

Epistemología, metodología y técnicas del análisis de contenido

José Luis Piñuel Raigada

Universidad Complutense de Madrid

Departamento de Sociología IV
Facultad de CC. de la Información
Universidad Complutense de Madrid
Ciudad Universitaria, s/n
28040 - Madrid, Spain
pinuel@ccinf.ucm.es

Title: Epistemology, methodology and content analysis techniques

Abstract

It is exposed, first, an epistemological review of the analysis of content, as scientific practice in social sciences that studies an object (v.g. a singular communication) for whose representation (scientific and refutable therefore) it is used the selection procedure and analysis of communicative products that belong objectively to that communication that are significant to know it from a theoretical perspective, and that they are valid or pertinent to put on approval the conditions, so much general as matters, for those that these communicative products have been elaborated, or that they can be it, in any communication type similar to the one that has produced them. In second place the method is examined (the procedure rules) that aspires to tie in a valid way the succession of performances that concern to the selection of the corpus, to its segmentation in analysis units and to its exam to elaborate, to register and to treat data of diverse logical orders, thanks to which it is elaborated a speeches able to represent scientifically (or in a refutable way) to the study object that is not the own corpus, but the human practice that has engendered the material object of the analysis: the analyzed communicative product. In third place, the last section exposes the technical alternatives that underlie to any analysis procedure, so much if it is carried out in all its steps by the investigators and analysts, as if it is carried out automatically by the available resources of software, and shortly described following the work of Harald Klein.

Key words: content analysis, communication theory, epistemology of knowledge, sociolinguistics, psycholinguistics.

Resumo

Neste traballo expoñemos, en primeiro lugar, unha revisión epistemolóxica da análise de contidos, que constitúe unha práctica científica das ciencias sociais que estudia un obxecto –unha comunicación singular– que representamos, de xeito científico e refutable, mediante os procedementos de selección e análise de produtos comunicativos que pertencen obxectivamente á comunicación, que cómpre coñecer desde un punto de vista teórico e que son válidos ou pertinentes para poñer a proba as condicións, para os que tales produtos comunicativos foron elaborados, en calquera tipo de comunicación similar á que os produciu. En segundo lugar, examinamos a metodoloxía –as regras de procedemento– coa cal aspiramos a unir, de modo válido, a sucesión de actuacións que compete á selección do corpus, á súa segmentación en unidades de análise e ó seu exame para elaborarmos, rexistrarmos e tratarmos datos de diversas ordes lóxicas. Gracias a iso somos quen de elaborar un relato apto para representar cientificamente –ou dun xeito refutable– un obxecto de estudio que non é o propio corpus, senón a práctica humana que enxendrou o obxecto material da análise: o produto comunicativo analizado. En terceiro lugar, na derradeira sección expoñemos as alternativas técnicas que subxacen no procedemento de análise, tanto se este está sendo conducido polo investigador e analista en tódolos pasos precisos coma se está sendo desenvolvido automaticamente mediante o software dispoñible, tal como se describe seguindo o traballo de Harald Klein.

Palabras clave: análise de contido, teoría da comunicación, epistemoloxía do coñecemento, sociolingüística, psicolingüística.

1. Epistemología del análisis de contenido

Se suele llamar *análisis de contenido* al conjunto de procedimientos interpretativos de *productos comunicativos* (mensajes, textos o discursos) que proceden de procesos singulares de comunicación previamente registrados, y que, basados en técnicas de medida, a veces *cuantitativas* (estadísticas basadas en el recuento de unidades), a veces *cualitativas* (lógicas basadas en la combinación de categorías) tienen por objeto elaborar y procesar datos relevantes sobre las condiciones mismas en que se han producido aquellos textos, o sobre las condiciones que puedan darse para su empleo posterior. El análisis de contenido, de hecho, se convirtió a finales del siglo XX en una de las técnicas de uso más frecuente en muchas ciencias sociales, adquiriendo una relevancia desconocida en el pasado a medida que se introdujeron procedimientos informáticos en el tratamiento de los datos.

Su propia denominación de análisis de “contenido”, lleva a suponer que el “contenido” está encerrado, guardado –e incluso a veces oculto– dentro de un “continente” (el documento físico, el texto registrado, etc.) y que analizando “por dentro” ese “continente”, se puede desvelar su contenido (su significado, o su sentido), de forma que una nueva “interpretación” tomando en cuenta los datos del análisis, permitiría un *diagnóstico*, es decir, un nuevo conocimiento (*gnoscere* “conocer”) a través de su penetración intelectual (*dia*, en griego es un prefijo que significa “a través de” en el sentido de “atravesar”). Pero una tal suposición, semejante, de hecho, a la acostum-

brada cuando los análisis químicos y biológicos exploran componentes de la materia y facilitan diagnósticos, lleva a una visión falsa. El análisis de contenido, aplicado a “continentes” materiales, busca fuera, no dentro, pues las dimensiones de los datos extraídos del análisis sólo existen fuera de los “continentes”, es decir, en la mente de los sujetos productores o usuarios de los mensajes, textos, discursos, o documentos que se analizan, es decir, en la mente de los participantes de los procesos singulares de comunicación en los que se han producido los documentos analizados.

No obstante, varios hitos en la evolución del análisis de contenido permiten comprender el desarrollo del análisis de contenido en el ámbito sociológico; son los siguientes hitos de la investigación, reproducidos por López Aranguren (y recogidos en Piñuel & Gaitán, 1995: 516):

- 1.- Décadas de 1920 y 1930. Análisis sobre el concepto de “estereotipo social” sugerido por Lippmann (1922) y sobre el concepto de “actitud”, de reciente aparición en psicología.
- 2.- Segunda Guerra Mundial. Análisis de Lasswell y colaboradores, sistematizados más tarde por George (1959): conceptualización de objetivos y procesos de comunicación.
- 3.- Décadas de 1950 y 1960. Codificación manual. Obras de Pool (1959), Lasswell et al (1965) y Holsti (1969).
- 4.- Aplicación de la informática, codificación electrónica automatizada, a partir de la obra de Stone y colaboradores (1966) sobre el sistema *General Inquirer* para el análisis contenido de mensajes por medio de ordenador. (Ver García Ferrando, M., Ibáñez, J. & Alvira, F., 1986).

Todos estos hitos revelan que el análisis de contenido se ha remitido a técnicas para elaborar, registrar y tratar datos más allá del ámbito social y cognitivo concretos en que se recurre a la comunicación, y dentro de ella, a la producción de textos. Es como si fuese justificable elaborar, registrar y tratar datos de un texto asirio porque creemos conocer su gramática, sin tomar en cuenta cuál podría ser su interpretación por parte de quienes generaron y usaron comunicativamente aquel texto. En todos los casos, sin embargo, los textos remiten a un universo de prácticas sociales y cognitivas donde cualquier interacción comunicativa cobra unos límites particulares que son mediados y mediadores de aquellas prácticas a las que sirve. Hoy día, cuando ya nadie duda que el análisis de contenido pueda sustrarse a los supuestos teóricos de los que, en cada caso, se parte, conviene además aprovechar su tradición para no perder la perspectiva en análisis de la comunicación; la técnica del análisis de contenido no producirá interpretaciones relevantes de los datos si éstos mismos no son relevantes para conocer condiciones de comunicación; y estas condiciones son de naturaleza sociocognitiva.

2. El significado del texto: contexto comunicativo y sentidos social y cognitivo

El análisis de contenido no debe perseguir otro objetivo que el de lograr la emergencia de aquel sentido latente que procede de las prácticas sociales y cognitivas que instrumentalmente recurren a la comunicación para facilitar la interacción que subyace a los actos comunicativos concretos y subtiende la superficie material del texto. Como señala Bardin (citado en Piñuel & Gaitán, 1995) el análisis de contenido se convierte en una empresa de des-ocultación o re-velación de la *expresión*, donde ante todo interesa indagar sobre lo escondido, lo latente, lo no aparente, lo potencial, lo inédito (lo no dicho) de todo mensaje. Pero esta nueva perspectiva no sólo amplía el campo de estudio del análisis de contenido hacia la dimensión no manifiesta del texto cuanto que, dada su complejidad, exige introducir nuevas variables en el análisis a fin de que el texto cobre el sentido requerido para el analista. Esto sólo es posible si tal texto se abre –teóricamente hablando– a las condiciones contextuales del producto comunicativo, al proceso de comunicación en el que se inscribe, y por tanto a las circunstancias psicológicas, sociales, culturales e históricas de producción y de recepción de las expresiones comunicativas con que aparece.

Los actores de una interacción comunicativa nunca se enfrentan normalmente por primera vez a la tarea de comunicarse con otro ser humano. Poseen, en este sentido, un determinado conocimiento acerca de las posibilidades expresivas del cuerpo humano (por ejemplo, si se trata de comunicación interpersonal cara a cara), acerca de la significatividad de ciertas expresiones y acerca de las pautas de acción y coordinación de acciones y operaciones que regulan una clase de interacciones comunicativas. El “saber hacer” acerca de la comunicación es un patrimonio cognitivo, es decir, es un producto de la práctica comunicativa, que se almacena en la memoria de los sujetos y que se activa cuando estos entran a tomar parte de una interacción comunicacional en forma de “esquemas previos” (“auto-referencia”), que guían la construcción recursiva de nuestras “representaciones e inferencias” (Piñuel & García-Lomas, 2001). Los esquemas que poseemos acerca de los demás (esquemas sociales), contienen datos que no sólo se refieren a las propiedades identitarias del sujeto (su edad, su sexo, su aspecto físico, su personalidad, su familiaridad –si es conocido o desconocido– su nivel de conocimientos, etc.), sino también a la posición que ocupan (estatus) y a la función que desempeñan (rol) dentro de los grupos y organizaciones sociales, es decir, en su calidad de agentes de un sistema social. Esos datos pueden hacer que nos comportemos de una u otra manera¹.

¹ Por ejemplo, no nos comportamos de la misma manera con un dependiente de un comercio al que acabamos de conocer (del que esperamos que cumpla ciertas funciones y se comporte de acuerdo a su rol), que con un amigo de toda la vida, ni tampoco esperamos que ellos se comporten de la misma manera.

En otros lugares he desarrollado con muy distintas aplicaciones un modelo teórico² que prefigura estas diferentes dimensiones (comunicativas, cognitivas y sociales) las cuales se encuentran analíticamente separadas porque tal operación formal es epistemológicamente rentable para encontrar una explicación científica de los cambios históricos de la sociedad, el conocimiento y la cultura. Sin embargo en el funcionamiento normal de nuestro sistema cognitivo, no existen tres procesadores modulares diferentes: los datos de tipo identitario, social y comunicacional, cuando se refieren a nuestros congéneres, forman estructuras unitarias de conocimiento que han recibido el nombre en psicología cognitiva de “esquemas sociales”.

Los “esquemas”, en tanto que estructuras de conocimiento, son estructuras de categorías o variables de naturaleza “difusa”, es decir, sin límites exactos y absolutos, cuya relación estructural impone una cierta covariación a los valores de cada una de las variables o categorías que constituyen específicamente un esquema determinado, dentro de un amplio y flexible umbral de valores posibles. Así, nuestro esquema de “hotel”, por ejemplo, puede arrojar múltiples representaciones o variantes concretas, pero al variar uno de los constituyentes tienden a covariar los valores restantes. Por ejemplo, un hotel de lujo y una pensión barata tienen en común ciertas variables como es el hecho de tener habitaciones y tener un precio. Al variar el precio hacia la baja, las calidades de la habitación también “bajan”. Ciertamente, existirá un límite hacia la baja y otro hacia la alta, pero el umbral es amplio y es históricamente flexible. Lo interesante en este sentido es que existe una tendencia hacia la covariación, y por tanto es difícil encontrar pensiones baratas con calidades y servicios de lujo, como es difícil encontrar hoteles carísimos con calidades y servicios extraordinariamente “bajos”. En el centro podemos obtener sin embargo hoteles

² Se trata del modelo de la MDCS (Mediación Dialéctica de la Comunicación Social) el cual formula un sistema conceptual especializado en la descripción, explicación y predicción de los cambios que experimentan o pueden experimentar los sistemas de comunicación propios de nuestras sociedades. Entre sus principales postulados o hipótesis de base se encuentra la idea de que dichos cambios no pueden conocerse si no se establece primero una clara diferenciación entre los componentes y relaciones que son específicos de los sistemas de comunicación (SC) y los componentes y relaciones que son específicos de otros sistemas. Una vez efectuada esa diferenciación, el modelo postula que los cambios de los sistemas de comunicación (SC) sólo pueden explicarse examinando las relaciones de apertura que dicho sistema mantiene con otros dos sistemas generales: el sistema social (SS) y el sistema ecológico-adaptativo (SE), cuyo específico inventario de componentes genéricos y de relaciones internas (estructurales y funcionales) es igualmente proporcionado por dicho modelo (cfr. Martín Serrano et al., 1981; Martín Serrano, 1989, pero especialmente Piñuel, 1989, y Piñuel & Gaitán, 1995). La concepción de los sistemas de comunicación (SC), de los sistemas sociales (SS) y de los sistemas ecológico-adaptativos (SE) como sistemas abiertos entre sí, nos permite describir la transformación de cada uno de esos tres sistemas en términos de una mediación dialéctica “intersistémica”, es decir, concibiendo la transformación interna de un determinado sistema como el producto de la influencia externa que ejercen sobre él los sistemas restantes y viceversa.

relativamente baratos con servicios y calidades buenos, o pensiones relativamente caras con servicios y calidades relativamente bajos, etc. Similarmente podríamos razonar para muchos de nuestros esquemas identitarios: el esquema de “catalán”, o el esquema de “vasco”, o de “español”, etc. ofrecen múltiples representaciones o variantes concretas (“el industrial catalán”, vs. “el payés catalán”; o el “industrial vasco” vs. “el paisano de los caseríos”; etc.) e imponen una cierta covariación: al variar uno de los constituyentes tienden a covariar los valores restantes.

Las meta-representaciones que construimos específicamente acerca de nuestro interlocutor, en cada proceso comunicativo, se encuentran guiadas por los conocimientos sociales previos, por nuestros “esquemas sociales”. Los esquemas activados y las meta-representaciones referidas a las intenciones, creencias y estados mentales de nuestro interlocutor (teorías de la mente), juegan un papel fundamental en la regulación de nuestra propia conducta, pues anticipan una expectativa acerca de los límites específicos de esa interacción, o para decirlo con otros términos: colaboran en la formación del sentido mismo de la interacción. Lógicamente si los esquemas sociales que activamos se corresponden a la realidad, es decir, si el sujeto que tenemos enfrente se comporta dentro del margen de variables previstas, entonces las posibilidades de acoplamiento interactivo de nuestras acciones son mayores. Pero para que el acoplamiento interactivo sea completo, también la otra persona (alter), debe activar esquemas sociales y meta-representaciones acerca de nosotros mismos y de nuestro comportamiento, que le permitan formar una expectativa que efectivamente se ajuste a la realidad. Ahora bien, también es posible “falsificar” el comportamiento y asumir un “esquema social” que no se adecua con el que realmente nos corresponde, como es igualmente posible “confundirse” y atribuir a los demás una identidad social (activar un esquema social inadecuado) que no es el que se ajusta a la realidad.

Tanto si se consideran situaciones interpersonales de interacción, del tipo de encuentros sociales “cara a cara” para los que siempre podríamos encontrar guiones (como por ejemplo el encuentro social entre un vendedor y el cliente en unos almacenes –donde dominan patrones sociales– o el encuentro sexual –donde dominan patrones afectivos y eróticos–, etc.), que son propios de una perspectiva “microsociológica”, como si aumentamos la escala y asumimos una perspectiva macrosociológica, es evidente que no podemos penetrar con el mismo detalle o granularidad en los esquemas y meta-representaciones individuales, y que tendremos que utilizar algún método que permita describir el patrimonio social y no ya meramente individual de esquemas cognitivos previos o disponibles.

Para describir ese patrimonio cognitivo es necesario recurrir, a mi entender, a un nuevo concepto: el concepto de “socio-esquema” (Piñuel & García-Lomas, 2001). Este concepto es una síntesis emergente entre el concepto de “modelos de representación” y el concepto de “esquema cognitivo”. Por esta razón, en el concepto sintético de “socio-esquema” se salva la idea de “modelo” en el sentido de que podemos des-

cubrir los límites de los “esquemas” cognitivos disponibles para un grupo social. Y este es un capital sociocognitivo que es previo, tanto para establecer entre los interlocutores acuerdos metacomunicativos capaces de integrar cualquier intercambio de expresiones en la conducta social y cognitiva de los sujetos, como para establecer variables de análisis de los discursos que el analista aborda sobre relatos de los mass media, por ejemplo. Si el acuerdo que los sujetos construyen deja de ser contingente a cada situación, y se torna en transcendente (Ibáñez, 1986), es decir común al grupo y a una comunidad, es porque las prácticas sociales y cognitivas se reproducen según patrones de sentido, que son los que están en la base del discurso social, el cual a su vez no tendría sentido sin grupo, sin comunidad y sin comunicación social.

Por lo dicho, puede ya comprenderse la definición de análisis de contenido, ya adelantada: “al conjunto de procedimientos interpretativos de *productos comunicativos* (mensajes, textos o discursos) que proceden de procesos singulares de comunicación previamente registrados, y que, basados en técnicas de medida, a veces *cuantitativas* (estadísticas basadas en el recuento de unidades), a veces *cualitativas* (lógicas basadas en la combinación de categorías) tienen por objeto elaborar y procesar datos relevantes sobre las condiciones mismas en que se han producido aquellos textos, o sobre la condiciones que puedan darse para su empleo posterior”. Para que aquel conjunto de “procedimientos interpretativos” puedan sostenerse como estrategias y técnicas de investigación científica en comunicación, se suele requerir la elaboración previa de un repertorio estructurado de categorías derivadas de un marco metodológico en que se fija como objeto de estudio la comunicación. De este marco metodológico derivan las hipótesis y objetivos que sostienen el procedimiento de normalización de la diversidad superficial del corpus textual o material de análisis, con vistas al registro de los datos, a su procesamiento estadístico y/o lógico y a su posterior interpretación. En cualquier caso, el análisis de contenido ha de entenderse como un metatexto resultado de la transformación de un texto primitivo (o conjunto de ellos) sobre el que se ha operado aquella transformación para modificarlo (controladamente) de acuerdo a unas reglas de procedimiento, de análisis y de refutación (metodología) confiables y válidas, y que se hayan justificado metodológicamente.

3. Metodología del análisis de contenido

De acuerdo con aquellos planteamientos se puede entender que un análisis de contenido incluiría necesariamente los siguientes pasos:

- a) selección de la comunicación que será estudiada;
- b) selección de las categorías que se utilizarán;
- c) selección de las unidades de análisis, y
- d) selección del sistema de recuento o de medida

Con el objeto de ofrecer un perfil lo más aproximado posible de esta metodología, cabe elaborar una reflexión provisional sobre la metodología del análisis de contenido según los procedimientos previos que los caracterizan, aunque sin adentrarse en las combinaciones, intersecciones o mezclas, que se puedan dar entre ellos.

3.1. Análisis de contenido según la selección de la comunicación estudiada

Es obvio que, por ejemplo, no puede plantearse de la misma forma un análisis de contenido del discurso producido por un grupo de discusión, cuyos integrantes conversan sobre un tema, que un análisis de contenido de la correspondencia que mantienen dos enamorados, o que un análisis de contenido de relatos mediáticos. Incluso para los ejemplos citados pueden advertirse formatos y situaciones muy diversas; no se da la misma situación en la discusión de grupo cuando éste se constituye como técnica cualitativa de investigación social, que cuando se trata de un grupo de trabajo; a su vez, los grupos de trabajo adquieren condiciones sociales y cognitivas muy diversas según los efectos que social o cognitivamente se esperen de la discusión: toma de decisiones socialmente vinculantes (por ejemplo, una reunión formal de departamento universitario sobre distribución de tareas), o toma de decisiones psicológicamente vinculantes (por ejemplo, grupo terapéutico para dejar de fumar). Tampoco se dan los mismos formatos según las situaciones; por ejemplo, en los relatos mediáticos, los formatos son muy diversos según se trate, por ejemplo, de publicidad, de espectáculos o de información de actualidad, hasta el punto de que cualquier individuo adulto de nuestras sociedades dispone de un capital de socioesquemas cognitivos en virtud de los cuales discrimina variables en los flujos mediáticos (v.g. de la TV) que le permiten, con una velocidad vertiginosa cuando “zapatea” con el telemando, poner en juego creencias y teorías de la mente gracias a las cuales atribuye sentidos a lo que aparece en pantalla: “quieren que conozca y sienta la necesidad de comprar un producto” (caso de la publicidad), o “quieren que conozca y valore un acontecimiento ocurrido en mi entorno social” (caso de un informativo), o “quieren que contemple y sienta alguna emoción al ver lo que son capaces de hacer personajes reales (deportistas, músicos, etc.) de un espectáculo, o personajes de ficción en un escenario construido ad hoc” (caso de los programas de entretenimiento), etc. (cfr. Piñuel & García-Lomas, 2001).

Según los objetivos de la investigación aplicada a esos formatos y situaciones de comunicación, pueden también diferenciarse análisis de contenido de carácter *exploratorio*, *descriptivo* y *verificativo* y *lo explicativo*. Los análisis *exploratorios* sólo tienen por objeto una aproximación al diseño definitivo de una investigación en la que el análisis de contenido sea una técnica elegida para elaborar, registrar y tratar datos sobre “documentos” en el sentido que antes hemos explicado. Estos análisis *exploratorios*, denominados *pre-test*, generalmente se desarrollan para testar por primera vez un material de estudio, de modo que pueda determinarse la elección del

corpus de documentos de que se dispone para plantearse un problema de análisis en torno a situaciones que han producido aquellos textos, la señalización de categorías relevantes para aquel estudio de procesos comunicativos y la prueba de los indicadores de medida y de criterios más adecuados al corpus, o para la delimitación muestral de éste. Fundamentalmente estos análisis se orientan, pues, a resolver los problemas relativos a la elaboración de las categorías más pertinentes que han de configurar un protocolo de análisis y su eventual articulación (ver infra), de modo que resulten productivas de cara a la explotación de los datos³.

Los análisis *descriptivos* tienen por objeto, en un marco de estudio dado, la simple identificación y catalogación de la realidad empírica de los textos o documentos, mediante la definición de categorías o clases de sus elementos. Como uno de los análisis de contenido descriptivos más específico puede citarse el *análisis documental* (o de recuperación de información) con muchas variantes.

Aparte de la identificación intra-textual de la estructura que adopta el producto comunicativo, lo más interesante es partir de los contextos donde se inscribe el material de análisis, es decir, de su estructura de producción o de recepción comunicativa, social y culturalmente integrada: así, por ejemplo, descripción de las características de las situaciones, entornos y contextos implicados en su producción y recepción (desde los históricos a los personales, desde los públicos a los privados), o también descripción de la estructura sociológica y psicológica que rige la conducta de los individuos, grupos o instituciones emisoras y receptoras en los procesos de comunicación que dieron origen a esos productos.

Finalmente, sólo los análisis *verificativos* y explicativos pueden dar cuenta de inferencias sobre el origen, naturaleza, funcionamiento y efectos de los productos comunicativos: v.g. extrapolaciones como, al estudiar el contenido, inducir una predicción (inferencias inductivas), o estudiando un efecto, comprobar un contenido, o, estudiando un producto, comprobar los antecedentes de la comunicación (inferencias deductivas). A veces algunos análisis se centran en las inferencias que se realizan a partir del estudio de un tipo de productos comunicativos sobre otros productos comunicativos, o de los resultados de otros estudios sobre el estudio en curso. Pero en este caso hay ya que entrar en la consideración de los tipos de análisis de contenido según la selección de las categorías que se utilizan.

³ Cuando se requiere una clarificación de estas categorías y no existen datos suficientes para establecerlas a priori, los análisis exploratorios a veces se confían a grupos que, a través de otras técnicas como el *delphi*, permiten que sean expertos o involucrados en el campo objeto de estudio, los que contribuyan a perfilar o depurar la interpretación de los documentos y, por ella, la estructura categorial de un protocolo de análisis que servirá propiamente hablando para abordar los análisis de contenido definitivos.

3.2. Análisis de contenido según la selección de las categorías

Ya decía Berelson, su padre fundador, que cualquier análisis de contenido se sostiene o se hunde por cuáles sean sus categorías de análisis.

La selección de la comunicación a estudiar permite enmarcar fenomenológicamente el objeto material de análisis (y en su virtud, también las fuentes del material que haya de configurar el corpus de estudio⁴), pero la selección de categorías construye la “mirada” del objeto.

Las categorías siempre derivan de las miradas, o lo que es más preciso, de las representaciones que permiten la mirada del objeto de análisis. La vigencia de estas representaciones en el conocimiento vulgar, comparada con la del conocimiento científico, reside en una confianza contingente que va pegada al conocimiento particular de una actividad o acontecer particular, (v.g. “cuando llueve y hace sol sale el arco del Señor”). La vigencia de sus representaciones en el conocimiento científico reside en la transcendencia de su refutabilidad, más allá del propio conocimiento de una actividad o acontecer particular (por ejemplo, cuando un haz luminoso incide en la superficie de separación de dos medios ópticos distintos, parte de él se refleja en tanto que otra parte se propaga a través del segundo medio óptico, si bien con velocidad distinta de la incidente; los rayos solares, al incidir sobre las gotas de agua en suspensión en la atmósfera, se reflejan en parte sobre la superficie de las mismas y en parte se refractan, dispersándose en el interior de éstas según diferentes longitudes de onda que son los colores del espectro, etc.). Pero en la investigación básica la refutabilidad concierne a todo el capital de representaciones cognitivas de que se dispone para extraer, a medio o largo plazo, posibles prácticas (por ejemplo, las leyes matemáticas de la refracción y dispersión de la luz en física facilitaron tanto la explicación del arco iris, como la construcción de prismas y sus aplicaciones ópticas). Es decir, en la investigación básica la refutabilidad concierne a las teorías, mientras que en la investigación aplicada, la refutabilidad concierne sólo a las prácticas. En su virtud, los objetos de estudio, para la refutación, adquieren singularidad si son representativos, y si no la pierden por su forma de elaboración. En consecuencia, el análisis de contenido nunca puede ser independiente, cuando se aplica, de una teoría

⁴ Éstas pueden diferenciarse en fuentes *primarias*, *secundarias* y *terciarias*. Como fuentes primarias se pueden haber recabado cartas, diarios, libros, publicaciones periódicas, documentos históricos, discos, y otros registros sonoros, visuales y audiovisuales como cassettes, fotografías, películas, vídeos, etc. Pero a partir de estos materiales que constiuyen fuentes primarias, otros textos de segundo orden o derivados de los originales (transcripción, traducción o interpretación de las fuentes primarias) pueden constituir también fuentes apreciables para la comparación. Podría también darse un tercer tipo de fuentes, las terciarias, o sea, las de aquellos materiales que son el producto de un proceso de experimentación comunicacional en el que se introducen y/o se observan variables controladas experimentalmente, con el objeto de constituir discursos a partir de interacciones entre los participantes, (por ejemplo, reuniones de grupo, laboratorios de teatro, y aquellas técnicas etnográficas y psicosociológicas –de dinámica de grupos– o dramatúrgicas, etc., propias de la conocida como Investigación-Acción-Participativa –IAP).

sobre la comunicación que ha originado los textos que se estudian o analizan, ya sea una teoría social, comunicativa, o psicológica, etc. Las unidades para el análisis adquieren entonces una refutabilidad que procede de la teoría en virtud de la cual se han decidido cuáles sean sus categorías.

En análisis de contenido, concretamente, pueden ser objetos del estudio productos singulares de comunicación social, ya hayan sido obtenidos de procesos generales de comunicación (por ejemplo, ejemplares de periódicos impresos y distribuidos, o informativos emitidos y grabados en Medios de Comunicación de Masas –MCM), y/o recabados de sistemas particulares de comunicación (por ejemplo, los “comunicados de prensa” efectivamente distribuidos y los publicados y/o citados según los “medios”, en interés de una organización); ahora bien, productos singulares pueden ser objetos del estudio sólo si son refutables para el diagnóstico etiológico y/o prospectivo de los procesos comunicativos a que pertenecen. De lo contrario, pierden su refutabilidad. Este hecho está en la base de las operaciones lógicas y estadísticas de selección muestral: el objeto de estudio es el que marca cuál es el universo del que los productos singulares pueden ser representativos o no serlo.

Resumiendo, la naturaleza singular de los datos de investigación (elaborados siempre, no encontrados en estado puro) y de los procesos que guían su elaboración, registro y tratamiento posterior, se da en función de la previa fijación de objetivos (fines) de conocimiento y práctica, y de la delimitación de objetos de estudio adecuados a ellos, significativos o representativos, y pertinentes para su refutabilidad.

3.3. El análisis de contenido según el diseño de análisis del objeto de estudio

La pertinencia de acudir a situaciones espontáneas y a la elaboración de datos mediante disección (o análisis) de productos singulares de comunicación social (por ejemplo, el análisis de contenido) procede de refutar la adecuación, significatividad y validez de la singularidad de los productos respecto a aquellas situaciones.

No hay que olvidar que en aquellas situaciones de comunicación que brindan la posibilidad de elaborar, registrar y después procesar datos a partir del tratamiento de los productos singulares guardados, grabados o conservados (ejemplares de periódicos puestos en circulación, programas de radio o TV que han sido difundidos, etc.) pertenecientes a aquellas situaciones, requieren una previa disección de los productos orientada a su vez a elaborar, registrar y tratar datos extraídos de ellos; ahora bien, es obvio que no todos los datos que pueden ser elaborados, registrados y tratados, a partir de la disección de productos comunicativos o “textos”, resultarán adecuados, y serán significativos y suficientes, para representar científicamente la situación comunicativa que integra supuestamente el acontecer socialmente previsto en que se recurre a la comunicación, y que ha de constituir el objeto científico de análisis; por ejemplo, este texto que estamos escribiendo y que en su momento será impreso en el número de la revista que la editorial hará circular, para que los lectores puedan compartir con nosotros algu-

nas ideas de metodología, gracias a la comunicación escrita y comercialmente difundida en el mercado de librerías, puede ser guardado, diseccionado en sus elementos físicos, gráficos, etc.; y gracias a una operación de disección, (como por ejemplo, separar las palabras que contiene), elaborar como dato, v.g., un número de ellas; en función de cómo se trate luego este dato, los registros serán diferentes (como veremos más adelante), pero también será relevante lo que se hace con ese número; este número puede servir para establecer una cantidad correspondiente de gasto (pagos al corrector tipográfico, o al de estilo), o puede servir como un total respecto al cual calcular la frecuencia de sustantivos y verbos (palabras llenas) comparándola con la frecuencia de determinantes y partículas (palabras vacías) que contiene, en el supuesto de que esta comparación representara un índice de viveza literaria en el estilo y que se supone igualmente propio de cada autor; es obvio que un mismo dato (número de palabras) se somete a una operación por la cual se hacen cálculos que permiten extraer nuevos datos que sólo serán adecuados, si están bien hechos, pero que sólo serán significativos y suficientes por su relevancia para representar un objeto que ya no es el propio texto, sino el proceso singular de comunicación del que el texto es su producto material. Evidentemente, el dato de “segundo orden” referido al gasto editorial en función del número de palabras, no remite directamente al proceso de comunicación que facilita la interpretación cognitiva y social del texto, sino al proceso social de producción previo a su circulación comercial, y no a su circulación comunicativa. En cambio, el dato de “segundo orden” referido a la proporción de “palabras llenas” y “palabras vacías” se acerca más al proceso singular de comunicación como objeto de conocimiento, pues sirve para considerar un aspecto en la forma de expresión de sus autores (emisores); pero no es significativo ni válido para representar la singularidad del texto respecto a la situación comunicativa en función de la cual se produce, pues resulta demasiado marginal e insuficiente en un producto comunicativo cuyo objetivo primordial no es la expresión literaria o artística, sino la claridad y el rigor argumentativo de la representación verbal y escrita de un objeto científico de conocimiento: la metodología de investigación en análisis de contenido.

En términos generales, por consiguiente, la pertinencia de la elaboración de datos mediante disección (o análisis) de productos singulares de comunicación social (como es el caso del denominado análisis de contenido) procede de someter a prueba (refutar) que las operaciones de disección del “texto”, y la elaboración, registro y tratamiento de los datos referidos a éste, resulten adecuados, significativos y suficientes para mostrar su singularidad respecto a la situación de comunicación que constituye el campo de estudio para representar, a su vez singularmente, un objeto de conocimiento científico: la comunicación social a la que se recurre en la agenda del acontecer colectivo.

Según el diseño de análisis del objeto de estudio, es decir, del proceso singular de comunicación donde tienen su origen los productos comunicativos, pueden diferenciarse también análisis de contenido *horizontales, verticales, transversales, longitudinales y triangulares*.

Una primera manera de diferenciar los análisis de contenido según aquel diseño es considerar los análisis de contenido que utilizan un corpus de documentos relativamente amplio, o aquellos que eligen corpus muy reducidos, en ocasiones de un solo caso. Aquellos análisis que se valen de un corpus documental extenso, son los propios de los diseños *horizontales* o extensivos, típicos en el análisis de contenido de carácter cuantitativo, que suelen ser objeto de un tratamiento estadístico a partir de una construcción muestral. Un ejemplo de este análisis de contenido horizontal, desmesurado por el voluminoso universo textual al que fue referido, es el estudio propuesto por Lasswell (1965) que pretendía probar que se había puesto en marcha una revolución mundial a partir del análisis de los símbolos políticos de los editoriales periodísticos de Francia, Alemania, Gran Bretaña, Rusia y los Estados Unidos y de los que aparecían en los discursos políticos de grandes líderes nacionales en todos estos países. Se trataba por consiguiente de analizar como objeto de estudio, más que procesos singulares de comunicación, toda una categoría abstracta de estos procesos, construida por la similaridad de sus productos: los editoriales de prensa, por una parte, y los discursos políticos, por otra. Pero las conclusiones no apuntaban sin embargo al objeto, sino a hipótesis previas supuestamente operativas tanto para los procesos de comunicación de masas, como para las ceremonias políticas donde se pronuncian discursos, como para cualquier otro proceso de interacción social donde a su vez serían operativos los valores manifestados a través de la comunicación de masas y los propios discursos políticos.

El otro tipo de análisis de contenido que cabe distinguir es el de los diseños *verticales* o intensivos, como se ha dicho caracterizados por corpus muy reducidos o de un solo caso, cuyo desarrollo, a diferencia del que es propio de diseños horizontales, en ocasiones es mucho más amplio que el texto analizado. Por lo general, se trata de análisis no cuantitativos, que no recurren a procedimientos de muestreo, en los que el significado deriva más de las relaciones, las oposiciones y el contexto, que de medidas cuantitativas. Este tipo de análisis apenas puede ser replicado y puede adolecer de falta de representatividad en sus hallazgos, pues investigadores diferentes, estudiando el mismo documento o caso, pueden llegar a conclusiones opuestas. Un ejemplo de este análisis de contenido vertical es el que ofrecen los análisis estructural-semiológicos del tipo del realizado por Eco en la *Estructura ausente* (1968) sobre *spots* publicitarios, o del realizado por Todorov (1966) sobre una obra literaria como *Las relaciones peligrosas*, o del realizado por Hall (1973) sobre un género cinematográfico como el “Western”.

El análisis de contenido con diseño *transversal* es aquel que consiste en seleccionar muestras de corpus textuales que difieren, por ejemplo, en cuanto a una toma de postura ante un tema, y formar con ellos grupos independientes que se analizan en el mismo momento histórico crucial reflejado en esos corpus. Si son representativos y se igualan las condiciones de esos grupos, las diferencias que se observen habrá que atribuir las a las posiciones diversas, pero siempre permanecerá la duda de

si son debidas también a diferentes situaciones de comunicación que les son propias. En 1965, Holsti, Brody & North (ver Holsti, 1969) analizaron la sucesión de declaraciones públicas efectuadas durante la crisis de los misiles cubanos en 1962 por importantes funcionarios del gobierno norteamericano y de la Unión Soviética, evaluándolas según la teoría de las actitudes de Osgood, de lo cual extrajeron un modelo de interdependencia dinámica de la guerra fría en situaciones de crisis. Evidentemente, los contextos particulares de las situaciones originarias de la producción de discursos, no se consideraron relevantes: la hipótesis de base ligaba directamente las declaraciones públicas a la teoría de las actitudes, menospreciando, o no tomando en consideración, las mediaciones propias de los procesos que daban a conocer tales declaraciones.

El análisis de contenido con diseño *longitudinal* consiste en analizar corpus en diferentes momentos de su trayectoria, ya sea aplicando medidas repetidas o sirviéndose de muestras independientes. Se trata de los análisis de tipo sistémico propugnados por Rapoport (1969), que desarrolla una teoría sistémica de los corpus textuales, según la cual hay que analizar siempre bajo los mismos parámetros el cambio o la evolución de un mismo corpus textual (por ejemplo, editoriales de un periódico) como sistema autónomo. Si el análisis abarca a un número representativo de medios de comunicación, el autor señala que los resultados pueden considerarse como indicadores temáticos del clima social. Es evidente que también aquí se parte de presuposiciones tales como que los procesos singulares de comunicación se pueden categorizar por la sistematicidad de sus productos (lo cual puede ser cierto a veces, pero no siempre) y que el clima social puede ser descrito por una sola variable, aunque sólo sea indicativa, como es una categoría artificialmente construida de productos de comunicación. En consecuencia, el problema no es conocer la comunicación por sus productos, sino en todo caso el pretendido clima social, por la comunicación, de la cual sólo se retiene una categoría de variables: el género predefinido de productos analizados.

En el diseño *triangular*, el término “triangulación” se utiliza de varias formas, pero fundamentalmente se refiere a la recogida y comparación de distintas perspectivas sobre una misma situación de comunicación. Se basa en la contrastación de la descripción, explicación y evaluación de los contenidos analizados en una investigación, con otras descripciones, explicaciones o evaluaciones de otras investigaciones independientes realizadas sobre el mismo objeto, o bien, dentro de una misma investigación sobre idéntico objeto, de una combinación de técnicas, entre ellas el análisis de contenido, como medio de dar validez externa a los datos. Un ejemplo de este tipo de diseños lo muestra la investigación expuesta en Piñuel & Gaitán (1995) y realizada por Piñuel en 1992 sobre el plebiscito chileno de 1988, en la que se combinaron y compararon los resultados de un análisis de contenido de la campaña electoral del plebiscito chileno en TV, y de un análisis a partir de una encuesta realizada entre los ciudadanos chilenos nada más acabar la campaña y realizada la consulta electoral.

3.4. Análisis de contenido según los parámetros de medición y evaluación

Según los parámetros de medición y evaluación pueden diferenciarse análisis de contenido *cuantitativos* y *cualitativos*. La distinción entre análisis de contenido cuantitativos y cualitativos es débil, sin embargo, porque los aspectos cualitativos se encuentran en toda investigación que parta de una teoría que sirve para construir el objeto científico de estudio, y los análisis de contenido no pasan a la cuantificación de las unidades de análisis hasta que no se ha definido previamente un repertorio de categorías o ítems provistos por el marco teórico, el planteamiento de problemas y por los objetivos del estudio aplicado a un objeto, siempre construido a priori. Pero si reparamos, en cambio, en los parámetros de medición, se observaría que pueden diferenciarse análisis de contenido *frecuenciales*, que contabilizan el número de ocurrencias o de co-ocurrencias de indicadores o categorías, y análisis de contenido *no frecuenciales* que tienen en cuenta su presencia o ausencia.

Entre los análisis de contenido *frecuenciales* cabe distinguir los análisis distribucionales de tipo estadístico y los análisis de relaciones. En los primeros, la interpretación de frecuencias se atiene a varias normas estadísticas, y en concreto: la norma de la distribución uniforme, la norma de la distribución estable y la norma de la representación no desviada; la norma de la distribución uniforme permite resaltar que la frecuencia de una categoría es mayor o menor que el promedio de todas las categorías, mientras la norma de la distribución estable permite apreciar modificaciones en las frecuencias con el transcurso del tiempo. La norma de la representación no desviada, finalmente, pone de manifiesto que las frecuencias observadas son mayores o menores de lo que sería previsible si la muestra fuera representativa de la población de referencia. Las frecuencias sirven entonces para medir estadísticamente datos descriptivos, o para establecer escalas bipolares que marcan la dirección (favorable, desfavorable, neutra, ambivalente), o la intensidad y la dirección simultáneamente (cfr. Piñuel & Gaitán, 1995).

Por otra parte los análisis de contenido *relacionales* pueden medir la co-ocurrencia, pero siempre a partir de tabulaciones previas frecuenciales del tipo de las tablas de contingencia (ver Piñuel & Gaitán, 1995: cap. 4), entre distintos elementos categoriales, estableciendo entonces relaciones de determinación, asociación, equivalencia, oposición, exclusión, proximidad, simultaneidad, secuencialidad u orden. La tabulación de frecuencias cruzadas o de co-ocurrencia de dos dimensiones, por su gran simplicidad, es de las más utilizadas. Pero se dispone también de la posibilidad de hacer uso de técnicas multivariadas para verificar estructuras más complejas con datos multidimensionales. En el análisis relacional, el esclarecimiento de estructuras de relaciones lógicas entre categorías, para constatar la discriminación o la consistencia de unas sobre otras, son también otros ejemplos de cómo pueden desarrollarse estas técnicas (Piñuel & Gaitán, 1995).

Cuadro 1

CATEGORÍAS UNIDADES	DISECCIÓN DE ANÁLISIS	DISCIPLINA DE INFLUENCIA	PROCESAMIENTO O MEDIDA	OBJETOS DE ESTUDIO (EJEMPLOS)
Unidades léxicas	Términos, palabras	Semántica Lexicología	Cuantitativo (estadístico)	Hábitos verbales Campos semánticos
Unidades temáticas	Conceptos, referencias	Psicología cognitiva Psicosociología Sociología de la comunicación	Cuantitativos (estadístico) y/o cualitativos (lógicos)	Productos singulares de comunicación interpersonal y de masas
Unidades temático / evaluativas	Valoraciones, actitudes	Psicosociología	Cuantitativos (estadístico) y/o cualitativos (lógicos)	Productos singulares de comunicación de masas, según formatos: periodismo, publicidad, espectáculo, etc.
Unidades lingüístico-proposicionales	Oraciones Proposiciones	Gramatología Filosofía del lenguaje Sociolingüística	Cualitativos (lógicos)	Enunciaciones referidas a contextos en productos singulares de procesos de comunicación
Unidades formales análisis semiológico	Relatos	Antropología Psicosociología del conocimiento Narratología	Cualitativos (lógicos)	Relatos ubicados en contextos de procesos singulares de comunicación
Unidades pragmáticas del discurso y análisis semiótico de acciones	Relación entre enunciación e interacción comunicativa	Sociolingüística Etnometodología Semiótica Socioanálisis	Cualitativos (lógicos y hermenéuticos)	Dramatizaciones entre los interlocutores por el uso de enunciaciones y relatos latentes
Macro-unidades psicobiográficas y análisis de “historias de vida”	Actores, personajes históricos, sujetos representativos de aconteceres	Crítica literaria Psicosociología Psicología clínica	Cualitativos (lógicos y hermenéuticos)	Productos comunicativos referidos a historias personales de comportamientos comunicativos

3.5. El análisis de contenido según las unidades de registro y de análisis

Las unidades de registro y las unidades de análisis en ningún caso son independientes del objeto de estudio, de las fuentes que se eligen estudiar, del diseño de análisis, de la técnica de medición y evaluación empleada, y, por supuesto, de su campo de aplicación. En el Cuadro 1 se ilustra de modo esquemático el ajuste metodológico entre el tipo de análisis de contenido, los objetos de estudio, la disciplina de influencia y el campo de aplicación citando algunos ejemplos.

4. Las técnicas del análisis de contenido

Dos aspectos fundamentales hay que considerar respecto a las técnicas para elaborar, registrar y tratar datos en análisis de contenido. Uno es la normalización de los procedimientos para llevar a cabo cada investigación particular. Otro es la elección del software que mejor pueda ajustarse a los planteamientos teóricos y metodológicos de los que se parte. Los consideraremos por este orden.

4.1. La normalización de los procedimientos de análisis: el protocolo

Cualquier análisis de contenido, en tanto que técnica para la elaboración, registro y tratamiento de datos sobre procesos singulares de comunicación, a partir de un corpus representativo de sus *productos* (mensajes, textos o discursos) singulares, o a partir de un corpus representativo de grabaciones registradas que a la postre constituyen un documento, se somete a un conjunto de procedimientos interpretativos y de técnicas de refutación, conocido como *protocolo*⁵.

Puede decirse que el *protocolo* es, por consiguiente, un procedimiento o conjunto de normas que guían tanto la *segmentación* del corpus según el establecimiento de criterios interpretativos para su lectura u observación, como para efectuar el *registro* más adecuado de datos, disponibles después para su tratamiento estadístico o lógico y para el procesamiento posterior (o transmutación) en datos de “segundo” y “tercer” orden, es decir, en datos referidos, a su vez, a datos previamente registrados y tratados.

Gracias a la existencia y manejo adecuado de un protocolo de análisis, suficientemente objetivado, si el mismo corpus es analizado por investigadores diferentes, las conclusiones del estudio deben ser coincidentes; más aún, gracias a la existencia y manejo adecuado de un protocolo de análisis, suficientemente objetivado, si el mismo o diferentes investigadores estudian corpus diferentes pero igualmente representativos de los mismos procesos y sistemas de comunicación, también concluirán en semejantes resultados; y, finalmente, gracias a la existencia y manejo adecuado de un protocolo de análisis, suficientemente objetivado, si el mismo o diferentes

⁵ Los programas de análisis de contenido automático por ordenador tienen cada cual su protocolo. En este apartado, sin embargo, se toman en cuenta los procedimientos generales sean cuales sean las rutinas particulares de cada programa.

investigadores estudian corpus diferentes pero igualmente representativos de diferentes procesos y sistemas de comunicación, podrán llegar a conclusiones válidas tanto para comparar los procesos por su naturaleza, como por su evolución temporal.

En efecto, como es sabido, el término griego *prôtokollon*, derivado del adjetivo *prôtos* (primero) y del verbo *kollân* (pegar, ligar, unir), dio origen a los términos latino *protocollum*, y al nuestro, “protocolo”, cuyas acepciones específicas en los ámbitos ceremoniales (tanto desde el punto de vista jurídico como diplomático), pero también en la investigación científica, comparten una misma idea: lo que se dispone para ligar de forma válida un procedimiento. En la investigación científica, concretamente, se entiende por *protocolo*, o conjunto de enunciados protocolarios, precisamente aquellas reglas de procedimiento necesarias para llevar a cabo un proceso de observación o experimentación. En todos los ámbitos persiste, entonces, la misma noción: “protocolo” es aquello que se dispone para ligar de forma válida un procedimiento. Esto es tanto como decir que si el procedimiento (manera de proceder, de seguir unos determinados pasos tras otros, etc.) es válido, lo será en la medida que no se podrán refutar sus efectos, o sus resultados; las conclusiones, por consiguiente, si existe y se maneja adecuadamente un protocolo de análisis, suficientemente objetivado, resultarán irrefutables mientras no cambie (por la teoría o por la praxis) el objeto de estudio. De aquí la importancia de que el protocolo sea válido, al menos tanto, como para asegurarse unos límites razonables al riesgo epistemológico de cualquier conocimiento: que el procedimiento de observación sustituya al objeto observado, y entonces nunca podrá el conocimiento trascenderse a sí mismo.

En general, los protocolos para el análisis de contenido constan de dos tipos de guías de procedimiento. Una es la guía para la interpretación y segmentación del corpus, que suele llamarse *libro de códigos*. Otra es la guía para el registro de datos, que suele conocerse como *ficha de análisis*.

4.1.1. El libro de códigos: guía para la segmentación del corpus

El *libro de códigos* es un manual de instrucciones en el que se establece:

- 1) Cuál es la naturaleza del corpus, qué contiene, para qué sirve, cómo interpretarlo o desde qué perspectiva abordar su lectura, escucha o visualización;
- 2) Cuáles son sus unidades de análisis, es decir, cómo segmentar el corpus de forma que a medida que se lee, escucha o visualiza, el analista sepa cuándo abrir una *ficha de análisis* y proceder a ir registrando datos re-leyendo, re-escuchando o re-visualizando los segmentos establecidos de ese corpus;
- 3) Cómo rellenar las *fichas de análisis*, sabiendo a qué código alfanumérico corresponden cada una de las apreciaciones de la interpretación del segmento analizado;
- 4) Cómo dar por concluida cada ficha de análisis, y cómo ordenarlas y archivarlas para su uso posterior en el tratamiento de los datos registrados;

- 5) Cómo construir o usar una base de datos que permita elaborar y procesar datos de segundo y tercer orden, es decir, cómo producir nuevos datos sobre los inicialmente registrados en las fichas y en la base a la que se han trasladado;
- 6) Eventualmente, aunque esto no suele fomar ya parte del *libro de códigos*, cómo obtener, usando la base, datos de segundo y tercer orden mediante un plan de explotación: recuento de frecuencias, por ejemplo, establecimiento de medidas estadísticas, correlaciones, plan de cruces para tablas de contingencia, etc.

Consideremos una a una estas operaciones para realizar de las cuales se dan instrucciones en el libro de códigos. Si estas se tienen en cuenta, podrán posteriormente mejor seleccionarse las técnicas de software que puedan ser usadas, pues mejor podrá conocerse que fases son incluidas en sus protocolos de usuario, y cuál es la rentabilidad que cada técnica permite en ahorro de tiempo. De hecho, en los primeros tiempos del análisis de contenido, todos los procesos señalados se hacían manualmente. Después, lo primero que se introdujo fue el procesamiento informático a partir del repicado de los datos, desde las fichas de análisis, a las bases de datos estadísticas disponibles. Más tarde, se llegaron a sustituir las fichas de análisis por bases de datos diseñadas para ser utilizadas y rellenadas directamente desde la lectura u observación del texto. Finalmente, se diseñaron programas informáticos que a partir del reconocimiento óptico de los caracteres visualizados por escáner sobre los propios textos, realizan ya todos los pasos en aquellos análisis exclusivamente apoyados en la segmentación lingüística de textos impresos.

No obstante, el libro de códigos es siempre un manual de instrucciones a la medida de la metodología de investigación concreta, nunca puede funcionar como una prenda *prêt-à-porter*. Es obvio que no todos los datos que pueden ser elaborados, registrados y tratados, a partir de la disección de productos comunicativos o “textos”, resultarán adecuados, y serán significativos y suficientes, para representar científicamente la situación comunicativa que integra supuestamente el acontecer socialmente previsto en que se recurre a la comunicación, y que ha de constituir el objeto científico de análisis. En términos generales, por consiguiente, la pertinencia de la elaboración de datos mediante *disección* (o segmentación) *de productos singulares* de comunicación social (como es el caso del análisis de contenido) procede de someter a prueba (refutar) que las operaciones de disección del “texto”, y la elaboración, registro y tratamiento de los datos referidos a éste, resulten adecuados, significativos y suficientes para mostrar su singularidad respecto a la situación de comunicación que constituye el campo de estudio, el cual tiene que representar, a su vez singularmente, un objeto de conocimiento científico: la comunicación.

Para cumplir este requisito en lo que atañe al primer género de instrucciones citadas, y una vez justificada la representatividad muestral del corpus (con exigencias similares a la representatividad requerida de una muestra de población a encuestar, cuando se define un universo), es necesario que las operaciones de disección del

“texto” resulten adecuadas a ese objetivo, y a su justificación se dedica el primer capítulo del manual de instrucciones, es decir, a explicar cuál es la naturaleza del corpus, qué contiene, para qué sirve, cómo interpretarlo o desde qué perspectiva abordar su lectura, escucha o visualización. Ello supone que el analista cuente con la suficiente competencia comunicativa para interpretar el texto, o al menos que la pueda adquirir situándose en la perspectiva exigida por esa competencia. En los programas de análisis automático de textos, este tipo de instrucciones es el que se lleva a cabo los *diccionarios* y *thesaurus* del programa, razón por la cual estos programas sólo pueden abordar análisis demasiado pegados a la segmentación semántica de los términos y enunciados, pero no al registro y tratamiento de datos extratextuales (como los pragmáticos de tipo retórico, argumentativo, narrativo, etc.).

El segundo capítulo del manual de instrucciones que constituye el libro de códigos está dedicado a exponer y justificar cómo segmentar el corpus de forma que a medida que se lee, escucha o visualiza, el analista sepa cuándo abrir una *ficha de análisis* y proceder a ir registrando datos re-leyendo, re-escuchando o re-visualizando los segmentos establecidos de ese corpus. Una *ficha de análisis*, como se expondrá posteriormente, es semejante a un cuestionario de encuesta; si concluida la aplicación de una encuesta, hay tantos cuestionarios rellenos como encuestados hayan respondido a las preguntas que contiene, al concluir el registro de datos en análisis de contenido, habrá tantas *fichas de análisis* rellenas como unidades se hayan encontrado en la segmentación del corpus. La diferencia está en que antes de aplicar una encuesta, ya se sabe previamente cuántos cuestionarios se necesitan: tantos como individuos, grupos familiares, etc., se hayan seleccionado como muestra representativa del universo a estudiar. Pero en el análisis de contenido la muestra es el *corpus*: un número de ejemplares de periódicos que representan al universo de ejemplares difundidos en un tiempo dado, cubriendo un cierto territorio de audiencias; o un número de películas, programas de TV, obras escénicas, etc., que representan al universo de obras o productos comunicativos relevantes para conocer un proceso de comunicación como objeto de estudio, etc. Se sabe por consiguiente con cuántos ejemplares (periódicos, películas, programas de TV) se cuenta, y se sabe de antemano cómo segmentar el corpus, por ejemplo, en unidades de redacción en cada periódico (editoriales, o, por secciones, noticias, artículos de opinión, crónicas, reportajes, etc.), o por secuencias en el flujo audiovisual de las retransmisiones por TV, de los spots de publicidad, o de las teleseries. Pero no se sabe muchas veces con certeza cuántas unidades de análisis aparecerán. Este es el tipo de unidades que pueden denominarse *unidades formales* y que son las unidades que componen fenomenológicamente al objeto material de estudio, y sirven para fijar alternativas metodológicas por las que se *transmutan en datos manejables*, según elección de una o más *técnicas* para procesarlos y analizarlos. Obviamente, las alternativas metodológicas tienen que ver con la elaboración y registro posteriores de otros datos que hemos dado en llamar de “segundo orden” (presencia o ausencia de determinados rasgos en

las unidades formales: por ejemplo, en los editoriales de los periódicos, o en las secuencias del programa de TV), e incluso de “tercer orden” (índices de frecuencia, relaciones estadísticas o lógicas obtenidas por comparación de éstas, etc.) que tendrán que ser procedentes del objeto de estudio, y que tendrán igualmente que ser significativos y suficientes para contrastar el objeto de estudio y los objetivos fijados en su análisis.

Por ejemplo, el libro de códigos puede exponer y justificar que cada vez que en el flujo audiovisual se cambia de *secuencia*, hay que volver atrás, revisualizar la secuencia concluida, y considerándola de principio a fin, abrir una ficha en la que se registrarán datos internos sobre temas, personajes, escenarios, acciones, etc. (véase un ejemplo de libro de códigos y ficha de análisis en Piñuel & Gaitán, 1995: cap. 6). O elegidos todos los editoriales de *El País*, aparecidos a lo largo de un determinado tiempo, abrir una ficha cada vez que aparece un enunciado genérico universal, y releyéndolo registrar datos internos sobre variables de la referencia, la predicación, la atribución, la modalidad, etc. (véase Gaitán, 1991).

El tercer capítulo de un libro de códigos se dedica a establecer cómo rellenar las *fichas de análisis*, sabiendo a qué código alfanumérico de ella corresponden cada una de las apreciaciones de la interpretación del segmento analizado (secuencia de TV, cuadro escénico, frases de los titulares que componen una unidad periodística, etc.) Estas apreciaciones, si el analista o equipo de analistas cuentan con la competencia comunicativa requerida, constituyen los datos que más tarde irán, ficha por ficha, a integrar la base de datos. Es éste quizás el capítulo central de cualquier libro de códigos, hasta el punto que por esta causa, el manual de instrucciones así llamado, se denomina “libro de códigos”. Su contenido está estrechamente ligado a la estructura de la *ficha de análisis*, o cuestionario que el analista rellena como si él fuese un encuestador que se hace preguntas a sí mismo y las responde a la medida de su apreciación de cada segmento analizado. Para que esta apreciación, entonces, resulte lo más objetivada posible, las instrucciones (claras, precisas, sin ambigüedades posibles) suelen ensayarse de antemano en las pruebas piloto, e incluso probarse experimentalmente recurriendo a “grupos de jueces” que experimentalmente sirven para contrastar las apreciaciones buscadas ante la lectura, escucha o visualización de segmentaciones representativas del corpus. Las apreciaciones sobre los aspectos relevantes de cada segmentación, se estructuran por variables y categorías, a cada una de las cuales se le hará corresponder con un número de columnas y filas en la base de datos. Por ello, este capítulo del libro de códigos es tan crucial, ya que resulta el más consultado por los analistas cada vez que ante una apreciación no sepa con certeza qué número de variable asignarle y qué número de categoría le corresponde.

El cuarto capítulo del libro de códigos se dedica a explicar cómo dar por concluida cada ficha de análisis, y cómo ordenarlas y archivarlas para su uso posterior en el tratamiento de los datos registrados. Normalmente las instrucciones referidas al cierre, así como al orden de archivo para uso posterior respecto a la base de datos, viene

determinado por aquellas variables que en la ficha de análisis se refieren a la numeración e integración de cada unidad formal en la estructura del corpus. Por ejemplo, si a cada enunciado genérico universal le corresponde una ficha de análisis a rellenar, en ella habrá que consignar datos referidos al editorial en cuestión, identificable por su tema, su orden de aparición en el periódico, su fecha de publicación, etc. Y si a cada secuencia de un *spot* de publicidad le ha de corresponder una ficha de análisis, en ella habrá que consignar a qué *spot* pertenece, cuál es el producto anunciado, a qué categoría de productos pertenece, etc. Incluso a veces una ficha de análisis (o una fila en la base de datos, especialmente en aquellas bases de datos informáticas que directamente remplazan a las fichas que utiliza el analista), incluye variables que contienen ya recuentos o cálculos sobre datos registrados en la misma ficha (o fila); por ejemplo, si se han consignado variables para operadores y conectores ligados a unidades léxicas, o referenciales, o predicativas, etc., los cálculos en función de los cuales evaluar actitudes, argumentos, etc., y que servirán para caracterizar a cada unidad formal o segmentación analizada en la ficha de análisis. En ocasiones, no todas las variables de una ficha de análisis presentan precodificadas todas las casillas posibles para dar respuesta a las apreciaciones de sus rasgos; puede haber variables cuya apreciación la expresa cada analista escribiendo una frase o una palabra clave que a su juicio puede resumirla. En ese caso, será necesario posteriormente postcodificar las respuestas, antes de registrarlas en la base de datos, de forma que sólo se considerará concluida una ficha, cuando tras la palabra clave recogida, se acompaña el registro correspondiente a la categoría de pots-codificación que resulte necesario consignar.

El quinto capítulo del libro de códigos sirve para explicar cómo construir o usar una base de datos que permita elaborar y procesar datos de segundo y tercer orden, es decir, cómo producir nuevos datos sobre los inicialmente registrados en las fichas y en la base a la que se han trasladado. Esto se establece dando instrucciones sobre el tipo de operaciones a efectuar con el conjunto de las fichas de análisis ya rellenas: por ejemplo, agrupándolas a veces según categorías relevantes por el objeto de estudio, por el corpus analizado, por las segmentaciones efectuadas en éste, etc. Cuando los procedimientos de análisis eran manuales, esta parte del libro de códigos era muy prolija, especialmente en los análisis de unidades temáticas y evaluativas. Hoy en día, en que prácticamente sólo se utilizan bases de datos informáticas, este tipo de operaciones son más fáciles de realizar, pues se efectúan mediante el teclado de funciones útiles para segmentar los archivos, una vez que todos los datos de las fichas (o registros del archivo) correspondientes, han sido introducidos. Normalmente estas instrucciones son las que dejan ya dispuestos los datos registrados en el conjunto de las fichas de análisis, para su tratamiento o procesamiento posterior, y que constituye a veces un apartado de este capítulo, o un nuevo capítulo, el sexto, de los citados anteriormente.

El sexto y último capítulo del libro de códigos se dedica a explicar cómo obtener, usando la base, datos de segundo y tercer orden mediante un plan de explotación: recuento de frecuencias, por ejemplo, establecimiento de medidas estadísticas, corre-

laciones, plan de cruces para tablas de contingencia, etc. También puede ocurrir que estas instrucciones no formen parte del libro de códigos, pero siempre formarán parte del protocolo de análisis, pudiendo entonces constituir una tercera guía (llamada *plan de explotación*). La razón por la cual el *plan de explotación* de datos no suele incluirse en el libro de códigos es que esto va destinado al equipo de analistas cuya misión termina con el relleno de todas las fichas de análisis, mientras re-leen, re-escuchan o re-visualizan el corpus, y con el archivado de éstas cumpliendo las instrucciones dadas para su agrupamiento y ordenación. Se supone que el trabajo con las bases de datos, tras haber repicado todos los registros en ella, no es misión de los analistas, sino de otro equipo de colaboradores en la investigación: los procesadores de datos. Es a ellos entonces a quienes se destina el *plan de explotación*. Siguiendo el plan de explotación, los procesadores terminan su trabajo proporcionando lo que se denominan *tabulaciones*, a partir de las cuales el investigador principal efectuará la interpretación de los datos de segundo y tercer orden, y elaborará los datos de orden posterior es decir, aquellos que sirven para exponer y describir las conclusiones del estudio. Son estas últimas unidades, y sus correspondientes interpretaciones relativas a variables confirmadas y no confirmadas, las que precisamente se llevarán al informe de investigación en calidad de conclusiones, y especialmente las que dotan de sentido a todo el proceso de elaboración, registro y tratamiento de datos. Por esta razón, cuando la investigación está concluida, y el problema inicial cobra una solución aplicable, ya sea a corto o largo plazo, el informe presenta, junto a las correspondientes unidades de cada fase de aplicación del protocolo, los anexos del *Libro de códigos*, *Ficha de análisis* y *Plan de explotación*. Muchos datos que sin embargo han sido extraídos en el proceso de investigación y que permanecen ocultos en las *tabulaciones*, serán silenciados o desechados. Las tabulaciones posiblemente han contemplado cruces entre variables (e incluso, puesto que su base informática lo permite, se han efectuado nuevos cruces no previstos en el plan de explotación) que sin embargo no llegan al informe final, ya sea porque sus datos resultan no suficientemente significativos, o porque se hicieron de forma exploratoria aún suponiendo que podían carecer de significatividad. En el informe, pues, sólo aparecerán aquellas *unidades de orden posterior*, o conclusiones, que sirvan para confirmar en uno u otro sentido, las relaciones entre el objeto de estudio analizado, y el problema planteado desde los objetivos inicialmente formulados. Pero el *protocolo de análisis* con sus correspondientes *libro de códigos*, *ficha de análisis* y *plan de explotación*, se presenta como si de tal objeto de estudio, dado el planteamiento del problema inicial, sólo se pudieran haber elaborado, registrado y tratado los datos previstos, cuando en realidad –todo investigador lo sabe– hay infinidad de datos de diverso orden, elaborados, registrados y tratados pero que sin embargo nadie conocerá, salvo el investigador y sus colaboradores.

4.1.2. La ficha de análisis: plantilla para el registro de datos

La *ficha de análisis* es una plantilla para el registro de datos al re-leer, re-escuchar

o re-visualizar cada una de las segmentaciones del corpus, contempladas en el *libro de códigos*. Como se explicó antes, se trata de un cuestionario que el analista rellena como si él fuese un encuestador que se hace preguntas a sí mismo y las responde a la medida de su apreciación de cada segmento leído, escuchado o visualizado. Pero a diferencia de los cuestionarios de encuesta, donde para cada variable se formula la pregunta correspondiente, más las casillas para registrar las respuestas previstas y precodificadas, en la ficha de análisis a veces no se formula ninguna pregunta, sino sólo el nombre de la variable y en todo caso su número de código (columna, en la base de datos). La razón es que el analista siempre tiene a mano el *libro de códigos* para consulta y, además, siempre es más manejable una ficha de análisis de sólo una cara (habitualmente en una cartulina del tamaño ficha) que si para rellenar cada una de ellas tiene que manejar un folleto (el lector puede consultar una ficha de análisis en el ejemplo de investigación expuesto en Piñuel & Gaitán, 1995: 463). Y, sobre todo en este caso, el de usar fichas de análisis para consignar manualmente el registro de datos, se gastará menos papel, pues lo normal es que un análisis supere las 5.000 fichas.

La estructura de la ficha, con sus correspondientes variables y casillas para consignar el registro de datos, debe servir para rellenarlas lo mejor posible, de forma que el analista pronto aprenda, sin tener que recurrir más que al principio al libro de códigos, a registrar los códigos alfanuméricos que se corresponden con cada variable, y con cada una de las apreciaciones personales de la interpretación del segmento observado (frases de los titulares que componen una unidad periodística, secuencia de TV, cuadro escénico, etc.). Estas apreciaciones, si el analista o equipo de analistas cuentan con la competencia comunicativa requerida, y que constituyen los datos que más tarde irán, ficha por ficha, a integrar la base de datos, son las que se consignan rellenando la casilla o casillas correspondientes para cada variable.

Las fichas de análisis suelen comenzar por aquellas variables que se utilizan para relacionar la unidad de análisis, o segmento del corpus visualizado u observado, con el resto de unidades que integran cada ejemplar del corpus, y con todo el corpus globalmente. Por ejemplo, *spot* a que pertenece la secuencia visualizada, o editorial al que pertenece la frase retenida, y número o categoría que dentro del corpus le corresponde al *spot*, o al periódico, etc.

En segundo lugar, las *fichas de análisis* suelen continuar por aquellas variables que atañen a los aspectos más generales del segmento leído o visualizado (por ejemplo, temas generales, o ámbitos de referencia), hasta llegar a aquellas otras variables que atañen a detalles más particulares (por ejemplo, en una frase, variables referidas a la expresión de *sujeto*, y a la expresión de *predicado*, o en una secuencia, variables que atañen al *escenario* donde ni los actores ni los objetos cambian, aunque se sucedan acciones diversas); finalmente se da paso, de la forma que resulte más fácil para los analistas, a las variables que requieren relecturas, o re-visualizaciones, repetidas para interpretar detalles (por ejemplo, categorías de la atribución, modalidad, relaciones, o, refiriéndose a los personajes de un relato, categorías referentes a las apre-

ciaciones de sus rasgos físicos, o sociales, o comunicativos, o a sus acciones, aspiraciones, actos, enunciaciones, etc.). Incluso a veces una ficha de análisis (o una fila en la base de datos, especialmente en aquellas bases de datos informáticas que directamente remplazan a las fichas que utiliza el analista), incluye variables que contienen ya recuentos o cálculos sobre datos registrados en la misma ficha (o fila); por ejemplo, si se han consignado variables para operadores y conectores ligados a unidades léxicas, o referenciales, o predicativas, etc., los cálculos en función de los cuales evaluar actitudes, argumentos, etc., y que servirán para caracterizar cada unidad formal o segmentación analizada en la ficha de análisis. Estos cálculos para cada ficha, cuando éstas se rellenan manualmente, pueden efectuarse antes de concluirla, o en un segundo momento que será aquel en que se revisan todas las fichas cumplimentadas y se disponen para su ordenación, archivo y traslado a la base de datos informática (repicado). Más aún, ya hemos dicho anteriormente que en ocasiones no todas las variables de una ficha de análisis presentan precodificadas todas las casillas posibles para dar respuesta a las apreciaciones de sus rasgos; en ese caso puede haber variables cuya apreciación la expresa cada analista escribiendo una frase o una palabra clave que a su juicio puede resumirla. Esto facilita enormemente el proceso de visualización o lectura y el cumplimiento de cada ficha, evitando así que una cierta inercia en el registro de categorías demasiado cualitativas, pero expresadas por números, se escaspille en la rutina del analista. Consignada pues una palabra clave, o juicio extraído del texto en su literalidad, será necesario posteriormente postcodificar las respuestas así expresadas, antes de registrarlas en la base de datos, de forma que sólo se considerará concluida una ficha cuando tras la palabra clave recogida, se acompaña el registro correspondiente a la categoría de post-codificación que resulte necesario consignar. Este proceso lo suelen realizar los analistas, siguiendo las instrucciones del libro de códigos, cuando revisan los cálculos internos de cada ficha si los hubiera y cuando proceden a la ordenación y archivo de las fichas, momento que aprovechan para aplicar sus instrucciones y rellenar las casillas numeradas que se corresponden con las categorías de la post-codificación.

A partir de aquí, el restante tratamiento de datos en análisis de contenido corresponde al *plan de explotación*, cuyas características ya hemos explicado y cuyo documento es hoy exclusivamente un archivo informático (habitualmente denominado *tabulación*) que por su formato (en general en lenguaje ASCII universal) puede ser importado a los programas estadísticos de tratamiento de datos (v.g. el SSPS) que efectúan con ellos las operaciones que el investigador, él mismo, o los colaboradores encargados del procesamiento, realizan siguiendo el *plan de explotación*.

Las técnicas basadas en software informático suelen ahorrar muchos de estos pasos. La lógica marcada por sus protocolos internos suele ser del mismo tipo. Ahora sólo queda, una vez tenido lo anterior en consideración, contemplar cuál es el universo de técnicas informáticas de análisis de contenido, tratando de clasificarlas y describirlas brevemente.

4.2. Clasificación de software para análisis de textos

La clasificación que se ofrece a continuación es la proporcionada por Harald Klein (Social Science Consulting, 07407 Rudolstadt, Germany, en *Text Analysis Info Page*, en <http://www.textanalysis.info/html>), como la supuestamente más clara de las existentes. Citaré solamente los programas más relevantes a mi juicio, aparecidos en la última modificación de esta página (23 de enero de 2002).

4.2.1 Software aplicado al lenguaje: análisis de usos de las palabras

4.2.1.1. Tratamiento lingüístico: análisis sintácticos y análisis semánticos de las palabras

CAPITO 1.0

Autor: Hans Messelken, Oliver Fink. Distribuidor: Semiotis GmbH, Rheinbach, Alemania. Sistema(s) operativos: Win98, WinNT. Descripción: estadística de textos, compara textos, calcula la legibilidad con fórmulas, genera listas de palabra y resúmenes automáticos (en alemán)

PLCA 2.0F 2

Programa para análisis de contenido lingüístico. Autor: Carl W. Roberts Distribuidor: IEC progamma Groningen, Países Bajos. Documentación: incluido en versión de demostración. Sistema(s) operativos: DOS, versión de Windows en desarrollo. Descripción: PLCA suministra el medio para trazar un mapa de textos en una matriz de datos que contiene la información sobre palabras y sus interrelaciones. Cada cláusula dentro de un texto es cifrada en un registro separado de datos.

SATO 4.0

Autor: François Daoust. Distribuidor: Universidad de Montreal, Canadá (en francés). Sistema(s) operativos: DOS Descripción: SATO permite la anotación de documentos multilingües, tiene una lengua de pregunta que asegura la posición sistemática de segmentos textuales definidos por el usuario, la producción de un índice; listas de palabra clasificadas alfabéticamente o por frecuencia; la clasificación de palabras, compuestos de palabras o frases; la definición de variables para realizar enumeraciones múltiples y análisis lexicométrico; funciones de diccionario, y según los casos un índice de legibilidad (GUNNING).

CETA 2.0 –

Análisis de Texto Automatizado evaluativo (en holandés) Autor: Ridder. Distribuidor: IEC progamma Groningen, Países Bajos. Sistema (s) operativos: DOS. Descripción: CETA apoya la codificación semántica y asume que un texto consiste en una colección de las declaraciones, que describe relaciones entre segmentos de significado. Analizando cada oración de un texto, CETA representa el texto como una red. El análisis subsecuente de esta red puede revelar el discurso que es la base del texto.

PROFILER +

Autor: Michael Young. Distribuidor: Red de Ciencia Social. Sistema(s) operativos: Win98, WinNT. Descripción: un motor de análisis de contenido de objetivo general diseñado para análisis de liderazgo. Profiler + busca una oración de la izquierda siguiendo un juego ordenado de señales predefinido (palabras y/o puntuación) que sirve como indicador de un rasgo, de una medida de interés o quizás de un tipo particular de comunicación. Profi-

ler + examina cada serie y consulta una base de datos para determinar si una señal determinada sirve como indicador de un síntoma. Si la señal realmente sirve como indicador del síntoma, se pone el programa a determinar si otras señales del conjunto se hacen también presentes según un orden determinado. Si todas las señales pueden ser emparejadas entonces las acciones indicadas por ellas son tomadas en consideración y llevadas a un archivo, por el que se define un perfil para el hablante.

4.2.1.2. Banco de datos: recuperación de información en textos, elaboración de índices, correspondencia de textos, listas de palabras, KWIC/KWOC (key-word-in-context, palabra clave dentro de contexto)

ASK SAM 4.0

Autor, Distribuidor: ASK SAM Software. Sistema operativo: Win3.1, Win98, WinNT, OS/2. Descripción: AskSam es un programa de recuperación de documentos rápido, mediante sistema de hipertexto, semejante a un editor de páginas Web.

ATA- ASHTON TEXT ANALYSER (WinATAMark2)

Autor y distribuidor: Peter Roe. Sistema (s) operativos: Win98, WinNT. Descripción: ATA genera listas de palabra, palabras clave y palabras clave en contexto. (KWIC, KWOC).

DBT 3.1 - Data base testuale. (DBT-Lexis).

Autor: Eugenio Picchi. Distribuidor: Lexis (Universidad de Pisa). Sistema de operaciones: Win9x, WinNT. Descripción: DBT puede hacer búsquedas de palabras, correspondencias, juegos de palabras siguiendo la lógica booleana (incluso aplicando lógica borrosa) tanto dentro del texto o en componentes accesorios (apuntes, apéndices, etc.). También permite hacer listas de palabra según ordenes de clase diferentes, incluyendo un índice (index) de tópicos, un diccionario, o también frecuencia de secuencias de caracteres y secuencias de palabra, manejo de imágenes, etc. que pueden ser asociadas a cada parte del texto.

ERIC JOHNSON'S PROGRAMS.

Los programas de Eric Johnson sobre todo han sido realizados para el análisis de juegos dramáticos y poesía. Requieren textos SGML etiquetados, según limitaciones de las obras a que se aplican (por ejemplo de Jane Austen o Shakespeare). Así, el programa ACTORES efectúa la lista de personajes simultáneamente en escena -generando un tratamiento cada vez que uno de los personajes entra o sale de la escena, siguiendo las co-ocurrencias de personajes en la acción, etc. y efectuando el tratamiento de sus intervenciones, según diversas variables dramáticas y literarias (por ejemplo, FINDLIST comparación de listas de palabra (más de dos), o PICKWICK: filtro de programa para escenas o lugares dramáticos siguiendo el etiquetado de los textos. Etc.

LEXA 7.0

Corpus Processing Software.

Autor: Raymond Hickey, Universidad de Essen/Alemania. Distribuidor: Universidad de Bergen, Noruega. Sistema(s) operativos: DOS. Descripción: LEXA es un sistema abierto basado en archivos. Puede realizar análisis semánticos, listas de palabras, tablas de frecuencias léxicas, comparación de archivos y funciones de recopilación, así como estadística a caracteres, palabras, y oraciones, buscando los grupos de archivos capaces de detectar secuencias de segmentos de texto, frass, etc. (DBF-files).

METAMORPH

Distribuidor: Thunderstone Software.

Sistemas operativos: DOS, Win98, WinNT, Unix. Descripción: Metamorph es un paquete de búsqueda, sin cualesquiera pasos de proceso previo. Metamorph tiene un vocabulario de lengua en inglés de 250,000 palabra y asociaciones de frases por conceptos para preguntas en lenguaje natural, que utiliza la lógica booleana, aunque suministra el control de proximidad entre palabras, búsquedas borrosas, correspondencias regulares de expresión verdadera, y búsquedas de valor numérico

MicroOCP

Autor: Susan Hockey, Iain Hart. Sistema (s) operativos: DOS, versiones de unidad central para VMS, UNIX, y VM/CMS CON UN RECOMPILADOR LENGUAJE FORTRAN. Descripción: OCP es un instrumento para generar correspondencias, listas de palabras, e incluye un índice de textos en cualquier lengua o alfabeto. OCP funciona sobre un archivo de ASCII del texto. El texto puede ser etiquetado en ficheros para incluir la información adicional como el número de línea, capítulos, notas, etc.

TACTO 2.1.5

Autores: Michael Stairs, John Bradley, Ian Lancashire, Lidio Presutti. Distribuidor: Universidad de Toronto. Sistema (s) operativos: DOS. Descripción: El TACTO es un sistema de 15 programas diseñado para la recuperación texto y el análisis de obras literarias. Los investigadores usan el TACTO para recuperar las presencias (las ocurrencias) de una palabra, un modelo de palabra, o la combinación de palabras. Los programas también pueden hacer las clases más simples de análisis, como frecuencias clasificadas de juegos de palabras o frases, o la clasificación de enlaces de palabras según su fuerza de asociación. El TACTO es utilizado para textos individuales literarios, o de pequeño grupos de texto de tamaño medio.

MONOCONC PRO 2.0

Autor: Michael Barlow. Distribuidor: Athelstan. Sistema (s) operativos: Win95, Win 98, etc. Descripción: MonoConc es un programa de concordancias, que puede crear correspondencias, listas de palabras, (con listas de exclusión, el caso sensible/insensible), convierte textos, y trabaja con textos etiquetados y con lenguas diferentes.

4.2.2. Software aplicado al contenido: tratamiento del contenido de comunicación humana, principalmente en textos. A menudo las bases de datos son características de estos programas.

4.2.2.1. Análisis cualitativo: buscar regularidades y diferencias en el texto, explorar el texto entero (QDA –Qualitative Data Analysis, análisis cualitativo de datos). Algunos programas permiten el procesamiento de la información de audio y de vídeo. No hay un paradigma común de QDA, existiendo muchos enfoques.

AQUAD FIVE 5.8

Autor: Günter L. Huber (en alemán; también existe versión española: AQUAD CINCO 5.8). Distribuidor: Verlag Inge Huber, Schwangau. Documentación: manual inglés, manual alemán y manual español. Sistema (s) operativos: Win98, WinNT. Descripción: interpretación de textos mediante codificación simple del texto sin guía previa de categorías, o mediante un proceso semiautomático de codificación a partir de listados de ocurrencia de términos prefijados, que aparecen destacados sobre la pantalla siguiendo el texto. Los programas de análisis permiten desde el recuento de palabras aisladas, hasta la confección de listas de palabras (“diccionarios”), diferenciación de hablantes, codificación de las partes de textos producidos por hablantes diferentes, etc. Los segmentos de texto pueden ser recuperados por número de archivos de texto, números de línea del texto, código, palabra clave, etc. (la búsqueda texto es completa). Permite la comparación de casos/textos mediante análisis booleanos de rasgos críticos (“minimización lógica de tablas de contingencia”).

ATLAS.ti

Autor: Thomas Muhr. Distribuidor: Sabio Scolari. Documentación: manual corto en formato de PDF. Descripción: Tratamiento de datos textuales, gráficos, de audio, y de vídeo mediante codificación intuitiva y fácil sobre la pantalla (arrastre con el ratón) con definición flexible de segmentos de datos. Permite un número prácticamente ilimitado de documentos, segmentos, códigos, y notas, así como “trazar mapas mentales” y asignación de anotaciones a todo el tipo de unidades, segmentos de datos, códigos, notas, etc. Los datos enlazados pueden ser manejados mediante enlaces de hipertexto. La generación de textos en HTML permite formatear la investigación para su publicación en la red.

CODE-A-TEXT MULTI MEDIA V4

Autor: Alan Cartwright. Distribuidor: Sabio Scolari. Sistema (s) operativos: Win98, WinNT. Descripción: es un instrumento flexible que ayuda almacenar y cifrar prácticamente cualquier clase de datos tanto para métodos cualitativos como para cuantitativos de análisis. Además de datos basados en tex-

tos, el también acepta el sonido y el vídeo. Para los que trabajan con textos basados en la palabra escrita, es capaz de hacer ver el texto y oír las palabras habladas. Esta funcionalidad es útil para quienes trabajan en las áreas de psicología clínica, psicoterapia, y otras profesiones que confían en el análisis a fondo de textos y datos orales registrados.

C-I-SAID - Code-A-Text Integrated System for the Analysis of Interviews and Dialogues

Autor: Alan Cartwright. Distribuidor: Sabio Scolari. Sistema (s) operativos: Win98, WinNT. Descripción: CISAID se ha desarrollado como sistema multimedia de Códigos para análisis de texto y ha sido diseñado para suministrar métodos e instrumentos tanto para el análisis cualitativo como para el cuantitativo conservando los enlaces a los datos de la fuente (texto, audio, o vídeo). En CISAID se ofrecen modos sofisticados de organizar y buscar datos de la fuente textual, mediante tabulaciones al margen y representaciones gráficas de esto datos. Se incluyen pruebas estadísticas que a menudo suministran un modo de indicar la fuerza comparativa de una relación.

KWALITAN 5.05

Autor: Vincent Peters. Distribuidor: Vincent Peters. Sistema (s) operativos: DOS (4.0), Win95, 98, WinNT. Descripción: Kwalitan es un programa para el análisis de datos cualitativos y es un programa de base de datos de objetivo especial y basado en prueba experimental de teorías. Hasta 25 códigos puede asociarse a un segmento de texto. Kwalitan da una descripción alfabética de los códigos ya asociados a los segmentos seleccionados, juntos con sus frecuencias y puede crear una estructura de árbol para establecer la estructura jerárquica de los códigos. Los códigos y sus combinaciones pueden ser buscados, también mediante la lógica booleana. Kwalitan posee algunos rasgos para el análisis de contenido, por ejemplo las frecuencias de palabras, las descripciones de 'palabra clave en contexto' y de 'palabra clave fuera del contexto'. Matrices de frecuencias de palabras o códigos pueden ser generadas y analizadas por su software estadístico

TEXT EZ 3.06C

Autor: Conwal Incorporated. Distribuidor: CDC Sistema (s) operativos: Win98, WinNT. Descripción: "CDC el TEXTO EZ" es un programa de software desarrollado para asistir a investigadores en creación, manejo, y análisis semi-estructurado de bases de datos cualitativos. La entrada de datos permite a los investigadores crear libros de códigos (codebooks) en línea, aplicar códigos a segmentos de respuestas abiertas, desarrollar estudios de caso, conducir búsquedas sobre bases de datos para identificar los segmentos de texto que muestran condiciones especificadas por el usuario, y exportar datos en una amplia serie de formatos para el análisis remoto con otros programas de software de análisis cualitativos o estadísticos. La capacidad de exportar e importar el codebook (libro de códigos) ayuda a coordinar esfuerzos de codificación cuando se trabaja simultáneamente con copias del mismo archivo de base de datos.

HYPERQUAL 3.0 version 1.1

Autor: Raymond V. Padilla. Distribuidor: Hyperqual. Sistema operativo: MacOS 8 o mejor. Descripción: HyperQual3 es un instrumento para el almacenaje, dirección, organización, y análisis cualitativo de datos de texto. Maneja con eficacia entrevistas estructuradas e inestructuradas, observaciones, documentos, y notas de investigación. Rasgos: marcación de texto semi-automática, y automática (codificación). Clasificación fácil de segmentos de texto codificados. Muestra categorías disponibles mientras se codifican los datos. Analiza notas de investigación como datos. Exporta fácilmente a cualquier procesador de texto.

HYPERRESEARCH 2.0

Autor: S. Hesse-Biber, P. Dupuis, T. S. Distribuidor: Sabio, Inc. Sistemas operativos: Win95, Win98, WinNT, MacOS 7.5, PowerPC. Descripción: Fácil de usar. Interfaz: destaca menús desplegables. Permite trabajar con texto, material gráfico, material de audio, y de vídeo. Los informes extraídos incluyen hiperenlaces al material subyacente de la fuente. Funciones comprensivas cifrar y recuperar: permite seleccionar cualquier segmento de texto (desde 1 carácter a un archivo entero) y aplicar sin limitaciones nombres de código al segmento de texto seleccionado. Recuerda códigos nominales y los selecciona basandose en la proximidad a otros códigos. Los informes pueden ser

presentados mediante enlaces de hipertexto al material original de la fuente. Permite autocodificación: cifra automáticamente fuentes múltiples a casos múltiples, buscando frases múltiples en un solo paso. Especifica igualmente una porción de caracteres, palabras, o líneas antes y/o después de las frases encontradas para ser incluidas en el segmento seleccionado para la autocodificación. Permite trazar un mapa de código: usando la ventana de mapa de código representa gráficamente relaciones entre sus códigos. Permite pruebas de hipótesis: posee un sistema experto para realizar análisis a fondo de los datos codificados para ver si la codificación apoya las hipótesis previstas.

MARTIN 2.0

Autor y distribuidor: Robert Schuster. Sistemas operativos: Win3.1, Win95, 98, no WinNT. Descripción: Los diferentes pasos de ejecución son seleccionados por el teclado o con un ratón. A partir de la selección de un paso, un marcador numerado es agregado al texto y automáticamente es copiado a un índice por separado que aparece en ventana: "la tarjeta de índice". Cada tarjeta puede ser identificada con una descripción opcional breve y con cualquier número de palabras claves opcionales. Las anotaciones pueden ser asociadas a cada paso sin salir del programa. La longitud de anotaciones es limitada principalmente por los recursos del ordenador. Aunque sea posible unir unos pasos a otros por el rasgo de palabra clave opcional, las tarjetas también pueden ser relacionadas de manera informal apilando las sobre la pantalla de ordenador. Como las relaciones entre pasos comienzan a aparecer, éstas pueden ser formalizadas moviendo tarjetas según una jerarquía de grupos y carpetas. Las tarjetas también pueden ser movidas entre carpetas y pueden ser duplicadas para la inclusión en carpetas múltiples. Los proyectos de MARTIN están basa-

dos en la asociación física entre las partes de un análisis más que sobre estructuras de codificación formales; todas las descripciones, palabras claves, y comentarios al margen puede ser modificados en cualquier momento siguiendo el progreso de análisis. Y como cada uno de los componentes de un análisis es manejado como un objeto separado y ocupa su propia ventana, "el escritorio" de MARTIN puede mostrar tanto elementos del análisis como deseos tenga el investigador dentro de los límites de las ventanas y los recursos del ordenador. Las carpetas y los grupos de carpetas de MARTIN son continuamente jerarquizados. Los estudios llevados a cabo por separado por diversos investigadores pueden ser unidos unos a otros importando cualquiera de sus componentes como textos para el análisis remoto. Textos originales pueden ser impresos con sus marcadores de texto, apuntes, y pasos extraídos. Los pasos pueden ser impresos con la información de la fuente, apuntes, anotaciones, y palabras claves. Todas las opciones de impresión generan un archivo imprimible en ASCII, WordPerfect, o formato de Microsoft Word.

THE ETHNOGRAPH 5.04

Autor: John V. Seidel. Distribuidores: Sabio Scolari. Sistemas operativos: Win3.1, Win95, 98, WinNT. Descripción: El Ethnograph v5.0 para ordenadores personales sistema operativo Windows es un programa del ordenador versátil diseñado para hacer el análisis de datos durante la investigación cualitativa más fácil, más eficiente, y más eficaz. Puede importar directamente datos cualitativos basados en documentos de texto, desde cualquier procesador de texto. El Ethnograph ayuda buscar y anotar los segmentos que sean de interés dentro de un proyecto, marcarlos

con palabras de código y controlar los análisis que puede luego ser recuperado para su inclusión en informes. Su primera versión fue lanzada en 1985, y fue uno de los primeros programas para promover el uso del ordenador para el análisis de datos cualitativos. Desde entonces el Ethnograph ha seguido siendo desarrollado por analistas de datos cualitativos para uso de analistas de datos cualitativos como científicos sociales, historiadores, críticos literarios, investigadores del campo de la salud, analistas de negocio y de mercado, investigadores legales y otros.

QMA Qualitative Media Analysis, former KIT

Autor y distribuidor: Carl Verner Skou. Sistemas operativos: Win98, WinNT. Descripción: QMA es un programa para el análisis cualitativo de grabaciones. Puede tratar un máximo de 300 entrevistas (de hasta 18 horas cada una). Puede implementar códigos en una estructura de árbol, y compilarlos en grupos lógicos. El número de códigos usados para el análisis puede ser 65535, que combinados puede llegar a varios millones (dependiendo del tamaño RAM). Cada nota puede ir compuesta de 32000 caracteres. El análisis puede efectuarse a partir de

una buena porción de criterios: por ejemplo miembros de un grupo o hablantes. QMA puede encontrar el modelo de análisis sobre la base de variables independientes, orden temporal de intervenciones, integración en el grupo, qué persona habla - y combinaciones de todo ello. Todo el conjunto, apuntes y resultados, puede ser exportado por separado en archivos para el análisis avanzado en Excel o SPSS. QMA puede trabajar con los formatos siguientes de multimedia: Grabaciones de audio: WAV, AIFF, MP3, y Grabaciones en vídeo: AVI, Apple QuickTime 1.0-2.0, MPEG-1, MPEG-2 y Archivos de música: MIDI.

QSR NUD*IST 5 (N5)

Autores: Lyn y Tom Richards. Distribuidores: Sabio. Sistemas operativos: Win95, 98, WinNT. Descripción: diseñado para investigadores que hacen uso de datos complejos, N5 ofrece un juego de herramientas completo para la codificación rápida, la exploración cuidadosa y la dirección rigurosa del análisis. Con las ordenes precisas para automatizar la codificación y la búsqueda, N5 sirve una amplia gama de métodos. Sus archivos de funciones y procedimientos de importación hacen del programa un sistema rápido, ligando datos cualitativos y cuantitativos. Los documentos

son importados por separado o en grandes cantidades, en texto simple o con formato automático ligado a la unidad escogida de texto. Codificación sobre pantalla, con acceso inmediato al sistema de códigos. El material cifrado es mostrado para la reflexión, la revisión de codificación y la codificación sobre categorías nuevas. Con las búsquedas de codificación, el investigador puede probar hipótesis, localizar un modelo explicativo o perseguir una línea de hipótesis ligadas a una conclusión prevista o alternativa.

QSR NVivo 1.3

Autores: Lyn y Tom Richards. **Distribuidor:** Sabio. **Sistemas operativos:** Win95, 98, WinNT. **Descripción:** NVivo puede manejar datos de texto enriquecido –con plena capacidad de corregir, visualmente, para cifrar, y unir documentos cifrados, filtrados, codificados y encontrados mediante búsquedas. En NVivo, se puede crear, corregir, cifrar y explorar documentos compuestos, encajados mediante hipervínculos a archivos de multimedia y webs encajados entre ellos según datos e ideas. Hay varios modos de automatizar la codificación de documentos. Los documentos y nodos de enlace pueden ser entrelazados por hipertexto para suministrar redes. De este modo los documentos “de anotaciones”, de comentarios pueden ser ligados a nodos y a otros documentos; y los nodos pueden ser contruidos para actuar como referencias “de extracto (resumen)” para los temas mencionados en los documentos. Los documentos son totalmente editables, y la corrección no trastorna o invalida la codificación

previa. Este rasgo de “revisión mientras usted codifica” permite a los usuarios aumentar sus documentos con el tiempo, y cifrar y unir el texto agregado– así como corregir el texto ya cifrado. Puede darse a los documentos y nodos de análisis (y lo que ellos representan) un número indefinido “de atributos” (propiedades) para describirlos y caracterizarlos y usarlos además como filtros. También la importación y exportación de datos de atributos, vía tablas de contingencia, puede llevarse a programas estadísticos. Los atributos pueden tener cualesquiera valores relevantes (fechas, números, composiciones lógicas, secuencias en cadena, etc.). Los documentos y nodos pueden ser agrupados en un número indefinido “de juegos” para cualquier objetivo que desee alcanzar el usuario, y pueden ser estudiados por sus rasgos, contenido, etc. Un integrado “instrumento de búsqueda” suministra un número grande de caminos para buscar rasgos, incluyendo resúmenes tabulares estadísticos.

MAXQDA, o German MaxQDA

Autor: Udo Kuckartz **Distribuidores:** VERBI GmbH Ana Kuckartz. **Sistemas operativos:** Win95, 98, WinNT, versión inglesa y alemana. **Descripción:** la versión anterior de este programa es Winmax. Con MAXQDA se puede: crear e importar textos en Formato RTF; crear grupos de texto y organizar material textual para el análisis temporal; crear códigos jerárquicos con hasta diez niveles; seleccionar segmentos de texto y asignar códigos o subcódigos ligados a un color como atributo de cada código. Muestra el texto con rayas de codificaciones que indican que los códigos han sido ligados entre sí. Permite importación pre-cifrada, por ejemplo de focus groups o entrevistas. Selecciona palabras o términos (condiciones) del texto o los inserta automáticamente en su sistema de código (“in-vivo-coding”) Busca la palabra o

las combinaciones de palabras en sus textos. Enlaza iconos especiales a los tipos diferentes de notas (“notas de teoría”, “notas metodológicas” etc.). Maneja un juego de atributos (“variables”) para cada texto, usando las variables como con filtro para la recuperación de texto o búsquedas léxicas. Exporte la tablas de variables a programas estadísticos como Excel o SPSS. Un juego de hasta 10 funciones analíticas permite la recuperación de texto (“cerca”, “rodeado por”, “seguido de”, “al menos” etc.). Es capaz de facilitar el manejo tanto de pequeños estudios así como conjunto grandes de texto.

4.2.2.2. Transcripción de software (audio y video)

SALT 6.1 - Systematic Analysis of Language Transcripts

Autores: Robin S. Chapman y Jon F. Molinero. Distribuidor: Laboratorio de Lengua, Universidad de Wisconsin-Madison. Sistemas operativos: Win95, 98, WinNT, MacOS. Descripción: el programa de SAL contiene todo un surtido de análisis estándar. Incluye: los tipos de expresiones que remiten a distribución de imitaciones, respuestas a preguntas, expresiones incompletas, ininteligibles, y no verbales; el cálculo del número total de palabras, la longitud de expresión, y la lingüística de Brown; número y longitud de pausas y tarifa de oratoria; listas y frecuencias de raíces de palabra, morfemas ligados, y códigos; las distribuciones de expresiones por longitud en los términos y criterios de palabras y morfemas; distribución de hablantes y turnos de palabra según criterios de expresiones consecutivas; frecuencias para juegos de palabras, incluyendo palabras de pregunta, formas negativas, conjunciones, modales y semi-verbos –auxiliares-, pronombres, y cualquier juego de palabras que se quiera definir; número y tipos de circunloquios (pausas llenas, repeticiones, revisiones). Los valores de estas variables pueden ser comparados con la Base de datos de Referencia de SAL que contiene datos empíricamente extraídos de los niños del Wisconsin según categorías de edad diferentes (3-13 años), género, probando el contexto, y la longitud de transcripción. Se contempla la estadística de desviación estándar para muchas de las variables de análisis.

TRANSCRIBER 1.4.3

Autor: Claudio Barras y muchos otros. Documentación: manual de usuario (en francés) o manual de referencia (en inglés) Sistemas operativos: MacOS, Linux (Win98 en desarrollo). Descripción: Transcriber es un instrumento para ayudar a la anotación manual de señales de discurso. Suministra un interfaz de usuario fácil de usar, gráfico, para segmentar grabaciones de discurso de larga duración, transcribiéndolas, y etiquetando turnos de discurso, cambios de tema y condiciones acústicas. Esta expresamente diseñado para la anotación de grabaciones de noticias, pero sus características podrían también ser útiles en otras áreas de investigación de discurso.

ESA - Event Structure Analysis

Autor y distribuidor: David Heise. Sistema operativo: JAVA. Descripción: ESA es un programa en línea que ayuda a analizar acontecimientos secuenciales. El análisis previamente necesario produce un diagrama que muestra cómo algunos acontecimientos (por ejemplo, noticias) son necesarios para que ocurran otros acontecimientos, y cómo acontecimientos en abstracto pueden ser representados como acontecimientos concretos. Se puede probar si el diagrama resultante trabaja como una gramática para explicar la secuencia de acontecimientos o de noticias. El análisis por codificación comprende al agente, la acción, el objeto, el instrumento, el ajuste, la alineación, el producto, y el beneficiario de cada acontecimiento. Esto permite determinar cómo la gente, las cosas que se hacen, y las consecuencias resultan unidos por los acontecimientos del discurso o por el discurso de los acontecimientos (noticias).

KEDS - Kansas Event Data System

Autor y distribuidor: Philip A. Schrodtt. Sistema operativo: MacOS. Descripción: el Sistema KEDS es un programa para la codificación de datos de noticias internacionales que usan el modelo y el simple análisis gramatical. Está diseñado para trabajar con artículos cortos de noticias como aquellos

encontrados en informes de servicio de telégrafo o cronologías. Hasta el momento, KEDS principalmente ha sido usado para cifrar noticias del servicio de telégrafo de Reuters, pero en principio puede ser usado para otro tipo de referencias a acontecimientos mediante cifrado de esquemas. Hay también una lista de correo electrónico.

4.2.2.3. Análisis de datos sobre acontecimientos de referencia: análisis de acontecimientos (o noticias) en datos textuales.

4.2.2.4. Análisis cuantitativo: se analiza el texto para probar selectivamente hipótesis y formular inferencias estadísticas. El resultado es una matriz de datos que representa los resultados numéricos de la codificación.

4.2.2.4.1. Empleando sistemas de categorías: ya sean estos sistemas proporcionados (como herramienta) por el programador de software, o por el propio investigador (como modelo de representaciones), a fin de ser utilizados como criterios de búsqueda en el texto y facilitar codificaciones. Los paquetes del software con diccionarios incorporados son a menudo restringidos a ciertos idiomas, algunos se encuentran limitados por el tamaño de la unidad de texto y se restringen a la tarea de procesar respuestas a preguntas abiertas, siendo incapaces de analizar textos de medios de comunicación. Las categorías pueden ser temáticas o semánticas; esto puede tener implicaciones sobre la definición de unidades de texto y variables externas.

PCAD 2000

Autor y distribuidor: Software de GIGAOCTETO. Sistema (s) operativos: Win95, 98. Descripción: su área primaria de interés es medir estados psicobiológicos relevantes como la ansiedad, la hostilidad, y la esperanza mediante escalas de análisis de contenido Gottschalk-Gleser. Estas escalas han sido empíricamente desarrolladas y probadas, y han demostrado ser confiables y válidas en una amplia gama de estudios.

PROTAN - Protocol Analyser.

Autor y distribuidor: Robert Hogendraad. Sistema (s) operativos: DOS, MacOS, OS/2. Descripción: lista de palabras, concordancias, frecuencias de categorías, secuencias de manuales de categorías en forma electrónica e impresa. PROTAN es el sucesor en francés, del GENERAL INQUIRER, con muchas utilidades que realizan numerosas tareas de análisis de texto. PROTAN es muy complejo y difícil de manejarse. La documentación está en francés, pero las utilidades están en inglés.

DICCIÓN 5.0

Autor: Roderick F. Venado. Distribuidor: Sabio Scolari. Sistema operativo: Win95, 98, WinNT. Descripción: Dicción 5.0 emplea diccionarios (listas de palabras) para buscarlas en un texto con las siguientes variables: Certeza: indica resolución, inflexibilidad, y entereza y una tendencia de hablar ex-cathedra; Actividad que destaca movimiento, cambio, la puesta en práctica de ideas y la anula-

ción de cualquier inercia; Optimismo que describe referencias tangibles, inmediatas, reconocibles según afectan diariamente a la vida de la gente de forma confiada; Concordancia destacando el acuerdo sobre los valores de un grupo y rechazando los modos de acuerdo idiosincráticos; Realismo que describe referencias tangibles, inmediatas, reconocibles según afectan diariamente a la vida de la gente de forma probada (realista)

LIWC 2.0 - Linguistic Word Count

Autor: James B. Pennebaker. Distribuidor: Erlbaum asociados. Sistema operativo: Win95, 98, WinNT. Descripción: el programa analiza archivos de texto sobre una base de sinónimos, interesándose por los porcentajes de las palabras que se emparejan en cada una de las dimensiones de lengua. El programa tiene 68 dimensiones predeterminadas (variables de salida) según inclusión de dimensiones lingüísticas, para categorías de palabras que dan un toque psicológico, y categorías de preocupaciones (interés) personales, pudiendo también acomodarse a dimensiones definidas también por el usuario.

INTEXT 4.1 - INhaltsanalyse von TEXTen (análisis de contenido de textos)

Autor y distribuidor: Harald Klein. Sistema operativo: DOS. Descripción: Intext es la versión de MS-DOS de Textquest. Este usa diccionarios para cifrar textos, con rasgos especiales como son la codificación interactiva, el modelo de búsqueda de co-ocurrencias de palabras, y la detección de la negación. Pero también analiza la legibilidad y los estadísticos de texto así como de secuencias de palabras. Está disponible en inglés y el alemán. La ayuda en línea está disponible. La Versión windows de TextQuest tiene un interfaz de usuario moderno.

TEXTPACK 7.0 – TextPackage

Autores: Peter Ph. Mohler, Cornelia Züll. Distribuidor: ZUMA Mannheim. Sistemas operativos: Win95, 98, WinNT, en inglés o español. Descripción de sus rasgos: las frecuencias de palabra en un texto entero o en sus sub-unidades, pueden ser filtradas por variables externas (identificadores) y/o por su frecuencia, clasificadas alfabéticamente o por frecuencia, tablas de orden (jerárquico) según clases. Analiza “palabras-clave-en-contexto” y “palabra clave del contexto” (KWIC/KWOC) Singulariza palabras, raíces de palabra (los principios

de cada cadena de palabras) y puede mostrar secuencias de palabra en su contexto. TEXTPACK categoriza/clasifica un texto según un diccionario de usuario. Esto genera tantos archivos como frecuencias de categoría y secuencias de categorías haya. La validez de la codificación puede ser comprobada por varias opciones (por ejemplo, la inserción de números de categoría o etiquetas de categoría a lo largo del texto). Selección de unidades de texto: filtrado sobre la base de las variables externas o usando un archivo numérico para seleccionar unidades de texto.

GENERAL INQUIRER

Autor y distribuidor: Philip J. Piedra. Sistema operativo: JAVA, cuyos sistemas de categorías son Archivos Excel (XLS). Descripción: el abuelo del software de análisis de contenido está ahora disponible para los ordenadores que controlan JAVA y son capaces de leer el sistema de categoría (archivos de Excel).

WORDSTAT 3.1c

Autor: Normand Peladeau. Distribuidor: Provalis Investigación. Sistemas operativos: Win95, 98, WinNT. Descripción: WordStat es un programa de estadística de objetivo general (comparable al SPSS por ejemplo). Ambos paquetes están integrados y es sobre todo útil para la codificación de respuestas a preguntas abiertas.

4.2.2.4.2. Sin emplear sistemas de categorías: basados en la utilización de las co-ocurrencias de palabras/frases y/o conceptos. Cada palabra/o segmento de texto se toma en consideración.

HAMLET

Autor: Alan Brier. Sistema operativo: DOS, Win3.1, Win95, 98, WinNT. Descripción: la idea principal de HAMLET es de buscar un archivo de texto para palabras en una lista de vocabulario dada, y contar frecuencias conjuntas dentro de cualquier unidad de contexto especificada. Frecuencias de palabra individuales (f_i), frecuencias conjuntas (f_{ij}) para pares de palabras (yo, j), tanto expresadas en los términos (condiciones) de la unidad escogida de contexto, como en correspondencia estandarizada de frecuencias conjuntas: $S_{ij} = (f_{ij}) / (f_i + f_j - f_{ij})$. Una opción remota permite la comparación de los resultados al aplicar el escalamiento multidimensional a matrices de frecuencias conjuntas sacadas de una porción de textos; ofrece funciones como KWIC (listados, "palabras clave en el contexto" para cualquier cadena de palabras dada; WORDLIST genera listas de palabras y frecuencias; COMPARE, compara palabras de listas comunes a pares de textos, y es útil para la generación de listas de vocabulario, para la inclusión de sinónimos, para el empleo en la comparación de una porción de textos, etc.

INTELLIGENT MINER FOR TEXT - TEXT ANALYSIS TOOLS 2.3

Autor y distribuidor: IBM. Sistemas operativos: AIX 4.2.1, Solaris 2.5.1, WinNT 4.0, OS/390. Descripción: instrumentos de análisis de texto que pueden ser usados para analizar todos los tipos de documentación en línea, de peticiones de cliente e informes técnicos en artículos de revistas y periódicos. Organiza la documentación creando una estructura de directorio de carpetas, con instrumentos de clasificación que pueden catalogar y clasificar artículos según categorías definidas por el usuario. Navega por la documentación buscando la información sobre un tema específico, y se pueden usar muchos documentos diferentes en cada búsqueda. Los instrumentos de Clustering pueden suministrar una descripción de toda la documentación que ha sido usada. Resume la documentación y detalla la información, mediante resúmenes de documento automáticamente creados, y aportando sumarios para determinar si el documento entero debería ser leído.

SEMIO 2.0

Autor: Claudio Vogel. **Distribuidor:** Semio. **Sistema (s) operativos:** Win95, 98, WinNT, Solaris 2.5. **Descripción:** La taxonomía de Semio combina la tecnología de análisis lingüístico y de clustering estadístico con exigencias de vocabulario definidas por el usuario para crear por intuición la estructura de categorías que suministra el acceso inteligente al espacio global de la información dentro de una masa de texto inestructurado. Frases importantes y palabras clave son extraídas de una variedad de fuentes de texto como sitios de intranet/Internet, Lotus Notes, Documentum, bases de datos ODBC-, XML, etc. Este proceso combina la detección de lengua, el análisis de proximidad y la extracción de texto más limpia, más informativa disponible. Una vez que este proceso ha seleccionado la información realmente relevante del original de texto inestructurado, puede serlo aplicado un número indefinido de estructuras de clasificación de alto nivel. Estas estructuras extraen derivados léxicos y los colocan en categorías. El resultado: una estructura de categorías que en realidad suministra ideas al usuario sobre el espacio de búsqueda sin recurrir al método 'hunt-and-peck' de búsquedas de palabra clave. Ya que la única exigencia de una estructura de clasificación es que esto refleja la informa-

ción que puede ser encontrada dentro del texto fuente, de forma que la configuración y la personalización de la estructura es prácticamente ilimitada. El cliente puede configurar sus taxonomías para reflejar un tesoro corporativo o un vocabulario controlado. La taxonomía de Semio es totalmente compatible con la ISO thesauri, y puede ser adaptada a cualquier iniciativa de terminología por parte del usuario. El poder de aplicar estructuras de clasificación múltiples al mismo texto de la fuente se hace claro cuando los usuarios ven por primera vez la prueba real textual que condujo en primer lugar a aquellas estructuras. **Pasos de Proceso:** La taxonomía de Semio realiza un proceso de tres pasos para clasificar el contenido de texto. El texto es recogido de fuentes diferentes aproximadamente de 500 formatos diferentes que pueden ser leídos. Permite la extracción de frases relevantes, informativas dentro el texto. Las frases son asociadas a un juego de categorías que pueden proceder de un tesoro, pre construido de la Semio, o de una estructura de encargo. Las estructuras de categorías entonces pueden ser validadas y modificadas en un proceso fácil, iterativo para asegurar la calidad y la consistencia.

SPAD-T

Autor y distribuidor: CISIA-CERES-TA. **Sistemas operativos:** Win95, 98, WinNT. **Descripción:** la SPAD-T analiza los textos automáticamente asociando la información numéricamente cifrada. La clasificación también puede tomar variables externas (por ejemplo la edad, el sexo, la profesión) en consideración la utilización de la SPAD-N. La SPAD-T cuenta palabras y secuencias de palabra (frases) que usan tablas de orden según clase y criterios de exclusión como la longitud o la frecuencia. Usando palabras caracterizadas por méto-

dos de probabilidad, se encuentran las secuencias de palabras, u oraciones más probables. También efectúa KWICs con una longitud de línea fija de 132 caracteres. Compara vocabularios de textos con diferentes tipos de análisis de factorial y análisis de correspondencia. También se pueden incluir variables externas. Son también posibles tablas de contingencia de palabras comunes o los segmentos repetidos dentro de los textos. Los análisis jerárquicos permiten por ejemplo una clasificación automática de respuestas a preguntas abiertas.

TEXTANALYST 2.0
(versión alemana)

Autor: Sergej Ananyan. Distribuidor: Megaputer. Sistemas operativos: Win95, 98, WinNT. Descripción: TextAnalyst es un instrumento de tratamiento de texto capaz de análisis semántico automatizado. Además, TextAnalyst puede ayudar a realizar clustering de documentos en textbase, a la recuperación semántica de documentos, y a enfocar (concentrar) una exploración de texto alrededor de un cierto tema.

T-LAB

Autor: Franco Lancia y Marco Silvestri. Distribuidor: Computer Data Sistemi. Sistemas operativos: Win95, 98, WinNT. Descripción: el T-LAB, compara y traza un mapa del contenido de clases diferentes de textos: transcripciones de discurso, artículos de periódico, textos legislativos, documentos (inclusión de transmisiones de Internet), libros, respuestas a preguntas abiertas, etc. Rasgos básicos son: las Asociaciones de palabras, trazar un mapa de correspondencias, especificidades, concordancias, nota, vocabulario. Sus rasgos avanzados son: marcación, selección de palabra, análisis de correspondencia, análisis de redes, mapa de co-ocurrencias, análisis de contexto elemental, análisis de secuencia

4.2.2.5. Codificación de respuestas a preguntas abiertas: estos programas no pueden analizar una cantidad grande de textos, sólo pueden servir para ciertos textos homogéneos y a menudo son limitados según el tamaño de una unidad de texto.

TEXTSMART 1.0

Distribuidor SPSS Inc. Sistema operativo: Win95, 98, WinNT, el programa está en lenguaje Java. TextSmart es un instrumento para la codificación automática de respuestas a preguntas abiertas, y no puede cifrar textos muy largos. La clasificación requiere preselección de palabras y la definición de listas (palabras y sus sinónimos), y trabaja sólo con textos ingleses.

VERBASTAT 3.0.

Distribuidor: SPSS Inc. Autor: DataStat, empresa que fue comprada por SPSS. Sistema operativo: Win95, 98, WinNT. Descripción: VerbaStat es el software para codificar respuestas a preguntas abiertas. Suministra una especie de máscara, la codificación tiene que ser hecha a mano.

VERBATIMBLASTER

Autor: David S. Walonick. Distribuidor: StatPac Inc, EE. UU. Sistema operativo: Win95, Win 98, WinNT, Win2000 Descripción: El programa VERBATIMBLASTER es una parte de StatPac y codifica respuestas a preguntas abiertas automáticamente.

5. Conclusiones

En las páginas precedentes se le ha dedicado, en primer lugar, un apartado a la reflexión epistemológica sobre el *análisis de contenido* como objeto de estudio caracterizado por constituir una práctica de investigación en ciencias sociales que consiste, a su vez, en construir un objeto de estudio (por ejemplo, una comunicación singular) para cuya representación científica (trascendente y por tanto refutable) se usa como procedimiento la selección y análisis de productos comunicativos que objetivamente pertenecen a aquella comunicación, que son significativos para conocerla desde una perspectiva teórica, y que son válidos o pertinentes para poner a prueba las condiciones, tanto particulares como generales, en virtud de las cuales estos productos comunicativos han sido empleados, o pueden serlo, en cualquier tipo de comunicación semejante a aquella que los ha producido.

En segundo lugar, le hemos dedicado un apartado a la metodología (reglas de procedimiento) que aspira a ligar de forma válida una secuencia de actuaciones y alternativas que atañen a la selección del corpus, a su segmentación en unidades de análisis y a su examen orientado a elaborar, registrar y tratar datos de diversos órdenes lógicos, gracias a los cuales se construye un metadiscurso capaz de representar científicamente (o de forma trascendente o refutable) al objeto de estudio que en ningún caso es el propio corpus, sino la práctica humana que ha generado el objeto material de análisis: el producto comunicativo analizado.

En tercer lugar, se le ha dedicado un último apartado a las alternativas técnicas que subyacen a cualquier procedimiento de análisis, ya sea este realizado en todos sus pasos por los investigadores y analistas, ya sea este realizado automáticamente sirviéndonos de los recursos de software disponibles, y brevemente repertoriados y descritos siguiendo el trabajo de Harald Klein.

Es de esperar que ahora cobre sentido una última reflexión con la que pretendo concluir esta exposición: en análisis de contenido, como en cualquier otro procedimiento de investigación científica, la verdad o falsedad nunca son imputables a los métodos, sino a las teorías a que ellos sirven; la precisión o el error de unos datos tampoco son imputables a los métodos, sino a las técnicas para su elaboración, registro y tratamiento; a los métodos (y el análisis de contenido es un método) le compete la adecuación o inadecuación de los procedimientos para servir a una teoría, y la adecuación o inadecuación de los procedimientos para elegir y usar unas técnicas capaces de proporcionar datos susceptibles, por una parte, de refutar una teoría, y, por otra parte, susceptibles de evitar errores formales o materiales.

Referencias bibliográficas

- Berelson, B. (1952). *Content analysis in communication research*. Nueva York: The Free Press.
- Eco, U. (1975). *La estructura ausente*. Barcelona: Lumen. [Edición original italiana, 1968].
- Gaitán, J.A. (1991). *Historia, comunicación y reproducción social en la transición española. Análisis de las expresiones generales y universales de la representación del acontecer en un diario de referencia dominante: El País, 1976, 1981*. Madrid: Universidad Complutense de Madrid.
- Gaitán, J.A. & J.L. Piñuel Raigada (1997). *Técnicas de investigación social. Elaboración y registro de datos*. Madrid: Síntesis.
- George, A.L. (1959). *Propaganda analysis: A study of inferences made from Nazi propaganda in World War II*. Evanston, IL: Row Peterson.
- Hall, E.T. (1973). *La dimensión oculta. Enfoque antropológico del uso del espacio*. Madrid: Instituto de Estudios de Administración Local. [Edición original: (1966). *The hidden dimension*. Nueva York: Doubleday & C].
- Holsti, O.R. (1969). *Content analysis for the social sciences and humanities*. MA: Reading Addison-Wesley.
- Ibáñez, J. (1986). *Más allá de la sociología*. Madrid: Siglo XXI.
- Klein, H. *Text Analysis Info Page*. URL: <http://www.textanalysis.info/html>
- Lasswell, H. (1965). "Why be quantitative?". In H. Lasswell & N. Leites (comps.), *Language of politics*. Cambridge: MIT Press, 142-44.
- Lippmann, W. (1922). *Public opinion*. Nueva York: Macmillan.
- Martín Serrano, M. (1989). *La producción social de comunicación*. Madrid: Alianza Universidad.
- Martín Serrano, M. et al. (1981). *Epistemología de la comunicación y análisis de la referencia*. Madrid: Visor.
- Piñuel Raigada, J.L. (1989). *La expresión. Una introducción a la filosofía de la comunicación*. Madrid: Visor.
- Piñuel Raigada, J.L. (1993). *Cultura política y TV en la transición en Chile*. Madrid: C.E.D.E.A.L.
- Piñuel Raigada, J.L. & J.A. Gaitán (1995). *Metodología general. Conocimiento científico e investigación en la comunicación social*. Madrid: Síntesis.
- Piñuel Raigada, J.L. y J.I. García-Lomas (2001). "Autopoiesis y comunicación". Research Group nº 51 (ISA), *III Congreso Internacional de Sociocibernética*. León, Guanajuato (México), 26-30 junio 2001.
- Pool, I. de S. (1959). *Trend in content analysis*. Urbana, IL.: University of Illinois.
- Rapoport, A. (1969). *Hous form and culture*. Nueva York: Prentice Hall.
- Todorov, T. (1982). "Las categorías del relato literario". In VV.AA., *Análisis estructural del relato*. Buenos Aires: Ediciones B. [Texto original, 1966].