se podrían llevar a cabo otras aplicaciones especialmente en investigación psicológica y educativa, incluyendo el análisis de respuestas a preguntas abiertas en cuestionarios, el análisis de protocolos de orientación (incluyendo datos cuantitativos sobre fases de silencio debido a intervenciones del orientador o de reflexiones del cliente), la interpretación de entrevistas biográficas y el análisis de diarios.

REFERENCIAS

- DREHER, E., DREHER, M., FINK, B. & HINKELMANN, R. (1985): *Dissens als methodologisches Konstrukt qualitativer Auswertung* (Manuskript), München, Institut für Empirische Pädagogik und Pädagogische Psychologie der Universität München.
- HUBER, G. L. (1989): *AQUAD. Auswertung qualitativer Daten, Version 2.2.*, Bericht Nr. 25 aus dem Arbeitsbereich Pädagogische Psychologie der Universität Tübingen.
- JORDELL, K. (1987): «Structural and Personal Influence in the Socialization of Beginning Teachers», *Teaching and Teacher Education*, 3 (3), 165-177.
- LECOMPTE, M. & GOETZ, J. (1982): «Problems of reliability and validity in ethnographic research», *Review of Educational Research*, 52, 31-60.
- LINCOLN, Y. & GUBA, E. (1985): *Naturalistic inquiry*, Beverly Hills, Sage.
- MILES, M. (1983): «Qualitative data as an attractive nuisance: The problem of analysis», in J. VAN MAANEN (Ed.), *Qualitative methodology* (pp. 117-134), Beverly Hills, Sage.
- MILES, M. B. & HUBERMAN, A. M. (1984): *Qualitative data analysis: A source book*, Beverly Hills, Sage.
- SHELLY, A. (1986): *Life after coding-moving to higher levels of abstraction*, Paper presented at the annual meeting of the A.E.R.A., S. Francisco.
- SHELLY, A. & SIBERT, E. (1985): *The QUALOG user manual*, School of Computer and Information Science, Syracuse University.
- SIBERT, E. & SHELLY, A. (1985): *Logic programming: Computer programming that complements qualitative research*, Paper presented at the annual meeting of the A.E.R.A., Chicago.
- WEBER, R. P. (1985): *Basic content analysis*, Sage University Paper series on Quantitative Applications in the Social Sciences, series n. 07-049, Beverly Hills, CA, Sage.

GÜNTER HUBER
 Profesor
 University of Tübingen
 Munzgasse 22-30
 Tübingen
 West Germany D-7400

CARLOS MARCELO GARCIA
 Departamento Didáctica
 y Organización Escolar
 Facultad Filosofía y Ciencias
 de la Educación
 Avda. S. Francisco Javier s/n
 41005 Sevilla