

Aproximación al análisis textual informatizado

Júlia Behar
Universidad de Barcelona

Este artículo pretende acercar al lector conocedor del análisis textual clásico, como el análisis de contenido, a alguna de las técnicas más recientes de análisis textual mediante programas informáticos. Para cada una de ellas se facilitan unas mínimas bases teóricas, información sobre el tipo de material verbal adecuado para ser analizado, el tipo de cuantificación del texto utilizado por el programa, los índices calculados y finalmente la utilidad y aplicación de dichos índices. En todos los casos, y para facilitar la comprensión, se ofrecen datos sobre investigaciones en las que se ha aplicado dicha técnica. Se hace referencia a la lexicometría, la lexicometría cronológica, al análisis factorial de correspondencias de Benzecri (1974), al análisis de correspondencias de protocolos verbales, y al sistema Alceste (Análisis léxico por contexto de un conjunto de segmentos de texto).

Palabras clave: Análisis textual, análisis textual informatizado, forma gráfica, unidades de contexto, variables intratextuales, variables extratextuales, variables categóricas, codificación.

The aim of this article is to inform the reader who is familiar with classical textual analysis, such as content analysis, of the most recent techniques of such analysis by means of computer programs. For each one of these techniques the following is provided: theoretical bases, information on the type of verbal material that is adequate for analysis, the way in which the program quantifies the text, the indexes that are calculated and lastly the usage and application of these indexes. In all cases, and in order to make understanding easier, information is included as investigations in which such methods have been put to use. Reference is made to lexicometry, chronological lexicometry, Benzecri's factor analysis of correspondence, the analysis of verbal reports and to the Alceste system («Analyse Lexical par Contexte d'un Ensemble de Segments de Texte»).

Key words: Computerized Textual Analysis, Graphic Units of Context, Intratextual Variables, Extratextual Variables, Categorical Variables, Encoding.

Importancia del material verbal en la investigación psicológica

En un número monográfico sobre discurso y análisis textual quizás no sea necesario resaltar la importancia del estudio de la producción verbal de la conducta, sea ésta hablada o escrita, en la investigación psicológica, puesto que su propia existencia es una prueba del interés acordado a este aspecto de la conducta. Sin embargo el número relativamente escaso de trabajos de investigación con este tipo de datos hace aconsejable insistir sobre su relevancia.

En el ámbito profesional la situación es distinta. Entrevistas clínicas, de selección, psicoterapia, etc. se fundamentan en el intercambio verbal entre psicólogo y sujeto, donde la palabra es soberana. Sin embargo para el estudio del material verbal resultante rara vez se procede a su sistematización y análisis riguroso. Lo habitual es que el psicólogo, de acuerdo a su opción teórica, deduzca, extraiga conclusiones, interprete, se sirva, en definitiva, de su subjetividad para dar un sentido al material verbal recogido.

Sin restar importancia a esta línea de trabajo es conveniente destacar la posibilidad, por no decir la necesidad, de utilizar todo tipo de material verbal como datos idóneos para su estudio científico, en tanto que es posible obtener a partir de los mismos información objetiva por medio del uso de análisis del texto, como el análisis de contenido o la semántica analítica. Asimismo es preciso tener en cuenta que en muchas ocasiones la conducta verbal, oral o escrita, ha sido producida espontáneamente por el sujeto, como los diarios personales, correspondencia, autobiografías y cualquier otro tipo de documento personal, por lo que los datos a analizar no incorporan ningún tipo de reactividad (Behar, 1991).

En los últimos años han cobrado gran importancia los análisis textuales realizados por medio de programas informáticos diseñados especialmente para este tipo de datos, de los cuales se ofrece en este artículo una panorámica no exhaustiva.

Análisis textuales no informatizados

El análisis cuantitativo de textos se remonta al siglo XVIII. Autores como Krippendorff (1981) y Rosengren (1981) sitúan el inicio del análisis de contenido en Suecia en donde en dicha época se procedió al recuento de los símbolos religiosos presentes en una colección de 90 himnos. Otros autores, como Berelson (1971), consideran que el análisis de contenido nace en Estados Unidos, en 1787, cuando un periódico cuantificó los elementos básicos del perfecto discurso antifederalista, con trece menciones a la libertad de prensa, una a la libertad, una a la esclavitud y dieciocho a la aristocracia.

Sin embargo hasta principios del presente siglo no es posible referirse al análisis textual como procedimiento científico. Actualmente existen diversas técnicas de análisis textual que difieren entre sí tanto en los objetivos como en los

procedimientos para alcanzarlos, así como en el mayor o menor grado de interpretación del texto y por lo tanto de subjetividad. Si se considera un continuo objetividad/subjetividad, en el primer polo se situaría el análisis de contenido (Bardin, 1986; Berelson, 1971; Krippendorff, 1981; Muchielli, 1977) por su carácter estrictamente cuantitativo. En el análisis de contenido se parte del supuesto de que las palabras utilizadas más frecuentemente por el autor del texto son indicativas de los intereses del autor por lo que se evita cualquier tipo de interpretación o de inferencia respecto al texto. El polo de mayor subjetividad correspondería a la hermenéutica, caso especial de la fenomenología general (Gadamer, 1960; Hirschs, 1967). La hermenéutica pretende captar los significados ocultos más allá de la apariencia inmediata. Se basa en la idea de que las experiencias humanas, el lenguaje, no son directamente accesibles y que para abordarlos es preciso realizar una lectura interpretativa. En la zona intermedia del continuo se encontrarían técnicas como la semántica analítica (Lindkvist, 1981) que presupone que el texto tiene una estructura abierta que posibilita la creación de nuevas perspectivas. Se procede estableciendo relación entre diferentes categorías textuales, para presentarlas posteriormente de forma más sistemática que el propio texto.

Actualmente cuando se utilizan las técnicas más objetivas para realizar análisis textuales, éstos se llevan a cabo mediante programas informáticos que además de proceder al recuento de la frecuencia de cada una de las palabras del texto, están en condiciones de establecer relaciones entre variables y de contrastar hipótesis. Por el momento no existen programas informáticos que realicen las interpretaciones propias de las técnicas de análisis textual más subjetivas y éstas siguen exigiendo la intervención personal del investigador.

Análisis textuales informatizados

La incorporación de la informática al análisis textual es relativamente reciente y ha supuesto un enorme avance en cuanto a la posibilidad de analizar textos de gran tamaño como las autobiografías o los diarios personales. Existen diversos programas y si bien cada uno de ellos posee características propias y persigue objetivos diferenciados, es posible referirse a algunos aspectos comunes, como la objetividad en el tratamiento del material verbal, la posibilidad de tener en cuenta simultáneamente variables intratextuales y variables extratextuales, y la capacidad para establecer relaciones entre las variables estudiadas.

Sin embargo este tipo de análisis no es aceptado por algunos investigadores que trabajan con datos verbales puesto que consideran que sólo algunos de los procedimientos no informatizados y más interpretativos, como la hermenéutica o el análisis del discurso, pueden aportar información sobre lo realmente importante, lo que está oculto en el texto, más allá de la aproximación cuantitativa.

Autores como Korosec-Serfaty (1985, 1989) utilizan indistintamente, según las ocasiones, procedimientos cuantitativos, como el análisis de contenido temático o interpretativos como el método fenomenológico. La elección de uno u otro procedimiento, por esta autora o por otros autores, depende del deseo o no de

desvelar significados ocultos del texto para aproximarse al mundo experimental del sujeto, y de la dosis de subjetividad que esté dispuesto a incorporar a la comprensión del texto.

Lexicometría. Programa LEXICLOUD

La lexicometría supone un enfoque cuantitativo del estudio de textos que permite establecer relaciones estadísticas entre las unidades léxicas. Se presenta aquí el programa LEXICLOUD desarrollado por el Laboratorio de Lexicometría y Textos Políticos de la Escuela Normal Superior de Fontenay-St. Cloud que trabaja directamente sobre el texto original sin que el investigador realice previamente ningún tipo de codificación ni de interpretación. Únicamente es preciso seguir ciertas pautas al introducir los datos en ordenador para que el programa pueda segmentar el texto automáticamente y reconocer las distintas partes del corpus.

La unidad de medida utilizada para proceder a la cuantificación del texto es la forma gráfica. No se trata de un simple recuento de palabras sino que se tiene en cuenta la forma de las mismas y su entorno con el fin de conocer su comportamiento textual. Por medio de programas específicos es posible recuperar la dimensión sintagmática de cada forma del texto, aislando las secuencias en que se encuentra cada forma.

Es posible diferenciar los procedimientos documentales y los estadísticos. Los *documentales* realizan reorganizaciones de las unidades textuales y los procedimientos que para ello ofrece el programa son el *índice de frecuencias*, las *concordancias*, y los *segmentos repetidos*. El *índice de frecuencias* informa sobre el número de ocurrencias de una forma en un texto y la información puede presentarse jerárquica o alfabéticamente. Las *concordancias* se refieren a la localización del conjunto de las ocurrencias de una forma en un contexto (Tabla 1). Cada concordancia puede presentarse por orden alfabético de la primera palabra que sigue o precede a la forma, o por orden cronológico de aparición en el texto. En tanto que las concordancias suponen una reorganización de la secuencia textual proporcionan una visión más sintética que la resultante de una lectura secuencial. Los *segmentos repetidos* son aquellos con frecuencia igual o superior a dos, entendiéndose por segmento la sucesión de secuencias no interrumpidas por caracteres delimitadores de frecuencia (comas) o de frases (punto, punto y coma, etc.). La información sobre cuáles y cuántos son los segmentos repetidos en el texto permite conocer la composición cuantitativa del texto y aislar configuraciones sintagmáticas recurrentes. Es posible obtener el inventario tanto jerárquico como alfabético de los segmentos repetidos.

Los procedimientos *estadísticos* realizan cálculos de índices estadísticos y comparaciones entre los distintos textos de un corpus, para lo cual el programa ofrece el *análisis factorial de correspondencias* y el *método de las especificidades*. Al primero se dedica un apartado en este mismo artículo y su aplicación en un contexto lexicométrico tiene como objetivo comparar las distintas partes

1604125	a aproximar el catalán con el alemán d' integració	lingüística	de l' estat, la lògica que els catalans parlants resideixen
1603957	quant al català i la procedència de la paraula normalització	lingüística	de la constitució dels organismes comunitats autònomes
1604046	evident que el problema de la plena normalització	lingüística	de tots els parlants, el caràcter participatiu flexiós de
1604113	o fomentar la unitat de la llengua i la llibertat	lingüística	de Catalunya en port, perquè definitivament per a la llengua
1503964	la significativitat de cadra col·lectiva, la identitat	lingüística	de Catalunya, són, doncs, objectius d' aquesta línia: "
1604104	l' del català i del castellà, i atesa la situació	lingüística	de Catalunya, respecte a la part central del país, i les 11
1004405	arza es un exponent, s' la línia de normalització	lingüística	del país. " El segon fet afecta el comportament lingüístic
1503958	no privilegiada de contribuir a la normalització	lingüística	del país i, finalment, veiem manifestar: " i que a la vegada
1503697	clau, sense la qual es impossibilita la recuperació	lingüística	del país valencian, si llenguas com el castellà, francès
1604171	o és fonamental per a la necessària recuperació	lingüística	del territori, aquest és un principi que s' ha de dir e
1604160	casos en el mateix: salvaguardar la personalitat	lingüística	del Principat e' sempre dos aspectes: un de costat
1504252	els autòctons i la integració lliure de normalització	lingüística	dels ciutadans adscrits a professions de la difusió de
1604182	e d' a més restrictiva per la perfecta competència	lingüística	dels Països Catalans. " La línia catalana i la línia
1504242	immediatament parlat de la totalitat de l' àrea	lingüística	de Catalunya. " i, doncs, s' ha de veure que sempre la llengua cat
1504241	comunicar-se entre si dels parlants d' una àrea	lingüística	de difusió e' hauria de basar en les següents línies: " enc
1604240	a comunicar, però - i condicions que una política	lingüística	de comptes de treballar per a l' ajustament de la norma
1504274	est obligat a intervenir en qüestions de política	lingüística	de l' definitiu, després del caràcter d' acceptar l' ajustam
1604255	en l' àrea de normalització i la recuperació	lingüística	en el país, s' ha de garantir amb a sempre la llengua cat
1604256	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604257	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604258	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604259	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604260	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604261	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604262	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604263	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604264	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604265	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604266	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604267	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604268	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604269	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604270	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604271	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604272	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604273	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604274	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604275	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604276	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604277	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604278	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604279	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604280	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604281	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604282	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604283	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604284	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604285	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604286	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604287	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604288	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604289	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604290	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604291	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604292	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604293	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604294	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604295	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604296	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604297	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604298	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604299	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604300	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604301	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604302	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604303	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604304	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604305	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604306	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604307	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604308	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604309	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604310	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604311	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604312	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604313	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604314	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604315	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604316	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604317	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604318	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604319	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604320	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604321	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604322	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604323	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604324	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604325	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604326	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604327	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604328	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604329	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604330	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604331	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604332	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604333	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604334	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604335	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604336	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604337	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604338	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604339	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat
1604340	o bélsip " s' ataca a realmentar la recuperació	lingüística	de l' llengua catalana i a garantir amb a sempre la llengua cat

Tabla 1. Fragmento de la concordancia de la forma lingüística de un corpus de textos catalanes.

del corpus según el vocabulario utilizado y su frecuencia de utilización para establecer similitudes o diferencias entre distintos emisores o entre distintos textos de un mismo emisor; y para establecer qué vocabulario tiende a ser infra o sobrerutilizado en determinadas partes del texto. El método de las especificidades permite evaluar si una forma gráfica tiene una presencia significativamente más alta en determinadas partes del corpus. El modelo estadístico utilizado para obtener la distribución de probabilidad es la ley hipergeométrica que toma como parámetros el tamaño del corpus, el tamaño de la muestra o parte del corpus considerado, la frecuencia de una determinada forma gráfica en el corpus y la frecuencia de la forma gráfica en la parte del corpus considerada. Es decir, que se toma el corpus en su totalidad como baremo para cada una de las partes, sin

que exista ninguna norma de referencia exterior. De este modo, para una forma gráfica determinada se compara su frecuencia en la muestra con la moda de la distribución en el corpus. Si el valor de la frecuencia no presenta diferencia significativa con el de la moda se trata de una *forma general* distribuida uniformemente en el corpus; si por el contrario se detecta diferencia significativa entre ambos valores se trata de una *forma específica* característica de una parte del corpus.

El *método de las especificidades* se utiliza también para completar los análisis globales realizados tras la utilización del análisis factorial de correspondencias.

Por último cabe referirse a la *lexicometría cronológica* (Salem, 1988) que analiza los datos de un mismo emisor, individual o colectivo, recogidos a través del tiempo. El objetivo de este estudio es poner en evidencia las variaciones que se han producido a lo largo del tiempo. Sirviéndose de las especificidades cronológicas es posible asociar la utilización de un término determinado con un periodo de tiempo, y conocer así la evolución de su uso en los distintos periodos. Esta técnica se ha utilizado básicamente para el análisis de textos políticos, pero sus características la hacen muy adecuada para el estudio de procesos psicológicos de los que se disponga de datos textuales de los sujetos implicados, recogidos a lo largo del tiempo, como diarios personales, correspondencia, entrevistas grabadas, etc.

Análisis factorial de correspondencias. Modelo de Benzecri

Benzecri (1974) es el primer autor en aportar una técnica estadística, fuera del contexto de la estadística inferencial clásica, capaz de analizar tablas de contingencia. La propuesta de Benzecri permite profundizar en el conocimiento de las relaciones de dependencia entre dos variables categóricas observadas, al explicar cuáles son las relaciones existentes entre los distintos valores de ambas variables. Se trata de una técnica exploratoria, capaz de describir el grupo estudiado, especialmente adecuada para investigaciones inductivas en las que no se parte de una hipótesis de trabajo.

Cuando se desea aplicar la técnica al material verbal es preciso realizar previamente un análisis textual, tipo análisis de contenido, que sintetice la información en forma de frecuencias de variables categóricas. Ello supone que antes de aplicar el programa de ordenador el investigador debe elaborar un sistema de categorías, a partir del material verbal, lo que siempre comporta un cierto grado de subjetividad.

La técnica parte de una tabla de contingencia en la que aparecen en línea las diversas formas gráficas y en columna las diversas categorías de sujetos, por lo que cada casilla recoge el número de ocurrencias de una determinada forma gráfica en una categoría de sujetos. Posteriormente, y a partir de esta tabla, se procede a la construcción y análisis de las nubes de puntos generadas por dichas variables. El programa utiliza la nube de puntos definida en el espacio de dimen-

sión más pequeño y extrae los ejes factoriales. Dado que el origen de los ejes corresponde a los perfiles medios, en la representación gráfica los puntos muy próximos al origen de los ejes tienen perfiles próximos a los medios, en tanto que los puntos con posiciones más periféricas tienen perfiles más diferenciados del medio y son más característicos. También es posible interpretar la posición de un punto línea en relación al conjunto de puntos columna y viceversa, pero no es aconsejable interpretar la proximidad entre puntos línea (formas gráficas) y puntos columna (categorías de sujetos) puesto que no están situados en el mismo plano.

El programa facilita asimismo información de la importancia de cada punto en la definición de un eje —contribuciones absolutas— y de la calidad de la interpretación de un punto sobre un eje —contribuciones relativas—. Los valores propios, los porcentajes de variancia correspondientes a estos valores propios, y la traza o suma de todos los valores propios son parámetros obtenidos en el análisis de gran importancia para la interpretación. También es posible hallar subespacios de representación de las proximidades entre perfiles y situar en los mismos líneas o columnas suplementarias que recogen información sobre variables extratextuales o ilustrativas. Estas variables ilustrativas no participan en la construcción del plan factorial sino que se proyectan posteriormente sobre el plano. De este modo la interpretación se realiza en dos fases y diferenciadamente, antes y después de la proyección de las variables ilustrativas.

Esta técnica fue utilizada por Rodríguez-Tomé, Behar, Martínez y Bariaud (1982) en una investigación que estudiaba los componentes cognitivo-afectivos de la identidad en la adolescencia por medio del análisis de las respuestas de los adolescentes a la pregunta abierta *¿quién soy yo?* Las unidades de análisis —enunciados elementales— fueron codificadas, mediante un análisis de contenido semántico, en referencias identificatorias. Realizada dicha codificación se procedió al tratamiento estadístico de los datos por medio del análisis factorial de correspondencias de Benzecri, utilizando como unidades de análisis las referencias identificatorias obtenidas en el análisis de contenido y mencionadas por lo menos por ocho de los sujetos de la muestra. Se introdujeron asimismo en el análisis elementos extratextuales; variables ilustrativas como edad y sexo, que sin intervenir en la etapa de extracción de los factores, se proyectaron posteriormente sobre los ejes. Los cinco factores obtenidos permitieron establecer unas dimensiones evolutivas en la adolescencia y asimismo poner en evidencia la existencia de algunos elementos culturales en la estructura de la identidad.

Como puede verse en la Figura 1 el primer factor diferencia a los sujetos de la muestra en función de la edad, oponiendo claramente los 12 y los 16 años. El segundo factor por su parte está asociado al sexo contraponiendo chicos y chicas.

Esta investigación y otras que en su momento se sirvieron de esta técnica estadística (Rodríguez-Tomé y Zlotowicz, 1972; Zlotowicz, 1972) ilustran la utilidad de la misma en el tratamiento de los datos verbales. Sin embargo el constante trabajo en el estudio de técnicas estadísticas a utilizar en el análisis de datos textuales ha permitido su optimización, tal como se describe en el siguiente apartado.

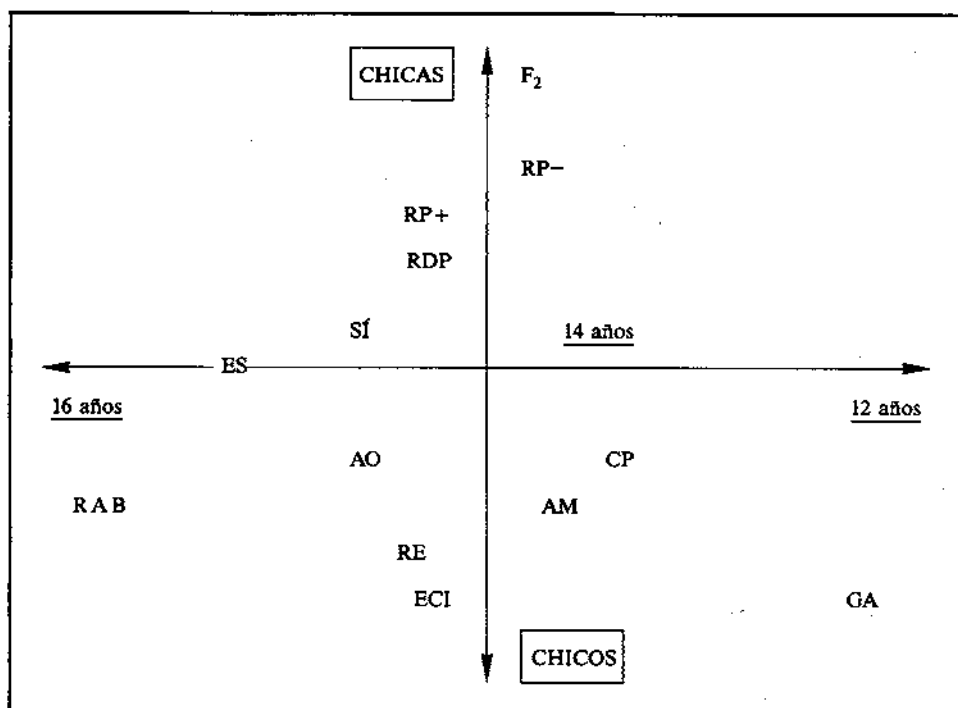


Figura 1. Espacio factorial de los dos primeros factores, en los que se han proyectado las variables edad y sexo. (Rodríguez-Tomé y otros, 1982, p. 523.)

Análisis de correspondencias. Análisis de protocolos verbales

Como se acaba de indicar, Benzecri (1974) realizó una aportación fundamental para el tratamiento estadístico de datos procedentes de la categorización de material verbal, al posibilitar mediante el análisis de correspondencias el conocimiento de las agrupaciones de categorías existentes y también las relaciones de las mismas con las variables ilustrativas. Sin embargo seguía siendo necesario proceder a la categorización de los protocolos verbales antes de realizar el análisis estadístico.

Lebart (1982), Lebart, Morineau y Warwick (1984), y Lebart y Salem (1988) amplían las posibilidades de aplicación del análisis de correspondencias al proponer una modalidad del mismo, SPAD.T, que permite el tratamiento directo, sin categorización previa, de los protocolos verbales procedentes de todo tipo de documentos: documentos personales, entrevistas, preguntas abiertas, procedimientos de asociación, auto-observación, etc. Los datos verbales se recogen en soporte informático y el análisis estadístico se lleva a cabo directamente sobre el texto, lo que permite considerar simultáneamente todo el material verbal disponible y

asimismo eliminar la subjetividad que en alguna medida se produce durante el proceso de categorización.

El programa SPAD.T parte de las tablas léxicas de base R y T y de las tablas léxicas agregadas Z y C. La tabla R tiene tantas líneas como individuos y las líneas contienen las respuestas literales de los individuos. El número de columnas es igual a la longitud de la respuesta más larga por lo que la tabla no es rectangular. La tabla T tiene también tantas líneas como individuos pero aquí el número de columnas coincide con el número de palabras del léxico, es decir con el número de formas gráficas utilizadas por el conjunto de los individuos. En cada casilla aparece por lo tanto el número de veces en que una forma gráfica ha sido utilizada por el conjunto de los individuos. En esta tabla se pierde la información sobre el orden de las formas gráficas en la respuesta. La tabla agregada Z describe las respuestas de los sujetos a una pregunta cerrada con varias alternativas de respuesta mutuamente excluyentes. La tabla C se obtiene del producto matricial $C=T'Z$ y permite comparar los perfiles léxicos de diferentes categorías de población.

El programa, al proceder a la decodificación automática para elaborar las tablas, no considera ni la información semántica ni la sintáctica y por lo tanto no se tienen en cuenta ni la sinonimia ni la homonimia. En cuanto al singular y al plural y a las distintas flexiones de un verbo pueden aparecer en el plan factorial próximos o separados. En este último caso es adecuado recontextualizar para comprender mejor el origen de las diferencias. Por su parte la proximidad supone una validación suplementaria de los resultados. Se conservan también las palabras herramienta (de, a, que...) pero sólo aparecen en el plan factorial si están irregularmente distribuidas en el texto, es decir si caracterizan diferentes partes del mismo. Desde la perspectiva del estudio de las frecuencias, si bien inicialmente no se considera el orden de las formas gráficas, la búsqueda de los *segmentos repetidos* y de las *respuestas modales* comporta la obtención de unidades de mayor riqueza semántica.

A partir de las distintas tablas el programa permite:

—Agrupar las respuestas en clases correspondientes a las diversas modalidades de una o varias variables nominales extratextuales lo que facilita la diferenciación de los grupos y su caracterización en función de las formas léxicas propias de cada uno de ellos.

—Situación las respuestas sobre planos factoriales y considerar entonces las respuestas como identificadores de cada uno de los sujetos. Este posicionamiento da lugar a la agrupación de palabras y por lo tanto a la asociación de temas en determinadas zonas.

—Que el investigador no deba realizar ninguna codificación previa dado que en una primera fase el programa SPAD.T identifica, cuenta y clasifica las formas léxicas. Posteriormente calcula la tabla de contingencia que cruza las clases y las formas léxicas y lleva a cabo el análisis de correspondencias.

Para ilustrar esta técnica se facilitan unos ejemplos. En una investigación, con niñas, sobre la relación entre el estado puberal y la estructura de la identidad (Behar, 1988) se utiliza el programa SPAD.T para analizar las respuestas a la pregunta abierta ¿quién soy yo? y las asociaciones a la palabra «regla». Los factores

obtenidos permiten constatar la diferencia, tanto en la estructura de la identidad como en la percepción de la regla, entre las niñas que ya han alcanzado la menarquia y las que aún no han tenido la primera regla.

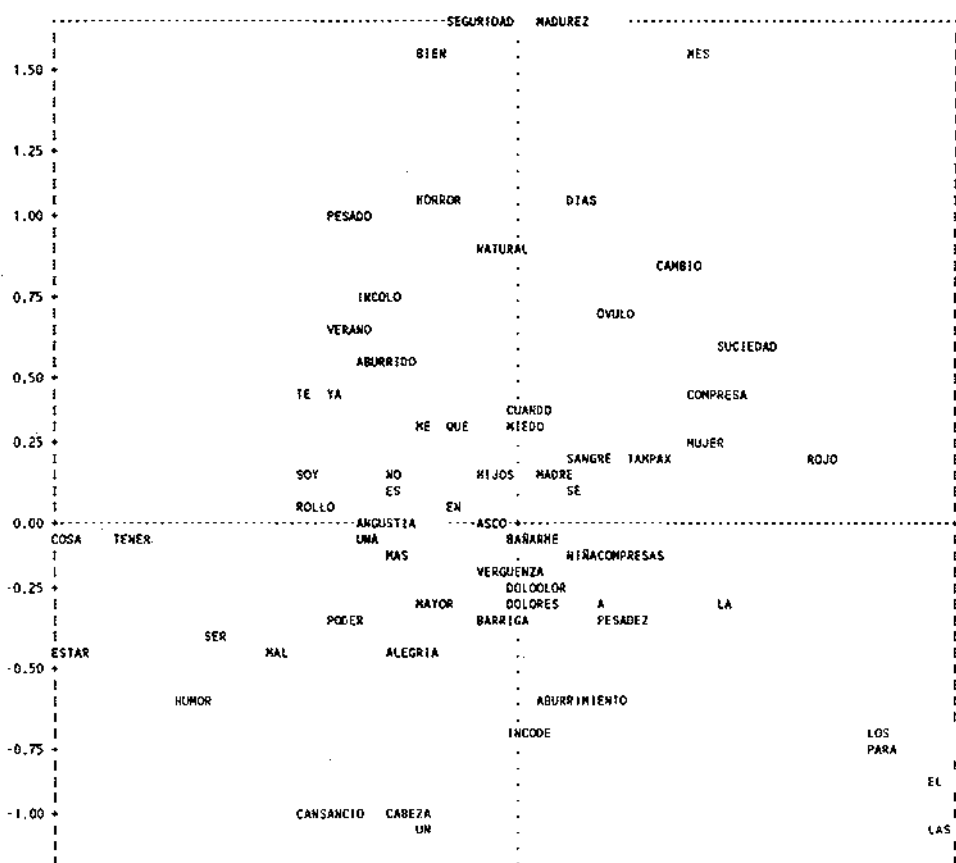


Figura 2. Espacio factorial de los dos primeros factores.

La Figura 2 permite visualizar cómo tras proceder al análisis de correspondencias mediante el SPAD.T, en la representación sobre los ejes, se dispone directamente de las palabras del texto y de las agrupaciones de las mismas, y no del número o de los símbolos adjudicados a las distintas categorías procedentes de una codificación previa, como ocurría en el análisis de Benzecri. Con palabras como «asco, vergüenza, dolor, pesadez», el polo positivo del primer factor puede interpretarse como una valoración negativa, física y psíquica de la regla. Para interpretar los factores es preciso tener en cuenta las coordenadas que indican, por medio del signo, a qué polo del factor se adscribe cada palabra, y el valor de las contribuciones absolutas de la palabra al factor (Tabla 2).

TABLA 2. VALORES DE LAS COORDENADAS Y DE LAS CONTRIBUCIONES ABSOLUTAS PARA CADA UNA DE LAS PALABRAS DEL TEXTO

<i>Coordenadas y contribuciones de las columnas</i>								
<i>Nombres</i>	<i>Coordenadas</i>				<i>Contribuciones absolutas</i>			
	F1	F2	F3	F4	F1	F2	F3	F4
a	-0.61	-0.10	0.38	-0.29	1.3	0.0	0.7	0.4
aburrido	0.16	-0.90	-0.56	0.55	0.0	1.5	0.6	0.6
alegría	0.54	-0.81	-0.42	-0.46	0.6	1.5	0.4	0.6
angustia	0.40	-0.80	-0.50	0.01	0.3	1.5	0.7	0.0
asco	0.52	-0.42	-0.09	0.01	2.2	1.7	0.1	0.0
barriga	0.15	-0.29	-0.13	-0.32	0.1	0.5	0.1	0.7
bañarme	0.61	0.36	0.04	-0.09	0.4	0.2	0.0	0.0
bien	-1.34	-0.50	-0.26	1.56	2.0	0.3	0.1	3.7
cabeza	0.50	-0.93	-0.43	-1.00	0.3	1.4	0.3	1.8
cambio	0.37	0.31	0.55	0.85	0.5	0.4	1.5	3.7
cansancio	-0.38	-0.81	-0.73	-1.00	0.2	0.9	0.8	1.6
compresa	0.48	0.36	0.68	0.43	1.6	1.1	4.3	1.9
compresas	0.70	-0.08	0.41	-0.13	1.3	0.0	0.6	0.1
cosa	0.28	2.42	-1.80	-0.04	0.1	10.7	6.5	0.0
cuando	-1.69	-0.67	-0.02	0.37	6.9	1.3	0.0	0.5
días	0.48	0.02	0.28	1.08	0.3	0.0	0.2	10.5
dolor	0.38	-0.25	0.02	-0.25	1.7	0.8	0.0	1.0
dolores	1.17	-1.30	0.08	-0.27	2.7	3.9	0.0	0.2
hijos	0.45	0.91	-0.13	0.16	0.5	2.4	0.0	0.1
horror	0.06	-0.13	-0.27	1.06	0.0	0.0	0.1	1.7
humor	-0.13	-1.62	-1.26	-0.62	0.0	3.4	2.3	3.6
incómodo	-0.93	-0.76	-0.58	0.74	1.7	1.4	0.9	1.5

Por su parte Cornejo (1988) utiliza el SPAD.T en un estudio sobre el diferencial de campos semánticos para conocer la modificación de actitudes a partir de la aplicación de técnicas creativas de grupo. El autor, tras la aplicación del programa aporta información sobre las dimensiones generales de las asociaciones verbales, lo que supone una primera aproximación al espacio semántico generado. En la Figura 3 puede verse cómo el primer eje, en uno de los polos, viene determinado por palabras como «inalcanzable, irreal, deseo, cielo, utopía y mañana». Este polo correspondería a la imposibilidad de alcanzar el deseo. En cambio en el polo opuesto del mismo eje destaca la presencia de verbos activos y positivos, como vivir, saber y reír.

Asimismo y desde una perspectiva intergrupar, el autor procede a caracterizar a cada grupo por medio de un conjunto de palabras características (Figura 4).

Aun cuando las clasificaciones obtenidas en los estudios presentados son muy útiles para una primera aproximación a los textos, el carácter exploratorio del análisis de correspondencias, y también el tipo de datos, pueden llevar a cuestionar la validez de los resultados obtenidos. Lebart (1992) aporta una posible respuesta a esta cuestión al proponer la aplicación del análisis discriminante tras la utilización del análisis factorial de correspondencias con objeto de validar las clasificaciones obtenidas. La matriz de confusión resultante proporciona el estadístico con la medida intrínseca de asociación entre la variable categórica y el contexto de variables, sea cual sea el número, el tipo de dichas variables, y el tamaño de la muestra.

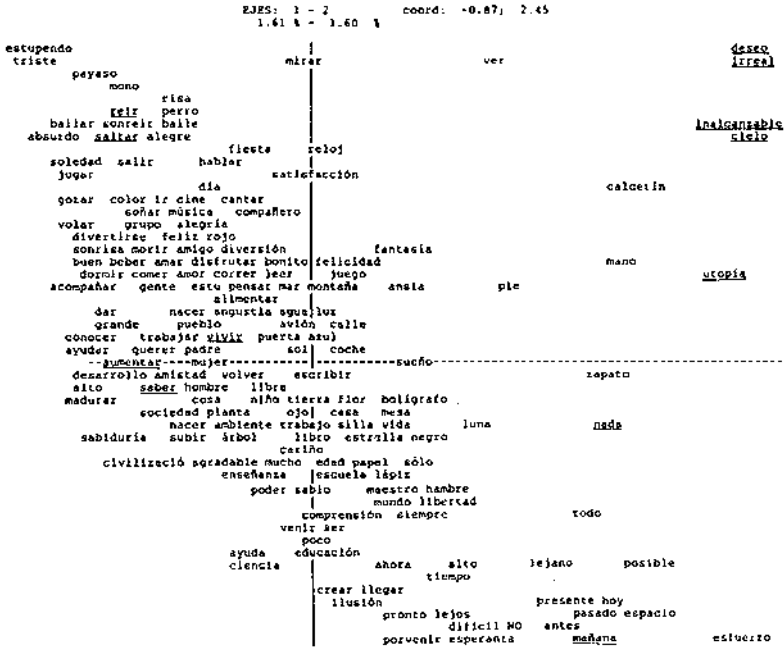


Figura 3. Espacio factorial de los dos primeros factores. Plano de proyección general inter-respuestas (Cornejo, 1988, p. 271).

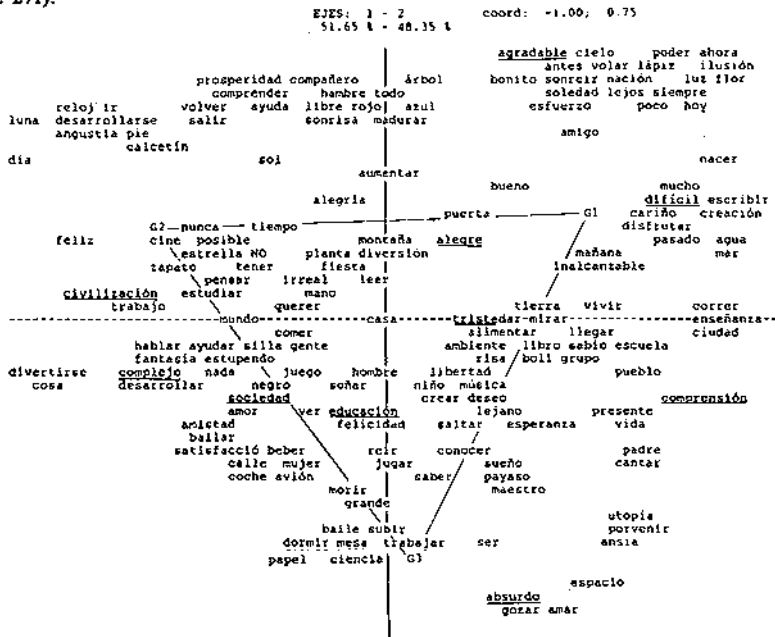


Figura 4. Plano de proyección intergrupos (Cornejo, 1988, p. 273).

Sistema ALCESTE (Análisis léxico por contexto de un conjunto de segmentos de texto)

El sistema ALCESTE es un programa desarrollado por Reinert (1990) con objeto de conocer la estructura semántica de un texto, diferenciando en el mismo los distintos campos conceptuales. No se trata de comparar grupos, como se ha descrito en los dos tipos de análisis de correspondencias presentados, sino de profundizar en el conocimiento de las relaciones existentes entre los segmentos de un texto.

Desde esta perspectiva el corpus se segmenta en unidades con sentido, frase, párrafo, etc. y cada una de ellas recibe el nombre de *unidad de contexto*. El objetivo, a partir de esta segmentación, consiste en hallar clases de *unidades de contexto* en las que se utilicen específicamente cierto tipo de palabras. A partir de aquí se considera que estos conjuntos de palabras, especialmente en razón de la redundancia, guardan relación con representaciones subyacentes que será posible explicitar.

Reinert diferencia dos aspectos de la noción de contexto, una en relación con lo que rodea a una palabra en un texto, operacionalizada con la noción de *unidad de contexto*; otra en relación con el uso, que se puede operacionalizar en un corpus determinado a través de la búsqueda de regularidades en la distribución del vocabulario en estas unidades de contexto, lo que supone una representación colectiva.

Para modelizar las leyes de distribución del vocabulario en un corpus se utiliza una tabla de doble entrada con las unidades de contexto en las líneas y el vocabulario generado por el mismo documento en las columnas. Las casillas con notación uno corresponden a palabra presente y las de notación cero [0] a palabra ausente. Este procedimiento supone la pérdida de la sintaxis de las frases pero ésta es relativamente pequeña dada la dimensión reducida, alrededor de unas 20 palabras, de las unidades de contexto. En los criterios para la segmentación del corpus se tienen en cuenta el tamaño de las unidades de contexto y la puntuación, y de este modo el final de una unidad de contexto corresponde al final de una frase, y su longitud es como máximo de algunas líneas. Considerando estos criterios se procede a una segmentación arbitraria, susceptible de modificación, con objeto de diferenciar los elementos estables de aquellos que no lo son en los resultados obtenidos después de cada modificación.

Los nexos a establecer entre las unidades de contexto deben permitir captar, tras el análisis estadístico, los *contextos tipo*, es decir las regularidades en la distribución del vocabulario en las unidades de contexto. Estos nexos se tienen en cuenta desde una perspectiva cognitiva. El enunciado es la marca de una representación mental individual y el contexto tipo la marca de una representación colectiva. La marca en la lengua de estas representaciones sólo puede conocerse a partir de un gran número de enunciados que desde cierto punto de vista sean parecidos. Por esta razón sólo se tienen en cuenta las palabras con contenido semántico, despreciando todos los rasgos morfosintácticos como artículos, preposiciones, conjunciones y pronombres. Es decir que en el análisis sólo se consideran los nexos entre unidades de contexto; los verbos, nombres, adjetivos y adverbios;

y también la distribución de las formas reducidas una vez suprimidas las desinencias gramaticales o ciertos sufijos. Se trata de descubrir de manera formal las leyes de distribución del vocabulario en el texto y a partir de las mismas los tipos de representaciones.

TABLA 3. UNIDADES DE CONTEXTO MÁS REPRESENTATIVAS DE LAS CLASES 1 Y 2
(Reinert, 1990, p. 43)

<i>Clase número 1</i>	
103 24	je chantais en marchant un hymne mysterieux dont je croyais me souvenir comme l'ayant entendu dans quelque autre existence,
870 18	je continuai ma route» et j'arrivai aux galeries du palais Royal.
876 18	de la, je sortis des galeries et je me dirigeai vers la rue saint-Honore
524 14	je me mis a parler avec violence, expliquant mes griefs et invoquant le secours de ceux qui me connaissaient.
847 14	j'allai ensuite visiter les galeries d'osteologie.
888 14	des medecins vinrent alors, et je continuai mes discours sur l'impuissance de leur art.
1015 14	une nuit, je parlais et chantais dans une sorte-d'extase.
470 13	les personnes les plus cheres qui venaient me voir et me consoler me paraissaient en proie a l'incertitude,
798 13	on termina ensuite la priere, et le pretre fit un discours qui me semblait faire allusion a moi seul.
987 13	on crie, on chante, on rit aux eclats; c'est gai ou triste a entendre, selon les heures et selon les impressions.
1129 13	tout-a-coup, o merveille! je me mis a songer a cette auguste soeur de l'empereur de russie, dont j'ai vue le palais imperial a weimar.
101 11	mon ami m'avait quitte, voyant ses efforts inutiles, et me croyant sans-doute en proie a quelque idee fixe que la marche calmerait.
658 11	je me trouvais dans une salle inconnue et je causais avec quelqu un du monde exterieur, l'ami dont je viens de parler, peut-etre.
83 10	«et pendant qu'il m'accompagnait, je me mis a chercher dans le ciel une etoile, que je croyais connaitre,
85 10	l'ayant trouvee, je continuai ma marche en suivant les rues dans la direction desquelles elle etait visible,
<i>Clase número 2</i>	
475 31	eh bien, me dis je, luttons contre l'esprit fatal, luttons contre le dieu lui meme avec les armes de la tradition et de la science.
969 19	cette pensee me rassura, mais ne m'ota pas la crainte d'etre a jamais classe parmi les malheureux.
395 17	quand au peuple, a tout jamais engrene dans les divisions des castes, il ne pouvait compter ni sur la vie, ni sur la liberte.
406 17	ici ma memoire se trouble, et je ne sais quel fut le resultat de cette lutte supreme.
541 17	c-est un des ces rapports etranges dont je ne me rends pas compte moi meme et qu'il est plus aise d'indiquer que de definir;

En el análisis se diferencian cinco etapas técnicas: la definición de las unidades de contexto, la búsqueda de las formas reducidas analizadas, la definición de las tablas de datos asociados, la búsqueda de clases de unidades de contexto características y la descripción de estas clases para ayudar a la interpretación. En este proceso se procede de forma que sea posible, de una parte, controlar la

estabilidad de las clases en función de una variación de la definición de la unidad de contexto, y de otra, elegir una partición aceptable en la jerarquía de particiones propuestas. Para conseguirlo se construyen dos tablas de datos con definiciones ligeramente distintas de las unidades de contexto; se efectúa la clasificación de las unidades de contexto de cada tabla; se comparan las clases obtenidas y se conservan únicamente las clases relativamente estables.

Para construir las clases se utiliza un procedimiento de clasificación descendente jerárquico que permite tratar cuadros lógicos de gran dimensión. La primera clase analizada comprende todas las unidades de contexto retenidas. Después, progresivamente, se busca la partición en dos de la mayor de las clases restantes, maximizando un cierto criterio, lo que conduce a la sucesión del análisis. Para la interpretación de las clases se analiza la estructura de las mismas, ya sea por medio del vocabulario más específico de la clase retenida, como por la extracción de las unidades más representativas de este vocabulario.

Reinert, en el artículo antes citado, utiliza el sistema ALCESTE para el análisis de la estructura de un texto de Gerard de Nerval, y tras el mismo se refiere a la existencia de tres mundos en el texto, el imaginario, el real y el simbólico (Tabla 3).

TABLA 4. ALGUNAS DE LAS UNIDADES DE CONTENIDO CLASIFICADAS EN LA PRIMERA CLASE (Villegas y Behar, 1992)

517	1201	El cas ES QUE AQUESTES persones QUE sempre tenen
517	1202	coses PER explicar ES PERQUE sempre ES fixen EN
517	1203	LES coses i HO retenen tot A LA memoria.
662	1181	M'HO passo tan be A L'institut. M'encanta anar hi.
662	1182	si NO HI hagues anat M'HO passaria tan malament.
190	1151	pero ja EM vindran. idea! pot ser UN noi i dues
190	1152	noies (res DE sexe) (o pot ser si) (no HO se) QUE
190	1153	descobreixen Un cas misterios, oi?
820	1151	haig D'estudiar musica PERQUE dema EL tisanet ENS
820	1152	fa preguntes involuntaries i COM QUE tinc molta
820	1153	mala sort se QUE EM preguntara.
793	1141	hola! senzillament avui NO HA passat res, pero mira,
793	1142	tinc ganes D'escriure. QUE vols QUE T'expliqui?
553	1131	si no, mai NO podre ser ho. HE anat A Hostalets.
553	1132	be, dema T'explicare EL QUE HA passat.
661	1131	res mes, D'acord? passo molt DE LES altres coses.
661	1132	COM TU ja saps. O NO HO saps? doncs ara ja HO saps.
998	1131	penso guardar AQUEST diari tota LA meva vida. I
998	1132	tindra UN especial significat quan EL llegeixi
998	1133	D'aquí A 20, 25, 30 o mes anys.
614	1121	tinc LES galtes totes vermelles i molles./ ja estic
614	1122	mes calmada. M'HE asserenat, ara estic be.
344	1111	M'agradaria ser LA olga PER ser maca, tenir novios
344	1112	I ser rica, NO PER ser UNA puta.
615	1111	NO entec PER QUE M'HA passat. mai NO M'havia
615	1112	passat. deu ser del cansament? NO HO crec, PER MI
615	1113	QUE estic enamorada, massa.

Otra aplicación más reciente (Villegas, Behar, 1992) se ha llevado a cabo a partir de un material textual de mayor interés psicológico, el diario de una ado-

lescente, a propósito del cual se había realizado ya un análisis anterior por medio de otra técnica (Behar, Villegas, 1990). El tratamiento por medio de ALCESTE de este documento personal pone de manifiesto la existencia de tres mundos claramente diferenciados, todos ellos propios de la adolescencia: el inicio de la reflexión sobre la propia vida, la referencia a las actividades propias de la edad, y el ámbito de las relaciones personales en forma de amistades y primeros amores. En la Tabla 4 se muestran algunas de las unidades de contenido clasificadas en la clase I, en la lengua original del diario, catalán. La primera columna de cifras corresponde al número de la unidad de contexto. En la segunda columna, la primera cifra indica el número de clase, las dos siguientes el valor del chi-cuadrado de asociación obtenido, y la última cifra, el número de línea, 1, 2 o 3, de cada una de las unidades de contenido.

La objetividad es una característica a tener en cuenta en esta técnica. El investigador no manipula el material verbal en ningún momento del proceso, material verbal que es considerado inicialmente en su totalidad por el programa.

Consideraciones finales

Tras esta somera revisión cabe destacar que únicamente uno de los análisis propuestos, el análisis de correspondencias de Benzecri (1973), precisa la codificación previa del material verbal por el investigador, puesto que en los restantes procedimientos dicha codificación la lleva a cabo el propio programa. Esto elimina cualquier tipo de subjetividad al no ser necesario actuar sobre el material verbal al principio del proceso de análisis. Una vez realizado el análisis es cuando el investigador deberá interpretar los resultados, incorporando en dicha interpretación los criterios propios de la teoría a la que esté adscrito, o las conjeturas establecidas al inicio o durante la investigación. Hasta ese momento las técnicas propuestas permiten la replicabilidad.

Otro aspecto importante es la dimensión de los textos a analizar que, por medio de estas técnicas, puede ser de gran tamaño. En este sentido, alguno de los ejemplos citados es una buena prueba, como el diario personal de una adolescente durante tres años (Behar y Villegas, 1990). Con el análisis de contenido clásico era impensable abordar un texto de tal longitud. Con las técnicas informatizadas es posible analizar, con gran rapidez, desde una autobiografía hasta las sesiones terapéuticas mantenidas día a día con un paciente durante una psicoterapia. Ello facilita la objetividad en el estudio de textos que habitualmente no se someten a ningún tipo de análisis objetivo.

Finalmente se trataría de proponer, para escándalo de los ortodoxos de ambos enfoques, cuantitativo-objetivo y cualitativo-subjetivo, una aproximación ecléctica al análisis textual. ¿Por qué no utilizar para un mismo texto una técnica informatizada y otra más subjetiva y cualitativa? Esto permitiría enriquecer la comprensión del texto, descubrir los posibles aspectos comunes entre ambos procedimientos y también los puntos de divergencia. Las diversas técnicas no tienen porqué ser mutuamente excluyentes sino que pueden ser complementa-

rias. Si en otro ámbito es posible aplicar a unos mismos datos y en igual investigación un análisis factorial, un análisis de *cluster* y un análisis discriminante con objetivos claramente diferenciados (Behar, 1988), ello debe ser también posible con las técnicas de análisis textual. Según Lindkvist (1981) «el potencial de una provechosa integración entre las aproximaciones existe, hasta donde yo puedo ver, no como un método general o modelo, sino como una batería de instrumentos con capacidad para iniciar y desarrollar la espiral de la construcción del modelo textual» (p. 40). Esta visión integradora, además de permitir profundizar en el conocimiento de las distintas técnicas, puede considerarse como un proceso para alcanzar un objetivo, dado que de experiencias sucesivas en este sentido podría llegar a generarse el modelo textual al que se refiere Lindkvist.

REFERENCIAS

- Bardin, L. (1986). *El análisis de contenido*. Ediciones Akal S.A.
- Bécue, M., Behar, J., Freixa, M., Rajadell, N., Salafranca, L. y Turbany, J. (1993). *Análisis estadístico de textos: una aplicación al estudio del sentimiento de identidad europea*. Comunicación presentada en el III Simposium de Metodología de las Ciencias Sociales y del Comportamiento. Santiago de Compostela.
- Behar, J. (1988). Dimensión social de la pubertad. En T. Ibáñez (Ed.), *Ideologías de la vida cotidiana* (pp. 205-245). Barcelona: Sendai.
- Behar, J. (1988). Relation de l'état puberal avec la structure de l'identité. Analyse au moyen de l'observation indirecte. *Actes du Colloque Européen: Construction et fonctionnement de l'identité*. Aix-en-Provence, Francia.
- 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 I (pp. 330-389). Barcelona: PPU.
- Behar, J. & Villegas, M. (1990, junio). Analyse de la construction de l'identité au moyen des documents personnels. Comunicación presentada en el Colloque: *Education familiale et reconnaissance sociale*. Toulouse.
- Benzecri, J.P. (1973). *L'analyse des données*. Vol. II. Paris: Dunod.
- Berelson, B. (1971). *Content analysis in communication research*. New York: Hafner.
- Cornejo, J.M. (1988). *Técnicas de investigación social: el análisis de correspondencias (teoría y práctica)*. Barcelona: PPU.
- Gadamer, H.G. (1960). *Wahrheit und Methode*. Tübingen: J.C.B. Mohr.
- Hirsch, E.D. Jr. (1967). *Validity and Interpretation*. New Haven, C.T.: Yale University Press.
- Korosec-Serfaty, P. (1985). Experience and use of the dwelling. In I. Altman & C.M. Werner (Ed.), *Home Environments* (pp. 65-86). New York: Plenum Press.
- Korosec-Serfaty, P. (1989). Demeure et alterité: mise à distance et proximité de l'autre. *Arch. Behav.*, 5, 2, 161-173.
- Krippendorff, K. (1980). *Content analysis. An introduction to its methodology*. London: Sage.
- Lebart, L. (1982). L'analyse statistique des réponses libres dans les enquêtes socio-économiques. *Consommation*, 1, 39-62. Paris: Dunod.
- Lebart, L. (1992). *Assesing and comparing patterns in multivariate exploratory analysis*. Comunicación presentada en el Congreso Compstat' 92.
- Lebart, L., Morineau, A. & Warwick, R.M. (1984). *Multivariate descriptive statistical analysis: correspondence analysis and related techniques for large matrices*. New York: Wiley.
- Lebart, L. & Salem, A. (1988). *Analyse statistique des données textuelles*. Paris: Dunod.
- Lindkvist, K. (1981). Approaches to textual analysis. In K.E. Rosengren (Ed.), *Advances in content analysis* (pp. 23-41). Beverly Hills, Cal.: Sage.
- Muchielli, R. (1979). *L'analyse de contenu des documents et des communications*. Paris: Editions ESF.
- Reinert, M. (1990). ALCESTE. Une méthodologie d'analyse des données textuelles et une application: Aurelia de Gerard de Nerval. *Bulletin de Méthodologie Sociologique*, 26, 24-54.
- Rodríguez-Tomé, H. & Zlotowicz, M. (1972). Peurs et angoisse à l'adolescence. *Enfance*, 3-4, 239-307.

- Rodríguez-Tomé, H., Behar, J., Martínez, C. y Bariaud, F. (1982). Dimensiones de la identidad personal en la adolescencia. Estudio comparativo. *Revista de Psicología General y Aplicada*, 37, 3, 507-527.
- Rosengren, K.E. (1981). *Advances in content analysis*. Beverly Hills: Sage.
- Salem, A. (1988). Approches du temps lexical. Statistique textuelle et séries chronologiques. *Mots*, 17, 105-143.
- Villegas, M. (1992). Análisis del discurso terapéutico. *Revista de Psicoterapia*, 10-11, 23-65.
- Villegas, M. & Behar, J. (1992, setiembre). The study of the adolescence identity using personal documents. An application of Alceste program. *V European Conference on Developmental Psychology*. Sevilla.
- Zlotowicz, M. (1972). Peurs et angoisse chez les enfants d'âge scolaire. *Enfance*, 3-4, 175-237.