

# El tratamiento de la información en la investigación educativa (una propuesta informatizada en entorno Pc)

Gregorio Rodríguez Gómez, Ana Corrales Pérez, Javier Gil Flores y Eduardo García Jiménez

Universidad de Sevilla

Durante la década de los ochenta se comienza a utilizar el ordenador en la investigación educativa de una forma generalizada. Muchas y diferentes han sido las herramientas informáticas que los investigadores han venido utilizando en sus tareas: desde los populares procesadores de texto (Word, WordPerfect, etc.) hasta los sofisticados programas para el análisis cuantitativo (BMDP PC-90 o SPSS/PC+), así como hojas de cálculo o sistemas de gestión de bases de datos. En este trabajo proponemos la utilización de unos determinados programas informáticos para el manejo de la información y el análisis de datos cualitativos de naturaleza textual, de cara a su utilización en el entorno PC. Por último realizamos algunas reflexiones sobre las limitaciones de la informática en este campo.

During the eighties, computers have become widely used in educational research. Many different software tools, from popular word processors (Word, WordPerfect, etc.) to sophisticated programmes for quantitative analysis (BMDP PC-90 or SPSS/PC+), as well as spreadsheets and data base management systems, have been used by researchers. In this paper we propose to use specific software tools for information management and qualitative data analysis of a textual nature with PC computers. Finally, we present some reflections about computer limitations in this field.

**DESCRIPTORES:** Software, Ordenador, Análisis de datos cualitativos. Análisis textual. Lexicometría. Bases de datos. NUDIST. AQUAD. Pro-Cite. Biblio-Links. SmarTerm. Procomm.

## 1. Introducción.

Al igual que en otras muchas áreas, la década de los ochenta supuso la generalización del uso del ordenador en todas las fases del proceso que exige cualquier investigación educativa, debido fundamentalmente a la irrupción del ordenador personal y a las prestaciones que los mismos ofrecen. A lo largo de estos años se han venido utilizando una serie de programas informáticos que, con mayor o menor acierto, han facilitado la labor de los investigadores en las distintas fases de su trabajo. Desde el popular procesador de textos (Word, WordPerfect, etc.), hasta los sofisticados programas de análisis de datos cuantitativos (BMDP PC-90 o SPSS/PC+), pasando por las hojas de cálculo y los sistemas de gestión de bases de datos, se han convertido en elementos cada vez más habituales en cualquier proceso de investigación educativa.

Llegados a este momento cabe plantearse cuáles podrían ser las herramientas informáticas que podríamos considerar básicas en cada una de las fases de la investigación, y proponer la utilización de programas concretos que a lo largo de estos años han demostrado eficacia. No obstante, vamos a centrar el trabajo en las fases de documentación (determinar el estado de la cuestión) y análisis de la información, concretamente en la fase de análisis nos ceñiremos a los programas que pueden facilitar las tareas de análisis de la información textual procedente de libros, textos legales, políticos, históricos, artículos de prensa, transcripciones de entrevistas, etc..

Así pues, las páginas que siguen son fruto de nuestra propia experiencia en el manejo y utilización de estos programas, y presentamos una breve propuesta de utilización de aquéllos que consideramos como más útiles en las dos fases señaladas del proceso de investigación educativa, sea cual fuere la naturaleza de ésta.

## **2. El acceso a las fuentes de información.**

Una de las primeras tareas con las que nos enfrentamos al realizar una investigación consiste en determinar "el estado de la cuestión"; es decir, debemos buscar y organizar toda la información posible en torno al problema sobre el que hemos centrado nuestro interés. Hasta hoy lo normal era ir rellenando fichas bibliográficas en distintos tamaños o formatos de papel, a partir de búsquedas por bibliotecas y archivos. Pero esta "cultura del papel" cada vez va quedando más atrás, y viene siendo sustituida por la "cultura magnética" de los discos magnéticos.

El ordenador, al igual que en otras áreas, ha irrumpido en este campo con un éxito enorme, ya que las funciones a desarrollar son plenamente asumibles por estas "nuevas máquinas". Sirviéndonos de las herramientas informáticas actuales, el acceso a las fuentes de información podemos realizarlo a través de medios como las bases de datos y, gracias a la red Internet, el correo electrónico y los servidores Gopher, entre otras posibilidades.

### *a) Las bases de datos.*

Con gran frecuencia se llega a confundir una "base de datos" con los programas de carácter informático que las manejan, esto es, con los "sistemas de gestión de bases de datos", pero debemos tener en cuenta que son dos realidades bastante diferentes, aunque estrechamente relacionadas. Podemos definir una base de datos como "...un conjunto de informaciones, organizada de tal manera que con el mínimo esfuerzo se puedan extraer los datos que interesan para poder realizar con ellos las operaciones oportunas" (De Lara, 1987), o bien como "...cualquier conjunto de información almacenada en cualquier tipo de soporte de forma que sea "legible" (léase accesible) y manipulable por un sistema informático" (Lizasoain, 1992). A pesar de las diferencias de matices, ambas definiciones se asemejan y nos ofrecen una idea lo suficientemente general como para no tener que incidir más en ellas, pero sí en los elementos que componen una base de datos: registros, campos y ficheros.

Una base de datos se compone de *registros*. Estos no son más que un conjunto de datos organizados en campos. Dicho de otra forma, son los elementos individuales de que

consta una base de datos: un libro, un artículo, un manuscrito, un texto legal, una cinta de vídeo, una película, o cualquier otra fuente de información. Los registros de una base de datos, a su vez, se componen de *campos*, cada uno de los cuales no son más que un ítem de información dentro de un registro. Por ejemplo, si hablamos de un libro (registro), un campo será el autor, otro el título, otro el año...etc. En fin, toda la información recogida en los distintos registros, es almacenada en memoria en forma de *ficheros* que son grabados en los dispositivos de almacenamiento de información (discos magnéticos).

Los "*Sistemas de Gestión de Bases de Datos*", por su parte, son un conjunto de programas informáticos que nos permiten crear, guardar, manipular y gestionar los registros de una base de datos; son los que nos van a permitir el acceso a la información que se encuentra archivada en las bases de datos. Si nos fijamos en la figura 1 podemos observar un resumen gráfico de lo expresado hasta el momento. Como podemos observar en el mismo, esta base de datos, a la que hemos denominado BIBLOS, la constituyen una serie de registros (libros y artículos), los cuales a su vez se componen de campos (autor, fecha, título...), y esta base de datos queda almacenada en el soporte magnético como un conjunto de ficheros, que podremos gestionar a través de un sistema de gestión de bases de datos.



**Figura 1. Estructura de una base de datos.**

Existen múltiples y diferentes tipos de bases de datos y llegar a una clasificación de las mismas va a depender de los criterios que utilicemos. El más usual suele ser establecer las distinciones sobre la base del contenido y así podemos hablar en un primer momento de bases de datos *referenciales* y bases de datos *fuentes* (Lizasoain, 1992). Las primeras son aquéllas que contienen tan sólo una referencia de la fuente original, mientras que las últimas se refieren directamente a la fuente original. En la figura 2 podemos contemplar de forma gráfica la clasificación de las bases de datos tomando como criterio el contenido de las mismas.



**Figura 2. Clasificación de las bases de datos.**

Las primeras son aquéllas que contienen tan sólo una referencia de la fuente original, mientras que las últimas se refieren directamente a la fuente original. En la figura 2 podemos contemplar de forma gráfica la clasificación de las bases de datos tomando como criterio el contenido de las mismas.

Una *base de datos bibliográfica* suele ser el tipo más frecuente de base de datos referencial. En este tipo se almacena tan sólo información de referencia de los documentos originales, los cuales pueden ser documentos impresos o en cualquier otro formato (audio, vídeo...etc.). La información que almacenan los *directorios*, por su parte, se refiere a personas o instituciones que se engloban por áreas de actividad. Los *bancos de datos factuales o numéricos* son bases de datos fuentes que contienen información de carácter numérico, organizada según diferentes criterios. Del mismo modo, las *bases de datostextuales*, también conocidas como de texto completo o integral, hacen referencia al conjunto de bases de datos en las que cada registro está formado al menos por la referencia bibliográfica del documento original. Por último, las *bases de datos mixtas* son aquéllas que incorporan información numérica y textual a la vez.

A pesar de lo expresado hasta el momento, podríamos haber tomado otros criterios de clasificación, y así, según los *temas* podríamos hablar de bases de datos científicas, tecnológicas, sociales, económicas...; según el *ámbito territorial* nos encontraríamos con bases de datos internacionales, nacionales, regionales o locales; de acuerdo a su *difusión* podrían ser restringidas o públicas; y, por último, según su *origen institucional* podrían ser públicas o privadas.

Hasta no hace mucho la forma de acceder a las grandes bases de datos era a través de los servicios de documentación especializados, que conectaban directamente a través de terminales con un ordenador central. En el caso concreto de nuestro país, para aquellos que se encuentran alejados de estos centros, el proceso se complicaba, pues se veían obligados a enviar por correo sus necesidades al servicio de documentación y éste remitía el resultado de las búsquedas realizadas, consistente en una serie de fotocopias de los registros encontrados. Estos servicios de documentación conectan con las bases de datos a través de lo que se conoce como conexión "en línea" (on-line), por medio de terminales, a través de redes de transporte como la red telefónica conmutada.

No obstante, en primer lugar, gracias a las mejoras habidas en los sistemas de comunicaciones, en la actualidad muchas de estas consultas pueden ser realizadas de forma directa a través del ordenador personal, siempre y cuando se disponga de acceso a las redes Internet o DECnet, una línea de conexión a la red de conmutación de paquetes (X.25) o, simplemente, una línea telefónica que nos permita la colocación de un modem. En el caso de la Comunidad Autónoma Andaluza, por ejemplo, gracias a los servicios prestados por parte del Centro de Informática Científica de Andalucía (C.I.C.A.), podemos acceder directamente a las siguientes bases de datos del Consejo Superior de Investigaciones científicas (CSIC): IME (Biomedicina), YCIT (Ciencia y Tecnología), ISOC (Ciencias Sociales y Humanidades), AMERICA LATINA (Ciencias Sociales y Humanas relativas a América Latina) y CIRBIC (Catálogo colectivo de libros y revistas de las bibliotecas del CSIC).

De esta forma podemos consultar fácilmente la base de datos PSEDISOC, en la que el CSIC recoge la producción científica de las revistas españolas en las

áreas de Psicología y Ciencias de la Educación, pudiendo consultarse la misma por los siguientes campos: autor, institución de trabajo, clasificación temática, año de publicación, materia (descriptores y descriptores secundarios) o identificadores. Desde nuestro ordenador personal, y una vez conectado con los ordenadores del C.I.C.A., podríamos localizar, por ejemplo, todos los artículos y trabajos que se hayan publicado durante 1992 sobre "informática" en las revistas españolas.

En segundo lugar, el incremento de la capacidad y velocidad de los ordenadores personales han posibilitado el acceso directo a las bases de datos, gracias a la incorporación de los discos compactos. Estos permiten tener las actualizaciones de las bases de datos más usuales, pudiendo consultarse las mismas con tan sólo disponer de un ordenador de tipo medio y un lector de discos compactos (CD-ROM). Esta fórmula es la que ha venido estableciéndose en las bibliotecas, de tal forma que suele ser generalizada la presencia de lectores CD-ROM para consultar las bases de datos especializadas. Sobre las posibilidades y uso de las mismas puede consultarse el texto de Lizasoain (1992) , y como un primera introducción el trabajo de Cabero y Duarte (1994).

Uno de los inconvenientes de este tipo de bases de datos reside en el acceso a las mismas, ya que puede que no encontremos en un entorno próximo aquéllas sobre las que nos interesa realizar una consulta. Este hecho se viene paliando desde hace algún tiempo gracias a la concentración de las bases de datos en los centros informáticos y bibliotecas, y la posibilidad de consultarlas a través de terminal. Esto nos permitirá el acceso a una gran cantidad de bases de datos, de una forma directa y desde el propio ordenador personal.

#### b) *El correo electrónico.*

La instalación de redes de comunicación y la propagación del correo electrónico está posibilitando la implantación de una nueva forma de acceder a fuentes de información de una forma más cercana y actualizada (Sanz, 1994). Gracias a las posibilidades que nos ofrece el correo electrónico empiezan a proliferar "grupos de discusión" sobre distintas áreas de conocimiento. Un grupo de discusión consiste en una serie de personas, inscritas en una lista, interesadas en un tema, que utilizan el correo electrónico como medio de discusión y diálogo. En este sentido, a partir de octubre de 1993 la *American Educational Research Association* (AERA and ASU Create, 1993; Pierce y otros, 1994), en colaboración con la Universidad de Arizona, ha facilitado la creación de once grupos de discusión, que recogen las distintas especialidades en las que se divide la asociación: administración, curriculum, aprendizaje e instrucción, medida y metodología de la investigación, orientación, historia, contexto social de la educación, evaluación y desarrollo de programas, educación profesional, educación postsecundaria y formación de profesores.

El inscribirse en una de estas listas permite al usuario formular preguntas a los compañeros de la lista, sugerir temas de discusión o solicitar informaciones

sobre determinados tópicos o temas, sabiendo que llegará a todos y cada uno de los miembros de esa lista. Además se utiliza como medio para anunciar distintos acontecimientos, tales como jornadas o congresos de interés, e incluso se llegan a presentar ofertas de trabajo dentro del campo de especialización de la lista concreta. En cualquier momento una persona puede darse de alta o de baja en una lista, con tan sólo enviar un mensaje a través del correo electrónico solicitándolo.

#### c) *Servidores Gopher.*

Gopher (McCahill, 1992) se desarrolla por primera vez en la Universidad de Minnesota, y permite a los usuarios buscar y navegar a través de la información distribuida a lo largo de la red Internet. Gopher organiza la información de forma jerárquica arbórea, de tal forma que los nudos intermedios son directorios, o índices, y las hojas son los documentos. Para navegar a través de Gopher los usuarios lo hacen a través de los menús del sistema.

Gracias a estos servidores, desde nuestro ordenador personal, previa conexión a red, podemos buscar información sobre determinados tópicos de nuestro interés. Entre algunas de las posibilidades que nos ofrecen están las de consultar los catálogos de bibliotecas, acceder a artículos de revistas electrónicas, actas de jornadas y congresos o las direcciones de los profesores de un determinado centro, por ejemplo.

Todas las tareas que suponen la conexión con los grandes centros informáticos exigen la utilización de programas de comunicaciones, que nos permitan hacer que nuestros ordenadores funcionen como si fueran terminales de un gran ordenador. Actualmente los que se vienen utilizando, entre otros, son *Procomm* o *SmarTerm* en el entorno PC, y cada vez se vienen utilizando con mayor profusión los programas emuladores de X-Windows, aunque por las necesidades de hardware de los mismos son todavía poco utilizados.

#### d) *Los sistemas de gestión de bases de datos.*

Los Sistemas de Gestión de Bases de Datos se definían en un principio como un método para la organización y gestión de una gran cantidad de información en un gran sistema informatizado; pero con la llegada de los microordenadores esta concepción ha dejado de tener sentido. Así, podemos definirlos actualmente como programas de ordenador que pueden gestionar eficazmente un enorme número de elementos dados. Es decir, son un conjunto de herramientas con las que se puede organizar y manipular datos de una manera simple y efectiva. Un Sistema de Gestión de Bases de Datos nos sirve para crear, organizar y/o manipular los registros de una base de datos.

En el campo de la microinformática, el primer sistema de este tipo fue el dBase II (Tsu-Der-Chu, 1987), que ha adquirido una gran divulgación, pero que

adolecía de una serie de defectos que le hacían poco útil de cara a su utilización como herramienta para gestionar bases de datos bibliográficas. Como principal limitación del dBase podemos señalar el carácter restringido de sus campos. Al crear una base de datos a partir de dBase debemos establecer campos limitados en el número de caracteres que pueden introducirse, lo cual supone un serio inconveniente, pues al trabajar con bibliografía venimos trabajando con grandes cantidades de información textual que no tienen cabida en los límites que el programa nos pone a cada campo.

Esta dificultad parecía haberse solucionado al aparecer otros programas tales como el KNOSYS, cuya principal virtud reside, precisamente, en el manejo de grandes cantidades de texto, disponiendo de campos casi ilimitados; pero por este carácter flexible, que le permite ser utilizado para múltiples propósitos, el programa a veces no se adapta a las necesidades planteadas por la gestión de bibliografía. No obstante debemos señalar que este es el programa que se viene utilizando en nuestro país para gestionar bases de datos como las desarrolladas por el Consejo Superior de Investigaciones Científicas.

En suma, si bien los Sistemas de Gestión de Bases de Datos han sufrido en los últimos años una gran mejora, a pesar de ello, el hecho de estar pensados como programas generales hace que a veces no respondan a las necesidades específicas de los investigadores que manejan una gran cantidad de información textual, sobre la cual han de realizar numerosas operaciones de manipulación y/o cambio.

Entre los sistemas de gestión de bases de datos, merece una atención especial el sistema SPIRS (Silver Platter Information Retrieval System) (Lizasoain, 1992), enfocado de forma exclusiva a la recuperación de la información y cuyo funcionamiento está basado en comandos o instrucciones. En el entorno de la investigación educativa este sistema se utiliza con una gran frecuencia, dado que el acceso a las grandes bases de datos de carácter educativo se realiza a través del mismo. Este sistema nos permitirá acceder a bases de datos como la ERIC (Educational Resources Information Center), que es la base de datos más consultada y más completa del campo de la educación, producida por el departamento de educación de los Estados Unidos; o la PSYCLIT que se constituye en la principal fuente de información en el campo de la psicología, y que es producida por la APA (American Psychological Association).

### **3. La gestión de la información.**

Una vez consultadas las bases de datos lo normal venía siendo llevarse esta información en disco y posteriormente visualizarla, o imprimirla, a fin de examinar su contenido de forma más detenida. Para aquellos registros que resultaban de interés podrían abrirse fichas bibliográficas en papel, o en bases de datos "caseras", teniendo que escribir cada uno de los datos. Pero esta desagradable tarea ya no es preciso realizarla, gracias a los sistemas de gestión que en la actualidad nos permiten recuperar la información de las grandes bases de datos e incorporarla a la base de datos personal.

La información que sacamos de estas bases de datos, unida a la que por otros medios alcanzamos, van a constituir nuestro fondo documental para iniciar el proceso de investigación. Mas todo ese cúmulo de información necesita ser ordenado, organizado, estructurado. Para ello nos serán de gran utilidad los sistemas de gestión de bases de datos, pero no cualquier sistema, sino aquéllos que hayan sido diseñados de forma específica para este fin y, a la vez, puedan resolvernos las dificultades con las que a lo largo del proceso de investigación nos encontramos al tener que manejar toda la información de la que disponemos.

En este sentido, en entorno PC contamos con unas excelentes herramientas como son los programas *Pro-Cite* y *Biblio-Links* (Personal Bibliographic Software, 1992). Pro-Cite es un sistema de gestión de bases de datos diseñado específicamente para manejar información bibliográfica, caracterizado por su potencia y flexibilidad que nos permite, entre otras posibilidades, construir bases de datos, manejarlas, crear bibliografías, personalizar normas de referencia y puntuación e importar y/o exportar registros. Este programa contempla un total de 20 tipos diferentes de registros predeterminados (libros, capítulos de libros, artículos de revistas, tesis, actas de congresos, programas de ordenador, videos...etc), con un total de 45 campos diferentes (autor, año, título, ISBN, descriptores, resumen...etc.), utilizando unos campos u otros en función del tipo de registro de que se trate. Además el usuario puede definir sus propios tipos de registros, lo cual dota al programa de una gran flexibilidad. Pro-Cite nos permite la ordenación de los registros de cuatro formas diferentes: alfabéticamente (por autor o por título), cronológicamente por el año de la publicación, y por el tipo de registro.

La introducción de la información se puede realizar de varias formas, y prácticamente sin limitación alguna de cantidad de texto en cada uno de los campos. La primera y más simple es la manual, tecleando cada uno de los registros; una segunda forma consiste en importar los registros desde otras bases de datos; y una tercera forma es a partir de la incorporación de los registros obtenidos en una consulta realizada en las bases de datos (a través del sistema SPIRS, por ejemplo), convirtiendo estos registros en una base de datos Pro-Cite. Esta última posibilidad es la que nos permite el programa Biblio-Links. Otras de la utilidades consiste en la posibilidad de exportar registros y la de copiar una bases de datos, completa o parcialmente, en otras ya existentes.

El manejo de las bases de datos creadas con Pro-Cite nos va a permitir encontrar la información a partir de la realización de búsquedas de texto completo, búsquedas truncadas o con operadores booleanos. Las búsquedas de texto completo nos permite acceder a los registros en los que se encuentre una determinada palabra o frase sobre la que estemos interesados. Por ejemplo, podríamos encontrar todos los registros en los que se encuentre la palabra "informática". Las búsqueda truncada nos permite acceder a todos los registros en los que se da alguna variación de una palabra. Por ejemplo, podríamos hacer una búsqueda truncada, a partir de "informatic\*" y nos aparecerían los registros que contienen palabras como informática, informático, informáticas o informáticos. Por último, las búsquedas con operadores booleanos nos permite combinar distintas formas de búsquedas con los



términos "y", "o", etc., así podríamos realizar una consulta en la que tan sólo aparecieran los registros en los que aparece el término informática y que se han publicado entre 1985 y 1990.

De poco nos serviría un sistema de gestión de bases de datos, si no nos permitiera disponer de toda su información. A tal fin, Pro-Cite nos permite realizar informes a partir de los registros de los que disponemos, de tal forma que podemos exportar a otros tipos de ficheros (ASCII, WordPerfect, Word...etc.) los resultados de las consultas realizadas, o bien directamente por la pantalla o la impresora.

Este programa nos permite realizar bibliografías, pero no como un listado de todos los registros, sino seleccionando los registros que se van a listar y el estilo de referencia con el que ha de presentarse la bibliografía -considérese la utilidad de esta posibilidad cuando tenemos que enviar trabajos a distintas revistas que exigen normas de puntuación diferentes-, contando para ello con los sistemas de puntuación más usuales (APA, CHICAGO, ANSI...etc.); permitiéndonos también la realización de nuestros propios ficheros de estilos de referencia. En esta misma línea, el programa cuenta con la posibilidad de realizar la bibliografía a partir del examen del texto manuscrito donde se hayan incorporado las referencias bibliográficas, seleccionando los registros citados y realizando un listado de todas las referencias utilizadas en el trabajo. Otras de las utilidades consiste en la posibilidad de exportar registros y copiar una base de datos, completa o parcialmente, en otra ya existente.

El programa Biblio-Links, como ya hemos adelantado, nos permite convertir las consultas realizadas "on-line" o en CD-ROM, en bases de datos Pro-Cite, así si hemos realizado una búsqueda en la base de datos ERIC, utilizando el programa SPIRS, a través de Biblio-Links podremos convertir el resultado de la consulta en una base de datos sobre la que podremos buscar, organizar, eliminar, añadir y todas las posibilidades que el programa Pro-Cite nos ofrece. Actualmente el programa Biblio-Links permite la transferencia de ficheros desde los sistemas de gestión más importantes, como son BRS, DIALOG, MEDLARS, SILVERPLATTER y STN.

#### **4. El análisis de la información textual.**

En un trabajo anterior (García, Gil y Rodríguez, 1993) presentamos una propuesta de proceso para el análisis de datos cualitativos, procedentes de distintas fuentes de información, pero que básicamente comparten el estar constituidos por textos, que se concretaba básicamente en la realización de tres tareas: reducción de datos, disposición y transformación de datos 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.

Las operaciones consideradas en este proceso de análisis se han venido realizando a través de procedimientos manuales, en los que intervenía básicamente la capacidad de cada investigador para aportar orden a la ingente cantidad de información de la que disponía. Uno de los procedimientos consistía en la utilización de *ficheros o carpetas*, donde se iban almacenando los textos según diferentes criterios como la fuente de los datos, el tema que se aborda, las categorías que contempla, etc. Este procedimiento exigía tener que cortar y pegar trozos de diarios o transcripciones, viéndose enormemente facilitado con la llegada de las fotocopiadoras, pues éstas simplificaban la labor de tener que realizar múltiples copias de un mismo material, pudiendo encontrarlas en varias carpetas o archivos a la vez.

"El investigador reúne los datos codificados pertenecientes a cada categoría. Hacemos esto manualmente: se recortan las notas de campo, las transcripciones y otros materiales y se colocan los datos de cada categoría en carpetas de archivo o en sobres de papel manila". (Taylor y Bogdan, 1986).

Otro procedimiento común ha consistido en utilizar *lápices de colores* con el fin de subrayar o marcar en los textos las distintas unidades y los tópicos a los que hacen referencia. Cualquiera de estos procedimientos se enfrentaba a la dificultad de tener que revisar centenares de páginas tratando de encontrar posibles relaciones entre los temas, y llegar así a descubrir conexiones entre los mismos.

La llegada del ordenador y, sobre todo, su popularización ha supuesto una auténtica revolución en múltiples campos de la actividad humana. Las posibilidades que el ordenador ofrece para manejar grandes cantidades de datos numéricos, verbales o visuales, y su rapidez en la realización de cálculos o tareas de búsqueda y organización le han situado en un lugar de privilegio en el ámbito de la investigación educativa. De hecho, con el acceso a las primeras bases de datos textuales y procesadores de texto son muchos los investigadores que comienzan a intentar aplicar estas nuevas tecnologías al terreno de la investigación cualitativa. En este sentido se van a utilizar estas herramientas a pesar de que no estuvieran diseñadas de forma específica para el análisis.

A través de los procesadores de texto se pueden introducir y almacenar grandes cantidades de texto. Las bases de datos permiten ordenar y buscar información específica. Los recuperadores confeccionan listas de ocurrencia de todas las palabras de un texto, ofrecen frecuencias de palabras o categorías (grupos de sinónimos), localizan palabras en varios documentos, crean índices para la localización y frecuentemente proveen todos los contextos en que aparece una palabra, lo cual es útil para inferir los distintos significados con que es empleada.

Actualmente, y dado el alto nivel de prestaciones que ofrecen los programas informáticos de uso general que existen en el mercado, aún se siguen utilizando éstos en las investigaciones cualitativas (Carro, 1993), con la consiguiente pérdida de tiempo a la hora de adaptarlos a las necesidades específicas del proceso de análisis de información textual. No obstante, sigue estando vigente la utilización de los procesadores de texto (Word, WordPerfect) a lo largo de todo el proceso de la investigación. Desde el inicio del planteamiento del problema, hasta la redacción final del informe de investigación, pasando por la preparación de los textos para su análisis a través de los programas específicos son tareas en las que el procesador de texto se convierte en una herramienta prácticamente imprescindible.

## **5. Programas para el análisis de la información textual.**

Debido fundamentalmente a las limitaciones que los programas comerciales (procesadores de texto, bases de datos...) ofrecían de cara a las tareas necesarias para realizar el análisis de datos cualitativos, poco a poco los distintos investigadores cualitativos comienzan a desarrollar, en colaboración con especialistas informáticos, una serie de programas que facilitarían este tipo de análisis cuya principal característica radicaba en enfrentarse a textos.

Durante la década de los ochenta, sobre todo por la popularización del ordenador personal, la utilización de éste para ejecutar las tareas de investigación alcanza tal auge que, en el contexto del análisis de datos, comienza a hablarse de «análisis asistido por ordenador». Tesch (1988, 1990, 1991), a través de sus distintos trabajos, da una gran popularidad a la utilización del ordenador en las tareas de análisis cualitativo y se convierte en impulsora de su utilización. En 1990, Tesch publica una obra en la que compendia sus aportaciones anteriores y nos presenta una excelente síntesis del estado de la cuestión, examinando cómo se desarrolla la investigación cualitativa en diversos campos (educación, psicología, sociología), los distintos tipos de investigación y el análisis cualitativo, y una pormenorizada relación de programas informáticos específicamente diseñados para el análisis de datos cualitativos.

A pesar de la proximidad temporal del texto, la rapidez con que el mundo informático se renueva hace que en ciertos aspectos el texto necesite ya algunas revisiones. Algunos de los programas que presenta han modificado su estructura y funciones, y lo que en 1990 se anunciaba como posibilidad en la actualidad se encuentra superado con creces. No obstante, el texto de Tesch sigue siendo básico para introducirnos en el terreno del análisis de datos cualitativos asistido por ordenador. Pero esta obra tan sólo se centra en un conjunto de programas que podríamos considerar "cualitativos", obviando otra serie de programas que enlazan con la corriente francesa de la lexicometría o estadística textual (Lebart y Salem, 1988). Desde esta corriente de análisis de los datos textuales se recurre a procedimientos que implican contar las ocurrencias de las unidades verbales básicas (generalmente palabras) y operar algún tipo de análisis estadísticos a partir de los resultados de tales recuentos.

Los programas que se han venido desarrollando en los últimos años intentan responder a las necesidades que los investigadores precisan cubrir cuando se enfrentan al análisis de la información. En el proceso general de análisis al que aludíamos con anterioridad, se contemplan una serie de operaciones que podemos denominar «mecánicas», como son separar unidades de texto, codificar, agrupar, disponer, transformar....etc. que pueden ser ejecutadas con la ayuda del ordenador. No obstante todas estas operaciones se realizan sobre la base de un «marco conceptual» que guía y orienta el proceso general. Este marco conceptual exige pensar, decidir, interpretar, etc., habilidades que todavía son realizadas exclusivamente por parte del investigador, pues el ordenador no es capaz de tomar decisiones conceptuales.

En definitiva, el ordenador y los programas de análisis tan sólo sirven como herramientas, de ahí que todavía estemos bastante lejos de poder decir "hemos realizado el análisis de datos con el programa X", y sería más acertado señalar que "en el proceso de análisis nos hemos auxiliado o servido del programa X". Es importante insistir en este hecho, pues a veces da la sensación de que los programas son los que realizan los análisis por sí solos, como si el analista no hubiera intervenido, presentando un cierto "aire" de independencia y objetividad, cuando en realidad sucede todo lo contrario.

Teniendo presente las consideraciones realizadas, a pesar de la multiplicidad de corrientes existentes hoy día en el terreno de la investigación cualitativa, en el proceso de análisis se dan una serie de necesidades que son comunes a todas ellas (Tesch, 1991). Así nos encontramos con que los investigadores tendrán que

- \*localizar palabras y frases en los textos objeto de análisis;
- \*crear listados alfabéticos de palabras, contabilizando la frecuencia de ocurrencia de cada palabra;
- \*crear índices (asignando fuentes de información a cada ocurrencia) y concordancias de palabras claves contextualizadas;
- \*asignar palabras claves a segmentos de textos;
- \*asignar códigos a segmentos de texto;
- \*conectar unos códigos con otros.

## **6. Tipos de programas.**

Richards y Richards (1991) consideran cuatro tipos diferentes de programas: los que investigan los textos, los que posibilitan almacenar datos, los que permiten la codificación y recuperación, y aquéllos que posibilitan el análisis de los conceptos. Por su parte, Tesch (1990), tomando como base las corrientes de investigación, divide a los programas existentes en tres grandes grupos: a)

los diseñados para la investigación orientada al lenguaje, b) los programas para el análisis cualitativo de carácter descriptivo/interpretativo, y c) los diseñados para la construcción de teoría. No obstante esta diferenciación es más bien de grado y no absoluta, pues las operaciones que realizan los distintos programas se solapan, y en las últimas versiones que están saliendo al mercado cada vez se dan menos diferencias entre unos y otros. Así, no resulta extraño que Richards y Richards (1994) lleguen a plantear una nueva clasificación, considerando dos grupos de programas:

- a. los programas de propósito general (procesadores de texto, sistemas de gestión de bases de datos y recuperadores de texto); y
- b. los programas específicos para el análisis de datos cualitativos, englobando en éstos a cinco tipos de programas: los que siguen el método de codificación-recuperación, los sistemas de construcción de teoría centrados en las normas, los sistemas basados en la lógica, los sistemas de redes conceptuales y los centrados en la indización.

Desde nuestra perspectiva vamos a considerar dos tipos de programas: los que se centran en los aspectos "cualitativos", y los que consideran básico el análisis "cuantitativo". Entre los primeros destacamos en el entorno Windows el programa NUDIS (QSR, 1994), y en el entorno MS-DOS el programa AQUAD (Huber, 1991). Y entre los cuantitativos caben destacar el programa SPAD-T (Lebart y otros, 1994a, 1994b).

#### *a) Los programas cualitativos.*

Este tipo de programas se insertan en los que Tesch (1990) denomina de carácter descriptivo/interpretativo o diseñados para la construcción de teoría. Tanto NUDIST como AQUAD permiten realizar las tareas básicas del proceso de análisis: reducción de datos, disposición y transformación de datos y obtención y verificación de conclusiones. Así, gracias a estos programas podremos:

- . dividir los textos en unidades de significado,
- . asignar códigos y metacódigos,
- . contar códigos (AQUAD) o unidades de texto codificadas (NUDIST),
- . establecer relaciones jerárquicas,
- . modificar los sistemas de categorías,
- . examinar e investigar los documentos,
- . recuperar textos a partir de la codificación realizada, o de palabras de interés,

- . construir matrices textuales,
- . recuperar los contextos de las palabras claves,
- . realizar anotaciones o memorándums,
- . conectar unos códigos con otros,
- . establecer relaciones entre códigos,
- . comprobar hipótesis de carácter cualitati- vo..., etc.

El programa AQUAD, en su versión 3.0, se ejecuta bajo MS-DOS y necesita un total de 1 Mb libre en disco duro. El proceso de codificación puede realizarse directamente sobre la pantalla, o bien en un proceso de dos pasos, imprimiendo el texto para realizar posteriormente la codificación. Además de los códigos, se pueden añadir variables cuantitativas o cualitativas (sexo, edad, profesión...) que posteriormente se pueden utilizar en combinación con los códigos normales para búsquedas selectivas.

El programa permite la formulación y análisis de hipótesis cualitativas. Para ello ofrece un total de 12 estructuras de hipótesis ya prescritas, que permiten definir diferentes códigos y distancias máximas entre ellos. Además permite formular hipótesis propias. La causalidad puede explorarse con el algoritmo de Quine-McClusky a través del componente «minimalización». A través de este componente se pueden llevar a cabo minimalizaciones de las configuraciones condicionales partiendo directamente de datos. El programa facilita la búsqueda de palabras y frases en los datos del texto (o partes de palabras como, por ejemplo, sufijos); asimismo permite contar frecuencias de aparición de las palabras y extraer palabras junto con su contexto. También posibilita la asignación de comentarios a los códigos y «memos» a los segmentos de texto.

En cuanto al programa NUDIST, actualmente se dispone de la versión 3.0 para su ejecución en entorno Windows, y existe una versión para grandes ordenadores. En la versión para PC, la codificación ha de realizarse en un procedimiento de dos pasos, no limitándose tan sólo a las líneas de texto, pues cualquier conjunto de datos (libros, cassettes, gráficos, fotografías...) puede ser dividido en unidades que son numeradas y posteriormente codificadas. En el programa se diferencia una base de datos originales y una base de datos indizada (codificación). En el sistema de indización cada categoría es un "nudo" que tiene la estructura jerárquica de árbol, pudiendo modificarse fácilmente.

Las búsquedas se pueden realizar a través de un conjunto de operadores que comprenden operadores booleanos, operadores no booleanos, operadores relacionales, operadores que excluyen o restringen una búsqueda a una serie de documentos indizados por un nudo particular y operadores que explotan la estructura jerarquizada del sistema de indización. Todas las búsquedas que se realizan se archivan en un nuevo nodo que el investigador puede visualizar en pantalla o imprimir tras su archivo en un fichero.

Por último reseñar la posibilidad que nos ofrece este programa en su versión para Macintosh de conexión con el programa C-Video (Envisionology, 1992), que nos permite la catalogación, anotación y localización de escenas grabadas en vídeo, posibilitando así el análisis de las grabaciones en vídeo, integrando imagen y texto en un solo entorno.

*b) Los programas cuantitativos.*

Dentro de la corriente de la estadística textual, actualmente contamos en entorno PC con el programa SPAD-T (Lebart y otros, 1994a, 1994b) que nos permite realizar las siguientes funciones:

- . descripción del vocabulario, a través de búsquedas de formas lexicales o de segmentos repetidos,
- . la selección de elementos característicos (palabras, segmentos o frases),
- . la construcción de tablas léxicas,
- . la realización de análisis factorial de correspondencias, clasificaciones jerárquicas ascendentes,
- . búsqueda de los contextos de palabras clave,
- . descripción de los contextos,
- . lectura y corrección de los textos.

**7. Propuesta de utilización de programas informáticos en la investigación educativa.**

En estos momentos consideramos que estamos en situación de poder ofrecer una propuesta de un conjunto de herramientas informáticas que serán de gran utilidad al investigador que se enfrente, sobre todo, con datos de naturaleza textual.

En el cuadro 1 podemos observar las dos fases del proceso de investigación en las que nos hemos centrado (estado de la cuestión y análisis de datos), así como los diferentes tipos de programas.

FASES DE LA INVESTIGACIÓN	TIPOS DE PROGRAMAS	Programas entorno PC
ESTADO DE LA CUESTIÓN	Comunicaciones	SmarTerm Procomm

	Sistemas de gestión de bases de datos	Pro-Cite Biblio-Links
ANÁLISIS DE LA INFORMACIÓN TEXTUAL	Análisis cualitativo	NUDIST (Windows) AQUAD (Ms-Dos)
	Análisis cuantitativo	SPAD-T

**Cuadro 1. Fases del proceso de investigación y tipos de programas.**

Lógicamente existen otras posibles opciones en el mercado informático, pero planteamos esta elección concreta considerando el entorno de trabajo PC, las prestaciones que ofrecen cada uno de los programas, las necesidades que se le plantea en su trabajo al investigador educativo y nuestra propia experiencia como investigadores en el manejo y uso de estos y otros programas en distintas investigaciones.

## **8. Limitaciones de la informática.**

A pesar de las enormes posibilidades que, en general, nos ofrece hoy en día la informática, como hemos podido comprobar a través de las páginas precedentes, no obstante debemos considerar que los programas nos facilitan las tareas mecánicas del proceso de investigación, y es por ello que se constituyen en «facilitadores del análisis», pero nunca en ejecutores del mismo.

Toda investigación educativa se realiza dentro de un marco teórico, y con un objetivo determinado, siendo éstos los determinantes fundamentales del proceso de búsqueda y análisis de la información. Los criterios para la selección, la supremacía de unos tópicos sobre otros, la elaboración de unas categorías o la asignación de unos códigos determinados, son operaciones que vienen mediatizadas por los determinantes señalados.

El ordenador tan sólo nos facilitará las tareas más mecánicas, pero incluso así está llegando a ejercer una gran influencia en los procesos de investigación (Richards y Richards, 1991). Las posibilidades que ofrece el ordenador para poder disponer en todo momento de la información que se precise, o realizar tareas de codificación en equipo, o modificar en cualquier momento los sistemas de categorías y de códigos, son posibilidades que están haciendo evolucionar las formas de trabajar de los investigadores.

No obstante, debemos tener en cuenta que también podemos encontrarnos con el lado oscuro de la utilización del ordenador. En este sentido De la Orden (1988) nos señala la "amenaza mecanicista y deshumanizadora", que en el caso del análisis cualitativo iría en contra de la flexibilidad y singularidad que le caracteriza. Tesch (1988) añade el peligro del distanciamiento entre el investigador y sus datos. Seidel (1991), por su parte, nos señala cómo un posible peligro de la utilización de los ordenadores es que puede llevar a



algunos investigadores a pensar que se trata de identificar y contar elementos, sin necesidad de escudriñar, analizar y evaluar críticamente los elementos contados.

Mas los riesgos que se corren, como señala Tesch (1990), no surgen del propio ordenador, sino de la actitud de quienes lo usan. El ordenador es una herramienta que está al servicio de un proceso general que es dirigido por el investigador. No se debe incurrir en la utilización rígida de un programa con independencia de los objetivos y características de la investigación, ni en atribuirles tareas inapropiadas por el mero hecho de encontrarse dentro del campo conceptual y decisional del investigador.

### **Referencias Bibliográficas.**

AERA and ASU create division-based INTERNET forums (1993): **Educational Researcher**, 22. 9, 21.

CABERO, J. y DUARTE, A. (1994): CD-ROM en la enseñanza e investigación: una tecnología en aumento. **Pixel-Bit. Revista de Medios y Educación**, 1, 83-101.

CARRO, L. (1993): **Posibilidades del WordPerfect 5.1. para el análisis de datos cualitativos.** Comunicación presentada en el VI Seminario de Modelos de Investigación Educativa, Madrid.

DE LA ORDEN, A (1988): Informática e investigación educativa, en **Aspectos metodológicos en la investigación educativa. II Congreso Mundial Vasco.** DENDALUCE, I. (Coord.). Madrid, Narcea, 276-295.

DE LARA GUIJARRO, E. (1992): **El ordenador en la investigación educativa.** Madrid. UNED.

ENVISIONOLOGY (1992): **CVideo User Guide.** San Francisco, CA. Envisionology Inc.

GARCÍA, E., GIL, J. y RODRÍGUEZ, G. (1993): **Análisis de datos cualitativos en la investigación sobre la diferenciación educativa.** Ponencia presentada en el VI Seminario de Métodos de Investigación, Madrid.

HUBER, G.L. (1991): **AQUAD. Análisis de datos cualitativos con ordenadores. Principios y manual del paquete de programas AQUAD 3.0.** Sevilla. Carlos Marcelo.

LEBART, L. y otros (1994a): **SPAD\*T Version 1.5. Système Portable pour l'Analyse des Données Textuelles.** Saint-Mandé. CISIA.

LEBART, L. y otros (1994b): **Introduction à SPAD\*T Intégré Version 1.5 P.C. Système Portable pour l'Analyse des Données Textuelles.**Saint-Mandé. CISIA.

LEBART, L. y SALEM, A. (1988): **Analyse Statistique des Données Textuelles. Questions ouvertes et Lexicométrie.** París. Bordas.

LIZASOAIN, L. (1992): **Bases de datos en CD-ROM.** Madrid. Paranifno.

McCAHILL, M. (1992): The Internet Gopher Protocol: A Distributed Server Information System. **Connexions- The Interoperability Report**, 6, 7, 10-14.

PERSOFT, Inc. (1990): **SmarTerm340. User Manual.** Madison, WI: Autor.

PERSONAL BIBLIOGRAPHIC SOFTWARE (1992): **Pro-Cite for IBM personal computers and compatibles.** Ann Arbor, MI. PBS, Inc.

PIERCE, J. y otros (1994): The Educational Research List (ERL-L) on BITNET/INTERNET. **Educational Research**, 23, 2, 25-28.

QSR (1994): **NUDIST Version 3.0 User Guide.** Victoria. Qualitative Solutions Research.

RICHARDS, L. y RICHARDS, T. (1991): The Transformation of Qualitative Method:Computational Paradigms and Research Processes, en **Using Computers in Qualitative Researchs.** FIELDING, N. y LEE, R. (Eds.), Londres. Sage, 38-53.

RICHARDS, T. y RICHARDS, L. (1994): Using Computers in Qualitative Research, en **Handbook of Qualitative Research.** DENZIN, N.K. y LINCOLN, Y.S. (eds.). Londres. Sage, 445-462.

RODRIGUEZ, G. y otros (1993): Las concepciones de los futuros profesores sobre la función docente. **Revista de Enseñanza Universitaria**, 6, 9-26.

SANZ SACRISTAN, M.A. (1994): A, B, C de Internet. **Red IRIS. Boletín de la red nacional de I+D, RedIRIS**, 28, 15-30.

SEIDEL, J. (1991): Method and Madness in the Application of Computer Technology to Qualitative Data Analysis, en **Using Computers in Qualitative Researchs.** FIELDING, N. y LEE, R. (Eds.), Londres. Sage, 107-116.

TAYLOR, S.J. y BOGDAN, R. (1986): **Introducción a los métodos cualitativos de investigación.** Barcelona. Paidós.

TESCH, R. (1988): **The impact of the computer on qualitative data analysis.** Comunicación presentada en la reunión Computers in Qualitative Researchs. FIELDING, N. y LEE, R. (Eds.), Londres. Sage, 107-116.

TESCH, R. (1988): **The impact of the computer on qualitative data analysis.** Comunicación presentada en la reunión anual de la American Educational Research Association, 1988.

TESCH, R. (1990): **Qualitative Research: Analysis Types and Software Tools.** Londres, The Falmer Press.

TESCH, R. (1991): Software for Qualitative Researchers: Analysis Needs and Program Capabilities, en **Using Computers in Qualitative Researchs.** FIELDING, N. y LEE, R. (Eds.), Londres. Sage,16-37.

TSU-DER-CHOU, G. (1987): **dBASE III. Guía del programador.** Madrid. Anaya Multimedia.