

El análisis contextual: aplicación CONTEXT (SOCIOS)*

José Manuel Cornejo
Universitat de Barcelona

Este artículo plantea el estudio del lenguaje como fenómeno de primera magnitud para avanzar en la investigación y el conocimiento en las ciencias humanas y sociales. Se revisa la tradición del Análisis de Contenido y se operativiza la noción de contexto psicosocial como elemento referencial necesario en la captación del conjunto de significados, personales y sociales, que vehicula. Se presenta la aplicación informática ConText, dentro de los módulos del sistema SOCIOS, especialmente diseñado en el Laboratorio de Psicología Social de la Universidad de Barcelona para el tratamiento de datos de interacción social, aplicándolo en esta ocasión a datos de contenido verbal, desde la perspectiva del “Análisis Textual” o “Text Mining”. Se aportan criterios teóricos y metodológicos de utilización y se detallan algunas de sus características y procedimientos principales, tratando de mostrar su utilidad para la investigación. Se realiza una ejemplificación sobre la base de un corpus de definiciones personales sobre la noción de ‘Felicidad’ en una muestra ocasional de 160 universitarios. Finalmente se reflexiona sobre la importancia de considerar el análisis del lenguaje desde una perspectiva multidisciplinar y crítica, en la línea desarrollada por Van Dijk y el “Análisis del Discurso”.

Palabras clave: software, análisis de contenido, análisis textual, text Minino, psico-socio-lingüística, contexto, análisis-de-correspondencias; redes-sociales.

*Agradecimientos: En la elaboración, durante estos últimos doce años, del sistema SOCIOS se reconoce la especial contribución del Laboratorio de Psicología Social de la Universidad de Barcelona. Enlace: http://www.ub.edu/dppss/lps/menu7_e.htm
Correspondencia: José Manuel Cornejo, Departament de Psicologia Social, Laboratori de Psicologia Social, Facultat de Psicologia, Universitat de Barcelona, P. Vall d'Hebron, 171, Campus Mundet, Barcelona 08035.
Correo electrónico: jcornejo@ub.edu

Contextual analysis: CONTEXT (SOCIOS), a data processing system

This article A first approach has already been published in this journal (see Anuario de Psicología, 37(3), 277-297, 2006), suggests that language must be studied as a first order phenomenon in order to advance research and knowledge in the humanities and social sciences. We review the tradition of content analysis and operationalize the concept of psychosocial context, as a necessary reference in the collection of the set of personal and social meanings it conveys. We present ConText, a computer application among the modules of the system SOCIOS, specially designed in the Laboratorio de Psicología Social of the University of Barcelona for the treatment of social interaction data. In this case ConText applies to contents of verbal data from the perspective of "Textual Analysis" or "Text Mining". We provide the theoretical and methodological criteria for its use and detail some of the software characteristics and main procedures, trying to show their usefulness for research. Exemplification is made on the basis of a corpus of personal definitions of the "Happiness" notion in a casual sample of 160 university students. Finally we reflect on the importance of considering the analysis of language from a multidisciplinary perspective and criticism, in line with Van Dijk's "Analysis of Discours".

Keywords: software, content analysis, textual analysis, text Mining, psycho-socio-linguistic, context, analysis-of-correspondence, and social networks.

En continuidad con el concepto de 'interacción', y en relación con el sistema SOCIOS, elaborado y experimentado durante los doce últimos años en el Laboratorio de Psicología Social de la Universidad de Barcelona, centramos nuestro interés en su aplicación al análisis de datos procedentes del lenguaje natural, a través del programa informático *ConText*.¹

El hecho lingüístico y su manifestación en los procesos de comunicación intra-inter personales, intra-inter grupales y societales es uno de los fenómenos más relevantes del comportamiento humano y la vía más fecunda de investigación en las ciencias humanas y sociales y en las neurociencias. En el interior de los fenómenos del lenguaje fraguan los esquemas cognitivos, las representaciones sociales, las actitudes; y en su entramado se encuentran insertas las principales claves de comprensión del ser humano. Los sistemas lingüísticos son sistemas de producción de significados socialmente construidos y compartidos que contienen todos los elementos, estructurales y de proceso, de cualquier sistema de interacción. El texto en su forma concreta constituye un universo de relaciones significativas entre los elementos léxicos que lo constituyen.

El análisis de datos aplicado al lenguaje natural, dentro de la rama de la Lexicometría, es actualmente un área en crecimiento exponencial. La espectacular difusión de la informática personal y la expansión de las TICs en todos los

¹ Una primera aproximación a este tema ya fue publicada en esta revista: *Anuario de Psicología*, 37(3), 277-297.

ámbitos de la actividad social, han impulsado el desarrollo de procedimientos automatizados para facilitar el acceso y la lectura eficaz del progresivo aumento de documentos textuales y gráficos disponibles en la red. Los progresos en la traducción automática, la indexación documental, la eficacia de los nuevos buscadores o las herramientas de resumen documental son la expresión de este reto donde el tratamiento del lenguaje natural ocupa una posición estratégica y cuyas posibilidades de aplicación son todavía en parte desconocidas.

Se han venido desarrollando dos tipos básicos de análisis; el más utilizado ha sido el “análisis textual”, que trata de las estructuras abstractas del código escrito, considerado como objeto o corpus delimitado, en la perspectiva del estructuralismo lingüístico. El otro, está relacionado con el ‘estudio del habla’ y se centra en aquellos aspectos más dinámicos de la interacción espontánea en la perspectiva de las ciencias sociales. A pesar de las diferencias de enfoques, ambos tipos provienen del tronco común de la ‘Lingüística Estructural’ iniciada por Ferdinand de Saussure (1980). Ambas tienen como objetivo común el descubrimiento de “reglas” y “regularidades” de “estrategias” y “estructuras”. En general tienen una orientación descriptiva y suelen ignorar contextos mayores como el “cognitivo” y el “social”. También podemos diferenciar dos modalidades predominantes de análisis: la más formal y abstracta del tratamiento del *lenguaje simbólico* (inteligencia artificial y gramática computacional) frente a los estudios más concretos, sobre textos reales, en los que los usuarios de una lengua se manejan como “actores sociales”, hablando, significando, y “*haciendo cosas con palabras*”.

En el análisis del lenguaje concurren muchas disciplinas desde sus propias ópticas: estudio de las formas, los significados, la interacción y la cognición; sin embargo es el contexto el que tiene un papel predominante en la descripción y explicación de los textos escritos y orales. Cada enfoque ha desarrollado sus propios conceptos, métodos y técnicas de análisis; una breve clasificación distinguiría entre *a)* los que se centran en el texto mismo y/o en su estructura, *b)* los que consideran el lenguaje como comunicación en el ámbito de la “cognición”, y *c)* aquellos que se centran en la estructura socio-cultural del fenómeno lingüístico. *Lenguaje, cognición y sociedad* forman los tres vértices que configuran el lenguaje como objetivo de investigación. Con diferentes denominaciones: ‘Análisis de Contenido’, ‘Análisis del Discurso’, ‘Análisis Conversacional’, ‘Pragmática’, ‘Análisis Sociosemántico’ y más recientemente, el ‘Análisis Textual’ existe una expansión de metodologías aplicadas a la investigación psicosocial cuyo objeto fundamental de análisis es el lenguaje.

ConText

En el ámbito de las múltiples herramientas existentes actualmente para el análisis de material verbal, ConText es una aplicación informática que ha sido diseñada para facilitar el análisis relacional de corpus textuales muy variados, desde perspectivas teóricas y metodológicas muy generales, aunque privilegia el punto de vista psicosocial sobre un análisis estrictamente lingüístico, y con-

ceptualiza el lenguaje como comportamiento individual/social en contextos situacionales socialmente compartidos. Está orientado a explorar la red de interrelaciones que subyace en su trama textual (etimológicamente ‘tejido’) y que ayudan a contextualizar y valorar críticamente el alcance de su significado. En general disponemos de un documento o un conjunto de documentos, más o menos complejos, que deseamos explorar con mayor profundidad que la derivada de una somera lectura y extraer de ellos aquellos elementos de su estructura susceptibles de aportar nueva comprensión de los procesos psico-socio-lingüísticos que están implicados en la determinación de sus significados latentes o potenciales.

ConText dispone de distintos procedimientos de recodificación y clasificación automática, semi-automática y manual que facilitan al usuario ensayar distintas estrategias de lematización y categorización, hasta conseguir la que estime más adecuada a sus necesidades y objetivos. En todo momento las decisiones son reversibles y los distintos procedimientos pretenden agilizar la toma de decisiones que en último extremo corresponden específicamente al investigador y que tendrán consecuencias directas sobre la calidad de los resultados obtenidos.

También aporta un sistema integrado de módulos de análisis que están fundamentados en la noción de contexto, en su sentido distribucional, y en la descripción ‘modo sociométrico’ de la red de relaciones entre los elementos del corpus, para configurar su espacio semántico. Utiliza como elemento básico el método comparativo sobre el recuento selectivo de frecuencias de las *formas léxicas* (FLs), con o sin lematizar, excluyendo, si se precisa, las llamadas ‘palabras vacías’ o gramaticales. La posibilidad de incluir en el análisis variables extratextuales dota al corpus de una fuerte estructura de contraste. El conjunto de resultados confluye en una serie de tablas de contingencia susceptibles, mediante el Análisis de Correspondencias y la Clasificación Automática, de poner de relieve los hechos de estructura, su valoración y su visualización mediante los ejes de proyección multidimensional (Cornejo, 1988). Más adelante se explicitarán con cierto detalle algunos de los procedimientos más específicos utilizados por los distintos módulos del programa.

Una de las primeras aplicaciones de ConText fué en la tesis doctoral de la UB sobre las “Relaciones de influencia entre grupos de mujeres: un análisis evolutivo de los discursos” (Berbel, 1997), a través de un análisis textual retrospectivo de documentos políticos de 1975 a 1996 (21 años) y la evolución de los distintos grupos políticos respecto a la defensa de temas asociados tradicionalmente al feminismo militante. En este estudio, se evidenció la influencia minoritaria de los grupos radicales de mujeres que trabajaban por un cambio político y social en defensa del colectivo femenino, sobre los grupos políticos, produciéndose lo que en Psicología Social se ha llamado el efecto retardado (*sleeper effect*) del cambio de actitudes. Se trata de un proceso mediante el cual, con el paso del tiempo, se produce una “criptomnesia” social por la que las personas o los grupos sometidos a una influencia minoritaria olvidarían el origen y la identidad de la influencia pero asumirían en cambio su contenido.

ConText se ha elaborado dentro de la tradición de la escuela francesa de los años 70 del análisis multidimensional de datos, conocida como “L’Analyse des données”, y en su terminología actual ‘Data Mining’ ‘Text Mining’ o ‘Web Mining’ para datos numéricos, textuales o simbólicos. En los años 60, Benzécri (1973, 1981) ya había propuesto un método estadístico para la resolución de los problemas fundamentales que interesaban al lingüista. Había desarrollado un método algebraico inductivo, al que denominó Análisis de Correspondencias (AC), como procedimiento multidimensional para el tratamiento de grandes tablas de datos, definiendo una distancia entre distribuciones condicionadas, conocida como distancia de chi-cuadrado o distancia de Benzécri (Behar, 1991, 1993). Este método efectúa una abstracción cuantitativa partiendo de tablas de datos cualitativos diversos, para obtener por medio del cálculo, nuevos parámetros que miden entidades a un nivel de abstracción superior al de los datos iniciales. A partir de aquí, se hacían posibles estudios de muy diversa índole como el establecimiento de parentesco entre textos de autores diferentes, aproximando los que tienen un vocabulario similar, o bien el establecimiento de tipologías entre los capítulos de una misma obra, o bien el establecimiento de los posibles cambios de vocabulario de una organización política o sindical en el curso de un período dado. Una propiedad fundamental de este método reposa sobre la noción lingüística de “equivalencia distribucional”, retomada de Harris (1954).

Es ilustrativo señalar que la lingüística estructural americana había trabajado en el estudio de las lenguas amerindias, de las cuales algunas habían sido recopiladas, antes de su extinción, de labios de personas incapaces de expresarse claramente en ninguna otra lengua conocida. Esta situación llevó al desarrollo de métodos especiales para tratar corpus lingüísticos sin tomar en consideración el sentido de las palabras. La noción de Harris de ‘equivalencia distribucional’ fue utilizada con éxito en este tipo de análisis; llevándole a afirmar, en una serie de artículos de gran repercusión, que todo o casi todo de una lengua puede obtenerse, sin recurrir al sentido, solamente atendiendo al análisis de los hechos distribucionales. Harris define el sentido o significado de una palabra como “la distribución de esa palabra en el conjunto de todos sus contextos de ubicación posibles”. Es una concepción que podemos considerar revolucionaria, en la medida en que permite una construcción inductiva del análisis textual a partir de los propios datos, sin supuestos ‘a priori’. De tal modo que, dado un corpus textual, se trata de observar las palabras, las frases y el texto en su conjunto como una sucesión de elementos, donde una palabra o un segmento quedará perfectamente definido por el conjunto de todos los contextos en los que pueda insertarse para obtener una frase correcta. En un corpus cerrado, por tal que no sea muy reducido, una frase correcta es la que tiene sentido en el contexto de ese corpus. Del mismo modo, dos o más elementos que tengan un mismo perfil distribucional serán necesariamente sinónimos (hablar, decir...).

La equivalencia distribucional confiere gran estabilidad a los resultados del AC. En la práctica no se dará el caso de una proporcionalidad perfecta entre dos columnas (o filas), pero pueden ser más o menos próximas, lo que

introduce un elemento de graduación de mucho interés. Así, liberados parcialmente de la ‘presión ontológica’ del significado, se abre la posibilidad de un tratamiento automático del material verbal, comparable y compatible con otras perspectivas alternativas de análisis.

Desde entonces, la conjunción entre el análisis textual y el análisis de datos multidimensionales está en continua evolución. A partir de 1990, el desarrollo de los programas informáticos se produce de forma paralela al de los métodos y a sus aplicaciones a grandes conjuntos de datos procedentes de encuestas socioeconómicas, entrevistas, investigaciones literarias, de textos políticos, archivos históricos, bases de datos documentales, etc. Actualmente la utilización de la clasificación ascendente jerarquizada (método Ward) de las coordenadas factoriales, como análisis complementario al AC es cada vez más frecuente. Esto es debido a que las técnicas de visualización de los primeros ejes factoriales, en el caso de grandes corpus, resultan a menudo poco esclarecedoras.

En este ámbito citaré, aunque sea brevemente, las aportaciones de Ludovic Lebart (1984), en el campo de respuestas libres a cuestiones abiertas en encuestas, y las de Mónica Bécue, que presentó en la Facultad de Informática de la Universidad de Barcelona su tesis: “Un sistema informático para el Análisis de Datos Textuales” (1988); y el programa SPAD.T (Lebart, L., Morineau, A., Bécue, M., Haeusler, L. 1992), que investiga de forma sistemática los métodos lexicométricos y estadísticos utilizados en el análisis textual y, en particular, el tratamiento de respuestas abiertas de encuestas y de su relación con respuestas de información cerrada.

El análisis de contenido

Pero los antecedentes inmediatos de los métodos actuales de Análisis Textual hay que buscarlos en el Análisis de Contenido. Bernard R. Berelson, sociólogo norteamericano, es reconocido como pionero en la metodología del análisis de contenido aplicado a los medios de comunicación. Especializado en demografía, politología y comunicación, publica, junto con Paul F. Lazarsfeld y H. Gaudet, “The people’s choice: How the voter makes up his mind in a presidential campaign” (1944), uno de los trabajos más influyentes sobre los factores que determinan la toma de decisión de los votantes y en particular sobre los efectos de influencia de la radio (Berelson y Lazarsfeld, 1948). En 1952, publica lo que podemos considerar el texto fundacional de la metodología del análisis de contenido: “Content analysis in communication research”. La define como una técnica “que tiene por objeto la descripción sistemática y cuantitativa de los mensajes de la comunicación”.

La idea de cuantificación diferencia un proceso intuitivo de uno científico. Todos realizamos de forma intuitiva análisis de contenido para encontrar el significado en los textos que leemos. El problema aparece cuando se desea extraer información de modo sistemático, fiable y válido (Holsti, 1969; Krippendorff, 1990). En la propuesta de Berelson, cuantificar significa crear un

sistema de categorías, adecuado a los objetivos del análisis, sobre las cuales se codificarán los diversos elementos del texto según su pertinencia. A su vez se contabilizarán las categorizaciones registradas para evaluar la importancia relativa de cada categoría en cada una de las partes del texto. Inicialmente esta metodología se utilizó casi exclusivamente para analizar textos producidos por los *mass media*, pero pronto se generalizó a otros tipos de material textual con el objeto de identificar y tratar de explicar las representaciones cognitivas y emocionales que dan sentido a todo relato, a través de la detección de sus elementos sintácticos y semánticos. En este sentido, Osgood había propuesto tempranamente varios procedimientos para la elaboración de índices cuantitativos: el análisis de las aserciones estimativas de un mensaje (*Evaluative Assertion Analysis*) o el procedimiento *Cloze*, al tiempo que iniciaba su trabajo sobre la medida del significado a través del ‘Diferencial Semántico’. Afirmaba que los significantes o enunciados de una comunicación poseen orientaciones estimativas, afectivas o connotativas, que esas orientaciones estaban polarizadas y que eran potencialmente mensurables a partir de escalas de valoración, siendo universales algunas de estas dimensiones. Según Osgood (1959), se pueden investigar las causas a partir de los efectos y algunas de las variables inferidas sobre el emisor podrían ser: la inteligencia, la facilidad de comunicación, el origen racial, la ansiedad, la agresividad, la estructura asociativa, las actitudes, los valores, etc. Para ello los índices que permitirían la inferencia serían las unidades léxicas, las estructuras sintácticas, las pausas, los errores, las expresiones idiosincrásicas, etc.

Uno de los escollos más difíciles de sortear se focalizó en la clarificación de la noción de “significado”, tratada de maneras muy diferentes por la Semántica Formal, la Sociología, la Psicología Cognitiva, y la Etnografía, aunque siempre con una actitud crítica hacia la cuantificación como procedimiento adecuado para acceder al significado.

La corriente estructuralista en Lingüística ofreció un marco amplio para el estudio de la narrativa, los mitos, la literatura, el cine y otras prácticas semióticas en el estudio de los medios. Pero pronto aparecieron críticas a estos enfoques por no dar cuenta, en sus análisis, ni de los procesos cognitivos, ni de los de interacción social, ni de los contextos sociales implicados en toda producción lingüística. Una mera descripción aséptica del sistema sintáctico sobre la base de oraciones aisladas de contexto, sin tomar en consideración el uso real de la lengua y de su necesaria dimensión social, fue creando en muchos la necesidad de ampliar el análisis puramente lingüístico a aspectos importantes de las relaciones semánticas y funcionales que producen la coherencia de un texto, asumiendo que el lenguaje expresivo, además de su función instrumental de comunicación, contiene información valiosa sobre la naturaleza de los estados psicológicos y los determinantes socioculturales que lo originan.

La Etnometodología surge a finales de los 60 como enfoque “micro-sociológico”, distanciado del macro-sociológico (a la búsqueda de estructuras sociales), y se interesa en temas como la “interacción diaria” o el análisis conversacional de la “vida cotidiana”, con especial atención a los cambios de tur-

no en las alternancias del diálogo o los efectos de la expresión en el tipo de interacción social que genera (Giles y Coupland, 1991; Stubbs, 1993).

El análisis del lenguaje va adquiriendo relevancia en Etnografía y en Antropología, produciéndose datos sobre las formas de hablar en los distintos contextos culturales. Se plantea la hipótesis de que los hablantes de una lengua no sólo conocen las reglas o principios que estructuran esa lengua, como sistema gramatical, sino que también son poseedores de una “competencia comunicativa” (saber qué decir y cuándo), como miembros de una cultura más amplia (Silva, 1988), y comparten un conocimiento de las reglas sobre cómo hablar correcta y apropiadamente en distintas situaciones; por ejemplo, “cómo hacer una advertencia”, “cómo contar un cuento”, o “cómo entablar una discusión o un debate” (Saville-Troike, 1982).

La Psicología Cognitiva surge en los 70 como corriente de investigación en el estudio de las funciones que desempeñan los “procesos mentales” (atención, memoria, percepción, concentración, solución de problemas, etc.) en la comprensión de “textos”, aportando valiosos elementos en aspectos lingüísticos, cognitivos, neurológicos y filosóficos. Junto con la Psicolingüística se avanza hipótesis sobre lo que ocurre en el “procesamiento” de la información verbal en textos orales y escritos, tanto para el emisor como para el receptor de los mensajes. Este desarrollo se da en estrecha relación con los avances de la “lingüística computacional” en el ámbito de la “inteligencia artificial”, que integra y enriquece las perspectivas de la “Lingüística Textual” (Greene, 1989). Asimismo la Sociolingüística y la Pragmática aportan nuevos enfoques sobre la naturaleza discursiva y secuencial del uso del lenguaje, los “actos de habla”, y la “interacción verbal” en contextos socio-culturales determinados.

Por su parte, la Psicología Social aborda el tema del análisis del lenguaje en los 80, a pesar de que en sus áreas de interés ya se habían tratado temas relevantes como los “procesos de socialización” y de “atribución” (Moscovici, 1967; 1972). Se va desarrollando la temática de la interacción social y de la construcción de representaciones sociales desde una perspectiva cognitiva, centrada en el estudio de la interacción como “fenómeno psicológico” y aborda la “comprensión”, la “explicación” y la “formación de opiniones e ideologías” (Giles, 1979).

Desde una perspectiva socio-cultural el interés por la “interacción social situada” ha favorecido el intercambio y la sinergia entre todas estas disciplinas convergiendo en el concepto de “cognición social” como elemento integrador de las dicotomías entre cognición e interacción, entre sociedad y cultura. Los estudios sobre comunicación en los 70 y 80 amplían su perspectiva al análisis de la estructura de los mensajes en los medios de comunicación masivos, la comunicación interpersonal, intercultural y empresarial y ponen las bases de una “teoría de la comunicación humana” (Brown, R. 1994).

En la actualidad, el análisis de contenido más que una metodología cerrada es un conjunto de procedimientos diversos cuyo objetivo común es conseguir, de modo controlado y sistemático, la verificación de significados basada en la deducción a partir de los indicios manifiestos. Se mueve entre el rigor de la objetividad y la fecundidad de la subjetividad (Bardin, 1977). Tiene una doble función:

heurística (descubrir lo que se escapa a una lectura superficial), y de validación de las interpretaciones realizadas sobre el contenido subyacente en el texto. Sus áreas de aplicación son muy diversas según se desee poner al descubierto las características temáticas, lexicológicas, las actitudes implícitas, los sistemas de valores, los estereotipos e ideologías relativos a las condiciones de producción y recepción de los mensajes comunicativos.

La noción de contexto

La noción de contexto está muy generalizada, pero de hecho es polisémica. En Psicología Social se utiliza generalmente para referirse a los diferentes ambientes o situaciones en los que se desenvuelve la actividad humana. Bunge (1989) considera un contexto al conjunto de proposiciones formadas por conceptos que tienen referentes comunes y Benzécri (1981), desde la Lingüística, entiende que una palabra o un segmento queda caracterizado por el conjunto de todos los contextos en los que puede insertarse formando una frase correcta.

De manera más genérica, podemos decir que un contexto es el conjunto de combinaciones posibles que determinan el significado específico de un elemento concreto por su posición referida a ese conjunto. Lo que implica que los elementos singulares quedan definidos por sus relaciones con el conjunto de otros elementos que operan en su campo, es decir que tienen un referencial común.

Hay tantos contextos particulares simultáneos como clases a las que un elemento puede pertenecer por sus relaciones de semejanza u oposición. Estos distintos contextos pueden estructurarse respecto a un elemento concreto formando jerarquías contextuales en función del punto de vista que se desee priorizar. En un momento determinado, cada elemento queda definido por el conjunto de contextos jerárquicamente estructurados que conforman su 'función de estado'.

Por ejemplo, una persona puede ser a la vez esposa, madre de familia y directora general de una empresa, etc., lo que significa que con respecto a ella existen una serie de contextos particulares identificables como los diferentes conjuntos de proposiciones referidas a las esposas, madres de familia, etc. donde la persona ocupa una posición relativa en cada uno de ellos. Dependiendo de los objetivos de cada investigación, el investigador favorecerá determinados contextos, subordinando los restantes. La noción de contexto debe entenderse en relación con la noción de sistema, cuya propiedad fundamental es la interdependencia de las partes, definición de Parsons (1951) que sigue siendo una de las más consensuadas. La jerarquía de contextos tiene su traducción operativa en distintos órdenes de segmentación y la construcción de modelos de realidad como representación formal de un sistema en términos de variables, parámetros constantes y funciones que determinan un conjunto de relaciones con vistas al estudio de un problema determinado. Esto nos lleva a considerar el estudio de la interacción como estudio de las estructuras de relación entre los elementos de un sistema.

Siguiendo el enfoque de Van Dijk (1977, 1993, 1997, 1999) desde el punto de vista crítico, el contexto puede definirse como “la estructura que involucra todas las propiedades o atributos de la situación social que son relevantes en la producción y comprensión del discurso”. Las características del contexto no sólo influyen en el significado denotativo del texto (escrito y oral), sino que es posible también la recíproca, ya que el texto, como comportamiento social de primer orden, también puede modificar los determinantes locales y globales del propio contexto. Van Dijk hace observar que en todos los niveles del discurso podemos encontrar las “huellas del contexto” en las que, junto a las características personales, juegan un rol fundamental las características sociales de los participantes (por ejemplo, “género”, “clase”, “etnicidad”, “edad”, “origen” y “posición”, u otras formas de pertenencia grupal).

Entre las características principales del contexto podemos considerar: el “ambiente” (tiempo, ubicación, circunstancias, etc.), los “participantes” y sus “roles socio-comunicativos” (locutor, amigo, presidente, etc.), sus “intenciones-metas o propósitos”. Un enfoque contextual del lenguaje involucra muchos aspectos de la sociedad y su cultura. Por ejemplo el uso pronominal “tú”-“usted”, como forma de cortesía, implica que los hablantes conocen en cierta medida el estatus del “otro” dentro de la situación de interacción social. Igualmente, las variaciones en la utilización del léxico expresan, junto al significado propiamente denotativo, muy diferentes opciones ideológicas, actitudes, emociones, etc. del hablante (lo que comúnmente se conoce como aspecto connotativo: “viejo-anciano-persona mayor”). Por otra parte, los actos de habla, como órdenes o imperativos, presuponen siempre un juego sutil de modulación de diferencias de estatus, de poder y de autoridad entre los hablantes. El contexto global se hace evidente en la identificación de las acciones de las instituciones y organizaciones (procedimientos, legislaciones, sentencias judiciales, educación, reportajes o informes). También se manifiesta cuando los participantes, además de sus características personales propias, se involucran en interacciones como miembros de un grupo, clase o institución social: (mujeres-hombres; anciano-joven; jefe-empleado; profesor-alumnos, funcionarios-ciudadanos).

Los procedimientos y módulos del ConText

ConText ha sido escrito en el lenguaje de programación Qbasic de Microsoft para el sistema operativo MSDOS entre los años 1994 y 1996 en el LPS de la Universidad de Barcelona, con actualizaciones periódicas hasta el presente. Como conjunto de procedimientos automáticos e interactivos, creados para facilitar el análisis contextual de material escrito o transcripciones de material oral, y en general de cualquier material simbólico, su utilización no exime al investigador de la responsabilidad de valorar las decisiones concretas que conlleva y los efectos que éstas tendrán necesariamente sobre la validez de sus resultados (Pêcheux, 1978).

Todos los procedimientos son reversibles y pasada una fase automática previa, el investigador puede modificar las propuestas de recodificación (lematización) que no considere pertinentes, así como elegir los sistemas de categorización apropiados o modificar las opciones por defecto para que se adapten mejor a los objetivos y características del análisis.

El programa trabaja con la *forma léxica* (gráfica) de la palabra, o FL, como unidad de análisis. Una FL es una cadena de caracteres separados por delimitadores: espacios, coma, punto y coma, punto interrogación e interjección. El texto puede contener conjuntos de números, de letras o combinaciones de ambos. El programa permite el uso de acentos y otros símbolos, como la ð o ç, vertidas a mayúsculas. También proporciona una exportación de los datos al formato SPSS y tablas que pueden utilizarse con otros programas de análisis.

Las operaciones básicas de ConText comportan:

1. La entrada de uno o varios archivos de texto en formato ASCII plano (es decir sin los caracteres de control que introduce típicamente un procesador de textos).
2. Una ordenación alfabética a la vez que un recuento de frecuencias de aparición de cada FL con adscripción de la misma a una Unidad Léxica secuencial (UL) (que identifica su posición en el texto original) y a uno o varios subtextos formados por las modalidades de las variables extratextuales definidas por el investigador.

Antes de la descripción de los distintos procedimientos y tipos de resultados que pueden obtenerse con el ConText, es importante detallar tres pasos previos que son particularmente decisivos para desarrollar con éxito un análisis.

Pasos previos al análisis

Efectivamente, existen tres momentos especialmente sensibles que condicionarán los resultados del resto de los módulos: la formalización del archivo de entrada de los datos textuales, la supervisión del procedimiento de lematización de las FLs y la fase de categorización extra-textual.

La formalización del archivo de entrada de los datos textuales

La formalización del archivo de texto que sirve de entrada a los datos requiere informar al sistema sobre la estructura de los textos a analizar. Como ya hemos precisado, el archivo deberá estar en formato ASCII de texto plano. Se escribe con un editor de texto o más sencillamente desde un procesador mediante la opción 'guardar como' y seleccionando el tipo 'Texto MSDOS'. ConText reconoce los acentos y los delimitadores como el punto, los signos de interrogación y de exclamación.

Se consideran 5 tipos de formato, en función de la menor o mayor estructuración del texto. Cada tipo quedará identificado con una extensión de archivo diferente (tabla 1).

TABLA 1. FORMATOS POSIBLES DE ESTRUCTURACIÓN DEL MATERIAL TEXTUAL

<i>.TXT</i>	Extensión genérica que sirve para cualquier tipo de archivo.	
<i>.TX1</i>	Para textos únicos, sin marcas especiales y sin formato previo.	
<i>.TX2</i>	Para conjuntos de textos ordenados secuencialmente.	<i>*TEX{n} título</i>
<i>.TX3</i>	Para un texto estructurado en variables identificatorias extratextuales, donde cada unidad léxica (UL) o frase está referida a un conjunto de variables categoriales que identifican la procedencia o un conjunto de identificaciones que definen sus características.	<i>*VARn{4,3,5} título</i>
<i>.TX4</i>	Para textos organizados en forma de intervenciones realizadas por los distintos participantes en un grupo (grupos de discusión, chats, debates, etc.) donde interesa personalizar las aportaciones de cada miembro y la secuencia de intervenciones.	<i>*SUJ{n} título</i>
<i>.TX5</i>	Para textos que unifican los formatos TX3 y TX4. Es el tipo de entrada léxica más estructurada y permite un análisis más profundo del material textual.	<i>*SARn{3,2} título</i>
<i>.TXP</i>	Es un archivo contenedor de otros archivos textuales que se utiliza para tratar un texto muy extenso que esté subdividido en archivos más manejables.	

Es importante señalar que la mayor estructuración de los datos textuales redundará en una más completa contextualización y por tanto en una mejor contrastación, riqueza y precisión en el análisis. Esta estructuración operativiza el concepto de contexto como conjunto de elementos referenciales que coadyuvan a la interpretabilidad del texto a analizar. Puede incluir una gran variedad de variables secuenciales, técnicas, demográficas, psicológicas y sociales que caracterizarán a cada unidad léxica (UL).

Excepto en el caso de texto único *.TX1*, caso que no requiere ninguna marca especial, cada archivo va precedido de una cabecera que indica de modo conciso el formato concreto de lectura. Se indica por las marcas

**TEX{n} *VARn{4,3,5} *SUJ{N} *SAR2{3,2},*

donde *n* ó *N* se refiere al número de textos, al número de variables o al número de sujetos. Los dígitos entre llaves en el caso de *.TX3* y *.TX5* se refieren al número de modalidades que posee de cada variable.

En la primera línea sigue un título general de identificación de los datos y del análisis. Y en las siguientes líneas figurarán las etiquetas correspondientes a los distintos textos, a los distintos sujetos del grupo o las distintas variables y modalidades de las variables, introducidas por marcas específicas. Existen dos tipos de etiquetas, unas breves (8 caracteres) y otras largas (20 caracteres) para facilitar su visualización en pantalla y la legibilidad de los archivos de resultados. En la tabla 2 se muestra la cabecera de un texto *.TX5* sobre el que ejemplifica-

remos las salidas más representativas. En el manual de utilización se añaden otros detalles más concretos para cada tipo.

TABLA 2. FORMATO DE LA CABECERA DE UN ARCHIVO DE ENTRADA DEL TIPO .TX5

```
*SAR3{2,4,3} Descripción de FELICIDAD en alumnos de PS ...2008
*MOD1 {H,M}
*MOD2 {E1,E2,E3,E4}
*MOD3 {F1,F2,F3}
*TIT1 {SEXO}
*TIT2 {EDAD}
*TIT3 {FELICIDAD GENERAL}
*T1M1 {hombre}
*T1M2 {mujer}
*T2M1 {18a}
*T2M2 {19a}
*T2M3 {20-23a}
*T2M4 {+23a}
*T3M1 {fel baja}
*T3M2 {fel media}
*T3M3 {fel alta}
*SS{A1,A2,A3,A4,A5,A6,A7,A8,A9,A10}
.....
*SS{A150,A151,A152,A153,A154,A155,A156,A157,A158,A150,A160}
```

En la tabla 3 se reproducen las 3 primeras UL del texto. Como formato general cada UL va precedida en su margen izquierdo de 4 elementos posibles dentro de las marcas siguientes: $\langle U \rangle \{L\} [S] \{2,4,1\}$. Sólo es necesario introducirlas directamente en el texto cuando cambia un elemento en relación con la UL anterior. En el ejemplo, cada UL es la respuesta libre dada por un sujeto distinto y que, además, está caracterizado por sus modalidades en cada una de las variables extratextuales.

TABLA 3. FORMATO GENERAL DE LAS UNIDADES LÉXICAS (UL).

$\langle U \rangle \{L\} [S] \{2,4,1\}$

- $\langle U \rangle$ Número de unidad léxica (por defecto es automática).
- $\{L\}$ Variable de secuencia o bloque (por defecto es automática).
- $[S]$ Número del sujeto que interviene en el grupo.
- $\{2,4,1\}$ Modalidades de las variables extratextuales que caracterizan a la UL.
- $[1]\{2,4,1\}$ Estar bien con uno mismo y quererse para poder estar bien con el resto de nuestras personas a las que queremos.
- $[2]\{1,2,1\}$ Aquella sensación que me produce un bienestar general cuando la amistad el amor la familia la salud y la economía van bien.
- $[3]\{2,4,1\}$ Estar satisfecho con tu vida con la gente que te rodea y con lo que has conseguido en el transcurso de los años.

Una vez establecido el archivo de entrada del texto, el sistema inicia la lectura, reconocimiento y recuento de las frecuencias de aparición de cada FL, así como su almacenamiento y organización para posteriores tratamientos.

Debemos señalar que una forma léxica no se corresponde con una palabra, sino con la forma gráfica. De modo que 'bueno' y 'buena', se contabilizan como dos formas léxicas distintas.

A la lectura del texto sigue una doble ordenación: alfabética y por orden de frecuencias, dando lugar a un listado o diccionario general, donde cada FL queda definida por un número de orden definitivo. Esto permite el proceso de 'numeralización' del texto, una forma de comprimirlo en una secuencia numérica, donde cada FL queda caracterizada por su número de orden.

Sigue un procedimiento complementario de detección de formas gramaticales o 'palabras vacías' mediante un archivo externo en el que figura un listado de las FL que se desea considerar como tales. Dicho archivo es editable por el investigador antes de iniciar el análisis.

Aunque en una primera consideración esta selección puede facilitar la claridad del texto, en la práctica la distinción entre las formas funcionales o instrumentales y formas lexicales parece difícil de establecer y, actualmente, se considera que las categorías gramaticales (pronombres personales, pronombres y adjetivos posesivos, demostrativos, etc.) constituyen una parte esencial del entramado formal de la enunciación, caracterizando la presencia del locutor en el texto, así como otros fenómenos de potencial interés para el análisis. La propiedad de equivalencia distribucional asegura en general una gran estabilidad en los resultados y minimiza el efecto de incluir estas formas en los AC.

La supervisión del procedimiento de lematización de las FL

En esta fase se realiza inicialmente y de modo heurístico una propuesta de lematización o agrupamiento de formas léxicas flexionadas o con similitud de significado, que el investigador deberá supervisar, sea mediante un procedimiento interactivo, sea mediante la modificación de un archivo externo generado por el programa. Este procedimiento suele dar como resultado una reducción considerable del volumen de FL diferentes. El porcentaje de reducción dependerá de tres factores: tamaño del texto, radicalidad de la recodificación y tipo de texto. La identificación de FL similares se realiza mediante un algoritmo de comparación entre FLs, que tiene en cuenta la longitud de la palabra y el número de caracteres idénticos en ella, así como su posición en la palabra. Si el nivel de semejanza supera un determinado umbral, la FL queda aglutinada, de modo que se condensan en un solo término todas las ocurrencias de los diferentes tiempos verbales y los sustantivos y adjetivos que contengan una misma raíz. Por ejemplo, el programa considerará ABANDONA, ABANDONABAN, ABANDONO, ABANDONADAS... como ABANDONAR o como la FL que sea más frecuente en el corpus, de modo que toma como frecuencia acumulada la suma de las frecuencias de las FL agrupadas bajo ella.

El sistema utiliza igualmente un diccionario externo con el listado de las formas irregulares para optimizar su recodificación (sería el caso de los verbos

irregulares). Este archivo también puede servir para entremezclar las FLs de dos o más lenguas utilizadas por los hablantes en el mismo corpus, como puede ser el caso del castellano y del catalán. En este caso cada FL se agrupará con la FL más frecuente en el texto. Las FLs recodificadas disponen de una etiqueta optativa, modificable por el investigador, para señalar su nueva adscripción. Por defecto adoptan la etiqueta de la forma aglutinadora. Propuesta la lematización ‘automática’, se genera un archivo editable que permite fácilmente al investigador eliminar adscripciones no adecuadas o reagrupar FLs de interés (tabla 4). Este archivo una vez modificado puede ser releído por el programa para reconfigurar todo el sistema de lematización; y también puede reutilizarse para la recodificación posterior de otros textos similares, aunque de hecho es difícil que un archivo de este tipo tenga un uso universal.

TABLA 4. ARCHIVO DE LEMATIZACIÓN O RECODIFICACIÓN

```
*A #1& [62,62] {G } A
AL #41& [5,5] {G } A
-A_DIARIO #2& [1,68] { } A_DIARIO
-A_PARTIR_DE #3& [1,1] { } A_PARTIR_DE
*A_VECES #5& [5,6] { } A_VECES
VECES #893& [1,1] {G } A_VECES
*ABSOLUTA #6& [2,3] {G } ABSOLUTA
ABSOLUTO #7& [1,1] { } ABSOLUTA
*ACEPTAR #12& [2,6] { } ACEPTAR
ACEPTADA #10& [1,1] { } ACEPTAR
ACEPTALOS #11& [1,1] { } ACEPTAR
ACEPTARSE #13& [1,1] { } ACEPTAR
ACEPTAS #14& [1,1] { } ACEPTAR
ACCEPTANT #8& [1,1] { } ACEPTAR
*ACONTECIMIENTO #16& [1,2] { } ACONTECIMIENTO
ACONTECIMIENTOS #17& [1,1] { } ACONTECIMIENTO
-ACORDE #18& [1,1] { } ACORDE
-ACTITUD #19& [1,1] { } ACTITUD
-ACTUAL #20& [1,1] { } ACTUAL
*ACTUAR #21& [2,7] { } ACTUAR
ACCIONES #9& [4,4] { } ACTUAR
ACTUEMOS #22& [1,1] { } ACTUAR
-ACUERDO #23& [1,1] { } ACUERDO
-ADAPTARSE #24& [1,1] { } ADAPTARSE
-ADELANTE #25& [3,3] {G } ADELANTE
*ADEMAS #26& [2,27] {G } ADEMAS
DEMÁS #225& [25,25] { } ADEMAS
```

* aglutinadora ‘ ‘ acumular - eliminar recodificación

La estructura de este archivo está pensada para una modificación fácil por parte del usuario. El archivo está hasta cierto punto organizado por orden alfabético. El asterisco en la primera columna muestra la FL aglutinadora y en líneas sucesivas las FLs que el programa ha detectado como similares, con un espacio en blanco en la 1ª columna. Si una FL no se desea aglutinar, basta con poner un ‘-’ en la primera columna. Por el contrario, si deseamos aglutinar una

FL aislada, basta con cambiarla de lugar y ponerla debajo de la FL aglutinadora, eliminando el signo '-'. Los números precedidos del símbolo '#' indican el número de orden de la FL en el diccionario general y no deben modificarse. Los números entre corchetes informan sobre la frecuencia de la FL y, en el caso de las aglutinadoras, sobre la frecuencia acumulada. Esta información ayuda a tomar decisiones. Entre paréntesis aparece si la FL está catalogada 'G', como forma gramatical o 'palabra vacía'; 'N', como forma numérica; 'X', como forma suprimida del léxico. En general, es mejor dejar solas las FL relativamente frecuentes y aglutinar únicamente aquellas con frecuencias bajas. La recodificación sirve para corregir formas mal escritas (por ejemplo 'ACCEPTANT' por 'ACEPTAR').

Aunque este procedimiento suele resultar útil para reducir el número de FL distintas y aumentar las frecuencias de FLs excesivamente atomizadas, en general se ha desarrollado el criterio de que, en corpus de un cierto tamaño, hay que mantener las variaciones flexivas. El procedimiento también resulta de interés para caracterizar modalidades lingüísticas de la expresión relacionadas con hechos diferenciados en partes del corpus, al menos para FLs cuya frecuencia sea mayor que 1.

Terminada esta fase, los análisis pueden utilizar cuatro versiones textuales distintas en función de una doble distinción: FLs originales/FLs lematizadas combinada con Supresión/No supresión de FLs 'vacías'. La comparación entre estas versiones puede aportar puntos de vista esclarecedores sobre el funcionamiento léxico. En cualquier momento el investigador, de modo interactivo, puede optar por el tipo de versión que desee visualizar para la presentación de los resultados.

La fase de categorización extra-textual

La etapa de categorización es, sin duda, la etapa más sensible y laboriosa del análisis. En ella el investigador define a voluntad un conjunto de variables de categorización que utilizará como elementos de recodificación abstractiva de segundo orden. Según los objetivos o intereses que se plantee el investigador, siempre existe la posibilidad de considerar un determinado número de informaciones como siendo equivalentes desde un determinado punto de vista y a un cierto nivel de abstracción. La síntesis de informaciones masivas sólo puede realizarse agrupando previa y progresivamente los elementos que puedan considerarse similares.

En términos informáticos el proceso equivale a una super-lematización complementaria de las FLs mediante su adscripción a alguna de las categorías propuestas. El funcionamiento del programa requiere que en cada sistema de categorización se reserve una categoría residual de no pertinencia, donde se incluirán las FLs de difícil adscripción. Se trata en definitiva de reducir las FLs a un sistema restringido de categorías predefinidas por el usuario y cuya selección, siempre problemática, deberá estar guiada por los objetivos e hipótesis de la investigación para garantizar su fiabilidad y validez.

La categorización permite distribuir los elementos del texto según distintos tipos de criterios posibles: léxico, sintáctico, semántico, temático, expresivo o varios a un tiempo. Se trata de identificar ciertas características comunes en los elementos del texto y agruparlos en categorías. Como requisito metodológico para garantizar su operatividad, cada sistema de categorías debe reunir un conjunto de características como son: la exclusividad, la homogeneidad, la pertinencia, la objetividad, la fidelidad y la productividad. Es habitual utilizar simultáneamente varios sistemas de categorías sobre un mismo corpus, atendiendo a distintos aspectos. Un posible procedimiento a seguir sería:

1. Lectura previa de los textos a analizar.
2. Análisis cualitativo de los aspectos relevantes que deseamos considerar.
3. Contraste de los textos con bibliografía referencial similar.
4. Definición de las variables categoriales en las que se basará el análisis.
5. Adscripción, a criterio del investigador, de cada FL a una de las categorías propuestas.

Conviene señalar que este procedimiento, en sí mismo, no es estrictamente necesario para conducir el análisis textual, a no ser por el propósito de mantener el estilo propio del Análisis de Contenido en su forma tradicional. Desde el análisis textual y de forma general la categorización del investigador introduce en el corpus nuevas variables extra-textuales de contexto 'ad hoc' para ilustrar sus puntos de vista específicos. El interés de las mismas sólo se podrá evaluar a posteriori, en función de su utilidad para aportar un plus de inteligibilidad al corpus analizado. Los restantes módulos de ConText pueden facilitar en gran medida la toma de decisiones sobre los sistemas de categorización potencialmente más productivos.

Aunque este procedimiento puede automatizarse en ciertos aspectos mediante la incorporación de diccionarios de clasificación elaborados previamente, requiere siempre la intervención crítica del investigador para determinar los sistemas de categorías que resultarán más pertinentes y los criterios objetivos más adecuados para realizar la adscripción de cada FL a cada categoría.

Por lo demás el programa permite utilizar hasta nueve variables de categorización independientes. Por defecto ofrece un sistema de tres variables:

1. Una basada en la valoración de la positividad / negatividad connotativa de cada FL.
- 2 Otra basada en las partes gramaticales de la oración.
3. Una tercera basada en las clásicas dimensiones EPA (evaluación, potencia y actividad) del diferencial semántico de Osgood.

Pero puede optarse por categorías que reflejen estilos del texto, contenidos temáticos, funcionales, cognitivos y emocionales de las FLs, u otros muchos aspectos de interés.

Una vez realizada la categorización se genera un archivo editable por cada variable, archivo que permite fácilmente al investigador revisar la bondad de

la categorización realizada, reclasificar FLs o suprimir las no pertinentes, como en el caso de la lematización. Este archivo, una vez modificado, puede ser releído por el programa para reconfigurar todo el sistema de categorización y puede reutilizarse igualmente para categorizaciones posteriores de textos similares.

Finalmente y en cuanto a los resultados, la categorización permite enriquecer la descripción contextual mediante el cruce entre sí de todas las variables de categorización junto con las restantes variables contextuales, así como ofrecer una representación multidimensional del espacio de proyección conjunto, mediante el AC y la Clasificación Automática. Una ejemplificación amplia del proceso de categorización se encuentra en Berbel (1997, pp. 257-263), dentro de su tesis realizada con ConText.

Tipos de resultados

Los procedimientos de los restantes módulos mostrarán interactivamente por pantalla distintos tipos de resultados. Es importante recordar que existen 4 modos de presentar los resultados de los análisis, teniendo en cuenta una doble diferenciación en la selección de los datos que se analizan:

Por un lado cabe diferenciar entre la utilización de FLs originales, sin ninguna transformación, y las FLs recodificadas o lematizadas, en las que las FLs desinenciales, del tipo bueno-buena-buenos-buenas, son agrupadas formando un único lema o FLr. Cada una de estas dos alternativas de selección de los datos puede desdoblarse a su vez en otros dos modos de selección en función de la inclusión o exclusión de las formas léxicas llamadas ‘formas vacías’ (formas gramaticales, sin significado propio, como artículos, pronombres, preposiciones, conjunciones, etc.) (tabla 5).

TABLA 5. VERSIONES DEL CORPUS PARA LOS RESULTADOS

	<i>FL originales</i>	<i>FL lematizadas</i>
<i>Texto íntegro</i>	ORIGINAL	RECO LEX
<i>Texto menos FL vacías</i>	ORIGINAL – Gr	RECO LEX -Gr

En el caso de corpus de gran tamaño, cabe la posibilidad de seleccionar (Opción F2-SELECT en pantalla) solamente aquellas FLs que superan una determinada frecuencia, con el fin de reducir el espacio semántico de resultados. El resto de los procedimientos muestran distintos tipos de resultados, que ilustraremos mediante ejemplos extraídos de un estudio sobre la noción de “Felicidad” en una muestra ocasional de 160 universitarios.

Información General

En la tabla 6 se muestra un resumen de las principales magnitudes del corpus.

Las ULs se corresponden con las frases o párrafos que forman una unidad de contexto lingüístico. Por defecto el programa, a medida que va leyendo el texto del corpus, lo va segmentado en ULs utilizando el ‘.?!’ como delimitadores. Puede forzarse otra segmentación secuencial del texto introduciendo los símbolos ‘<<’ ‘>>’ para señalar una unidad más amplia, formada por dos o más frases o un párrafo entero en una única UL.

El (VL) se refiere al cómputo total de FLs en el corpus e indica su tamaño.

Las (FL) es el número de FLs distintas en el corpus (formas gráficas).

Sujetos emisores: en los textos tipo .TX4 y .TX5 señala el número de sujetos que intervienen.

(Grm) es el número de FLs que han sido etiquetadas como formas ‘vacías’ o gramaticales.

(Num) es el número de FLs que han sido etiquetadas como formas numéricas.

(Xpr) es el número de FLs del corpus suprimidas por el investigador.

Variables textuales: en los textos tipo .TX3 y .TX5, se refiere al número de variables contextuales extratextuales que estructuran el texto.

Modalidades textuales: el número de subtextos o partes del corpus en que éste se puede particionar en función de cada modalidad de cada variable extratextual.

Variables de Categorización: Número de variables utilizadas para categorizar el texto

Modalidades de Categorización: total de categorías de clasificación del texto en función de las modalidades de cada variable de categorización.

TABLA 6. RESULTADOS ESTADÍSTICOS GENERALES

(UL) Unidades léxicas	160
(VL) Volumen léxico	4318
(FL) Formas léxicas	916
Sujetos emisores	160
(Grm) Formas gramaticales	257
(Num) Formas numéricas	0
(Xpr) Formas suprimidas	0
Variables textuales	3
Modalidades textuales	9
Variables de Categorización	3
Mod. de Categorización	18

Diccionario de FLs

La tabla 7 muestra el listado de las FLs del corpus ordenado por orden alfabético y por orden ascendente de frecuencias. Cada FL va precedida por su número de orden en el Diccionario y por su frecuencia absoluta de aparición.

En caso de lematización, aparecería una tercera columna ordenada por la frecuencia descendente de las formas lematizadas (FLr).

Los procedimientos automáticos de tratamiento del material verbal se basan en el análisis de las frecuencias de aparición de las formas léxicas. Se considera que las formas más frecuentes son demasiado comunes y no aportan significado semántico específico; suelen clasificarse como ‘palabras vacías’ o palabras gramaticales (artículos, conjunciones, preposiciones, etc.). Georges K. Zipf (1935, 1949) formuló su famosa ecuación de relevancia, para determinar cuándo un término es representativo de un corpus: listando las palabras en orden decreciente de frecuencia de aparición, el producto de la frecuencia relativa de cada palabra por su número de orden se aproxima a una constante, de modo que la palabra que ocupa la posición r -ésima se distribuirá según la ecuación:

$$f(r) = k / (r+w)b$$

donde $r = 1, 2, 3, \dots$, con w y b parámetros variables y donde el valor de k sólo depende del tamaño del texto. Se verifica de modo aproximado que los términos con mayor contenido semántico de un documento se encuentran en la zona de transición entre las palabras de frecuencia muy alta (comunes) y las de frecuencia muy baja (raras, características idiosincrásicas del vocabulario de autor). Puede establecerse un umbral superior y otro inferior de frecuencias de ocurrencia de un término en un texto, con el fin de separar los que se considerarían elementos poco relevantes para el sistema. Estos filtrados son muy útiles para reducir los términos representativos de un texto, cuando se utilizan métodos estadísticos que suelen operar con un número limitado de elementos.

TABLA 7. DICCIONARIO FELI. TX5.

1	62	A		694	202	QUE
2	1	A_DIARIO		913	188	Y
3	1	A_PARTIR_DE		220	182	DE
4	1	A_QUE		455	141	LA
5	5	A_VECES		158	125	CON
6	2	ABSOLUTA		282	95	EN
7	1	ABSOLUTO		871	94	UN
8	1	ACCEPTANT		270	85	EL
9	4	ACCIONES		112	74	BIEN
10	1	ACEPTADA		521	64	MISMO
11	1	ACEPTALOS		477	63	LOS
12	2	ACEPTAR		1	62	A
13	1	ACEPTARSE		830	62	TE
14	1	ACEPTAS		473	61	LO
15	1	ACONSEGUEIXES		304	56	ES
16	1	ACONTECIMIENTO		899	54	VIDA
17	1	ACONTECIMIENTOS		320	50	ESTADO
18	1	ACORDE		874	46	UNO
19	1	ACTITUD		773	42	SENTIRSE

Descripción de unidades léxicas (UL) prototípicas y atípicas en los textos

En la tabla 8 se calculan índices de representatividad de cada UL en función de la suma de frecuencias de las FLs que la componen y se ordenan en función de su frecuencia descendente. Se selecciona un número de UL que muestre las características más abundantes en el texto. Es un modo rápido de resumir los aspectos más característicos del corpus y disponer de una visión global. Junto a las ULs prototípicas se muestran también, como contrapartida, las ULs más atípicas para dar cuenta de aspectos particulares del corpus. Dado el modo de calcular el índice de representatividad, se muestran dos formas: la forma larga privilegia las ULs que contienen más FLs, y la forma corta, que divide el total por el número de FLs en cada UL, viene a ser una forma ponderada (Lebart, 1982). El módulo presenta este tipo de tablas para el corpus completo tomado como una unidad y para las ULs que forman cada texto del corpus (tabla 9).

TABLA 8. SELECCIÓN ORDENADA DE UL (TEXTO COMPLETO) PROTOTÍPICAS Y ATÍPICAS

<p><i>UL TIPICAS TEXTO COMPLETO Textos largos</i></p> <p><56>(2)[56]{1,2,2} Creo que la felicidad es un principio la ausencia de dolor cualquier dolor pero prefiero decir que la felicidad es un estado en el que uno se siente bien en paz enteramente tranquilo y los de su alrededor gozan de una vida con pocas preocupaciones o de bienestar general.</p> <p><52>(2)[52]{2,2,2} Sentirse bien con uno mismo y con los demás seguir el día a día orgulloso de aquello que se tiene y sabiendo valorar en cada momento las cosas buenas y las malas y mirar a la cara a la vida y afrontarla con valor y fuerza.</p> <p><118>(3)[118]{2,1,2} Estar y sentirse bien contigo mismo con tus cosas con tu vida sentirse satisfecho y saber encauzar los distintos aspecto de tu vida y tus problemas de forma positiva y productiva y poder encauzar tu vida de la manera que mas te gusta.</p>
<p><i>UL TIPICAS TEXTO COMPLETO Textos cortos</i></p> <p><48>(1)[48]{2,3,1} Sentirse bien con uno mismo y con lo que le rodea.</p> <p><54>(2)[54]{1,1,3} Sentirse bien con uno mismo y con los demás.</p> <p><3>(1)[3]{2,4,1} Estar satisfecho con tu vida con la gente que te rodea y con lo que has conseguido en el transcurso de los años.</p> <p><53>(2)[53]{2,1,3} Sentirte bien contigo mismo y con el mundo que te rodea.</p> <p><88>(2)[88]{1,1,2} La alegría a la hora de vivir.</p>
<p><i>UL ATIPICAS TEXTO COMPLETO Textos largos</i></p> <p><70>(2)[70]{1,2,3} La sensación producida por ver cumplidas todas aquellas cosas buenas según la propia escala de valores y a partir de ahí mayor cuantas más ilusiones creadas se vayan satisfaciendo.</p> <p><84>(2)[84]{2,1,1} La estabilidad por dentro fijarte y disfrutar del día a día estar en paz contigo misma intentar superar los problemas y no maximizarlos es la media entre la tranquilidad y la fiesta.</p> <p><62>(2)[62]{2,1,2} Sentirte bien contigo mismo y con la gente de tu entorno es para mí querer todo lo que tienes apreciar cada instante y cada cosa estar feliz es sentirte bien puedes tener un mal día pero puedes seguir siendo feliz.</p>
<p><i>UL ATIPICAS TEXTO COMPLETO Textos cortos</i></p> <p><87>(2)[87]{1,2,2} Equilibrio.</p> <p><29>(1)[29]{2,1,1} Un sentimiento por el cual os encontramos mas o menos alegres eufóricos contentos y realizados.</p> <p><62>(2)[62]{2,1,2} Sentirte bien contigo mismo y con la gente de tu alrededor para mí es querer todo lo que tienes apreciar cada instante y cada cosa estar feliz es sentirte bien puedes tener un mal día per puedes seguir siendo feliz.</p> <p><86>(2)[86]{2,2,1} Una utopía que jamás alcanzamos pero que muchas veces creemos poseer.</p>

TABLA 9. UNIDADES LÉXICAS PROTOTÍPICAS DE LOS TEXTOS
CORRESPONDIENTES A LAS MODALIDADES DE LA VARIABLE SEXO

<p><i>UL TIPICAS TEXTO 1 VARI_MOD1 hombre Textos largos</i></p> <p><56>(2)[56]{1,2,2} Creo que la felicidad es un principio la ausencia de dolor cualquier dolor pero prefiero decir que la felicidad es un estado en el que uno se siente bien en_paz enteramente tranquilo y los de su alrededor gozan de una vida con pocas preocupaciones o de bienestar general.</p> <p><59>(2)[59]{1,1,1} Estar en un momento de la vida en una situacion en que las cosas verdaderamente positivas vengan de casa y que los problemas no te afecten y puedas solucionarlos gracias a que lo verdaderamente importante para ti es positivo y te ves capaz de todo.</p>
<p><i>UL TIPICAS TEXTO 2 VARI_MOD2 mujer Textos largos</i></p> <p><52>(2)[52]{2,2,2} Sentirse bien con uno mismo y con los demas seguir el dia a dia orgulloso de aquello que_se tiene y sabiendo valorar en cada momento las cosas buenas y las malas y mirar a la cara_a la vida y afrontarla con valor y fuerza.</p> <p><118>(3)[118]{2,1,2} Estar y sentirse bien contigo mismo con tus cosas con tu vida sentirse satisfecho y saber encauzar los distintos aspecto de tu vida y tus problemas de forma positiva y productiva y poder encauzar tu vida de la manera que mas te gusta.</p>
<p><i>UL ATIPICAS TEXTO 1 VARI_MOD1 hombre Textos largos</i></p> <p><157>(3)[157]{1,4,3} Sentirse bien con uno mismo y con las otras personas que me rodean, así como con el ambiente que me rodea aceptando todo lo positivo y megativo que me sale al paso.</p> <p><111>(3)[111]{1,3,1} Que la gente que quiero estén bien y me quieran es lograr mis objetivos o simplemente intentar que se hagan realidad es sentirte útil necesario en la vida.</p>
<p><i>UL ATIPICAS TEXTO 2 VARI_MOD2 mujer Textos largos</i></p> <p><84>(2)[84]{2,1,1} La estabilidad por dentro, fijarte y disfrutar del día a día, estar en paz con tu misma, intemar superar los problemas y no maximizarlos es la media entre la tranquilidad y la fiesta.</p> <p><62>(2)[62]{2,1,2} Sentirte bien contigo mismo y con la gente de tu entrono es, para mí, querer todo lo que tienes; apreciar cada instatne y cada cosa es estar feliz, es sentirte bien; puedes tener un mal día pero seguir siendo feliz.</p>

Listado ordenado de frecuencias descendentes de las FLs predominantes en el corpus completo y en cada texto

En la tabla 10 se listan de modo compacto las FLs más frecuentes en cada texto. Por ejemplo las de la Variable 3, 'Felicidad general'. Pueden observarse distintos matices en el orden en que aparecen. Las frecuencias directas pueden convertirse en frecuencias ponderadas en función del tamaño de cada texto.

Descripciones de FLs comunes para cada variable contextual

Se calcula una chi-cuadrado para la distribución de cada FL en el conjunto de modalidades de cada variable contextual extratexto y se seleccionan ordenadamente aquellas FLs con una chi-cuadrado no significativa, es decir aquellas FLs que se distribuyen regularmente entre todos los textos de las modalidades de cada variable. Se muestran tantas tablas como variables contextuales.

FLs características de cada texto

Se seleccionan por orden de significación, en cada modalidad de cada variable contextual, aquellas FLs cuya chi-cuadrado es significativa para esa modalidad textual. Es decir, se muestran aquellas FLs que están sobre-representadas o infra-representadas en el texto de esa modalidad. De modo que, para cada texto, disponemos del conjunto de FLs que lo diferencian de los demás.

TABLA 10. FORMAS LÉXICAS DE LOS TEXTOS DE LA VARIABLE 'FELICIDAD GENERAL' ORDENADAS POR FRECUENCIAS DECRECIENTES. FL ACTIVAS: 916

<p><i>TEXTO 7 VAR3 Felicidad general MOD1 felicidad baja</i></p> <p>57 Que 49 Y 49 La 48 De 48 Con 33 En 29 El 25 Un 24 Lo 23 Bien 22 Te 19 Los 16 Mismo 16 Uno 15 A 14 Es 14 Vida 14 Estado 13 Sentirse 13 Las 13 Estar 13 Una 11 No 10 Felicidad 10 Tu 10 Su 8 Persona 7 O 7 Como 7 Para 7 Alrededor 6 Demás 6 Mas 6 Hace 6 Sientes 5 Se</p>
<p><i>TEXTO 8 VAR3 Felicidad general MOD2 felicidad media</i></p> <p>65 De 64 Que 57 Y 49 La 37 Un 35 En 33 El 31 Con 22 Bien 22 Vida 21 Mismo 20 Te 19 Los 19 A 19 Es 16 Estado 14 Lo 14 Las 13 Uno 13 Se 11 Sentirse 11 No 11 O 10 Una 10 Felicidad 10 Momento 9 Bienestar 9 Mi 9 Si 8 Estar 8 Contigo 8 Día 8 Por 8 Cuando 7 Sentimiento 7 Hace 7 Nos 7 Rodea 7 Me 7 Pero 7 Cada 7 Sentir 6 Demás 6 Como 6 Para 6 Que_se 6 Cosas 5 Tu</p>
<p><i>TEXTO 9 VAR3 Felicidad general MOD3 felicidad alta</i></p> <p>82 Y 81 Que 69 De 46 Con 43 La 32 Un 29 Bien 28 A 27 En 27 Mismo 25 Los 23 El 23 Lo 23 Es 20 Te 20 Estado 18 Vida 18 Sentirse 17 Uno 14 Las 13 No 13 Estar 13 O 13 Demás 13 Mas 12 Sentimiento 12 Ser 11 Una 11 Se 10 Como 10 Día 10 Nos 10 Que_se 9 Felicidad 9 Contigo 9 Rodea 9 Todo 8 Bienestar 8 Para 8 Persona 8 Por 7 Tu 7 Hace 7 Animo 7 Feliz 7 Del 7 Objetivos 6 Su 6 Cosas 6 Me 6 Misma 6 Puede 6 Vivir 6 Positivo 6 Ganas 6 Sin</p>

FL características en cada categoría, por cada variable

Listado ordenado por frecuencias en cada categoría de las FLs adscritas. Permite supervisar la corrección de las FLs atribuidas a cada categoría.

Listados específicos o Multitextos

El programa permite listar textos específicos que reúnan una o varias características determinadas, ya sea por modalidades de variables, por la presencia o ausencia de ciertas FLs, etc..

Descripciones contextuales intratexto

Este módulo está especialmente dedicado al análisis contextual intratextos. Permite mostrar tablas ordenadas del contexto en que aparece una determinada FL. El programa selecciona las ULs donde aparece la FL cuyo contexto deseamos ver y selecciona (-3) las 3 FLs anteriores y (+3) las 3 FLs posteriores formando texto seguido. En la forma extensa (tabla 11; FL 'BIEN') el signo '=' representa la FL cuyo contexto se describe. Hay una forma acumulada en la que, por cada columna, se agrupan las FLs repetidas (tabla 12 'VIDA'), generando estructuras de significados posibles. El signo '&' representa la FL cuyo contexto se describe. Estas tablas pueden generarse para el corpus completo tomado en tanto que unidad y para cada texto formado por las variables extratextuales en cada modalidad.

TABLA 11. CONTEXTO EXTENSO DE UNA FL ===== BIEN ==== (FREC.= 74)

PARA PODER ESTAR = CON EL RESTO
 QUE TE SIENTAS = Y CONTENTO CONTIGO
 ETC ES ESTAR = EN TODOS LOS
 HECHO DE SENTIRSE = ANIMICAMENT PORQUE TE
 PERSONA SE SIENTE = CONSIGO MISMO Y
 ME HACE SENTIR = Y QUE_SE ALCANZA
 DEL CUERPO MAS = POSITIVO CUANDO YO
 DIA PARA UN = AJENO Y PERSONAL
 UNO SE SIENTE = CONSIGO MISMO Y
 BUEN HUMOR CONTENTO = EN TI MISMO
 PORQUE TE SIENTES = EN MUCHOS SENTIDOS
 UNO SE SIENTE = EN_PAZ ENTERAMENTE TRANQUILO
 CUANDO TE ENCUENTRAS = CONTIGO MISMO Y
 LA PERSONA SENTIRSE = CONSIGO MISMO Y
 M12M M6M SENTIRSE = CON UNO MISMO
 RESPETARME ES ESTAR = CONMIGO MISMA CREO
 PARA MI ESTAR = CON UNO MISMO
 M12M M9M SENTIRSE = CON UNO MISMO
 TE HACE ESTAR = CONTIGO Y CON
 SOÑAR SI ESTAS = CONTIGO MISMO Y
 TE RODEAN ESTAS = MORALMENTE Y ESO
 TODO LO VES = INCLUSO LO MALO

TABLA 12. CONTEXTO ACUMULADO DE UNA FL: “VIDA” (-3,+3) FREC 54.

6 de 19 de 35 la & 10 y 5 un 2 lo
 4 disfruta 7 con 7 tu & 5 en 3 la 2 situacio
 3 que 4 en 5 su & 4 de 2 por 2 y
 3 satisfac 2 durante 3 nuestra & 2 con 2 una 1 a
 2 y 2 presenta 1 lo & 2 es 1 afrontar 1 adelante
 1 aconteci 2 vivir 1 mi & 2 me 1 apreciar 1 atributo
 1 alcanzar 1 a 1 propia & 2 que 1 cada 1 bien
 1 aspecto 1 brinda 1 una & 1 aceptar 1 como 1 como
 1 aspectos 1 cara_a - - - - & 1 bajo 1 con 1 con
 1 avanzamo 1 componen - - - - & 1 bien 1 conocert 1 cosas
 1 buenas 1 comprend - - - - & 1 cada 1 ella 1 debe
 1 circumst 1 cuando - - - - & 1 en_gener 1 es 1 dia
 1 cosas 1 encauzar - - - - & 1 expresa 1 esta 1 el
 1 dificult 1 enfocar - - - - & 1 la 1 estas 1 ello
 1 entorno 1 gusta - - - - & 1 no 1 felicida 1 engloba
 1 ese 1 la - - - - & 1 para 1 forma 1 eres
 1 es_decir 1 mejorar - - - - & 1 que_se 1 general 1 es
 1 etapa 1 momento - - - - & 1 recorrie 1 incluso 1 estado
 1 fluir 1 ofrece - - - - & 1 saber 1 las 1 felicida
 1 gozan 1 toda - - - - & 1 sentirse 1 llevas 1 fundamen
 1 la 1 todo - - - - & 1 si 1 lo 1 gente
 1 m6m 1 ver - - - - & 1 sin 1 muestra 1 individu
 1 me 1 viendo - - - - & 1 tanto 1 persona 1 la

El programa permite presentar también los contextos de una FL como tabla radial. La FL descrita queda en el centro de la tabla y a su alrededor se van posicionando aquellas FLs que le son próximas en el conjunto de las UL, siguiendo una rotación radial del centro a la periferia (tabla 13). Las FLs que aparecen con más frecuencia entre las que anteceden se situarán a la izquierda de la tabla y las que con mayor frecuencia siguen a la FL descrita se situarán a la derecha. Las FLs que están tanto antes como después se situarán en el centro. En el margen izquierdo de la tabla se muestran las FLs más frecuentemente asociadas por orden decreciente de frecuencia. Esta forma de presentar el contexto de una FL tiene su utilidad para el investigador.

Estadísticas generales

En la tabla 14 se muestra una tabla de Burt, con el cruce de todas las modalidades de las variables contextuales. Esta tabla puede ser objeto de un Análisis de Correspondencias Múltiples (ACM) con líneas o columnas suplementarias de las formas léxicas y su frecuencia de aparición en cada modalidad de cada variable.

La tabla 15 muestra de modo condensado los parámetros principales del material léxico del corpus distribuido por modalidad de las variables contextuales. Por cada modalidad se indica el número de ULs, el volumen total de FLs que contiene y, entre paréntesis, el porcentaje de FLs que representa en cada modalidad. Bajo el epígrafe ORIGINAL se presentan el número de FLs distintas, ente paréntesis el porcentaje de FLs distintas sobre el total de FLs

distintas, seguido del porcentaje que representa sobre el volumen total de FLs de cada modalidad. Bajo el epígrafe RECODIFICADO se muestran los mismos resultados pero referidos a las FL lematizadas en cada modalidad de cada variable. Bajo el epígrafe PROMEDIOS aparece el promedio de FLs originales por ULs de cada modalidad, y lo mismo para el caso de FLs lematizadas o recodificadas.

TABLA 13. CONTEXTO RADIAL DE LA FL: "FELICIDAD" (-6,+6)

ORIGINAL (-6, +6)

recorriendo amado ganas otro
 67+la | tenemos para_ que vive dificil muchísimos
 11+de | contento sino que_es mi completa
 10+no todo_el_mund se muy_bien vida bien
 8+y | solo quienes laboral los mismo
 8+el personalidad limpiar etc muy tal
 8-es | mismas gente conductas encontramos optimiza
 7+las | animado manifieste buenas puede compartir
 7 un | viviendo estudiada cada a amar
 6+del | niveles consigo te ser si
 6+sentirse | alto punto personal y mas
 5+mas |conseguir tanto un de del
 5+rodean | tipo es FELICIDAD la las
 5+ser | esto_es positivo tener no sentirse
 4+en | a±os son sabiendo el rodean
 4+a posiblemente energía uno en una
 4+con | absoluta infelicidad alcance con como
 4+puede | metas dependiendo busca dentro_de nos
 4+si | persona ideal demás mayor quererse
 4+una | peso llegar existe nivel tardaríamos
 3+amar | sus sabe me que ausencia
 3+como | aspiramos seguir pero dia camino
 3+compartir | superar fuerzas rodea cosas meta
 3+dentro_de | todos prohíbe al esa negativo
 3+encontramos | siente familiar impulsa para

TABLA 14. CRUCE DE VARIABLES

H	22	0	4	7	8	3	7	7	8
M	0	138	56	34	37	11	40	42	56
E1	4	56	60	0	0	0	15	17	28
E2	7	34	0	41	0	0	10	14	17
E3	8	37	0	0	45	0	18	15	12
E4	3	11	0	0	0	14	4	3	7
F1	7	40	15	10	18	4	47	0	0
F2	7	42	17	14	15	3	0	49	0
F3	8	56	28	17	12	7	0	0	64

TABLA 15. RECUENTOS DE FORMAS LÉXICAS POR VARIABLES

TEXT	TEXTO		ORIGINAL		RECODIFICADO		PROMEDIOS	
	UL	Formas Léxicas	Distintas	D/T	Distintas	R/T	morig	mreco
H	22	632 (15%)	295 (32%)	47%	152 (33%)	37%	13.4	6.9
M	138	3686 (85%)	810 (88%)	22%	413 (90%)	17%	5.9	3.0
E1	60	1601 (37%)	523 (57%)	33%	275 (60%)	27%	8.7	4.6
E2	41	1079 (25%)	374 (41%)	35%	203 (44%)	29%	9.1	5.0
E3	45	1248 (29%)	388 (42%)	31%	191 (41%)	23%	8.6	4.2
E4	14	390 (9%)	195 (21%)	50%	103 (22%)	41%	13.9	7.4
F1	47	1264 (29%)	434 (47%)	34%	228 (49%)	28%	9.2	4.9
F2	49	1377 (32%)	432 (47%)	31%	237 (51%)	26%	8.8	4.8
F3	64	1677 (39%)	519 (57%)	31%	267 (58%)	25%	8.1	4.2
TOTAL	160	4318 (100%)	916 (100%)	21%	461 (100%)	11%	5.7	2.9

Longitud de FLs por Variables

Sabido es que el número de caracteres de una forma léxica está relacionado con el grado de concreción/abstracción del lenguaje (tabla 16). Así, en la expresión espontánea de emociones y hechos de la vida cotidiana suelen utilizarse palabras cortas y directas, mientras que la utilización de palabras largas se asocia con el lenguaje abstracto. Esta característica permite detectar un efecto del estilo general del corpus y caracterizar la expresión en cada modalidad de cada variable. En esta tabla se presenta la frecuencia de aparición de FLs de 1 o más caracteres (hasta 12 letras), los totales y el promedio de caracteres por cada texto. En este caso los sujetos masculinos del corpus tienden a utilizar en promedio FLs más largas que el resto. El texto con FLs más cortas es el formado por aquellos que, en las pruebas que evaluaban el grado de felicidad general, se sitúan en el medio (4.43). En conjunto se muestra un texto de estilo espontáneo y familiar (4.55).

TABLA 16. LONGITUD DE FLS POR VARIABLES

N_1	1L	2L	3L	4L	5L	6L	7L	8L	9L	10L	11L	12L	Total	Media
H	34	160	120	87	65	48	42	42	39	13	16	6	683	4.66
M	259	813	771	465	430	282	248	256	193	85	63	35	3930	4.53
E1	117	334	349	219	179	115	120	109	78	36	29	14	1713	4.52
E2	75	255	203	113	127	95	70	79	64	25	22	10	1148	4.61
E3	77	299	251	170	143	93	76	78	78	24	21	14	1336	4.52
E4	24	85	88	50	46	27	24	32	12	13	7	3	416	4.56
F1	75	304	261	165	147	88	80	80	72	34	25	15	1357	4.56
F2	90	339	270	177	161	113	85	100	62	25	19	8	1464	4.43
F3	128	330	360	210	187	129	125	118	98	39	35	18	1792	4.64
TOTAL	293	973	891	552	495	330	290	298	232	98	79	41	4613	4.55

Raíces y desinencias

Desde un punto de vista semántico, el modo de vehicular el significado de las FLs se asocia al modo en que se forman las palabras mediante radicales y desinencias, o prefijos y sufijos (tabla 17). El programa extrae del conjunto de FLs grupos de 1,2,3... caracteres iniciales y finales, calcula su frecuencia de aparición en el corpus y da una lista en orden decreciente del porcentaje sobre el total de las combinaciones. Esta tabla muestra un resumen para el corpus analizado. Es un modo muy sintético de captar, más allá de las formas concretas, ciertas características generales de la expresión. En este caso la aparición de las iniciales MI, CO, FE, TI, VI sugiere una expresión de la autoreferencia, la referencia al grupo, la felicidad, las relaciones diádicas y la vida como *leit-motivs* de la expresión del corpus. Como elemento de comparación mostramos también la tabla 18 para grupos de 3 caracteres.

TABLA 17. RADICALES Y DESINENCIAS. TEXTO ORIGINAL
GRUPO DE 2 LETRAS: % SOBRE 19652 FLS

	Ord-Radical			Ord Desinen	
	RA-	DES-		-RA	-DES
QU	2.34	0.00	AS	0.06	5.81
MI	1.48	0.45	UE	0.00	5.71
PE	1.46	0.00	DE	1.28	5.24
CO	1.41	0.22	OS	0.01	5.12
DE	1.28	5.24	EN	0.70	4.89
FE	1.27	0.00	ES	1.02	4.50
PR	1.22	0.00	ON	0.00	4.35
CU	1.21	0.00	DO	0.03	4.07
DI	1.19	0.00	TE	0.44	3.93
AM	1.16	0.00	LA	0.94	3.73
TI	1.16	0.12	AR	0.03	3.20
VI	1.16	0.00	NO	0.30	2.93
PU	1.15	0.00	MO	0.19	2.73
BI	1.14	0.00	EL	0.48	2.63

TABLA 18. RADICALES Y DESINENCIAS. TEXTO ORIGINAL
GRUPO DE 3 CARACTERES: % SOBRE 64616 FL

	Ord-Radical			Ord Desinen	
	RA-	DES-		-RA	-DES
CON	0.45	4.10	QUE	0.38	7.44
MIS	0.43	0.10	CON	0.45	4.10
PER	0.43	0.07	IEN	0.00	2.79
FEL	0.38	0.00	ADO	0.00	2.52
QUE	0.38	7.44	TAR	0.00	2.42
COM	0.37	0.00	NTE	0.00	2.29

Análisis de Segmentos textuales

La probabilidad de que 2, 3, 4... FLs aparezcan seguidas en un corpus un número x de veces decrece casi exponencialmente, sobre todo en formas no instrumentales (tabla 19). Estos segmentos de texto vehiculan modismos o formas de expresión en cierta forma estereotipada. El inventario de segmentos repetidos (Lafon y Salem, 1983) también denominado método de ‘vecinos recurrentes’ aporta encadenamientos de FLs especialmente útiles para el análisis textual y ayuda a mejorar la descripción de los ejes factoriales cuando éstos se utilizan como elementos suplementarios en el AC.

En la tabla 20 se presenta un extracto de los 440 segmentos repetidos de 2 FLs por orden de frecuencia decreciente y por orden alfabético. Su consideración por parte del investigador es de gran utilidad para validar hipótesis auxiliares sobre el significado subyacente. La tabla 21 muestra un extracto de los 252 segmentos de 3 FLs donde se van perfilando mejor los significados que conllevan. En el corpus el segmento repetido más largo es de 9 FLs y su contenido es “*sentirse bien con uno mismo y con los demás*”, el cual aparece 2 veces.

TABLA 19. SEGMENTOS TEXTUALES. TEXTO ORIGINAL
HISTOGRAMA DE LA FRECUENCIA DE APARICIÓN DE SEGMENTOS DE 2 FLS

Nrep	frec	frec .acum
46 ->	1	1 .
42 ->	2	3 .
40 ->	1	4 .
37 ->	1	5 .
36 ->	1	6 . .
30 ->	1	7 . .
.....		
9 ->	8	42
8 ->	8	50
7 ->	10	60
6 ->	18	78
5 ->	18	96
4 ->	44	140
3 ->	58	198
2 ->	242	440

TABLA 20. SEGMENTOS REPETIDOS DE _2 FLS FREC = 440

46 y con	: 25 los demás	: 2 a mi
42 que te	: 24 la felicidad	: 2 a nivel
42 un estado	: 23 bien con	: 2 a si
40 lo que	: 23 estado de	: 2 alcanzar un
37 mismo y	: 22 con uno	: 2 alegre y
36 la vida	: 8 a día	: 2 alegría y
30 uno mismo	: 11 a gusto	: 3 algo que
28 sentirse bien	: 8 a la	: 2 algo que_se
27 de la	: 2 a las	: 2 all que
25 con los	: 4 a los	: 2 allò que

TABLA 21. SEGMENTOS REPETIDOS DE _3 FLS FREC = 252.

29 mismo y con	· 10 m6m un estado	· 6 que te rodea
22 y con los	· 10 m12m sentirse bien	· 5 que te rodean
21 con uno mismo	· 9 en el que	· 4 que te encuentras
20 un estado de	· 9 es un estado	· 5 que te sientes
20 bien con uno	· 9 todo lo que	· 2 que te vas
16 sentirse bien con	· 5 y con lo	· 5 que te hace
16 con lo que	· 3 y con las	· 20 un estado de
15 estado de ánimo	· 22 y con los	· 4 un estado emocional
13 uno mismo y	· 4 y con la	· 2 un estado anímico
13 con los demás	· 4 y con el	· 4 un estado en
11 de la vida	· 2 y con su	· 2 un estado permanente
11 contigo mismo y	·	·

Contraste de patrones léxicos

Complementario al AC, el contraste entre perfiles de distribución de las FLs en los distintos subtextos permite utilizar lambda, el coeficiente de predicibilidad de Guttman para variables nominales, como medida del grado en que dos textos comparten su FLs (tabla 22). Se trata de un coeficiente que no impone restricciones sobre el número de categorías, ni requiere de hipótesis especiales, y es fácilmente interpretable. Fluctúa entre 0 y 1, no tiene valores negativos y es de doble sentido asimétrico (un coeficiente lambda A y uno B) o simétrico (un coeficiente C para mutua predictibilidad). Estos coeficientes resumen el grado en que el conocimiento de los valores de una variable ayuda a predecir los valores que toma la otra (Freeman, 1965, pp. 102-110; Guttman, 1968):

$$\text{Lambda asimétrica} = (\text{Sum } f_i - F_d) / (N - F_d),$$

donde F_i es la frecuencia máxima de las correspondientes a las de la variable independiente.

F_d es la frecuencia máxima correspondiente a los totales de la variable dependiente.

N es el número de casos.

Dadas las FLs de los dos textos, se comparan sus perfiles de distribución. Se contabiliza la frecuencia de presencia exclusiva en cada texto, y la presencia mixta en una tabla 2x2. La descripción se completa con un extracto de las FLs exclusivas de cada texto (A y D) y de las FLs compartidas (B y C) ordenadas por sus frecuencias decrecientes.

En la tabla 23 se presenta la matriz de coeficientes que estiman la asociación entre todos los textos, tomados de dos en dos. Cabe señalar que existe una buena asociación (0.88) entre el texto de las mujeres y la edad más joven (E1: 18a), y entre las mujeres y “Felicidad alta” (0.86).

TABLA 22. CONTRASTE DE PATRONES LÉXICOS (COEF. DE ASOCIACIÓN DE GUTTMAN).
ORIGINAL FORMAS LÉXICAS COMPARTIDAS
ASOCIACIÓN ENTRE TEXTOS T1:VAR1_MOD1 HOMBRE->H - T2:VAR1_MOD2 MUJER->M
COEF. PREDICTIBILIDAD GUTTMAN= (A) 0.471; (B) 0.527; (C) 0.501

	A 18%	B 82%	C 71%	D 29%
3 absoluta	52 de	48 que	5 contigo	
3 em	49 la	46 y	5 día	
3 escala	24 un	31 con	4 del	
3 les	21 el	22 en	4 misma	
3 según	21 es	18 bien	4 puede	
3 simplemente	19 vida	16 mismo	4 sientes	
3 sobre_todo	14 o	16 a	3 consigo	
3 tranquilo	11 no	16 te	3 disfrutar	
3 valores	11 una	16 lo	3 gusto	
3 verdadera	9 por	15 los	2 entorno	
2 a_partir_de	8 camino	12 estado	2 mundo	
2 a_que	8 se	11 uno	2 rodean	

TABLA 23. MATRIZ DE COEFICIENTES DE PREDICTIBILIDAD SIMÉTRICOS DE GUTTMAN.

TEXTOS	T1	T2	T3	T4	T5	T6	T7	T8	T9
H		50	46	54	60	46	58	55	56
M			88	77	77	54	79	83	86
E1				52	54	35	66	71	77
E2					57	30	62	66	71
E3						36	73	70	65
E4							41	50	49
F1								55	55
F2									60
F3									

En la tabla 24 se presenta el análisis de contraste entre los textos pertenecientes a “Felicidad Baja” (F1) y “Felicidad Alta” (F3) para FLs lematizadas, donde se han suprimido las formas gramaticales. Pueden observarse los bajos coeficientes de predictibilidad de Guttman (0.07), lo que viene a indicar que ambos textos están muy diferenciados.

El mismo módulo puede ser utilizado para comparar un texto con un perfil de distribución externo y verificar el grado de asociación entre ambos. Para ciertos propósitos podemos considerar que cada texto contiene su propia “huella de identificación”, tomando su perfil de distribución léxica. De modo que un texto puede compararse con otros perfiles conocidos. Es un método aplicable a la determinación de autorías textuales o a la proximidad-distancia ideológica entre distintos corpus, entre otras aplicaciones.

Tablas para el Análisis

Se pueden crear numerosas tablas que cruzan la información textual del corpus y son susceptibles de un Análisis de Correspondencias con objeto de

generar los ejes de proyección que mejor den cuenta de las estructuras relacionales entre las diversas FLs. Los perfiles distribucionales de las FLs en los textos formados por las modalidades pueden dar lugar a tablas de contingencia por separado o apiladas.

TABLA 24. CONTRASTE DE PATRONES LÉXICOS (COEF. DE ASOCIACIÓN DE GUTTMAN).
PARA FLS LEMATIZADAS Y FLS GR SUPRIMIDAS
ASOCIACIÓN TEXTOS T7:VAR3_MOD1 FEL BAJA->F1 – T8:VAR3_MOD2 FEL MEDIA->F2
COEF. PREDICTIBILIDAD DE GUTTMAN=(A) 0.059 (B) 0.087 (C) 0.073

A 47%	B 53%	C 56%	D 44%
7 aunque	50 bien	43 vida	14 sentir
7 personalidad	29 sentirse	18 bienestar	8 preocupación
7 positivas	22 felicidad	16 cuando	6 dolor
7 sensación	18 persona	16 día	6 llevas
7 transitorio	13 demás	14 rodea	6 sentimos
4 años	13 sientes	14 sentimiento	4 agradables
4 conseguido	11 satisfecho	12 que_se	4 ambiente
4 encontrar	11 sentirse	10 ánimo	4 apreciar
4 intentar	9 camino	10 personal	4 aquellas
4 meta	9 consigo	10 personas	4 ausencia
4 negativos	9 problemas	10 poder	4 be
4 posibilidad	7 alegre	10 puede	4 buena

Las distribuciones más evidentes son: FLs x Textos de modalidades (tabla 25) o FLs x Categorías; pero también puede ser útil considerar otras tablas como FLs x sujetos (en los casos de .TX4 y .TX5) o FLs x FLs. Esta última tabla tiene particular interés para determinar la asociación interna entre FLs.

TABLA 25. TABLA DE CONTINGENCIA DE FLS X MODALIDADES DE VARIABLES (EXTRACTO)

	H	M	E1	E2	E3	E4	F1	F2	F3	Total
Que	25	177	76	49	60	17	57	64	81	606
Y	19	169	75	53	46	14	49	57	82	564
De	33	149	60	48	59	15	48	65	69	546
La	31	110	46	40	46	9	49	49	43	423
Con	11	114	47	28	33	17	48	31	46	375
En	13	82	26	26	33	10	33	35	27	285
Un	15	79	33	22	28	11	25	37	32	282
El	13	72	32	19	22	12	29	33	23	255
Bien	9	65	23	18	25	8	23	22	29	222
Mismo	4	60	22	15	20	7	16	21	27	192
Los	6	57	22	20	14	7	19	19	25	189
Total	317	2103	869	621	727	203	708	788	924	7260

Su construcción puede realizarse de varios modos:

1. Contabilizando en cada UL las veces que cada FL es co-ocurrente con otras.
2. Restringiendo la proximidad entre FLs a las 3 o 6 anteriores y posteriores.
3. Calculando un índice de proximidad entre dos FLs en el espacio de los textos.

El programa dispone de un procedimiento denominado 'Unitexto' que permite la elaboración automática o interactiva de una tabla donde las FLs más frecuentes, o potencialmente más significativas, son rastreadas por el corpus para contabilizar qué otras FLs aparecen como co-ocurrentes en sus entorno respectivos.

El Análisis de Correspondencias

No es propósito de este artículo tratar *in extenso* las características y propiedades del AC en el análisis de grandes tablas de contingencia. Su uso en general, y específicamente en el tratamiento lexicológico, está suficientemente documentado (Benzécri, 1973; Bécue, Lebart y Rajadell, 1992). Destacaremos solamente su adecuación a este tipo de tablas, ya que permite, mediante la operacionalización de la distancia de chi-cuadrado, factorializar en un espacio multidimensional y dual, y por tanto cuantificar lo cualitativo de las variables nominales. Al mismo tiempo facilita una visualización plástica muy efectiva de las interrelaciones e interdependencias entre los elementos que constituyen el universo descriptivo de los elementos línea y columna de la tabla de datos. Esta técnica de cálculo permite comparar las distintas partes de un corpus para mostrar las similitudes o diferencias entre distintos emisores, distintos textos, distintos conjuntos de campos semánticos, distintos sistemas de categorización, y poner de relieve la infra o sobrerrepresentación del vocabulario en determinadas partes del texto.

A modo de ejemplificación presentaremos algunos de los resultados más representativos del AC de la tabla 25 en las tablas subsiguientes, a través de algunos de sus resultados más representativos. La tabla 26 da cuenta de los valores propios de la matriz diagonalizada, indicando el porcentaje de información (inercia) transmitido por cada factor o eje de proyección, así como el porcentaje acumulado. Aunque en los resultados (p -valor 0,987) no se puede descartar la hipótesis nula de independencia entre líneas y columnas, tiene interés ilustrativo realizar la proyección de los dos primeros ejes que, juntos, aportan un 53.608 % de la información contenida en la tabla analizada.

La tabla 27 muestra las coordenadas factoriales de las modalidades de las variables contextuales de los elementos en columna, señalando en negrita sus contribuciones absolutas más destacadas. Para ilustrar la proyección dual se añaden algunas de las FLs que contribuyen a los dos primeros ejes (tabla 28). Una interpretación rápida de la tabla de coordenadas nos indica que el primer eje diferencia claramente el texto producido por los hombres (H) y la medida

de Felicidad Alta (F3). Se observa en este primer eje una secuencia ordenada que gradúa desde Felicidad Baja (G1: 0,686) a Felicidad Media (F2: 0,359) y, en el otro polo, Felicidad Alta (F3: -0,834). En función de esto vemos que los sujetos cuyos textos se aproximan más a la Felicidad Alta son tanto los más jóvenes (E1: -0,683) como los de mayor edad (E4: -0,825). Algunas de la FLs más asociadas el texto de los que puntúan en Felicidad Alta sería por orden SER, NOS, (LOS) DEMÁS, TODO, SENTIMIENTO, CONTIGO, etc.; y en el polo de Felicidad Baja destaca el uso de FLs como VIDA, BIENESTAR, CUANDO, etc.. Esta información aportada por el AC ayuda al investigador a la elaboración de hipótesis sobre el contenido del corpus analizado. Señalemos que lo que aquí se ha presentado es sólo una pequeña muestra del tipo de resultados que pueden obtenerse.

TABLA 26. VALORES PROPIOS Y PORCENTAJES DE INERCIA Y PRUEBA DE INDEPENDENCIA

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>	<i>F6</i>
Valor propio	0,015	0,009	0,007	0,006	0,005	0,003
Las filas dependen de las columnas (%)	33,066	20,543	15,558	14,022	11,119	5,694
% acumulado	33,066	53,608	69,166	83,188	94,306	100,000
<i>Chi-cuadrado</i> (Valor observado) 369,1408						
<i>Chi-cuadrado</i> (Valor crítico) 481,4586 (gl. 432)						
<i>p-valor</i> 0,987, <i>alfa</i> 0,05						
Como el p-valor calculado es mayor que el nivel de significación $\alpha = 0,05$, se puede aceptar la hipótesis de independencia entre líneas y columnas.						

TABLA 27. COORDENADAS ESTÁNDAR (COLUMNAS), PRINCIPALES CONTRIBUCIONES (NEGRITA). UN EXTRACTO DE LAS FLs (LÍNEAS) CON MAYORES CONTRIBUCIONES EN LOS DOS PRIMEROS EJES

	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>F5</i>	<i>F6</i>
<i>H</i>	3,645	-0,253	-1,119	0,996	1,831	0,424
<i>M</i>	-0,572	0,040	0,175	-0,156	-0,287	-0,066
<i>E1</i>	-0,683	0,285	0,035	1,538	0,229	1,535
<i>E2</i>	0,200	-2,600	0,730	-1,074	0,361	-0,377
<i>E3</i>	0,882	1,190	-0,132	0,137	-1,349	-1,716
<i>E4</i>	-0,825	2,416	-1,894	-3,818	2,737	0,688
<i>F1</i>	0,686	0,860	2,320	-0,450	0,300	0,498
<i>F2</i>	0,359	-0,486	-1,379	-0,611	-1,553	1,127
<i>F3</i>	-0,834	-0,257	-0,636	0,864	1,076	-1,335

TABLA 28. EXTRACTO DE LAS FLS (LÍNEAS) CON MAYORES
CONTRIBUCIONES EN LOS DOS PRIMEROS EJES

<i>Con</i>	-0,668	1,018	1,029	-0,861	0,971	0,584
<i>Vida</i>	1,013	0,369	-1,248	0,991	-0,329	1,830
<i>Sentirse</i>	-0,450	-1,123	1,222	0,680	-0,265	-1,266
<i>Estar</i>	-0,411	2,448	0,449	-1,528	0,689	-1,413
<i>Demás</i>	-2,017	-1,788	0,317	-0,972	2,099	-0,474
<i>Sentimie</i>	-1,939	0,191	-0,342	2,414	-0,423	0,568
<i>Bienestar</i>	0,776	-1,941	-0,906	-1,600	0,011	-0,648
<i>Mas</i>	-0,623	2,149	-1,362	-0,251	4,623	-1,837
<i>Tu</i>	-0,314	3,644	1,531	1,486	-0,736	1,070
<i>Contigo</i>	-1,896	-0,686	-0,995	0,486	-2,174	-0,669
<i>Ser</i>	-2,289	1,365	-1,359	3,818	0,039	-0,424
<i>Nos</i>	-2,223	0,540	-2,709	0,295	0,763	0,908
<i>Rodea</i>	-1,482	-1,731	-1,116	-1,106	0,085	-0,798
<i>Cuando</i>	0,407	-2,520	0,020	-0,458	-2,628	0,245
<i>Todo</i>	-1,998	-1,164	-0,254	3,266	-0,072	1,215

Finalmente, la figura 1 presenta el mapa de proyección de los dos primeros ejes factoriales (la información más destacada), mapa que permite visualizar simultáneamente (proyección dual) los elementos columna y los elementos línea.

No exploraremos en este artículo, a pesar de su importancia, la posibilidad de poder completar el análisis con líneas y columnas suplementarias (o ilustrativas), para enriquecer su contexto de interpretación. Añadamos que tiene mucho interés acompañar el AC con una Clasificación Jerarquizada Ascendente (Ward) sobre las coordenadas factoriales de los elementos, para una mejor descripción de los hechos de estructura, a través de los clusters o racimos de elementos que se forman.

A la búsqueda del significado

Como conclusión de esta presentación del programa ConText para el análisis contextual de material simbólico y, especialmente, para la descripción de corpus verbales, sobresale el hecho de que detrás de todas las técnicas de análisis subyace el esfuerzo por desentrañar el significado último del lenguaje. El problema reside en que, aunque no sea lícito decir que cualquier sentido es igualmente posible y adecuado, también es cierto que no existe un único significado que sea más verdadero que otros.

Formas léxicas x Variables contextuales

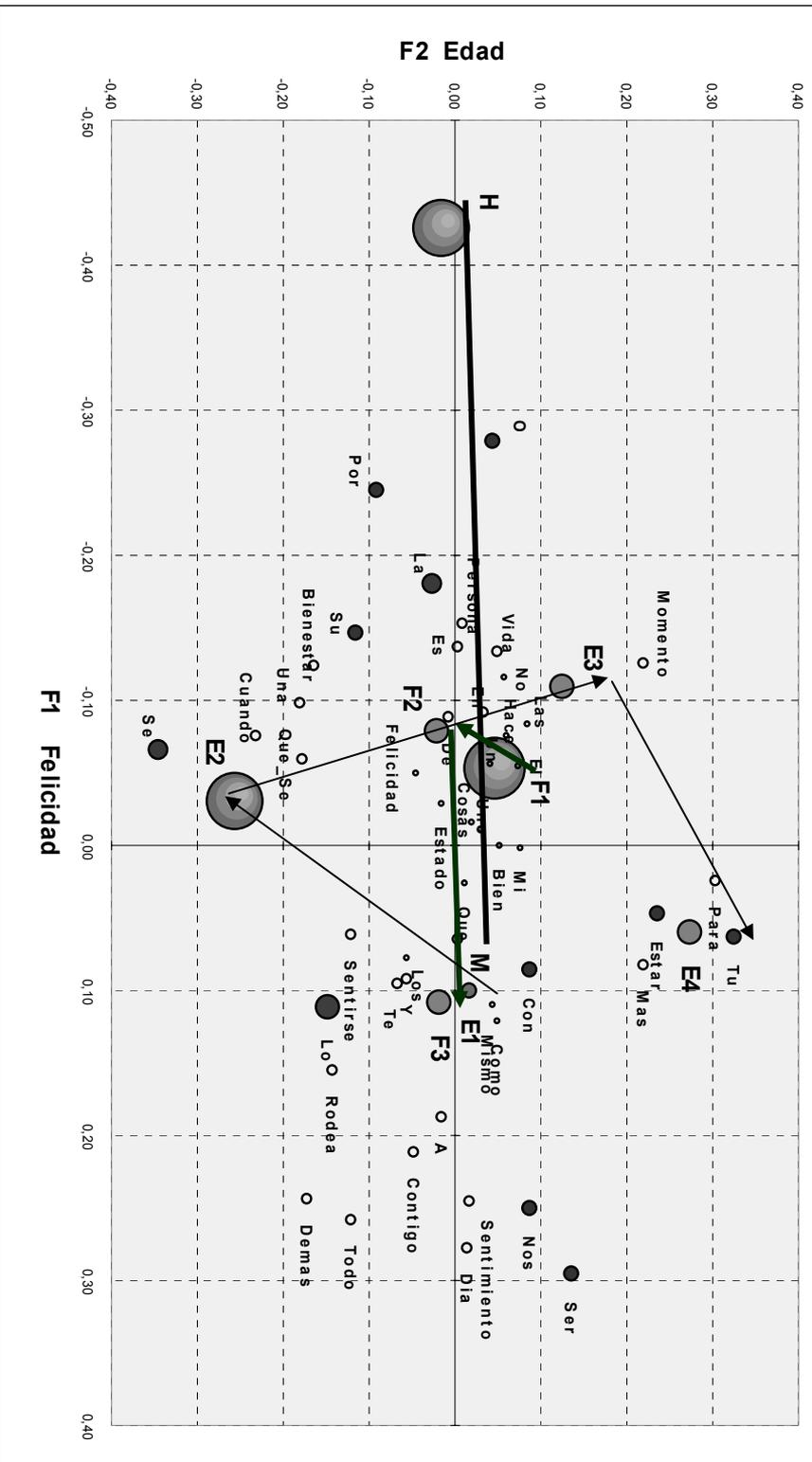


Figura 1. Plano de proyección del A. de Correspondencias de la tabla 25.

En el lenguaje se encajan racimos de significado, compatibles desde distintas perspectivas, que vehiculan la gran complejidad de un fenómeno cuyo control escapa a los propios hablantes. En el comportamiento lingüístico se incrustan, además de los significados particulares que el hablante tiene voluntad y consciencia de comunicar, todo un universo concomitante de significados que expresan la cosmovisión de los grupos y de la sociedad a la que se pertenece.

Podemos afirmar que, tanto o más que el hablante moldea su ‘habla’ en cada situación concreta, el espíritu de la lengua, como construcción de la realidad social, habla a través de él. Es lo que en la literatura especializada se denomina ‘contenido latente’ concepto desarrollado desde distintas corrientes como:

1. El determinismo lingüístico que considera el lenguaje como variable independiente, causante de la realidad social y de los esquemas cognitivos y perceptivos. El principio de que “las palabras moldean nuestras percepciones” (hipótesis Sapir-Whorf) establece la existencia de una cierta relación entre las categorías gramaticales del lenguaje que una persona habla y la forma en que la persona entiende y conceptualiza el mundo (De Bustos, 2004).

2. La pertenencia a una clase social determinada conlleva que la estructura familiar, la división de roles dentro de ella y su sistema de comunicación dé lugar a la existencia de una serie de códigos de los que se apropiará el niño en su socialización. (Bernstein, 1973).

3. La lingüística rusa y parte de la americana defienden que el lenguaje no es más que un reflejo de la realidad social (Labov, Cohen, Robins, y Lewis, 1968).

Tanto los datos expresos (lo que el autor dice) como los latentes (lo que dice sin pretenderlo) cobran sentido y pueden ser captados dentro de un contexto. El contexto es, por tanto, un marco de referencias que contiene toda aquella información que el lector puede conocer de antemano, o inferir a partir del texto mismo, para captar el contenido y el significado de todo lo que se dice en el texto. Texto y contexto se presentan como dos aspectos en continua interdependencia.

Desde una perspectiva psicosocial y dentro de la tradición crítica del “análisis del discurso”, cabría destacar la definición de Potter y Wetherell (1987), basada en la teoría lingüística de la escuela de Oxford, según la cual el discurso es “*un conjunto de prácticas lingüísticas que mantienen y promueven ciertas relaciones sociales*” regulando así la acción social (Iñiguez y Antaki, 1998). En el proceso de socialización y mediante el aprendizaje social se genera, a través del lenguaje, la adquisición de modelos normativos, valorativos y comportamentales en el grupo familiar. Aprendizaje de los roles sociales necesarios para la participación en la vida social conforme a las expectativas de los otros individuos (Blanco, 1980; 1981). La transmisión lingüística está en estrecha relación con las características asociadas al emisor, con el contenido del mensaje, con el contexto ecológico-situacional y con las características del receptor. Realidades sociales diferentes conllevan diferentes formas de ser

transmitidas y una cierta visión del mundo (mapas sociocognitivos, representaciones sociales: Farr y Moscovici, 1984) desde el punto de vista normativo, de los valores y de los modelos de comportamiento.

Van Dijk (1977) hace notar que en todos los niveles del discurso podemos encontrar las “*huellas del contexto*” en las que, junto a las características personales, las características sociales de los participantes juegan un rol fundamental (por ejemplo, “género”, “clase”, “etnicidad”, “edad”, “origen” y “posición” u otras formas de pertenencia grupal). A través del lenguaje y el control del contexto con el uso de tópicos, estilos y turnos de palabra en las situaciones comunicativas, se reproducen y se refuerzan subrepticamente sistemas de dominación y discriminación (como el sexismo) que prevalecen en un grupo social dado (Giles, 1979; Turner y Giles, 1981; Billig, 1985). Grupos raciales o étnicos desarrollan habitualmente formas específicas de hablar que dan origen a una influencia y adaptación mutua y que funcionan como refuerzo de su identidad social. Como efecto colateral, es casi inevitable la presencia de problemas de comunicación y de comprensión intercultural entre esos grupos, por los mecanismos de favoritismo endogrupal y discriminación exogrupal puestos de relieve por los estudios experimentales de H. Tajfel (Tajfel, 1982; Tajfel, y Turner, 1986). Por ello las relaciones interculturales o interétnicas suelen adoptar formas de “dominancia” que refuerzan, a través del uso “prejuiciado o discriminatorio” del lenguaje, el etnocentrismo y el racismo subyacentes (Giles y St. Clair, 1979; Stubbs, 1987; Bernete, 1992).

De este modo, los comportamientos lingüísticos habituales pueden incorporar grandes diferencias en el modo expresivo, dependiendo de los valores, normas, relaciones sociales, o instituciones que los inspiran y de la forma que adoptan o tienen determinados contextos socioculturales. En el contacto intercultural las diferencias discursivas pueden aceptarse de un modo tolerante y cooperativo o dar origen a incomprendiones o conflictos traducidos en formas de dominación, exclusión u opresión (Stubbs, 1993) hacia los de menor poder. El estudio y análisis de la comunicación intra e intercultural es sin duda un campo de investigación importante en el Análisis contextual multidisciplinario. Se trata de pasar de una interpretación del mero significado lingüístico y sintáctico de las distintas combinaciones “discursivas”, cambios de tópicos, turnos conversacionales o coherencia textual a la puesta en evidencia de los substratos psicosociales más profundos que vehicula. De ahí que, para explicar lo que representa el lenguaje, resulta insuficiente analizar sus estructuras internas, las acciones que se logran, o los procesos cognitivos que ocurren en su utilización. Es necesario dar cuenta del discurso como acción social que ocurre en un marco de comprensión, comunicación e interacción el cual, a su vez, forma parte de estructuras y procesos socioculturales más amplios (Van Dijk, 1977).

Lógicamente este tipo de análisis requiere de estudios más integrados en todos los niveles y disciplinas, incluyendo evidentemente el propiamente lingüístico. A menudo las relaciones entre el texto y las estructuras sociales resultan ser indirectas y difíciles de detectar a simple vista. Hay cosas de las que se habla sin hacer mención explícita de ellas, algunas se entienden no sólo por lo que se dice, sino por lo que se omite, y las connotaciones y significados

siguen unas leyes que no siempre coinciden con las meramente gramaticales. Los hablantes hacen referencia a una serie de elementos (histórico-biográficos, deseos, expectativas, etc.) que carecen de sentido sin unas mínimas pautas de referencia (Garfinkel, 1967, 1972). Por ejemplo, en el uso de los pronombres, antes de relacionarlos con diferencias de poder entre los miembros de un grupo o con estructuras y rutinas institucionales, deben contrastarse con la propia estructura sintáctica de la oración, su significado, los esquemas cognitivos y modelos mentales que lo informan, su función en la interacción, la identidad de los hablantes y sus creencias compartidas.

Se trata de desarrollar instrumentos, fundamentados teóricamente, que permitan apoyar empíricamente una mejor comprensión crítica de la “desigualdad social” implícita en contextos originados por razones de origen, género, clase, religión, lengua, u otro criterio que pueda definir las diferencias humanas. Es lo que Van Dijk denomina “Análisis Crítico del Discurso” desde una posición social y política explícita, comprometida con el cambio desde la responsabilidad académica. El lenguaje que se utiliza es una parte inherente de la construcción social y está implicado, tanto en todas las “injusticias” sociales que se cometen, como en su lucha contra ellas.

El esfuerzo y las aportaciones técnicas del programa ConText se inspiran en el deseo de ayudar a los investigadores, a partir de análisis recursivos de un determinado corpus textual, a detectar ‘las huellas’ de ese conjunto de posibles interpretaciones coherentes, las cuales se inscriben en la enorme complejidad de los fenómenos que operan en el lenguaje.

REFERENCIAS

- Bardin, L. (1977). *El análisis de contenido*. Madrid: Ediciones Akal, 1986.
- Behar, J. (1991). Observación y análisis de la producción verbal de la conducta. En M.T. Anguera (Ed.), *Metodología observacional en la Investigación Psicológica*, vol. 1, 330-389). Barcelona: PPU.
- Behar, J. (1993). Aproximación al análisis textual informatizado. *Anuario de Psicología*, 59, 61-78.
- Benzécri, J.P. (1973). *L'Analyse des données. Tome 1: La taxinomie. Tome 2: L'analyse des correspondances*. Paris: Dunod (2ª ed., 1976).
- Benzécri, J.P. (1981). *Pratique de l'analyse des données. Tome 3, Linguistique et lexicologie*. Paris: Dunod.
- Berbel, S. (1997). *Relaciones de influencia entre grupos de mujeres: un análisis evolutivo de los discursos*. Tesis doctoral no publicada. Universidad de Barcelona. Extraíble de http://www.tesisexarxa.net/TDX-0421108-113718/index_cs.html
- Berelson, B. (1952). *Content Analysis in communication research*. New York: Free Press (Hafner Publ. Co., 1971).
- Berelson, B. & Lazarsfeld, P.F. (1948). *The analysis of communications content*. Chicago (ILL) and New York (NJ): University of Chicago Press and Columbia University Press.
- Bernete, F. (1992). El estudio de los estereotipos a través del análisis de relatos. *REIS*, 57/92, 123-135.
- Bernstein, B. (1973). A Sociolinguistic Approach to Socialization with some reference to educability. En F. Williams (Ed.), *Language and Poverty*. Chicago: Aldine.
- Billig, M. (1985). Prejudice, categorization and particularization: From a perceptual to a rhetorical approach. *European Journal of Social Psychology*, 15, 79-103.
- Blanco, A. (1980). Psicología Social del Lenguaje I: Presupuestos básicos. *Estudios de Psicología*, 3, 109-119.
- Blanco, A. (1981). Psicología Social del Lenguaje II: Socialización, Interacción y Grupalidad. *Estudios de Psicología*, 5/6, 185-200.
- Brown, R. (1994). *Words and things*. New York: The Free Press.
- Bunge, M (1989). *La investigación científica*. Barcelona: Ariel.
- Bécue, M., Lebart, L. & Rajadell, N. (1992). El análisis estadístico de datos textuales. La lectura según los escolares de enseñanza primaria. *Anuario de Psicología*, 55, 7-22.

- Cicourel, A. V. (1973). *Cognitive sociology: language and meaning in social interaction*. New York: Free Press.
- Cornejo, J.M. (1988). *Técnicas de investigación social: el análisis de correspondencias (teoría y práctica)*. Barcelona: PPU.
- Cornejo, J.M. (2006). El análisis de las interacciones grupales: las aplicaciones SOCIOS. *Anuario de Psicología*, 37(3), 277-297.
- De Bustos Guadaño, E. (2004). *Lenguaje, comunicación y cognición: temas básicos*. Madrid: UNED.
- Farr, R.M. & Moscovici, S. (Eds.) (1984). *Social representations*. Cambridge: Cambridge University Press.
- Freeman L.C. (1965). *Elementary Applied Statistics For Students in Behavioral Science*. New York: John Wiley and Sons. (traducción española en Madrid: Euramérica, 1971).
- Giles, H. & St. Clair, R. N. (Eds.) (1979). *Language and Social Psychology*. Londres: Basil Blackwell.
- Giles, H. & Coupland, N. (1991). *Language: Context and consequences*. Milton Keynes: Open University Press.
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs (NJ): Prentice Hall.
- Garfinkel, H. (1972). Studies of the Routine Grounds of Everyday Activities. En D. Sudnow (Ed.), *Studies in Social Interaction*. New York: The Free Press.
- Greene, J. (1989). *Memory, Thinking, and Language*. Londres: Menth.
- Guttman, L. (1968). A general non-metric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 33, 469-506.
- Harris, Z. (1954). Distributional Structure. *Word*, 10 (23), 146-162.
- Holsti, O.R. (1969). *Content analysis for the social sciences and humanities*. Reading (MA): Addison Wesley Publishing Company.
- Ibáñez, J. (1985). Análisis sociológico de textos o discursos. *Revista Internacional de Sociología*, 43 (1), 119-160.
- Íñiguez, L. & Antaki, C. (1998). Análisis del discurso. *Anthropos*, 177, 59-66.
- Krippendorff, K. (1990). *Metodología del análisis de contenido. Teoría y Práctica*. Barcelona: Paidós.
- Labov, W., Cohen, P., Robins, C. & Lewis, J. (1968). *A Study of Non-standard English of Negro and Puerto Rican Speakers in New York City*. Washington, D.C.: US Office of Education and Welfare.
- Lafon, P. & Salem, A. (1983). L' Inventaire des Segments Répétés d'un Texte. *Mots*, 6, 161-177.
- Lazarsfeld, P.F., Berelson, B. & Gaudet, H. (1962). *El pueblo elige. Estudio del proceso de formación del voto durante una campaña presidencial*. Buenos Aires: Ediciones 3 (Original, 1944).
- Lebart, L. (1982). L'analyse statistique des réponses libres dans les enquêtes socio-économiques. *Consommation*, 1, 39-62.
- Lebart, L. & Salem, A. (1988). *Analyse statistique des données textuelles. Questions ouvertes et Lexicométrie*. Paris: Dunod.
- Lebart, L., Morineau, A., Bécue, M. & Haeusler, L. (1992). *SPAD.T. Système portable pour l'analyses des données textuelles*. Paris: CISIA.
- Moscovici, S. (1967). Communication process and the properties of language. *Advances in Experimental Social Psychology*, 3, 225-270.
- Moscovici, S. (1972). *The Psychosociology of Language*. Chicago: Markham Publishing Co.
- Osgood, C.E. (1959). The representation model and relevant research methods. En I. de Sola Pool (Comp.), *Trends in Content Analysis*, (pp. 33-88). Urbana (ILL): University of Illinois Press.
- Pinto, R. & Grawitz, M. (1967). Analyse de contenu et theorie. En *Méthodes des sciences sociales*, (pp. 456-499). Paris: Dalloz.
- Parsons, T. (1978). *Action theory and the human condition*. New York: Free Press.
- Potter, J. & Wetherell, M. (1987). *Discourse and Social Psychology*. Londres: Sage.
- Pêcheux, M. (1978). *Hacia el análisis automático del discurso*. Madrid: Gredos.
- Saussure, F. (1980). *Curso de Lingüística General*. Madrid: Akal.
- Saville-Troike, M. (1982). *The Ethnography of Communication: An Introduction*. Oxford: Basil Blackwell.
- Stubbs, M. (1987). *Analysis del discurso*. Madrid: Alianza.
- Stubbs, M. (1993). *Discourse analysis: A sociolinguistic analysis natural language*. Oxford: Basil Blackwell.
- Tajfel, H. (1981). *Human groups and social categories*. Cambridge (UK): Cambridge University Press
- Tajfel, H. (1982). *Social identity and intergroup relations*. Cambridge (UK): Cambridge University Press.
- Tajfel, H. & Turner, J. C. (1986). The social identity theory of intergroup behavior. En S. Worchel & W. G. Austin (Eds.), *The Psychology of Intergroup Relations*, (pp. 94-109). Monterey (CA): Brooks-Cole.
- Turner, J.C. & Giles, H. (Eds.) (1981). *Intergroup Behaviour*. Oxford: Blackwell.
- Van Dijk, T. A. (1980). *Texto y contexto. Semántica y Pragmática del discurso*. Madrid: Cátedra.
- Van Dijk, T. A. (1984). *Prejudice in discourse: An analysis of ethnic prejudice in cognition and conversation*. Amsterdam/Philadelphia: J. Benjamins.
- Van Dijk, T. A. (1997). *Racismo y análisis crítico de los medios*. Barcelona: Paidós.
- Van Dijk, T. A. (1999). El análisis crítico del discurso. *Anthropos*, 186, 23-36.
- Zipf, G. K. (1935). *The psychobiology of language, an introduction to dynamic philology*. Boston: Houghton-Mifflin.
- Zipf, G.K. (1949). *Human behaviour and the principle of least effort*. Cambridge (MA): Addison-Wesley.