UNA APLICACION INFORMATICA PARA EL ESTUDIO DE CASOS

Julio Cabrera Varela

«Debe entenderse por sociología: una ciencia que pretende entender, interpretándola, la acción social para de esa manera explicarla causalmente en su desarrollo y efectos. Por "acción" debe entenderse una conducta humana siempre que el sujeto o sujetos de la acción *enlacen* a ella un *sentido* subjetivo. La "acción social", por tanto, es una acción en donde el sentido mentado por su sujeto o sujetos está referido a la conducta de *otros,* orientándose por ésta en su desarrollo.»

(Max Weber)

I. DEL DISCURSO

La misión de la sociología, siguiendo a Weber, es la comprensión *(Verstehen)* de los fenómenos sociales atendiendo al «sentido», explícito o implícito, consciente o no, que los propios actores dan a sus acciones. Con esta intención, ciertas escuelas sociológicas han desarrollado procedimientos metodológicos de carácter cualitativo en los que el «punto de vista del actor» constituye el principal foco de atención. Se trata de comprender el sentido de la acción social, pero reconstruyendo la realidad [1] en la que el actor se inserta, la cual no podemos suponer coincidente con la concepción de la realidad dominante en un momento determinado, ni con una particular pre-

[1] H. SCHWARTZ y J. JACOBS, *Qualitative sociology a method to the madness,* The Free Press, Nueva York, 1979.

concepción del investigador. Los diversos procedimientos (entrevistas, entrevistas en profundidad, estudio de casos, historias de vida, biografías, etc.) deben, pues, esforzarse en la reconstrucción del universo de sentido que constituye la «realidad» para el actor.

Los hombres, los actores, actúan en el mundo de la vida cotidiana [2], que se les presenta siempre como una realidad coherente, con sentido, interpretada intersubjetivamente y, por ello, «objetiva», «real». Ahora bien, como señala Schütz, el conocimiento del que dispone el actor, dentro del mundo intersubjetivo del sentido común, es un conocimiento incoherente, sólo parcialmente claro, y lleno de contradicciones. Los actores sociales no son sujetos clónicos, sino que la comprensión que cada quien tiene del mundo en que vive, así como la orientación dentro del mismo, viene determinada por su particular situación biográfica. El mundo, que el investigador puede tipificar objetivamente, es vivenciado como «mi mundo»; esto es, yo, en cuanto actor en el mundo social de la vida cotidiana, defino desde mi particular situación biográfica la realidad que habito. Como dice Schütz: *«En cualquier momento de su vida diaria, el hombre se encuentra en una situación biográficamente determinada, vale decir, en un medio físico y sociocultural que él define y dentro del cual ocupa una posición, no sólo en términos de espacio físico y tiempo exterior, o de su estatus y su rol dentro del sistema social, sino también una posición moral e ideológica. Decir que esta definición de la situación está biográficamente determinada equivale a decir que tiene su historia; es la sedimentación de todas las experiencias previas del hombre, organizada en el patrimonio corriente de su acervo de conocimiento a mano, y, como tal,* es su posesión exclusiva, dada a él y sólo a él» [3].

En esta perspectiva, un estudio de caso o una historia de vida realizados en profundidad y con rigor requiere una explicitación de esas coordenadas biográficas que determinan la comprensión y vivencia de «la realidad» como «mi realidad» en cada actor analizado, así como el universo de sentido que caracteriza y orienta su vivencia actual, en un «Aquí» y un «Ahora» particular [4].

De todos los medios posibles para acceder a este conocimiento, uno destaca especialmente por su inmediatez y complejidad: el lenguaje. Ciertamente, el lenguaje, escrito u oral, es el intermediario habitual entre investigador y actor; digamos que es la manifestación más evidente que el primero tiene del segundo, pero, además, el lenguaje constituye algo así como el depósito de los elementos del universo de sentido que el actor emplea en la «construcción» de «su» realidad. *«El lenguaje usado en la vida cotidiana* —escriben Berger y Luckmann— *me proporciona continuamente las objetivaciones in-*

[2] A. SCHÜTZ, *Estudios sobre Teoría Social,* Amorrortu, Buenos Aires, 1974.
[3] A. SCHÜTZ, *El problema de la realidad social,* Amorrortu, Buenos Aires, 1974, p. 40. El destacado es nuestro.
[4] A. SCHÜTZ, ídem, p. 138.

dispensables y dispone el orden dentro del cual éstas adquieren sentido y dentro del cual la vida cotidiana tiene sentido para mí» [5].

Pero el lenguaje es algo más que eso: es el medio tipificador a través del cual se transmite el conocimiento originado socialmente. Sus tipificaciones se refieren al sistema de significatividades de la sociedad que lo generó, haciendo convivir el pasado histórico de esta comunidad con el presente, mediante su actualización dentro de las coordenadas que definen los nuevos intereses y, por lo tanto, las nuevas significatividades. *«El lenguaje habitual precientífico* —señala Schütz— *puede ser comparado con un depósito de tipos y características ya hechos y preconstruidos, todos ellos de origen social y que llevan consigo un horizonte abierto de contenido inexplorado»* [6].

El lenguaje de la vida cotidiana es el lenguaje que comparto con mis semejantes y a través del cual construimos las objetivaciones de nuestro «mundo de la vida», esto es, «nuestra realidad». Por ello, cualquier intento de comprender el mundo de la vida cotidiana tiene que ser primeramente un esfuerzo por comprender el lenguaje. Porque el lenguaje no sólo objetiva el ser del que lo emplea para manifestarse, lo hace presente a los demás, sino que también, y sobre todo, tipifica experiencias, esto es, las categoriza, de manera que trascienden el «Aquí» y el «Ahora» del que las nombra, y se tornan patrimonio común, anónimas.

> *«Como resultado de estas trascendencias, el lenguaje es capaz de "hacer presente" una diversidad de objetos que se hallan ausentes —espacial, temporal y socialmente— del "aquí y ahora".* Ipso facto, *una enorme acumulación de experiencias y significados puede llegar a objetivarse en el "aquí y ahora". Más sencillamente, en cualquier momento puede actualizarse todo un mundo a través del lenguaje»* [7].

Ahora bien, la comprensión del lenguaje y su interpretación no puede consistir solamente en la fácil referencia a la codificación de cada palabra en el diccionario y su articulación siguiendo determinadas reglas. El lenguaje empleado por un actor en un momento particular, en un «Aquí y un Ahora», esto es, una actualización particular de aquél en un habla, constituye una acción social realizada en un marco social determinado, en una cultura determinada. Así, pues, cada unidad discursiva implicará, además de su significación lexicográfica y gramatical, el momento cultural desde el cual se genera y el momento biográfico de quien la genera.

Toda unidad, sea palabra u oración, está recubierta por lo que William James llamaba «orlas», o universos finitos de significado, que la conectan con

[5] BERGER y LUCKMANN, *La construcción social de la realidad,* Amorrortu, Buenos Aires, 1968, p. 39.
[6] A. SCHÜTZ, ídem, p. 44.
[7] BERGER y LUCKMANN, ídem, p. 58.

la historia, el presente y el futuro del universo social y discursivo al que pertenece; pero, al mismo tiempo, esas orlas pertenecen también a la intimidad más exclusiva del que las enuncia, llenándose por ello de connotaciones emocionales e irracionales. Esa esfera del discurso, solamente presente en quien lo pronuncia, está siempre presente en la materialidad del discurso, pero oculta a la conciencia reflexiva.

Cualquier contexto cultural genera y/o integra una gran diversidad de subculturas y modalidades que mantienen con él una dialéctica referencial viva, sea conflictiva o armónica. En realidad, es la vida de éstas lo que constituye la cultura en un momento particular. Por ello, el lenguaje adquiere las connotaciones propias de la subcultura en que se usa y del momento particular en que se emplea. Pero, aún dentro de este marco, cada palabra, cada giro, está enriquecido por la experiencia común de quienes, con su empleo reiterado, lo actualizan constantemente. Esta experiencia común, restringida a grupos sociales específicos, es la que hace evolucionar al lenguaje de la vida cotidiana a través de sus diversos momentos semánticos. Podemos, así, decir con Schütz que toda la historia del grupo lingüístico se refleja en su manera de decir las cosas.

Es a esta particular riqueza semántica del lenguaje a la que el investigador de casos debe estar especialmente despierto. Sólo captando, en la medida de lo posible, las connotaciones de lo que se nos dice en una entrevista podremos captar plenamente a quien nos lo dice. Ciertamente, son estos campos semánticos generados socialmente los que hacen posible la *«objetivación, retención y acumulación de la experiencia biográfica e histórica»*. En otras palabras, los campos semánticos nos señalan el estado presente del conocimiento social, al que el individuo tiene acceso en su vida cotidiana, así como su particular ubicación en la sociedad.

Se trata, pues, de poder dar respuesta a las preguntas ¿qué quiere decir el entrevistado cuando emplea esta o estas palabras?, ¿por qué éstas y no otras?, ¿cuál es su «sentido»? Intentamos captar, al menos en este momento, el sentido de la acción, materializada en forma de narración, a través del sentido de la palabra, del discurso. El discurso suplanta a la acción, la oculta bajo su reproducción ideológica, la reifica. Pero el discurso también nos proporciona el «sentido subjetivamente mentado» que la constituye como acción social. En el discurso, el actor nos habla del sentido de su conducta, de «su» sentido y, por tanto, del universo de sentido en el que se mueve. El «sentido subjetivamente mentado» de la acción es el acto de pensamiento por el cual la conducta se constituye en acción social. Pensamiento que se articula lingüísticamente empleando el lenguaje que el actor habita, el lenguaje de la vida cotidiana, que por ello recupera y actualiza el universo de sentido social en el que actúa. El discurso reproduce para otro la acción (la narra) en el modo de la reflexión (del recuerdo) sobre el pensamiento que le proporcionó un sentido. El acceso al sentido del discurso nos abre el camino, tortuoso y

lleno de trampas, a la comprensión del sentido de la acción, a su definición. *«Si los hombres definen las situaciones como reales* —decía Thomas—, *éstas lo son en sus consecuencias»;* por ello, definir significa actuar, y la interpretación del mundo es un modo de actuar en él.

Buscar el sentido de la definición de la «realidad», del habla sobre la realidad, escudriñando en el interior de las palabras que conforman el discurso en que se formula, no otro es el objetivo en este momento del análisis. Pero *«el sentido de una forma lingüística se define por la totalidad de sus empleos, por su distribución y por los tipos de relación que de ello resultan»* [8]. Desvelar el sentido de una palabra implica el análisis de sus contextos, de sus empleos. Pero las formas lingüísticas, las palabras, no son elementos aislados y yuxtapuestos; el léxico constituye un sistema de elementos coordinados u opuestos que se definen precisamente por esa relación [9]. Por ello, el análisis implica la desestructuración y reestructuración del discurso, un trabajo de segmentación y reestructuración significativa.

Desentrañar los contextos, establecer las redes de asociación y oposición, las antinomias y las identidades, requiere como labor previa el establecimiento de las redes lexicales que articulan el discurso. Labor para la que puede ser de gran utilidad la lexicología.

II. DEL METODO

> «La lexicología es una disciplina [que] estudia grupos de palabras consideradas estadísticamente desde el punto de vista nocional... La lexicología, pues, tiene por objeto, *como* la sociología, el estudio de hechos sociales... *Partiendo del estudio del vocabulario intentaremos explicar una sociedad.* Así podremos definir la lexicología como una *disciplina sociológica* que utiliza el material lingüístico que son las palabras.»
>
> (G. Matoré)

Lexicometría, lexicología y análisis del discurso son métodos que se complementan; en el segundo se trabaja sobre los datos del primero, y ambos abren caminos para el último. La lexicología es un método útil para el sociólogo, pero no agota su labor. La estadística lexical basada únicamente en la frecuencia de empleo del vocabulario no puede concluir una investigación; por el contrario, su misión consiste en facilitar la construcción de hipótesis que el sociólogo tratará con su metodología específica. La estadística lexical ha de entenderse como descriptiva y no inferencial; nos puede describir, descubrir, relaciones en las que no hubiésemos reparado auxiliados solamente de la intuición, pero ahí termina su cometido. En esta perspectiva, análisis

[8] E. BENVENISTE, «Les problèmes sémantiques de la reconstruction», en *Problèmes de la linguistique générale,* NRF, 1966.

[9] J. DUBOIS, *Le vocabulaire politique et social en France, de 1869 à 1872,* París, 1962, p. 188.

de contenido (cuantitativo) y análisis de discurso (cualitativo) no se oponen, sino que se complementan: lo cuantitativo se pone al servicio de lo cualitativo.

Los métodos lexicológicos constituyen un conjunto de técnicas estadísticas diseñadas para medir y analizar el vocabulario que conforma un discurso y su particular estructuración. Se trata de detectar las constantes léxicas, así como los diversos campos semántico-nocionales que generan, del discurso sobre la realidad; intenta describir fielmente los términos en que una realidad social es «construida». Se trata, por tanto, de un estudio cuantitativo y cuantitativo-nocional del discurso tomado en su materialidad, pero ello no excluye la perspectiva cualitativa, antes bien la implica.

Podemos considerar la metodología lexicométrica bajo dos perspectivas diferenciadas, aunque complementarias: la lexicometría fuera de contexto y en contexto. Ambas se implican: la lexicometría en contexto sólo puede realizarse sobre la base de los datos estadísticos de la lexicometría fuera de contexto, mientras que esta última queda incompleta sin la primera.

La lexicometría fuera de contexto consiste en técnicas que se mueven en la superficie del discurso, en su materialidad. Aportan una exhaustiva descripción estadística de los componentes de un discurso mediante diversos análisis distribucionales realizados sobre las tablas de frecuencias de las formas lexicales que constituyen el *corpus* de los diversos discursos. Son, pues, técnicas de tratamiento de datos, esto es, descriptivas. Sin embargo, no carecen de gran interés. Estas técnicas, tal como han sido establecidas por Guiraud, Muller o Benzécri y desarrolladas por Dubois, Prost o la ENS de Saint-Cloud [10], pueden aportar datos de gran valor, de entre los que destacan:

El vocabulario común a varios discursos, con los análisis de su distribución frecuencial en cada uno de ellos, pudiendo determinar el alcance del «vocabulario trivial» y el propio de las condiciones en que se produce el discurso.

El vocabulario específico (original o característico), esto es, el recuento de formas presentes en un emisor (o emisores, si se los considera como homogéneos), mediante el análisis de la acumulación de formas originales.

Los coeficientes de repetición general, repetición funcional, repetición léxica y originalidad, entre otros, para los diversos emisores.

La determinación de la complejidad estructural del discurso, atendiendo a la extensión y complejidad de las frases (número de frases, número de segmentos, extensión media de las frases, de los segmentos y número medio de

[10] P. GUIRAUD, *Problèmes et méthodes de la statistique linguistique*, París, 1960; Ch. MULLER, *Initiation à la statistique linguistique*, Larousse, París, 1968; J. C. BENZÉCRI, *Analyse des données*, París, 1973; J. DUBOIS, *op. cit.*; A. PROST, *Vocabulaire des proclamations électorales de 1881, 1885, 1889*, París, 1974; Laboratoire de Lexicologie, ENS de Saint-Cloud, ERA 56 (DEMONET, GEFFROY, GOUAZE, LAFON, MOUILLAUD y TOURNIER), *Des Tracts en Mai 68. Mesures de vocabulaire et de contenu*, Champ Libre, París, 1978.

segmentos por frase, etc.), datos de los que el grupo ERA 56 de Saint-Cloud ha extraído importantes elementos descriptivos sobre los grupos políticos del Mayo del 68 en París.

La estructura léxica del vocabulario, análisis distribucional que permite detectar las «palabras clave» o «polos» en torno a los que se articula el discurso.

El estudio de los parentescos léxicos de los diversos emisores, mediante el recurso al Análisis Factorial, según fue empleado por Benzécri y Prost, entre otros.

Y, para terminar este breve resumen, las matrices de frecuencias construidas permiten su tratamiento por cualquier técnica de análisis distribucional.

Los datos así obtenidos son valiosos por sí mismos y pueden aportar abundante y rica información sobre los procesos discursivos, como queda patente en los trabajos ya clásicos de Dubois y de Prost. Pero constituyen también los datos básicos que permitirán realizar los análisis propios de la lexicometría en contexto.

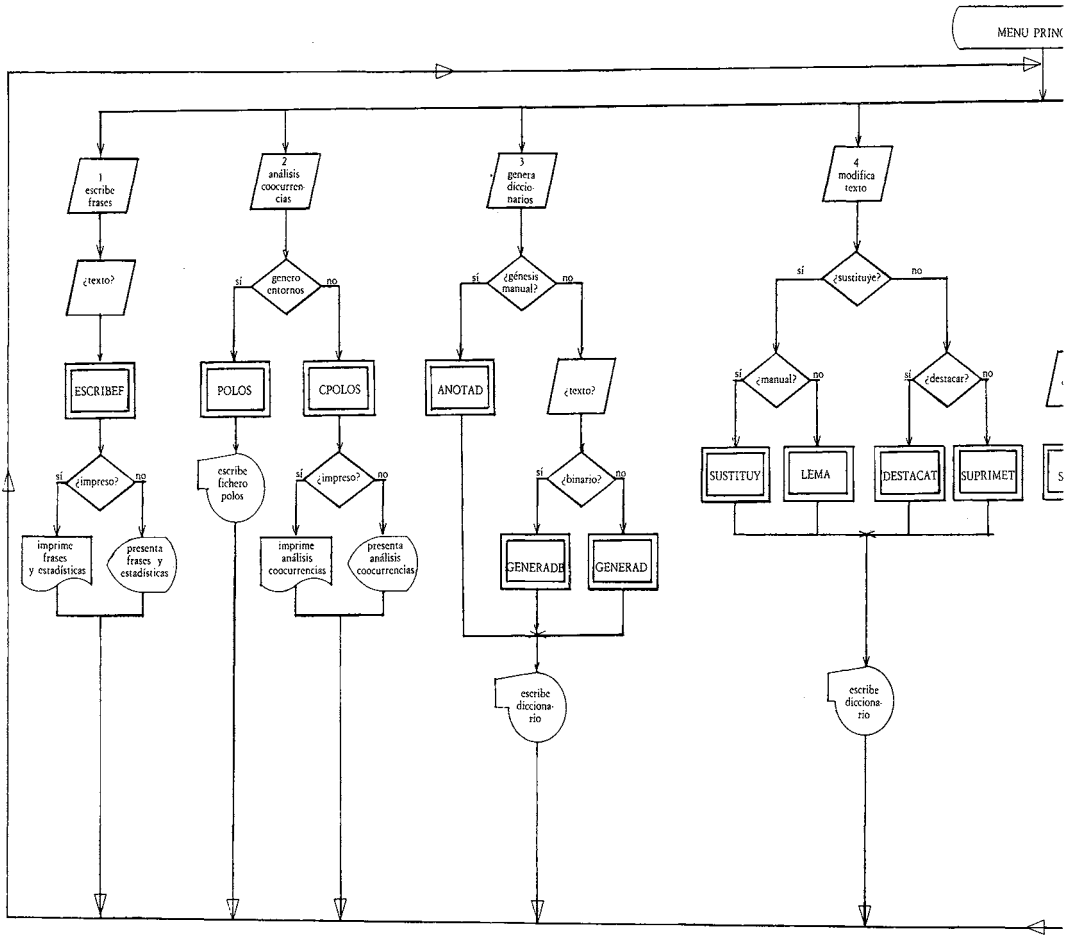
La lexicometría en contexto abarca dos ámbitos que, una vez más, se implican: el análisis de co-ocurrencias y el análisis semántico categorial.

Primeramente se tratarán los datos lingüísticos atendiendo a su posición material en las redes de atracción estadísticas de las formas de frecuencia relevante, de las «palabras clave» o «polos». Los enunciados se componen de significantes que se suceden, estando cada uno precedido, seguido o encuadrado por otros. A este hecho puramente material se le llama co-ocurrencia [11]. El sentido no es aún el objeto; se trata, por el momento, de la determinación de un orden de sucesión perseguido a lo largo de todo un texto o de un discurso. La presencia asociada de dos o más términos a lo largo del discurso (co-ocurrencia) puede responder a propiedades morfológicas (se trata, pues, de un caso de sintaxis), o bien puede deberse a que se reclaman uno al otro por alguna capacidad funcional. En el primer caso estamos ante la descripción del estado físico de la presencia simultánea de *n* ítems gráficos en la misma unidad de significación: es la co-ocurrencia *sensu estricto;* en el segundo diremos que esos dos o más términos están *«correlacionados»,* lo que en cierta medida ya nos habla del sentido (ERA 56). El análisis de las co-ocurrencias permite el de las correlaciones.

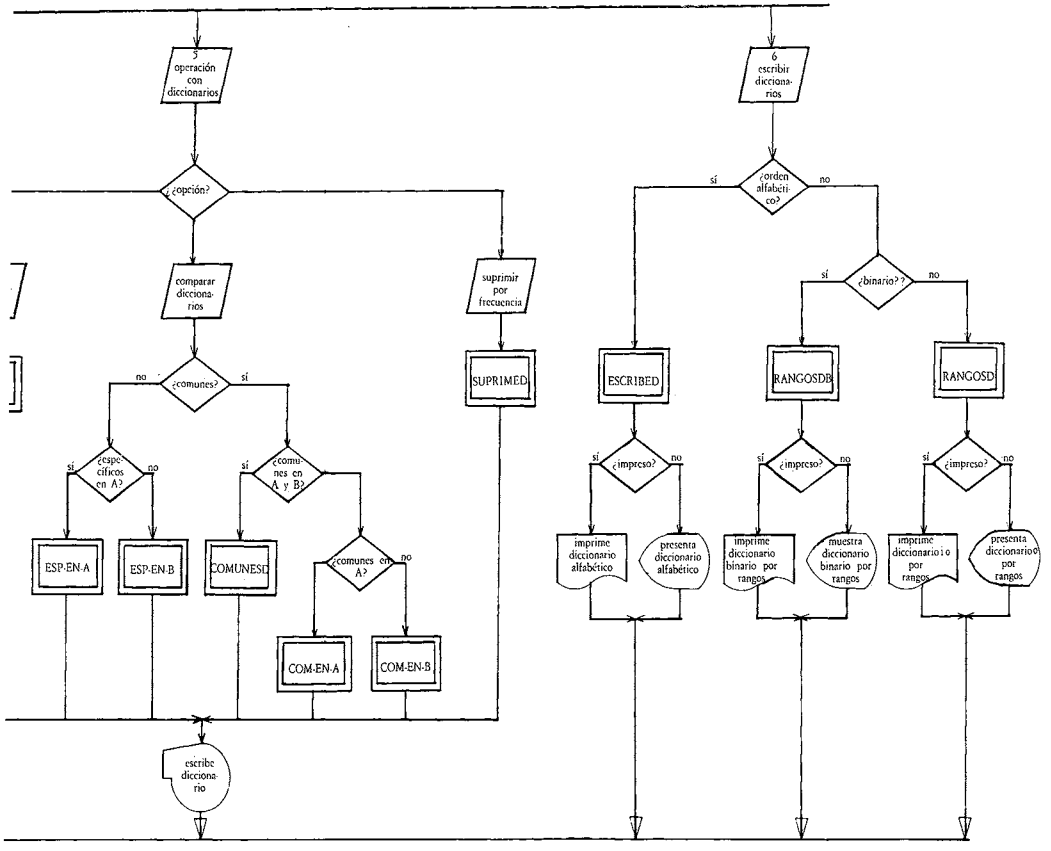
El primer paso consistirá, por tanto, en establecer las co-ocurrencias de determinado ítem lexical *(«polo»);* por tanto, hay que distinguir ocurrencia y co-ocurrencia. El universo de ocurrencias de un polo constituye su *«vecindario»* (también llamado expansión del polo —expansión derecha e izquierda—, y que se fija de antemano en número de ítems), esto es: *«Todo ítem observado en la unidad de significación en la que el polo está implicada pue-*

[11] R.-L. WAGNER, *Les Vocabulaires français,* Didier, París, 1970.

DIAGRAMA PRINCIPAL DEL SISTEMA LEXICOMETRICO Y LEX



INTEGRADO PARA EL TRATAMIENTO
ICOLOGICO DE TEXTOS



de ser llamado su "vecino"» [12]. Para que esta aparición, esta ocurrencia, pueda ser llamada co-ocurrencia del polo es preciso una prueba estadística que la destaque de entre las demás formas vecinas. Ello se obtiene mediante el Coeficiente de Vecindad, consistente en dotar a cada ocurrente de un índice resultado de la correlación de su frecuencia en la expansión del polo (co-frecuencia observada), su frecuencia relativa (*Fr*) para el conjunto del *corpus* (que se toma como frecuencia teórica de aparición en la expansión), el sumatorio de las distancias (medidas por número de ítems interpuestos entre la forma y el polo) y la distancia teórica de dicha forma al polo. Dicho índice es corregido posteriormente por el Coeficiente Medio de Co-ocurrencia, que reduce todos los polos a un coeficiente medio.

Se obtiene así un grupo restringido de formas significativas que pueden ser consideradas co-ocurrentes (derecha o izquierda) del «polo» analizado. A cada una de ellas le corresponde un índice conforme al cual pueden ser ordenadas bien jerárquicamente, bien describiendo una red de co-ocurrencias lexicales, cuya representación consistirá en un grafo de la red lexical del «polo» estudiado [13].

Un determinado «polo» presenta una red lexical propia en la que cada uno de sus co-ocurrentes es estadísticamente significativo. Ahora se podrá determinar cuáles de aquellos co-ocurrentes han de ser destacados como «polos» para realizar su correspondiente análisis componencial. El final del proceso será una completa descripción lexical del texto en torno a determinadas palabras clave aisladas, de entre el conjunto de las estadísticamente relevantes, siguiendo los intereses particulares de la investigación.

Las redes lexicales obtenidas mediante el análisis de co-ocurrencias constituyen el material base para comenzar la reconstrucción del sentido del discurso. Se tratará de reconstruir los campos semánticos de aquellas palabras consideradas como «polos»; por tanto, no nos moveremos ya en la superficie del texto; entramos, pues, en el campo del análisis cualitativo.

Esta nueva operación parte de los siguientes postulados:

— La no transparencia del texto, lo que implica la necesidad de un trabajo de desestructuración de la cadena lexical para reconstruirlo según una legibilidad significativa. Se necesita, por tanto, un cierto trabajo hermenéutico sobre el orden del discurso.

— Como ya se señaló, buscar el sentido de una palabra implica el análisis de la totalidad de sus empleos y contextos.

— Por último, la consideración del léxico no como una simple yuxtaposición de ítems, sino como un sistema en el que todas las unidades están

[12] A. GEFFROY, P. LAFON y M. TOURNIER, «Analyse lexicometrique des co-occurrences et formalisation», en *Les Applications de l'informatique aux textes philosophiques,* Documentation CNRS (Coloquio de 1970), pp. 8-23.

[13] C. BERGE, *La Théorie des graphes et ses aplications,* Dunod, 1958.

coordinadas u opuestas las unas a las otras. Ello implica un estudio detallado de las diversas relaciones establecidas en el texto: asociaciones (ligazón positiva), oposiciones (ligazón negativa), cualificaciones (unidades que califican el ser o la manera de ser de la unidad de análisis) y sus cualidades pragmáticas («acción de» y «acción sobre»). Así obtenemos también los diversos equivalentes (palabras que mantienen las mismas ligazones de asociación y oposición y que, por ello, pueden ser reemplazadas unas por otras en el texto sin alterar su sentido).

Mediante este procedimiento se construyen los diversos campos semánticos de las palabras «polo», puestos de relieve sintagmáticamente por las funciones de cualificación y de acción y paradigmáticamente por las de equivalencia y asociación. Una vez establecidos los diversos campos semánticos de las palabras centro de nuestra atención, podremos comenzar la tarea de reconstrucción del texto, realizando una lectura temática del mismo y tomando como base la estructura semántica de los diversos discursos. En otras palabras, disponemos de una sólida base para realizar un profundo análisis de contenido.

Una última nota referente al proceso anteriormente diseñado. Todo el proceso (comenzando por los primeros análisis lexicométricos —estadísticos— sobre el universo de formas lexicales que conforman un discurso, que nos permite generar las primeras hipótesis sobre las palabras clave —«polo»— de los análisis posteriores; continuando con el análisis co-ocurrencial —o componencial— de dichos polos y la construcción de las redes lexicales que han de servir como base, finalmente, para la reconstrucción de los diversos campos semánticos) es realizado mecánica o semimecánicamente, lo que permite que cualquier momento del análisis pueda ser revisado, corregido o ampliado en cualquier momento de la investigación y para el conjunto total del *corpus* tratado, con gran agilidad y economía de medios. Todo ello posibilita una constante retroalimentación entre los diversos momentos del análisis, lo que resultaría impensable empleando los procedimientos manuales propios del análisis de contenido clásico.

El análisis de contenido constituye la última fase del trabajo sobre el texto. Tres son los elementos básicos empleados para su diseño:

— En primer lugar, los resultados del análisis factorial realizado sobre las tablas de frecuencias nos aportan unos primeros datos sobre la agrupación de los emisores en función de su lexicalidad explícita, superficial.

— Partiendo de los datos obtenidos de la reconstrucción de los campos semánticos, y por comparación entre los diversos emisores, obtendremos sus parentescos y oposiciones. Tenemos, así, un segundo criterio clasificatorio de los discursos en función del «sentido», de la semanticidad de sus discursos.

— Los resultados obtenidos de los primeros análisis distribucionales combinados con los propios del análisis de co-ocurrencias nos fijarán las unidades

categoriales significativas, esto es, las unidades que constituirán el foco de atención del análisis categorial para todo el *corpus*.

Sobre esta base se comenzará la reconstrucción total del contenido de los discursos, partiendo de sólidas hipótesis sobre sus características diferenciales y de acuerdo con el código diseñado al comienzo de la investigación. Este código de referencia se verá enriquecido e incluso modificado en función de los descubrimientos de los análisis lexicológicos, de forma que sus elementos focales se estructurarán en función del tipo ideal de sentido subjetivo de la definición de la realidad presente en los actores construido a partir de dichos descubrimientos.

III. DESCRIPCION DEL SISTEMA INTEGRADO PARA EL TRATAMIENTO LEXICOMETRICO Y LEXICOLOGICO DE TEXTOS

El Sistema Integrado para el Tratamiento Lexicométrico y Lexicológico de Textos [14] es el resultado de casi dos años de investigación y desarrollo de técnicas originales para el proceso de textos y diccionarios, durante los cuales he podido beneficiarme de la estrecha y desinteresada colaboración de espe-

[14] El sistema ha sido desarrollado con la colaboración de Juan Rodríguez Marticorena. Originalmente, el programa, construido en Basic, ha sido concebido e implementado en un Ordenador Personal y Procesador de Textos Amstrad PCW-8512, con las siguientes características: Microprocesador Z80A a 4 Megahertzios, 512 K de RAM, de las cuales 368 K se corresponden al Disco Virtual de Memoria. Unidad de Disco Principal de 173 K de capacidad, una cara y densidad simple. Unidad de Disco Secundaria de 706 K de capacidad, doble cara, doble densidad. Teclado completamente castellanizado. Pantalla de texto de 32 líneas por 24 columnas. Sistema Operativo CP/M 3.1, de Digital Research ©1985. Procesador de Textos Locoscript, de Locomotive Software ©1985. Intérprete de Mallard Basic versión con JETSAM (rutina de indexación de ficheros incorporada), también de Locomotive Software ©1985. Este intérprete se caracteriza por su compatibilidad jerárquica con el MBASIC de Microsoft ©1983, pero incorpora sustanciales mejoras en cuanto al tratamiento de datos y la interacción con el sistema operativo subyacente. Existen versiones de este intérprete para los Sistemas Operativos CP/M 3.0, CP/M 3.1, CP/M-86 y DOSPLUS de Digital Research, así como para versiones del MS-DOS de Microsoft iguales o superiores a la 2.1. Todo ello garantiza al máximo la transportabilidad del programa entre una amplísima gama de Ordenadores Personales, que incluye todos los aparatos en los que arranque cualquiera de los Sistemas Operativos anteriormente mencionados (toda la gama Amstrad con 128 K de RAM o más, todos los compatibles IBM PC, etc.). Adicionalmente, este programa es trasladable a otros lenguajes de programación, revisable, modificable, ampliable y adaptable a configuraciones de Hardware disímiles de aquella para la que fue inicialmente concebido.

Este equipo reunía tres características esenciales, que fueron consideradas al comienzo de la investigación: que dispusiese de una memoria suficiente para almacenar y procesar ágilmente un volumen de texto suficiente, que los programas en él diseñados y la información generada pudiesen ser fácilmente trasladables a otros equipos y, lo que no es menos importante, que resultase asequible no sólo al investigador, sino a futuros posibles usuarios. En definitiva, salvar en alguna medida las barreras de carácter económico y tecnológico en el empleo de estas sofisticadas metodologías.

cialistas en áreas diversas (Ciencia Política, Sociología, Estadística, Filología, Análisis y lenguajes de programación, etc.) tanto de la comunidad universitaria como del mundo de la informática comercial.

Se ha optado por el desarrollo independiente de los diversos algoritmos que conforman el programa, cada uno de los cuales puede llegar a funcionar por sí solo, y compactarlos posteriormente en un único sistema. Ello responde a una concepción multiforme de los mismos, de forma que puedan ser compactados en el futuro, total o parcialmente, de manera diferente a fin de cubrir las necesidades de investigaciones de otra índole que decidan emplear estas metodologías.

Los algoritmos pueden ser agrupados en cuatro bloques, según las operaciones que realizan.

Operaciones con textos

En este apartado es necesario reseñar que los textos fuente pueden ser generados por cualquier procesador de textos de los existentes en el mercado, con la única restricción de que los ficheros generados lo estén en código ASCII. Tal es el caso del que se ha empleado, que, aun cuando utiliza un sistema operativo particular, sus ficheros pueden ser traducidos por el mismo procesador en ficheros ASCII. Este requisito se ha adoptado pensando en la conveniencia de facilitar el traslado de información entre las diversas configuraciones y no restringir su empleo a la presente.

Podemos decir, por evidente que parezca, que el primer programa a emplear consiste en un proceso de textos de las anteriores características; en este caso optamos por el Procesador de Textos Locoscript, de Locomotive Software.

Es importante señalar que el conjunto del sistema, en cualquiera de las funciones que a continuación se describen (tanto en las operaciones con textos como en la génesis y operaciones con diccionarios), posibilita el empleo de palabras acentuadas, lo que no es habitual en este tipo de programas, evitando así engorrosos problemas de codificación.

Una vez se han generado los textos, con los diversos ficheros existentes se pueden realizar operaciones de tres tipos:

A) MODIFICARLOS de manera mecánica, manteniendo la misma norma para todos ellos, o bien de manera semimecánica, atendiendo a las peculiaridades de cada fichero. A estas necesidades responden los programas:

«SUSTITUY»: El objetivo consiste en sustituir determinadas palabras por otras en el interior del texto. Puede ser útil tanto para lematizar un texto (reducir las diversas formas a sus radicales, plurales a singulares, feme-

ninos a masculinos, formas verbales conjugadas a sus infinitivos, etc.), requisito para realizar las diversas operaciones estadísticas propias de la metodología lexicométrica, como para unificar conceptualmente las diversas unidades lexicales sobre las que se ha de realizar el análisis posterior.

«LEMA»: Su misión específica consiste en realizar, de manera totalmente mecánica, sustituciones en un texto de acuerdo con un diccionario especializado previamente construido («ANOTAD»). Su aplicación permite mantener la norma de sustitución constante para todos cuantos textos se quieran procesar, unificando completamente el análisis y evitando los problemas característicos de ambigüedad propios de las modificaciones manuales y semimecánicas.

«SUPRIMET»: Este algoritmo permite suprimir de un texto original aquellas palabras que no interesa considerar a la hora de realizar el análisis lexicológico.

«DESTACAT»: Su misión y funcionamiento es semejante a «SUPRIMET», pero en lugar de suprimir en un texto las formas contenidas en un diccionario, las destaca aislándolas entre dos símbolos ($\Rightarrow \Leftarrow$), de manera que a los efectos del análisis lexicométrico su comportamiento es semejante a «SUPRIMET», pero, manteniendo las formas nos permite visualizar o imprimir el texto en su integridad, con las subsiguientes ventajas de control y comprensión.

B) AISLAR determinados segmentos del mismo, operación que ha sido desarrollada en una doble vertiente: aislar las diversas frases que componen el texto, numerándolas y realizando diversos cálculos en cuanto a su longitud y complejidad, y aislar los diversos entornos que acompañan a aquellas palabras (o raíces) que interesa explorar, realizando, si se requiere, diversos análisis lexicológicos sobre los mismos (análisis de co-ocurrencias o contingencias). Para la primera operación se empleará «ESCRIBEF»; para la segunda, «POLOS».

«ESCRIBEF»: Este algoritmo permite listar en pantalla o impresora un texto fuente aislando sus diversas frases, que aparecerán numeradas de manera consecutiva a partir del valor que se introduzca como número de la primera; la posibilidad abierta de introducir el primer valor numérico de la primera frase es interesante, por cuanto permite mantener la numeración consecutiva en caso de que se deseen analizar varios ficheros de texto consecutivamente.

Al finalizar el proceso de aislamiento y numeración de frases, el programa presenta datos estadísticos referentes al número de frases, longitud media de las mismas, complejidad media de las frases (número medio de segmentos por frase), longitud del texto en número de palabras, número de formas empleadas, frecuencia máxima (moda), frecuencia promedio y

tasa media de repetición, todo lo cual nos da una pormenorizada descripción estadística del texto en cuestión.

«POLOS»: La finalidad de este algoritmo consiste en facilitar los datos base para el análisis de co-ocurrencias (también llamado análisis componencial). Consta de dos elementos centrales: «POLOS», cuya misión consiste en aislar los diversos entornos de las palabras «clave» que han de ser analizadas, y «CPOLOS», encargado de realizar aquellos cálculos estadísticos necesarios para determinar qué ítems lexicales, presentes en dichos entornos, son relevantes.

Generación de diccionarios

Se trata de cumplir una doble función: la primera, que constituye, junto con el análisis de co-ocurrencias, el «grueso» del programa que nos ocupa, consiste en generar el *index* de un texto, esto es, el conjunto ordenado de los ítems que lo conforman; la segunda permite generar manualmente diccionarios de carácter auxiliar para el analista (diccionarios de formas funcionales, homónimos, diccionarios especializados, etc.).

«GENERAD»: El procedimiento es totalmente mecánico, registrando el diccionario construido todas aquellas formas presentes en un texto. Estos diccionarios constituyen la base sobre la que se han de realizar las comparaciones estadísticas propias de la metodología lexicométrica. El resultado final es un listado alfabético de dichas formas acompañadas de su frecuencia absoluta y relativa con respecto al *corpus* total, así como de su longitud; al final del mismo se aportan también los siguientes datos estadísticos: número de formas, número de palabras (extensión del *corpus*), frecuencia máxima, frecuencia promedio y tasa de repetición, así como una relación del número de palabras según longitudes y frecuencias, acompañadas de sus correspondientes diagramas de barras.

«GENERADB»: Este algoritmo es semejante al precedente en su funcionamiento y presentación de resultados, con la particularidad de que, en lugar de considerar las palabras como elemento de análisis, considera la presencia de dos formas contiguas como unidad, permitiendo de esta manera detectar la presencia, así como su relevancia, de las asociaciones lexicales persistentes en el *corpus*.

«ANOTAD»: Como se ha dicho, el presente algoritmo permite generar diccionarios de manera manual; éstos tienen como misión auxiliar a la depuración de los diversos textos (diccionario de formas funcionales, homónimas, no relevantes, etc.), así como para la corrección de los diversos *index* generados. Más relevante es la posibilidad de crear diccionarios especializados destinados a las diversas variantes del análisis de contenido.

Operaciones con diccionarios

Los diversos algoritmos encuadrados en el presente apartado sirven para diversas finalidades:

«SUMAD»: La creación de diccionarios por acumulación de otros ya existentes; en otras palabras, la suma de diccionarios —lo que supone la inclusión en el nuevo de todas las formas de los precedentes, así como la actualización (acumulación) de las frecuencias absolutas y relativas de aquellas comunes a dos o más diccionarios de los precedentes—. Todo ello posibilita tanto la obtención de un *corpus* general a partir de *corpus* particulares como el tratamiento fraccionado de un *corpus* determinado en fragmentos reducidos, lo que, sin duda alguna, facilita y agiliza el tratamiento de grandes masas de datos. De esta forma, un texto amplio puede ser «troceado» sin que por ello se tenga que renunciar a un análisis global; dicho troceamiento hace más manejables los textos, tanto para su almacenamiento como para sus posibles modificaciones y correcciones.

«SUPRIMED»: La modificación de diccionarios existentes, a fin de suprimir de los mismos aquellas formas que por su frecuencia sobrepasen cierto umbral de significatividad o no lo alcancen (tal es el caso de determinados verbos en función auxiliar o de las formas de frecuencia 1, o Hapax, cuya relevancia estadística se considera nula).

«COMUNESD», «COM-EN-A», «COM-EN-B», «ESP-EN-A» y «ESP-EN-B»: La comparación entre diccionarios, pudiendo obtener las formas comunes a dos *index* correspondientes a dos *corpus* distintos, con la posibilidad de distinguir las correspondientes formas comunes en cada uno de ellos, así como las formas específicas de un *index* frente a otro, sea este último el diccionario correspondiente a un *index* particular o a la acumulación de varios con respecto a los que se quiere hacer la comparación.

CRITICA DE LIBROS