

El análisis de contenido: apuntes de estudio.

Prof. Dr. J. Andrés Domínguez-Gómez
Dpto. Sociología, Trabajo Social y Salud Pública
andres@uhu.es

Apuntes de estudio para la asignatura 'Métodos y Técnicas de Investigación Social', 2º Curso, Grado en Trabajo Social. Universidad de Huelva.

Análisis de contenido: definición y usos.

Como es habitual, la literatura ofrece un enorme **repertorio de definiciones** del AC. Se citan habitualmente como clásicas:

Berelson (1962): “Es una técnica de investigación para la descripción objetiva, sistemática y cuantitativa del contenido manifiesto de la comunicación”.

Holsti (1969): “Cualquier técnica de investigación que sirva para hacer inferencias mediante la identificación sistemática y objetiva de características específicas dentro de un texto”.

Krippendorff (1990): “Es una técnica de investigación destinada a formar, a partir de ciertos datos, inferencias reproducibles y válidas que puedan aplicarse a su contexto”.

Weber (1990): “Es un método de investigación que utiliza un conjunto de procedimientos para realizar inferencias válidas de textos”.

Sierra Bravo (1992): “La técnica, sin duda más elaborada y que goza de mayor prestigio científico en el campo de la observación documental”.

De estas definiciones podemos deducir:

La evolución del AC hacia una técnica que **trasciende lo manifiesto**, para centrar la observación también en lo tácito de la comunicación.

También que el **'texto'** es el objeto de observación (aunque los datos en bruto pueden tener diversos formatos: texto, imagen, vídeo, audio...), pero que el **contexto** es fundamental para su comprensión.

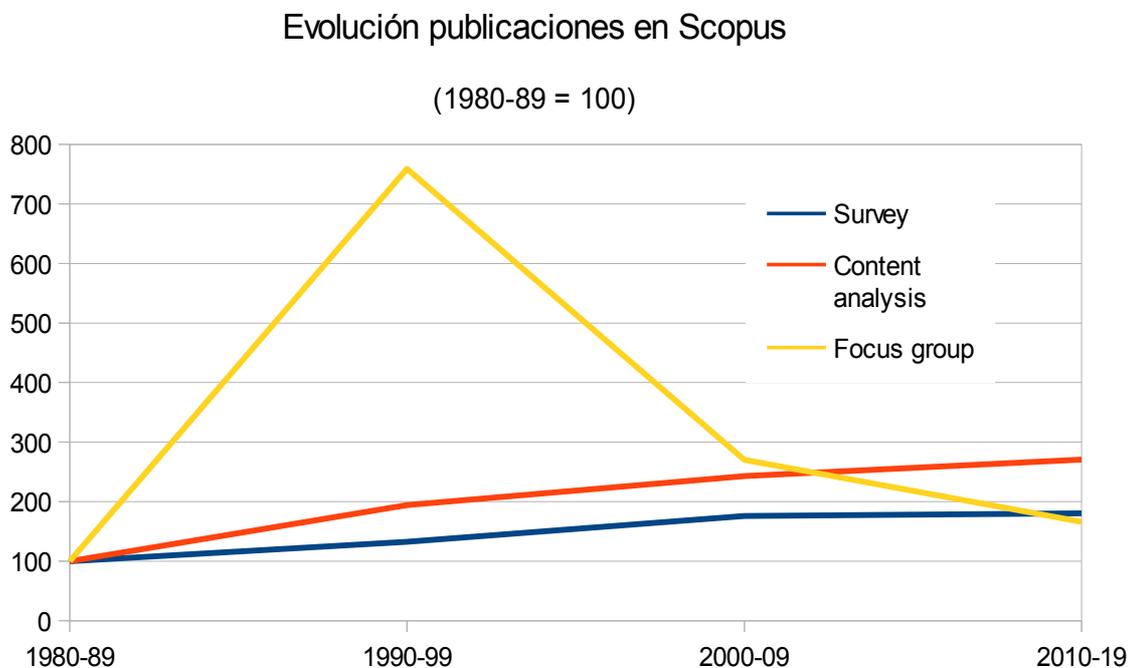
Las definiciones cargan tintas en la 'objetividad', con énfasis en conceptos como 'inferencia', 'reproducibilidad'..., cuya importancia, en la actualidad, puede relativizarse debido a la ampliación del uso del AC y la aceptación de su flexibilidad.

Desacuerdo sobre la consideración del AC **como método o como técnica** de investigación. Como método, el AC implicaría la adopción de una visión interpretativa, que aprovecharía su potencialidad de enfoque mixto en el procesamiento y análisis de datos. Como técnica,

podría definirse de forma clásica como cualquier otra herramienta a emplear en una metodología más amplia, definida ex-profeso para un proyecto de investigación determinado. El enfoque de método supone más bien la articulación de la metodología de un proyecto de investigación en torno a las diferentes fases del AC. Así, tendríamos el AC como método (enfoque, afrontamiento del objeto) o metodología (pasos a dar, desde la recogida de datos, hasta la elaboración de resultados, pasando por el procesamiento de datos y su análisis).

El empleo del AC como método y/o técnica de análisis va en aumento. De hecho, incluso tomando como referencia la producción de artículos científicos en todas las áreas de conocimiento (con indexación por Scopus) que tengan en su título “content analysis”, observamos que desde 1980 su frecuencia relativa ha ido en aumento (ver gráfico XXX), llegando a más de 3500 papers durante la década 2009-2019.

Figura 16. Evolución publicaciones sobre AC en Scopus



Fuente: elaboración propia.

Comparada con otras técnicas de investigación de las que tomamos datos, el AC mantiene una evolución relativa constante siempre al alza, de un modo estable, superando, en porcentajes de evolución sobre la referencia de 1980-89, a las otras dos referencias (encuesta y grupo focal).

En términos generales, podemos afirmar que el AC está **especialmente indicado para desentrañar la carga simbólica de la comunicación**, yendo más allá de 'lo aparente'. La historia del AC comienza claramente vinculada al análisis de la comunicación en la primera mitad del siglo XX, con investigaciones vinculadas al estudio de los estereotipos sociales, las actitudes o incluso la propaganda de la II Guerra Mundial.

Esta vinculación con el análisis comunicacional, aún manteniéndose, está siendo **completada por**

un enfoque holístico, más ambicioso, al tiempo que realista. Si adoptamos una interpretación puramente sociológica de la realidad social, como es el marco teórico *construccionista* ('la realidad social es construida socialmente'), el AC se posiciona como uno de los enfoques metodológicos de mayor interés en la actualidad, con mayor potencia analítica y de mayor aspiración o alcance. Su capacidad de abordaje mixto de la realidad ofrece un abanico de opciones en las distintas fases del proceso investigador. Desde un punto de vista cuantitativo, el AC ofrece outputs de interés descriptivo, con énfasis en objetivación de criterios de análisis, como por ejemplo índices de frecuencia, de coocurrencia... Desde un punto de vista cualitativo, ofrece la posibilidad de comprensión profunda de los discursos que articulan la comunicación, su estructura, su carga simbólica o de significados en general; se presta a la categorización, clasificación y codificación del contenido...

De este modo, cerramos este apartado introductorio con una actualización del concepto metodológico. El AC, que comenzó centrado en un enfoque analítico positivista, **se sitúa hoy en un punto claramente mixto**, donde apoya todas sus fortalezas. El software para realizar AC tiene bastante culpa de ello, pues está facilitando herramientas que se adaptan cada vez mejor a las necesidades del proceso analítico y a la emisión de resultados visuales, con buena capacidad de resumen, sin menoscabo del detalle numérico o simbólico. La demanda investigadora de abordaje holístico (cualitativo + cuantitativo) está creciendo, y cada vez es más habitual no renunciar a uno de los dos outputs, sino generarlos de forma complementaria.

A la vista de lo anterior, nuestra propuesta concreta de definición del AC quedaría como sigue:

Metodología de investigación de carácter mixto (cualitativo y cuantitativo), que aunque se desarrolla sobre datos primarios de carácter cualitativo (textos, imágenes, audiovisual...), posibilita la combinación flexible de procedimientos, técnicas y análisis en pos de resultados de máxima adaptación a la realidad compleja del objeto de estudio.

Características del AC

Los inicios positivistas del AC pretendían **objetivizar criterios de análisis**, que se reflejasen en posibilidad de contrastación de hipótesis y, en general, de afrontamiento científico-clásico de un objeto de estudio tan cargado de simbolismo, intencionado o no (carga social subconsciente) como es la comunicación. Hoy día, aún habiendo ampliado su enfoque con la dimensión cualitativa, podemos seguir afirmando lo mismo.

El AC pretende, por la vía de la **operacionalización**, establecer criterios objetivos para la sistematización en el análisis y la obtención de resultados válidos y fiables. El AC facilita la conceptualización, dimensionamiento, definición de indicadores y, a menudo, combinación en índices numéricos que describen la realidad analizada, que relacionan entre sí variables definitorias de la misma..., y que al mismo tiempo pueden estar referidos a micro-teorías contextualizadas propias de la Grounded Theory; como sabemos esta teoría es el apoyo clave de la mayoría de los enfoques 'soft' y cualitativos del análisis sociológico.

Así, podemos establecer otra característica propia del AC: su **flexibilidad y adaptabilidad** a las necesidades del analista. La función objetivizadora se añade a la de comprensión de la realidad.

Esta característica, en nuestra opinión, ha relativizado gran parte de las funcionalidades del AC que la literatura especializada nos ofrece, en formato de clasificación sistemática de las mismas. Por ejemplo, podemos leer que las aplicaciones del AC se distinguen según las formas de inferencia que pueden realizarse en el análisis, sobre la base de que el objetivo del AC es realizar inferencias (cuestión que, como sabemos, podemos, a su vez, relativizar: no sólo es realizar inferencias). Y a continuación leemos que también se distinguen por la ligazón de dichas inferencias a 'sistemas' (básicamente, formas de entender la realidad socialmente construida). Es claro que, en este punto, nos quedamos a un paso del análisis del discurso social, que tradicionalmente ha estado vinculado, como objeto, al grupo de discusión, focal u otras formas de dinámicas grupales orientadas a poner de manifiesto cómo un grupo social expresa su forma de entender la realidad (objeto del análisis). En el fondo, estamos hablando de análisis de la expresión y del contenido de la comunicación, donde toda la literatura del AC encuentra un punto en común en cuanto su objeto analítico...

Este enfoque contemporáneo del AC hace que, sin desdeñar el origen de su objeto -el análisis comunicacional-, se haya **ampliado enormemente su funcionalidad** a casi cualquier objeto de estudio; idealmente, aquellos cuyos objetivos apunten hacia una exploración profunda de temas complejos o desconocidos; también aquellos que requieran específicamente de la complementariedad de técnicas analíticas.

Por último, es cada vez más habitual tener que trabajar en equipos multidisciplinares, con especialistas de otros ámbitos científicos, como las ciencias clásicas o naturales. El AC se ha demostrado como una herramienta interesante de **puesta en valor** de las habilidades y posibilidades analíticas de las ciencias sociales sobre objetos de estudio complejos. Por ejemplo, en evaluaciones de impacto ambiental, científicos naturales cooperan con científicos sociales para estudiar los impactos de proyectos de desarrollo en el entorno socio-ambiental. Los clásicos modelos y herramientas de diagnóstico ambiental (basados en la detección de cambios en flora, fauna, suelo, aire y agua) suelen excluir la presencia de comunidades humanas en el entorno afectado. La complejidad de las relaciones socio-ambientales es un campo idóneo para el AC, y mostrará a nuestros compañeros de proyecto cómo somos capaces de aproximarnos al conocimiento de esa complejidad mediante técnicas de alta fiabilidad y validez, siguiendo criterios comunes a cualquier área científica: sistematicidad, objetividad, triangulación, definición y estandarización de herramientas y criterios de análisis, etc.

Procedimientos y conceptos clave en el AC

Los datos y el 'muestreo'.

Presentamos en este apartado los principales elementos y conceptos que hay que conocer para poder afrontar un proceso analítico basado en AC. Se ofrecen definiciones de los mismos, y a continuación se los relaciona con los diferentes momentos del citado proceso, utilizando para ello un ejemplo de investigación basada en AC.

En cuanto a **los datos** sobre los que trabajaremos con AC, se suelen definir como **unidades de información** grabadas en un medio resistente y duradero (papel, film, disco..., hoy día ya, casi siempre en formato de archivo digital). Estos datos, como en cualquier técnica, tienen que ser

significativos, es decir, de interés para nuestro objetivo de investigación. Han de ser identificados, separados y definidos, para determinar las 'unidades de análisis' (concepto clave que veremos más adelante). En el caso del AC los datos adquieren formato original cualitativo: texto, imágenes, audio...

El AC tradicional pretende **cubrir 'razonablemente' el equivalente al universo poblacional** en la técnica de encuesta, por lo que habitualmente se define el proceso de selección de casos de estudio como '**muestreo**'. No obstante, el hecho mismo de esa 'razonabilidad' puede eliminar su base fundamental positivista, en sentido estricto. En sentido estricto, 'muestreo' se refiere a un procedimiento sistemático que sirve para la selección de las unidades muestrales, es decir, los casos específicos que participarían en la investigación (objetos de estudio), a partir de un universo poblacional o población de base cuyas características y distribución en el ámbito de estudio han de ser perfectamente conocidas y controladas. Sin este conocimiento y control, no es posible realizar un proceso de muestreo convencional.

Tomamos en este punto un ejemplo real, que sirva como referencia a la hora de entender los conceptos vinculados al AC, y las diferencias entre la formulación clásica del AC y la contemporánea. A un equipo de investigación se le encargó definir y comprender la situación de la 'investigación e innovación responsable' (RRI de sus siglas en inglés, como objeto de reflexión académica) en la totalidad de la producción científica. La pregunta general de investigación fue la siguiente: ¿Cómo es la RRI a nivel global? A esta pregunta general se respondería, por ejemplo, con especificaciones basadas en zonas geográficas o regiones del mundo, temáticas concretas en las que se centra la RRI, evolución temporal de los estudios en RRI, etc. El muestreo en el AC habría de realizarse sobre la totalidad de la producción científica publicada desde una fecha determinada. Por razones prácticas, y bajo criterios de razonabilidad, se definió como universo la producción científica indexada en Scopus, y sobre esta base de datos se lanzaron un total de diez búsquedas con sus correspondientes diez *keywords* (expresiones clave), a localizar en el resumen de las publicaciones indexadas en Scopus. La base de datos quedó configurada como una matriz de 3521 casos (referencias bibliográficas) y un conjunto de variables o campos (autor, año, título, resumen, etc...) de cada uno de los casos. Obsérvese que los criterios establecidos para el muestreo, si bien responden plenamente al objetivo de la investigación, se trata de **criterios de carácter cualitativo** (*keywords*), que fueron previamente definidos por un estudio de la bibliografía especializada en RRI (siglas en inglés de investigación e innovación responsable).

En el AC, las unidades que componen la muestra se denominan **unidades de muestreo**. Son las diversas **partes de la realidad sometidas a observación**, que el investigador considera separadas entre sí y, por tanto, como susceptibles de ser analizadas independientemente unas de otras. En nuestro ejemplo, las unidades de muestreo serían las **publicaciones** indexadas en Scopus.

Las unidades de análisis y su relación entre ellas.

Cada parte de la unidad de muestreo que puede ser considerada como analizable separadamente es denominada '**unidad de registro**'. En cada unidad de registro aparece alguna de las referencias en las que el investigador está interesado: símbolos, palabras, números, etc. Por ejemplo, los temas vinculados al objetivo de investigación. En nuestro ejemplo, las unidades de registro fueron los diferentes **campos de la base de datos** (autor, año, título, resumen, etc...).

Otro de los conceptos claves para la práctica del AC es la **unidad de contexto**. Se trata de la porción informativa donde se enmarca la unidad de registro, concediéndole el sentido necesario para ser correctamente comprendida. Si no se define adecuadamente, se perderá contenido objeto de análisis, pues el contexto es pieza fundamental para cualquier AC. En nuestro ejemplo, se definió una unidad de contexto diferente para cada unidad de registro. Así, para el año de publicación se definió el resumen completo de la publicación, dado que la fecha de publicación está referida a la publicación al completo. Pero para la unidad de registro “título de la revista” se definió el conjunto de palabras que, enlazadas, definía el título de la revista donde estaba publicado el artículo. O en el caso de los diferentes temas a explorar (salud, educación, desarrollo, etc.), para los que se definió como unidad de contexto la frase (es decir, el conjunto de palabras entre dos puntos del resumen) en la que se encontrase la palabra clave ('salud', 'educación', etc.).

Las unidades de contexto no tienen por qué ser independientes unas de otras, y a menudo se superponen o incluso coinciden. Por ejemplo, también se definió el resumen como unidad de contexto para el caso de cada una de las keywords utilizadas en la búsqueda: la lógica aplicada fue que para entender por qué el buscador había marcado esa publicación como unidad muestral (es decir, objeto de nuestra investigación, con contenido de interés para cubrir nuestros objetivos) habría que tener en cuenta la totalidad del resumen, pues a priori no podríamos entender la referencia a una determinada *keyword* considerando tan sólo la frase donde se situase la expresión de búsqueda.

De este modo, la unidad de contexto es siempre, en términos de extensión, mayor o, como mínimo, igual a la unidad de registro. Cuando los datos tienen formato texto, las unidades de contexto más habituales son la frase o el párrafo. Según cada estudio y sus objetivos, los criterios para definir las unidades de análisis (de muestreo, registro y contexto), obviamente, pueden ser diferentes. Y también es obvio que la definición de las unidades de análisis es fundamental para la correcta consecución de los objetivos de investigación. La primera tarea del equipo de investigación, idealmente en la fase de diseño del estudio, es definir claramente dichas unidades y atenerse a ellas estrictamente durante todo el proceso investigador.

Tabla 10. Relación entre las unidades de análisis en el AC (ejemplo).

Unid. muestreo	Unid. de registro	Unid. de contexto
Paper 1	- Autor - Año - Título - Resumen - País de referencia del estudio - Tema 'salud' - Tema 'educación' ... - Keyword 1 ...	Apellidos, nombre del autor Resumen completo Texto completo del título Resumen completo Resumen completo Frase (de punto a punto) Frase (de punto a punto) ... Resumen completo ...
Paper 2
...

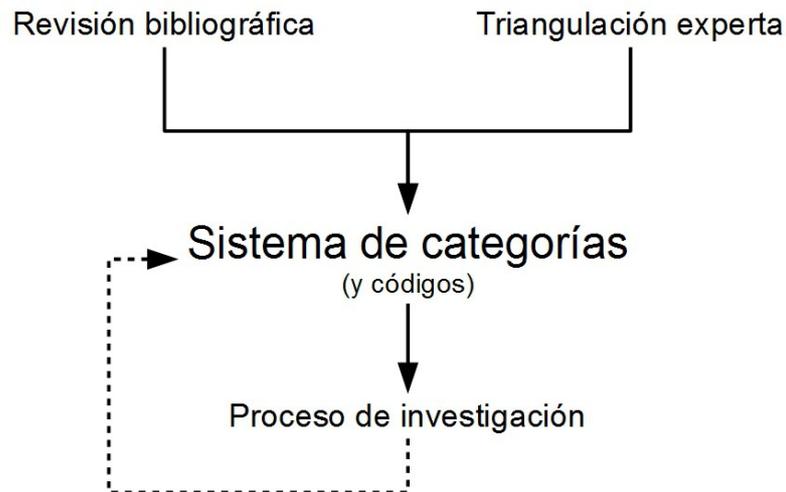
Fuente: elaboración propia.

Categorización y codificación.

Las unidades de registro suelen agruparse en categorías, bajo criterios coherentes con nuestros objetivos de investigación. A estas categorías se las define con un determinado código, por lo que se denomina proceso de **codificación**. El **proceso de registro y codificación** es un elemento crucial del AC. El investigador ha de tener en cuenta diversos factores. Por ejemplo, los codificadores deben estar familiarizados con la naturaleza del material que se va a codificar, y deben ser capaces de entender y manipular las categorías que se utilizan en la codificación y los diversos términos y formas que se emplean para registrar los datos. Es necesario establecer una estrategia de preparación o **entrenamiento de los codificadores**; o también, si es el caso, llevar a cabo reuniones con el equipo de investigación que vaya a dedicarse a la categorización y codificación, cuyos resultados queden fijados en un documento de referencia con las reglas de codificación.

El equipo de investigación ha de **desarrollar y definir un sistema de categorías**, elemento crucial del análisis. Este sistema de categorías surge, en primera instancia, del propio esfuerzo del equipo de investigación, que ha de realizar un estudio bibliográfico especializado sobre la/s temática/s que aborda su proyecto de investigación y , concretamente, las que están directamente vinculadas con los objetivos del mismo. No obstante, esta primera lista de categorías (o categorías 'a priori'), se completa durante el proceso de investigación. El AC, como decíamos arriba, es una metodología que se caracteriza por su flexibilidad; en este punto dicha flexibilidad se aplica a la posibilidad de modificar y/o completar el sistema de categorías conforme se avanza en el proceso de investigación. En cualquier caso, el sistema de categorías actualizado ha de ser conocido, comprendido y compartido por todo componente del equipo investigador que desarrolle tareas vinculadas con él.

Figura 17. Generación y retroalimentación del sistema de categorías.



Fuente: elaboración propia.

Esta flexibilidad es facilitada, además, por las herramientas informáticas de análisis. El software de referencia, Atlas.ti, ofrece la posibilidad de añadir comentarios a cada uno de los códigos que se asignan a las categorías para, a posteriori, poder obtener un reporte de comentarios de todos o de un conjunto determinado de códigos (en el argot Atlas.ti, una 'familia' de códigos). El inventario de los códigos asignados, sus explicaciones pertinentes y las relaciones que mantienen entre sí es lo que se denomina 'libro de códigos'. Atlas.ti permite, en el momento que sea necesario, obtener un informe con las diferentes citas, relaciones o cualquier otra información relacionada con los códigos. En buena medida, el éxito o fracaso de un AC depende del sistema de categorías establecido, y del modo en que se hayan definido, por lo que el libro de códigos es una herramienta clave para un análisis fiable.

Un sistema de categorías debería reunir las siguientes características:

Reflejar los objetivos de investigación, y por tanto los diferentes conceptos, dimensiones y variables implicados en los mismos.

Exhaustividad. Debe ser posible colocar cada unidad de registro en una de las categorías del sistema.

Las categorías deben ser excluyentes, es decir, sus significados deben estar claramente definidos y no pueden superponerse.

Criterios para la definición de categorías.

En el proceso de categorización y codificación propio del AC, es aconsejable el **recurso a criterios objetivos y homogéneos**, aunque la frecuente complejidad del AC requiere el apoyo en el criterio experto del equipo de investigación. Por ejemplo, es un criterio objetivo la presencia de la palabra “salud” dentro de la unidad de significado (contexto). Atlas.ti posibilita codificar automáticamente (con el código 'salud' en este caso), todas aquellas unidades de significado (frases, en nuestro caso) en las que aparezca la palabra 'salud'. Por tanto, 'salud' se asignará, de forma homogénea, a todas las frases donde aparezca.

Los criterios objetivos han de completarse con **acuerdos del equipo de investigación**. En nuestro caso, una vez codificadas todas las unidades de contexto en las que aparecía 'salud', el equipo llegó al acuerdo de asignar el mismo código también aquellas citas donde se hacía referencia a cuestiones relacionadas con la salud, siempre que la publicación en cuestión (unidad muestral) se centrara en una temática de estudio vinculada a la salud biofísica, psicológica o socio-ambiental. Este acuerdo quedó reflejado en un comentario del código 'Salud'.

Como en cualquier diseño de investigación, pueden establecerse categorías en base al estudio previo de la literatura sobre el tema a investigar. Pero en el caso del AC, la flexibilidad del diseño (propio de su adecuación a estudios exploratorios) permite la incorporación de nuevas categorías, y adaptación de las existentes, a lo largo de prácticamente todo el proceso de preparación y análisis de los datos.

En nuestro ejemplo, las categorías fueron definidas por su vinculación a los objetivos específicos, donde intervino el citado “criterio experto” del equipo de investigación. Esto significa que, a la hora de definir qué categorías corresponden, inicialmente, a cada objetivo de investigación, el trabajo en equipo de los investigadores es fundamental, porque ofrece un apoyo de **triangulación** en la definición de categorías y criterios de análisis; la triangulación, como sabemos, está entre los métodos más usados en investigación para aportar fiabilidad y validez a los trabajos. Así, se establecieron categorías como los diferentes temas sobre los que se desarrollaba la RRI, países y áreas geopolíticas donde se llevaban a cabo los estudios, años de publicación de los trabajos (análisis longitudinal), las diferentes subtemáticas en RRI (equivalentes a las keywords utilizadas en la búsqueda bibliográfica), etc. Algunas categorías definidas, con su código correspondiente, fueron “Salud”, “Educación” y otros temas. “Salud” (la palabra) sería el código asignado a una categoría que agrupa todas aquellas unidades de registro (fragmentos de texto) que incluyen la palabra 'salud'. Específicamente en este ejemplo, “Salud” sería el código asignado a cada unidad de registro, que en nuestro caso hemos hecho coincidir con la unidad de contexto (ver Figura XXXX).

Figura 18. Códigos de diferentes categorías en una cita (unidad de contexto)

<p>"In 2000, cancer health indicators for Native Hawaiians were worse than those of other ethnic groups in Hawaii, and Native Hawaiians were under-represented in research endeavors. To build capacity to reduce cancer health disparities, 'Imi Hale applied principles of community-based participatory research (CBPR) and empowerment theory. Strategies included: 1) engaging Native Hawaiians in defining cancer priorities; 2) developing culturally appropriate processes and</p>	<p>Journal Article 2008 3139 T...</p> <ul style="list-style-type: none"> ◇ Community ◇ Community-based ◇ Health ◇ Research
--	--

Fuente: elaboración propia.

Otros criterios objetivos para la categorización son los resultados que ofrezca el análisis de frecuencias de palabras, o también el análisis de coocurrencias. La **frecuencia de palabras** es un criterio objetivo en cuanto que, una vez realizada la selección de los casos de estudio (los 'datos brutos'), puede pedirse al sistema de análisis que obtenga un listado con todas las palabras (en el caso de datos en formato texto) y cuántas veces se repiten, en el conjunto de los datos, cada una de ellas. Previamente se le indica al sistema la 'lista de exclusión' a tener en cuenta, es decir, qué palabras no es necesario considerar (generalmente, preposiciones, artículos, conjunciones, etc.). A continuación, una simple jerarquización del listado nos ofrece una visión general del contenido semántico de los datos, y posibilita un filtro de objetividad para seleccionar categorías y sus códigos. Al referir un 'filtro', significamos la necesidad que a menudo encontramos en el AC de establecer límites en los datos a considerar. Y al referir objetividad queremos indicar que es el propio sistema el que, automáticamente, obtiene la frecuencia de las palabras. Podemos suponer que, cuanto más frecuente sea una palabra en ese listado, más centralidad semántica tiene en el micro-universo de significación cuyo análisis afrontamos con el AC.

Así, en nuestro ejemplo la tabla de frecuencias mostró un total de 26764 palabras diferentes, se seleccionaron como significativas tan solo 522, es decir, un 1,95% del total. Se trataba de palabras significativas con relación a los objetivos del proyecto, y fueron definidas así, discutidas y acordadas por el equipo de investigación. Estas 522 palabras mostraron una frecuencia total de 95985. Al objeto de realizar un análisis lo más eficiente posible, se optó por limitar la codificación inicial hasta los tres primeros quintiles de la distribución de palabras (criterio objetivo de filtro). Esto significa que, en primera instancia, se incluyeron en la codificación las palabras cuya frecuencia alcanzó el 60% (tres quintas partes de la distribución) del total de las palabras seleccionadas como significativas. A continuación, se procedió a definir familias de palabras afines con cada una de las categorías que, a priori, iban a encontrarse en los datos brutos. Por ejemplo, se codificaron con "Salud" unidades de registro que incluían palabras como *cuidados*, *enfermedad*, *clínico*, *médico*, etc... (y sus variantes en género y número).

El **análisis de coocurrencias** también ofrece criterios objetivos para la categorización y la relación entre las categorías. Las herramientas informáticas posibilitan la obtención de una tabla de doble entrada, en la que podemos definir, en filas y columnas, los elementos sobre los que deseamos información. En nuestro caso, para explorar posibles estructuras internas de los datos, situamos los diferentes códigos empleados en el proceso inicial (automático) de codificación tanto en filas como en columnas (matriz simétrica). En la Figura XXXX se muestra un extracto de la tabla, en columnas códigos como 2008 y 2018, en filas algunos de los temas vinculados a RRI y codificados en categorías (desarrollo, educación, etc.). Cada celda de la tabla indica: 1) la frecuencia absoluta de

citas (unidades de contexto) donde coinciden el código de filas y el de columnas; 2) el coeficiente de coocurrencia, es decir, el número de citas donde coinciden ambos en relación al total de citas de ambos códigos. Por ejemplo, a la vista del extracto, podemos afirmar que 'mujer' es un tema de estudio más frecuente en 2018 que en 2008 dentro de la literatura especializada en RRI, ya que el coeficiente de coocurrencia mujer-2018 es mayor que mujer-2008.

Figura 19. Extracto de tabla de coocurrencias.

	A	B	C	V	W
1		2008		2018	
2		count	coefficient	count	coefficient
3	Development	84	0,04	416	0,18
4	Education	52	0,04	228	0,13
5	Environment	50	0,03	357	0,18
6	Female	30	0,03	139	0,1
7	Health	150	0,05	714	0,22

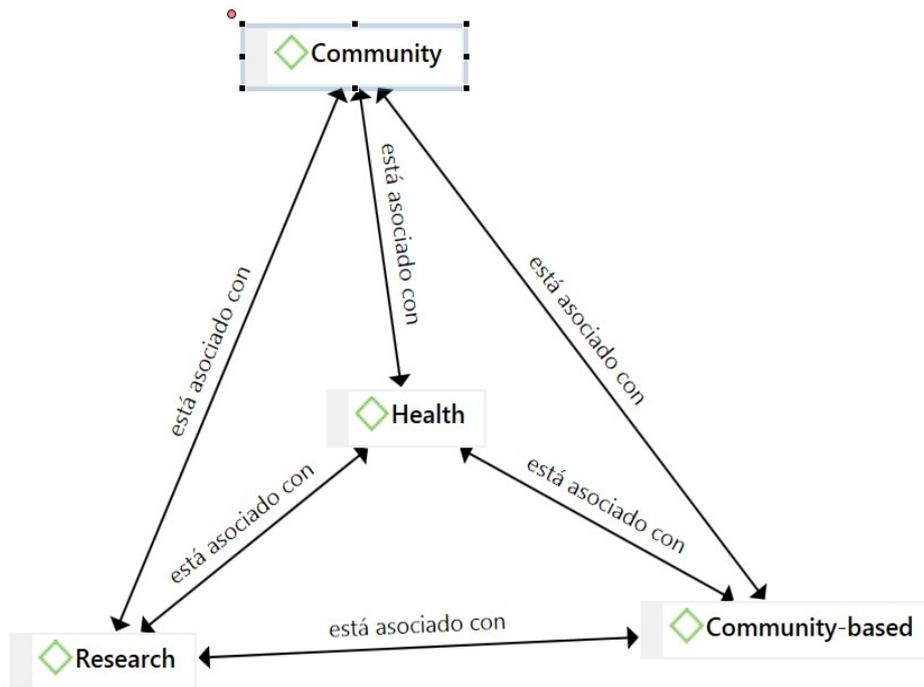
Fuente: elaboración propia.

Complementariedad de técnicas de análisis en el AC

Tanto las tablas de frecuencias de palabras como las tablas de coocurrencias ofrecen, en sí mismas, un input interesante para subsecuentes **análisis cuantitativos** en general, o estadísticos en particular. En el ejemplo anterior, los coeficientes de coocurrencia pueden ser considerados, en sí mismos, indicadores de conglomeración. Tomando otro ejemplo, podríamos mostrar en qué sentido las coocurrencias pueden emplearse como input para un **análisis estructural o de redes**. Por ejemplo, en la figura XXX podemos observar que la misma unidad de registro (fragmento marcado en azul) ha sido codificada para cuatro categorías diferentes (Community, Community-based, Health y Research). El hecho de que estas cuatro categorías 'ocurran' dentro de la misma unidad de contexto significa que están relacionadas entre sí de alguna forma, es decir, co-ocurren.

Gráficamente, podríamos simbolizar la **relación entre estas categorías mediante una red**, donde cada nodo sería una categoría (simbolizada por su código) y cada relación indicaría una coocurrencia. El software de análisis entiende la existencia o no de relación en código binario, es decir, coocurrencia (existe relación entre los códigos) = 1, no coocurrencia = 0. Obsérvese que la información “el código A se relaciona con el código B” ha sido generada mediante el hecho de la coocurrencia, con el soporte objetivo de que A y B están comprendidas dentro de la misma unidad de contexto (una frase, en esta ocasión; en la terminología del software de análisis cualitativo Atlas.ti, esa frase sería denominada 'cita').

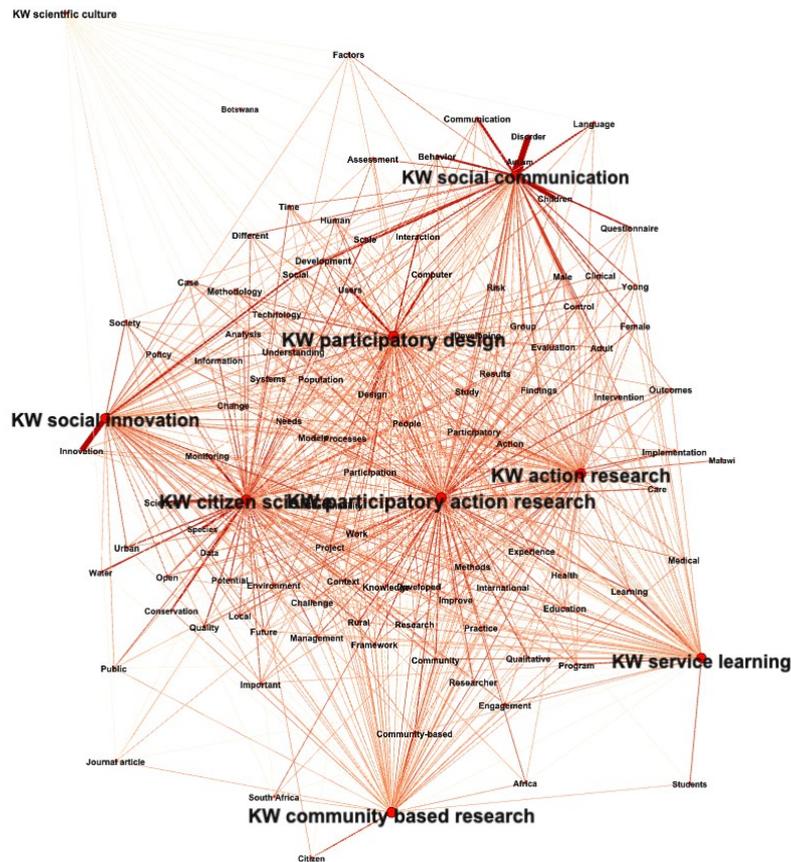
Figura 20. Ejemplo de visualización en red de códigos en coocurrencia.



Fuente: elaboración propia.

Si a este análisis estructural simple añadimos otros elementos, como por ejemplo intensidad en las relaciones entre códigos, podremos obtener análisis estadísticos y gráficos de mayor sofisticación. La intensidad en la relación entre códigos viene dada por el coeficiente de coocurrencia: cuanto más alto sea este coeficiente, se entiende que mayor intensidad tiene la relación entre ambos. De este modo, una tabla de coocurrencias puede ser, a su vez, el input para un software especializado en análisis estructural o de redes, que nos podrá ofrecer, por ejemplo, análisis estadísticos de centralidad, intermediación o cercanía; y también vistosos productos gráficos, como el expuesto en la Figura 21.

Figura 21. Ejemplo: estructura general de relaciones entre conceptos basada en coocurrencia (con Gephy)



Fuente: elaboración propia

En la literatura especializada en AC encontramos un amplio repertorio de técnicas adicionales. Podríamos clasificarlas en tres grupos, según el nivel de análisis al que recurren:

Nivel sintáctico

Análisis de la expresión. Se basa en métodos automáticos para el cómputo de distribución y frecuencias de palabras. Las diversas herramientas informáticas existentes facilitan multitud de posibilidades analíticas. Podemos citar índices e indicadores como el TTR (Type Token Ratio -número de palabras diferentes de un discurso dividido entre el total de palabras), que mide la riqueza del vocabulario, los cocientes gramaticales (ratios adjetivos / verbo, nombre + verbo / adjetivos + adverbios, ...), la longitud o la estructura de la frase, etc.

Análisis automático del discurso. Consiste en el apoyo en herramientas de software para la descomposición proposicional y la detección de relaciones de asociación, dependencia, causalidad, etc., entre las diferentes ideas que componen cada discurso social

Nivel semántico.

Análisis categorial. Es el más utilizado, y consiste en la descomposición temática de los datos en función de una estructura prevista, relacionada con los objetivos de la

investigación, espontánea (estudios exploratorios) o una mezcla de las dos. Podríamos decir que es el análisis más empleado.

Análisis evaluativo. Consiste en la detección de actitudes y valoraciones en el discurso, distinguiendo entre realidades u objetos de actitud (sustantivos o pronombres) y sus términos evaluativos (adjetivos, adverbios, verbos). Permite desde valoraciones binarias (positivo-negativo) hasta la definición de escalas de intensidad en las actitudes, o las valoraciones de los objetos de actitud.

Análisis de contingencias. Se denomina al análisis de relaciones entre los diversos elementos del discurso, que ofrezcan resultados para contrastes de hipótesis. ¿Se dan los temas X ó Y en los sujetos de estudio de tipo A ó B? ¿Son similares o diferentes las causas de los fenómenos estudiados, según A ó B? ...

Análisis discriminante. Conceptualmente, es similar a su homónimo estadístico, y se orienta al contraste de grupos de casos, de temáticas, de categorías..., según diferentes criterios, por ejemplo la fuente de los datos, las características de los informantes, etc.

Análisis estructural. Busca la organización subyacente de los datos brutos, las reglas que configuran las relaciones entre los elementos discursivos, la lógica o estructura interna de los mismos. Un reflejo gráfico de este enfoque se obtiene mediante el análisis de redes.

Análisis socio-semántico. Se orienta a desvelar la estructura social mediante la estructura semántica del discurso de los sujetos. Relaciona las personas y sus posiciones sociales mediante el análisis de la comunicación.

Nivel pragmático-instrumental

Análisis conversacional. De enfoque etnometodológico, busca comprender la organización social mediante el análisis de los recursos comunicativos de cada posición social. Por ejemplo, análisis del liderazgo mediante observación y codificación de turnos de palabra, imperativos, comunicación no verbal de respeto o imposición...

Análisis de la enunciación. Por oposición al nivel semántico, centrado en el fondo de la comunicación, el análisis de la enunciación se centra en 'la forma', en el modo en que cada emisor transmite sus mensajes

AC: resumen de procedimiento

A lo largo de estos apuntes se han descrito los diferentes pasos a dar en la realización de un AC. Se trata de unos pasos que ponen de manifiesto la naturaleza mixta de esta metodología, así como también su flexibilidad y adecuación para objetos de estudio de especial complejidad. De forma resumida, las fases generales de esta metodología podrían plantarse del siguiente modo:

1. **Diseño de la investigación.** El esquema inicial del diseño de la investigación basada en AC suele adoptar formas más parecidas a las positivistas o 'cuantitativistas', pues

se admiten, por ejemplo, procesos de muestreo bajo ciertos criterios objetivos, formulación y contrastación de hipótesis, definición de variables, etc.

2. **Muestreo.** En el caso del AC, el proceso de muestreo ha de realizarse, generalmente, sobre un conjunto de casos que ofrecen datos en un formato más propio para un enfoque metodológico cualitativo (textos, habitualmente, pero no exclusivamente). Los criterios de selección de última unidad muestral pueden incluir el azar (muestreo aleatorio de cualquier tipo) pero también la consideración de 'razonabilidad suficiente'. Del proceso de muestreo resulta el conjunto de casos objeto de estudio, que comprenden en su seno los datos brutos. En la práctica, a menudo se realiza una lectura superficial de los materiales seleccionados bajo criterios de muestreo, materiales que pueden provenir de entrevistas en profundidad, grupos focales, material gráfico o audiovisual, etc. En cualquier caso, la selección definitiva de materiales (que serán las unidades muestrales), respetará criterios de exhaustividad (todos los objetivos podrán cubrirse con la muestra de casos que sea seleccionada), representatividad (en el caso en que deseemos una muestra estadísticamente representativa del universo poblacional, con los requisitos que le son de aplicación), homogeneidad (empleo de los mismos criterios para seleccionar cada unidad muestral) y pertinencia (todas ellas se referirán al menos a uno de nuestros objetivos de investigación).

3. **Definición de unidades de análisis y vinculación a categorías.** Para la definición de la base de datos a analizar, hay que definir las diferentes unidades de análisis (casos de estudio -unidades muestrales-, unidades de registro y unidades de contexto). Dicha definición supone tomar decisiones en base a diversos criterios, que pueden ser de carácter objetivo o 'subjetivo' (cualitativo). Los subjetivos suelen vincularse a decisión experta bajo método de triangulación (habitual en ambos enfoques, pero principalmente en el cualitativo). Los objetivos se basan en características numéricas que ofrece el software de análisis, como por ejemplo la jerarquización de importancia de conceptos según frecuencia de palabras, y el empleo de dicha jerarquía para definir categorías de análisis. Por otra parte, ambos tipos de criterios suelen necesitarse entre sí en pos de un análisis posterior más ajustado a la realidad. Por ejemplo, la codificación automática, con instrucciones al sistema sobre la/s palabra/s o expresión/es a localizar en los datos, la unidad de contexto y el código a asignar, se simultanea con el conocimiento y comprensión contextual detallado, por ejemplo en las rondas sucesivas de codificación, para añadir palabras de la misma familia a un determinada categoría, o para expandir o contraer las unidades de contexto, a la vista del contenido y significado real de las mismas. El proceso de definición de categorías de análisis, y su correspondiente codificación, está entre las claves del éxito de la investigación basada en AC.

4. **Análisis de datos y obtención de resultados.** El repertorio de técnicas de análisis abarca prácticamente la totalidad de las habituales en ciencia social: análisis categorial, mapas cognitivos, ejemplificación..., pero también análisis estadístico -descriptivo e inferencial-, estructural o de redes. Por lo tanto, los resultados que podamos obtener son, también, de tipo mixto. Según necesidades, podríamos centrarnos más en resultados cuantitativos (por ejemplo, para la contrastación estadística de hipótesis) o cualitativos (por ejemplo, mapas cognitivos significantes de una estructura subyacente de ideas en el conjunto de los datos).

Bibliografía básica

- Cruz Beltrán, F., Gualda, E. (Coord.) (2001). *Métodos y técnicas de investigación social*. Huelva: ESEIS Estudios Sociales E Intervención Social. Pp. 233-270.
- Ferrando, M. G., Ibáñez, J., & Martín, F. A. (2015). *El análisis de la realidad social: métodos y técnicas de investigación* (5ª). Alianza Madrid. Pp. 594-633

Bibliografía complementaria

- BARDIN, L.: Análisis de Contenido, 1977, Ed. Akal, Madrid.
- SIERRA BRAVO, R.: Análisis de Contenido. Origen, noción e importancia (pag.286-291) en Técnicas de Investigación Social, 1992, Ed. Paraninfo, Madrid.
- CLEMENTE, M. y SANTALL, Z.: El documento persuasivo. Análisis de contenido y publicidad, 1991, Ed. Deusto, Bilbao.
- NAVARRO, P. y DÍAZ, C.: Análisis de contenido (pag.177-224) en Métodos y Técnicas cualitativas de investigación en ciencias sociales (coord: Delgado, J.M. y Gutiérrez, J.), 1994, Ed. Síntesis, Madrid.
- WEBER, R.P.: Basic content analysis, 1991, Sage Publications, California