

FORMULACIÓN Y COMPROBACIÓN DE HIPÓTESIS CUALITATIVAS CON AYUDA DEL PROGRAMA «AQUAD»

Javier Gil, Eduardo García, Gregorio Rodríguez y Ana Corrales
Universidad de Sevilla

Las operaciones que normalmente se asocian a los procesos de análisis de datos cualitativos (García et al., 1993) se han visto enormemente facilitadas por la incorporación del ordenador, cuyas posibilidades de manejo de gran cantidad 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 también en el ámbito de la investigación educativa y, en particular, en el análisis de datos cualitativos textuales. Las aplicaciones informáticas han venido a reducir la enorme cantidad de esfuerzo y tiempo que consumen las tareas de análisis de este tipo de datos. Las ventajas que la utilización de procesadores de texto o bases de datos ofrecen —permitiendo la entrada y el almacenamiento de los datos textuales, la ordenación de una serie de ítems o la búsqueda de informaciones específicas— han sido superadas por programas específicamente diseñados para el análisis del material cualitativo. Este tipo de programas se inscriben en las líneas cualitativas de investigación, y su contribución es importante de cara a realizar el proceso de análisis de datos cualitativos de un modo que pueda ser controlado, reconstruido y comunicado.

En este trabajo, hemos seleccionado el programa AQUAD —Analysis of Qualitative Data— (Huber, 1988) que, además de facilitar las funciones habitualmente recogidas en el software diseñado para esta tarea, tales como la codificación, recuperación de unidades o el recuento de ocurrencias y coocurrencias, permite formular y comprobar condiciones lógicas subyacentes a potenciales relaciones entre los códigos. Nuestra atención se dirigirá especialmente al estudio de sus posibilidades para la formulación y comprobación de hipótesis en el marco de la investigación cualitativa, entendiendo por tales hipótesis *relaciones entre tópicos identificados en un texto a partir de las cuales es posible reconstruir un todo estructurado que nos aproxima a una comprensión global de los datos.*

EL ANÁLISIS MEDIANTE AQUAD

El programa AQUAD ofrece un amplio abanico de posibilidades a los investigadores que se enfrentan a la tarea de analizar datos cualitativos, contribuyendo en las fases de reducción de datos, presentación de los mismos y extracción de conclusiones (Huber y Marcelo, 1990). Aunque aquí nos interesa especialmente la fase de comprobación de hipótesis (fase con la que AQUAD se distancia de otros programas para el análisis de datos cualitativos), no podemos perder de vista el proceso completo de análisis que permite desarrollar este programa. El procedimiento de análisis que aquí presentamos queda esquematizado en la Figura 1.

1. Codificación de los textos

Una operación habitual en el análisis de datos cualitativos es la reducción de datos, que suele concretarse en la codificación de los datos textuales. AQUAD permite al analista señalar, sobre el texto visualizado en pantalla, el inicio y el final de cada fragmento considerado como una unidad de análisis y etiquetarlo con un código. La codificación puede ser múltiple: un mismo fragmento puede ser codificado por varios códigos, los textos etiquetados con códigos pueden superponerse parcialmente, o quedar incluidos unos dentro de otros.

Durante la codificación puede resultar útil revisar el contenido de las categorías, examinando el contenido de los fragmentos incluidos en cada una de ellas con objeto de asegurarnos de la consistencia en la codificación. La recuperación de fragmentos etiquetados con un determinado código podría realizarse de forma automática mediante AQUAD. En cualquier caso, el ordenador cumple tareas mecánicas, como la búsqueda de fragmentos codificados, pero no conceptuales; no puede sustituir al investigador en la toma de decisiones sobre la asignación de tales fragmentos a determinadas categorías.

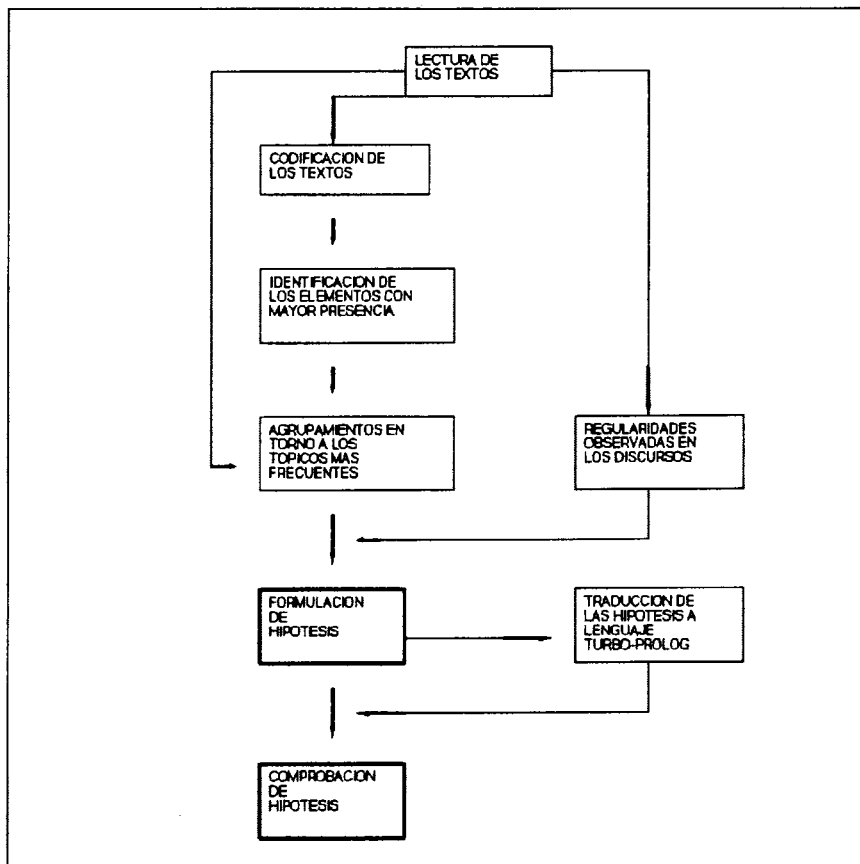


Figura 1. Proceso seguido en el análisis de los datos usando el programa AQUAD.

2. Identificación de los elementos con mayor presencia

A partir de la codificación, podemos identificar los temas que aparecen con mayor frecuencia en el conjunto de textos codificados. Tales tópicos serán los más relevantes de cara al tema investigado o, cuando menos, los aspectos sobre los cuales se posee mayor información. Puesto que nuestras conclusiones se basarán en la presentación de evidencia que apoye ciertas afirmaciones, es necesario centrar la atención sobre los tópicos de mayor presencia en los textos. Identificar los códigos más frecuentes es una operación bastante fácil de realizar mediante AQUAD, que puede facilitarnos un recuento de las apariciones de cada código en cada uno de los textos sometidos a análisis.

3. Agrupamientos en torno a los códigos más frecuentes

Identificados los tópicos más frecuentes, es necesario buscar qué otros tópicos se encuentran relacionados con éstos. Un primer paso para alcanzar tal objetivo consiste en analizar el agrupamiento físico de códigos que se da en torno a ellos. Estudiar los agrupamientos consistiría en localizar códigos cuyos fragmentos asociados coaparecen en el texto junto a los de los códigos más frecuentes. Como criterio de agrupamiento podríamos adoptar el no distanciarse en más de un determinado número de líneas del código considerado. Es decir, entenderíamos que se agrupan en torno a un código aquellos códigos cuyas áreas de texto correspondientes anteceden, se superponen o suceden al área de texto de un código determinado, sin que entre ambas áreas exista una distancia, en número de líneas, superior a un valor fijado.

En lugar de revisar una gran cantidad de páginas buscando tales agrupamientos, AQUAD podría facilitarnos este tipo de coapariciones, a partir de las cuales, como se verá en el próximo apartado, se intuyen regularidades y estructuras de significado conducentes a la formulación de hipótesis.

4. Formulación de hipótesis

Para entender la naturaleza de la posible relación entre los códigos que coaparecen en mayor número de ocasiones, es preciso volver a los textos codificados a fin de revisar algunos de los fragmentos correspondientes. Esta fase del proceso supone, por tanto, una vuelta a los datos para establecer si la coaparición de los códigos obedece a una simple yuxtaposición temática, a una secuenciación de tópicos, a una relación causa-efecto, a una inclusión de determinados temas en otros, a una contraposición de elementos, etc. A la formulación de hipótesis también contribuyen las ideas que pudieran haber surgido en el analista a partir de la lectura inicial de los textos.

Las hipótesis formuladas se refieren a las posibles regularidades de patrones de significado —constelaciones de códigos— presentes en los datos; plantean la existencia de una relación entre dos o más códigos, entendiendo que esta relación se traduce en proximidad espacial dentro del texto, inclusión entre unidades o secuencia de aparición. De este modo, el programa permite extraer conclusiones respecto a configuraciones de unidades de significado que se repiten a lo largo de un corpus de datos, cuya comprobación manual resultaría una tarea de complejidad considerable.

Para ilustrar el modo en que AQUAD contribuye a la formulación de hipótesis cualitativas, traemos aquí un ejemplo de hipótesis en la que se establece una relación entre dos códigos. Los datos utilizados corresponden a una investigación reciente en la que se pedía a grupos de profesores que opinaran sobre la reforma educativa (Gil Flores, 1992). Los datos textuales producidos en dos de estos grupos han sido utilizados para ejemplificar el proceso descrito en el presente trabajo.

Codificado el texto y realizado el recuento de códigos, hemos destacado el código «COP» (consecuencias de la aplicación de la reforma para el profesorado) como uno de los más frecuentes. El

posterior examen de agrupamientos en torno a este código, revela que el código «STR» (aspectos estructurales de la reforma) es el que en mayor medida coaparece junto a éste. Para determinar el tipo de relación que pudiera existir entre ambos, revisamos algunos de los fragmentos de texto implicados. Así, por ejemplo, citamos aquí una de las coocurrencias de los códigos COP y STR en nuestro corpus de datos, localizada con la ayuda del programa AQUAD:

1254 Yo, por lo que conozco de la LOGSE, el ^STR-1263
1255 diseño de la Primaria yo lo encuentro
1256 bien. Ahora, para mí, el meollo de la
1257 reforma es precisamente la Secundaria. A
1258 mí me parece... yo veo unas lagunas tan
1259 gordas, que yo es que no entiendo cómo se
1260 va a implantar la Secundaria. Es que no lo
1261 sé. Quién va a dar el primer ciclo, quién ^COP-1263
1262 el segundo, los licenciados, los
1263 profesores de la Segunda Etapa.

La lectura de las áreas de texto referidas en nuestro ejemplo permite establecer que las consecuencias para el profesorado se derivan de la nueva estructuración del sistema educativo. Esto concuerda con la impresión obtenida durante la lectura global de los textos, en la cual advertíamos una cierta preocupación por el modo en que se verán afectados los profesores por los aspectos estructurales, pero no por los curriculares (los fragmentos en los que se aludía a aspectos curriculares fueron codificados con CUR).

Consecuentemente, la hipótesis planteada afirmaría que *los profesores contemplan las consecuencias que tendrá para ellos la aplicación de la reforma como una derivación de los aspectos estructurales y no de los curriculares.*

5. Comprobación de hipótesis

Las hipótesis serán finalmente contrastadas en el conjunto de los datos. Debe tenerse en cuenta que esta comprobación no es un contraste de hipótesis en sentido teórico, puesto que no existe una distribución teórica de referencia a la que comparar los resultados obtenidos para un determinado estadístico ni existen criterios fijados para la aceptación o el rechazo de las mismas. Se trata de un contraste en sentido empírico, consistente en explorar la frecuencia con que determinadas «constelaciones» de códigos se dan en el corpus de datos, indicando tendencias que se ven, de este modo, apoyadas por un mayor o menor número de ocurrencias dentro del mismo. En este sentido, en lugar de formulación y comprobación de hipótesis, quizás sería más adecuado hablar de relaciones entre códigos y búsqueda de evidencia para apoyar la existencia de tales relaciones.

Para que las hipótesis puedan ser comprobadas con la ayuda de AQUAD, las relaciones que establecen deben ser trasladadas al lenguaje de programación turbo-prolog, de modo que el ordenador proceda sistemáticamente a comparar todos los elementos de significado presentes en el conjunto de textos, de acuerdo con condiciones lógicas especificadas para ellos. Retomando el ejemplo que presentábamos en el apartado anterior, para comprobar la hipótesis formulada en nuestro corpus de datos, habría que examinar si los códigos COP van precedidos por códigos STR, pero no por códigos CUR. Entenderemos que COP está precedido por STR o CUR cuando el fragmento correspondiente a COP no empieza antes que el de STR o CUR ni a una distancia mayor de 5 líneas después de su final. La comprobación de estas condiciones en cada pareja formada por un código COP y un código CUR

o STR constituye el eje del programa en turbo-prolog que permite comprobar la hipótesis de nuestro ejemplo. El programa completo quedaría del siguiente modo:

```

hy(1,1):-write("APARICIONES DEL CODIGO COP","\n").
hy(1,1):-C="COP", k(_,A1,E1,C), write(A1,E1,C,"\n"),fail.
hy(1,1):-write("ASPECTOS ESTRUCTURALES PRECEDIENDO A COP:", "\n").
hy(1,1):-C1="COP",C2="STR",D=5,
    k(N,A1,E1,C1), str_int(A1,IA1),
    k(N,A2,E2,C2), str_int(A2,IA2),str_int(E2,IE2),
    IA1-IA2>=0,IA1-IE2<=D,
    write(A2,E2,C2," - ",A1,E1,C1,"\n"),fail.
hy(1,1):-write("ASPECTOS CURRICULARES PRECEDIENDO A COP:", "\n").
hy(1,1):-C1="COP",C2="CUR",D=5,
    k(N,A1,E1,C1), str_int(A1,IA1),
    k(N,A2,E2,C2), str_int(A2,IA2),str_int(E2,IE2),
    IA1-IA2>=0,IA1-IE2<=D,
    write(A2,E2,C2," - ",A1,E1,C1,"\n"),fail.

```

Las instrucciones presentadas aquí y escritas al final del fichero AQD-H.PRO, se ejecutarían sobre los ficheros de códigos creados por AQUAD para cada texto codificado. En tales ficheros, cada código presenta la estructura $k(N,A,E,C)$, donde N es la extensión con que se nombra el fichero de texto a analizar, A y E son las líneas en que comienza y termina el área marcada por un código, y C la etiqueta de tres caracteres que identifica al código.

De acuerdo con lo formulado, en las dos primeras líneas se indica al ordenador que busque los códigos COP y escriba las líneas de comienzo y final de los párrafos correspondientes. El siguiente bloque de instrucciones comienza asignando valores a las variables código (C1 y C2) y distancia (D). Las líneas quinta y sexta implican la búsqueda de una pareja de códigos C1 y C2 y la transformación en números enteros (IA1, IA2, IE2), con las funciones «str_int», de los parámetros que indican principio y final de un área de código (A1, A2, E2), expresados inicialmente como cadenas de caracteres. Los números obtenidos deberán cumplir las condiciones establecidas en la línea siguiente, es decir, de acuerdo con la condición $IA1-IA2_0$, al restar la línea en que comienza el fragmento de STR a la línea en que comienza COP deberíamos obtener 0 ó un número positivo. Ello significaría que el número IA1 es mayor o igual que IA2, y por tanto, el fragmento de COP comienza después o al mismo tiempo que el de STR. Por otro lado, $IA1-IE2_5$ impone la condición de que la diferencia entre el inicio de COP y el final de STR no supere las 5 líneas, o lo que es igual, que COP no comience a una distancia superior a las 5 líneas después del final de STR. El programa escribirá los parámetros que definen a los fragmentos que cumplan tales condiciones, de acuerdo con lo especificado en la séptima línea. La instrucción «fail» indica que se vuelva a desarrollar el proceso hasta que hayan sido considerados todos los casos y soluciones posibles. El proceso seguido con la pareja de códigos COP y STR sería repetido con COP y CUR, como indican las restantes líneas del programa.

El resultado de ejecutar este programa es la localización en el texto de las parejas de códigos que presentan la estructura exigida por la hipótesis, para que en función de la evidencia empírica podamos valorar la plausibilidad de ésta. En el ejemplo que desarrollamos, los resultados fueron:

***** textouno.cod

APARICIONES DEL CODIGO COP

110	126	COP
142	155	COP
252	269	COP
289	307	COP
423	424	COP
694	699	COP
1586	1592	COP
1837	1845	COP
1860	1861	COP

ASPECTOS ESTRUCTURALES PRECEDIENDO A COP:

110	118	STR	-	110	126	COP
252	269	STR	-	252	269	COP
289	293	STR	-	289	307	COP
421	422	STR	-	423	424	COP
675	699	STR	-	694	699	COP
1832	1838	STR	-	1837	1845	COP
1857	1861	STR	-	1860	1861	COP

ASPECTOS CURRICULARES PRECEDIENDO A COP:

1586	1592	CUR	-	1586	1592	COP
------	------	-----	---	------	------	-----

***** textodos.cod

APARICIONES DEL CODIGO COP

36	38	COP
1156	1163	COP
1184	1189	COP
1230	1234	COP
1240	1244	COP
1261	1263	COP
1505	1507	COP
1732	1746	COP

ASPECTOS ESTRUCTURALES PRECEDIENDO A COP:

1139	1156	STR	-	1156	1163	COP
1168	1184	STR	-	1184	1189	COP
1218	1234	STR	-	1230	1234	COP
1236	1244	STR	-	1240	1244	COP
1254	1263	STR	-	1261	1263	COP
1732	1738	STR	-	1732	1746	COP

ASPECTOS CURRICULARES PRECEDIENDO A COP:

En el primer texto analizado, la hipótesis no se cumple, pues en una de las ocasiones el código COP estuvo precedido de un código CUR. En el segundo caso, seis de las ocho apariciones de COP estuvieron precedidas por STR, y nunca por CUR, lo cual permitiría afirmar que en este texto se confirma la hipótesis de partida.

CONCLUSIÓN

Si bien la utilidad de los programas para el análisis de datos cualitativos de cara a la codificación, recuperación de fragmentos codificados y examen del contenido de las categorías es manifiesta, AQUAD permiten ir más allá de estas funciones añadiendo la posibilidad de comprobar hipótesis formuladas por el investigador. Con ello, se consigue una cierta objetividad y sistematización en el proceso, que podría ser desarrollado por distintos autores de modo que los resultados obtenidos para cada una de las posibles hipótesis serían coincidentes.

A pesar de todo, es preciso destacar que el tipo de hipótesis que permite formular y comprobar es limitado; las relaciones que pueden establecerse entre los elementos de significado se basan exclusivamente en la proximidad dentro del texto, la secuenciación, la inclusión o superposición de las áreas de texto codificadas. Otro tipo de relaciones no detectables por la disposición espacial de las unidades sino a partir del significado contenido en las mismas quedarían excluidas. Además, podría cuestionarse el que exista realmente una relación entre dos códigos que aparecen secuenciados, próximos o superpuestos. De ahí que se restrinja la formulación de hipótesis a los códigos que presentan cierta frecuencia, para que la reiterada aparición en el texto de determinados patrones de códigos apoye la existencia de una cierta relación entre los códigos implicados. Por otra parte, podrían existir dificultades para establecer cuándo una hipótesis ha sido confirmada. Como afirmábamos anteriormente, el contraste de hipótesis en este tipo de análisis no es un contraste en sentido estricto, y por tanto no tenemos criterios teóricos para rechazar o aceptar una hipótesis. Las decisiones pueden tomarse a veces sobre la base de una valoración del investigador acerca de la relevancia del número de apariciones de un código dentro de una estructura buscada frente al número total de apariciones de ese código. Los resultados de comprobar una hipótesis indican únicamente la evidencia encontrada para apoyar la plausibilidad de lo formulado.

En cualquier caso, consideramos que el programa AQUAD representa una interesante herramienta en manos del analista, de cara a la formulación de relaciones entre tópicos y a la búsqueda de evidencia que las confirme.

REFERENCIAS BIBLIOGRÁFICAS

- BOGDAN, R. C. y BIKLEN, S. K. (1982): *Qualitative research for education. An introduction to theory and methods*. Boston: Allyn and Bacon Inc.
- GARCÍA, E. et al. (1993): *Análisis de datos cualitativos en la investigación sobre la diferenciación educativa*. Ponencia presentada al VI Seminario de Métodos de Investigación, El Escorial.
- GIL FLORES, J. (1992): *Análisis de datos cualitativos. Aplicación al caso de los datos procedentes de grupos de discusión*. Sevilla: Tesis Doctoral inédita.
- HUBER, G. (1988): Análisis de datos cualitativos: la aportación del ordenador. En C. MARCELO (Coord.): *Avances en el Estudio del Pensamiento de los Profesores*. Sevilla: Servicio de Publicaciones de la Universidad de Sevilla. (pp. 77-85).
- HUBER, G. y MARCELO, C. (1990): Algo más que recuperar palabras y contar frecuencias: la ayuda del ordenador en el análisis de datos cualitativos. *Enseñanza*, (8), 69-84.
- LOFLAND, J. y LOFLAND, L. H. (1984): *Analyzing social settings*. Belmont, CA: Wodsworth Publishing Company.