



## ANÁLISIS DE TEXTOS ASISTIDOS CON PROGRAMAS COMPUTACIONALES\*

Martín Bauer\*\*

David Maldavsky: El Dr. Martín Bauer dirige en LSE un seminario de metodología de la investigación cualitativa avanzada en psicología social, y ha escrito varios libros. Uno de ellos (Bauer MW & G Gaskell [2000] (eds) *Qualitative researching with text, image and sound: a practical handbook*, London, Sage), escrito con otros autores, está dedicado a la investigación cualitativa; es un texto de consulta, en el cual se advierte un esfuerzo por sintetizar muchos de los hallazgos recientes en el tema. Cuando expuse en su seminario mi método, sus comentarios me resultaron muy útiles. Martín va a exponer sobre las diferentes estrategias de empleo de la computadora en análisis del discurso.

Martín Bauer: Para nosotros (David y yo), que nos interesamos en el análisis de texto asistido por computadora, resulta evidente que la disciplina está floreciendo en cuanto a cantidades de técnicas. Si ustedes se conectasen, como lo hago yo, en un grupo de discusión por mail, se darían cuenta de que hay un anuncio cada semana de algún nuevo software. Es como estar en el mercado de la verdura: hay uno gritando tomates, otro también, otro dice: mis tomates son mejores, y hay que tener un buen ojo para ver cuál es el valor de cada uno. Una cosa es la perspectiva del marketing del producto, y otra la del consumidor, quien necesita tener una idea de lo que hay en el mercado y comparar las cosas. Yo les voy a presentar ahora algunas nuevas técnicas.

Comenzaré explorando brevemente la metodología cualitativa. Se da una controversia entre investigación cualitativa y cuantitativa. Como se suponía que la investigación cuantitativa es la investigación por excelencia, la investigación cualitativa tenía que venderse de alguna manera. Hay tradicionalmente una polémica entre los investigadores cuantitativos y los cualitativos. Los cuantitativos dicen que los cualitativos cuentan anécdotas y que presentan pobres evidencias en sus investigaciones. Y los cualitativos dicen que los cuantitativos usan un positivismo obsoleto y que están a la saga de los tiempos. En nuestro libro tratamos de analizar estas posiciones y de demostrar que la polémica no lleva a nada y que hay otras cosas por considerar. Siguiendo la típica tradición inglesa, estamos tratando de evitar los oscurecimientos epistemológicos, teóricos, y nos concentramos en los aspectos pragmáticos: cómo

---

\*\* London School of Economics  
E-mail: M.Bauer@lse.ac.uk

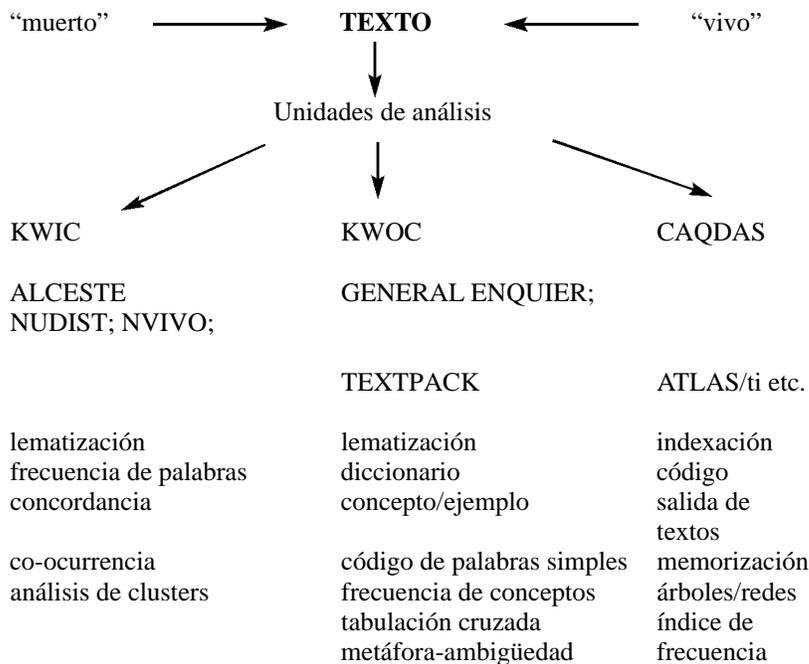


hacer las cosas bien. Desde esta perspectiva pragmática tenemos que concentrarnos en qué es lo que hay que hacer, en las indicaciones, cómo hay que juntar material, si hay que hacer encuestas, entrevistas. Hay una lista de cosas por hacer. En el libro van a encontrar todo esto. Tenemos una clase de actividades que implican técnicas cualitativas. La cuestión es determinar qué actividad sirve para qué propósito. Entonces podemos hacer un análisis de texto, un análisis de imagen, u otros tipos de análisis. La cuestión es determinar para qué lo vamos a hacer, el propósito, y después cómo lo vamos a hacer. Un paso ulterior podría ser determinar cómo hacer un buen análisis del discurso o un buen análisis de entrevista narrativa. El problema es evaluar los criterios. Entonces tenemos los indicadores de calidad de la investigación. El primer problema con el cual tropezamos es que los criterios tradicionales para la investigación cuantitativa: confiabilidad, validez y repetibilidad, no son funcionales, no son válidos para la investigación cualitativa. Así que buscamos criterios alternativos. Creo que en este sentido la cuestión básica es que la investigación cualitativa tiene que ser transparente y responsable en cuanto al público. La investigación tiene que ser cuestionable y tiene que comprobarse ante el público que la recibe, es decir otros investigadores. En este sentido, comparo este criterio con el concepto de transparencia y responsabilidad pública en otros ámbitos. Nuestros criterios de evaluación difieren de los criterios tradicionales de confiabilidad, validez y repetibilidad.

En este sentido, la computadora saca a la técnica de análisis de los datos de la dependencia de un intérprete genial y la sustituye por un nivel más transparente para el público que recibe esta información. Su actividad es de alguna manera estandarizable, la máquina no es una persona que hace tales operaciones. Así, tenemos en la computadora un documento que puede ser cualquier tipo de texto, puede ser una poesía, puede ser un artículo del diario, puede ser cualquier cosa que esté expresada en palabras. En cuanto a la relación entre el texto y su intérprete, el resultado es otro texto. La cuestión es cómo se produce esto y si no se analiza la transparencia de este proceso, tradicionalmente se da una relación de autoridad. Ciertos intérpretes tienen permiso de decir las cosas correctas y de aquí surge una ideología. En cambio, en ciencia tenemos propósitos más mundanos, tenemos la computadora y tenemos otros obstáculos. Con la computadora uno no está solo, tiene al diseñador del software, que de alguna manera transmite ciertas ideas sobre cómo organizar las cosas, y el resultado es un texto, pero con orígenes más complejos. Sabemos que pasar de un texto a otro es un terreno resbaladizo, porque hay que introducir, para que esto no sea un "vale todo", ciertas restricciones. Estas reglamentan el paso del texto original al texto interpretado. El problema en este momento es que, como están saliendo muchos paquetes de computación nuevos, si uno no es un profesional especializado, pierde de vista la diferencia entre lo que es bueno y lo que no lo es. Les voy a mostrar tres tipos de paquetes de computación que representan tres diferentes filosofías.



**Gráfico 1**  
**Tres tradiciones en programas de análisis de textos**



Esto es lo que pasaba antes de la computadora (Gráfico 2). ¿Qué hacíamos con los textos para entenderlos? Los alumnos aprenden esto (1, en Gráfico 2) en primer año como lo básico. El otro polo (2, en Gráfico 2) es la interpretación de textos por una elite. Creo que la mayoría de los paquetes de computación tratan de emular estas dos vertientes.

**Gráfico 2**  
**Qué hacemos tradicionalmente con los textos:**

- 1)
  - clasificarlos e indexarlos (biblioteca)
  - marcarlos, subrayarlos
  - tomar notas
  - anotaciones en los márgenes
  - resumirlos en tarjetas
  - cortar y pegar partes de textos
  - archivar tarjetas en cierto orden



2)

- contar palabras
- caracterizar el vocabulario
- sinopsis (Biblia)
- concordancia (Biblia)
- índice

Primero uno hace una tarjetita y después la clasifica. Lo principal en esto es poner una etiqueta para poder encontrar de vuelta el material. El próximo paso es marcar y subrayar, es decir, reducimos el texto a las cosas importantes y las secundarias. Todos sabemos que se usan diferentes colores o métodos para distinguir los niveles de importancia o las temáticas. Otro método es tomar apuntes y hacer un resumen. También están las notas al margen, los bibliotecarios odian esto, a menos que haya sido una persona famosa quien lo hizo, porque si es así realizan una exposición con los libros indebidamente marcados por aquella persona. En la universidad de Berna hicieron una exposición con las notas al margen de Lenin; supongo que habrán elegido todos los libros que él sacó de la biblioteca y concluyeron que las notas eran de él. Otra técnica es cortar y pegar, otra es archivar las tarjetas. Además, pueden contar las palabras, o diferentes tipos de palabras. Por ejemplo, en un seminario sobre la teoría marxista-leninista estaban discutiendo el concepto de dictadura del proletariado, y un oyente dijo: no se preocupen tanto por eso, ya que en los veinticuatro tomos Marx se refiere al tema sólo seis veces. Así que puede realizarse un estudio sobre la significancia. Se puede caracterizar el vocabulario en cuanto a las palabras diferentes que posea. Se pueden organizar sinopsis, diferentes versiones de la misma historia que se colocan en columnas comparativas. En la técnica de la concordancia, que se usa en la Biblia, se toma una palabra y se buscan todos los contextos en los que aparece y se los compara. Esto es para recordarles que no hay nada de misterio en la cuestión de la metodología de la organización del material. En esta perspectiva entra la computadora. Primero nos ayuda a construir los datos, por ejemplo el escaneo del material. El escaneo es laborioso, pero está cada vez mejor hecho y resulta más barato. También se puede encontrar el material en formato de archivos computarizados. Ahora hay enormes bases de datos públicos que están emergiendo, no sé si ustedes conocen Lexis y Nexis. Creo que Lexis es para los abogados, es una colección de leyes. Pero ahora hay también diarios on line. En Gran Bretaña consigo todos los diarios desde los '80 en forma computarizada, así que si quiero analizar un tema en particular, puedo buscar en el Daily Telegraph y me dan todos los artículos que han aparecido desde los '80, y voy a la base central, no al diario en sí, y puedo comparar textos. Otros materiales son transcripciones de grabaciones, entrevistas, conferencias.

Los paquetes de computación tienen estos tipos de procesos: codificación, indexación, lematización, análisis de redes, que es una graficación de interconexiones e interrelaciones, y análisis que son más tradicionalmente estadísticos: frecuencias y aso-



ciaciones. Cada paquete reproduce diferentes tipos de graficaciones o representaciones gráficas de contenidos, y en este punto es donde hay mayores progresos. Este es el núcleo de mi conferencia, es lo que yo llamaría la geografía o topología del análisis computarizado. Estos son los diferentes tipos de progresos que se produjeron en cuanto a análisis de textos. Podemos distinguir entre tres alternativas y tipos de programas. Uno se llama “palabra clave en contexto” (KWIC: key word in context), el otro “palabra clave fuera de contexto” (KWOC: key word out of context), y el tercero, análisis de datos cualitativos asistidos por computadora (CAQDAS: computer assisted qualitative data analysis software). Estas son las marcas registradas de los paquetes de computación. Alceste tiene una versión anterior que se llama Spat. Entre los KWOC se encuentra el programa del ADL (Algoritmo David Liberman). Voy a referirme a algunas características de estos paquetes. La cuestión es cuánto les puedo decir y si lo podemos llegar a discutir.

El KWOC es un proyecto histórico, que creó mucho entusiasmo en los '60. Un grupo de gente se juntó para tratar de resolver cómo la computadora puede eliminar el problema de interpretación, cómo hacerla confiable eliminando básicamente al intérprete. El problema ha sido la confiabilidad, porque los codificadores no pueden ponerse de acuerdo entre sí. Voy a ser un poco polémico, si me lo permiten. Había una esperanza de llegar al verdadero texto eliminando el problema de la confiabilidad. El diccionario tiene una estructura básica de conceptos y ejemplos, está formado por muchos conceptos, y cada concepto tiene varios ejemplos. El problema es reducir las palabras de un texto a una cantidad de conceptos, cada uno de los cuales tiene ejemplos. Todo texto tiene una distribución: algunas palabras tienen una incidencia muy alta y otras aparecen muy pocas veces. La idea es contraer esto para tener pocos conceptos con una distribución más interesante. Es una especie de truco mágico, pero puede ayudar al proceso de interpretación. Un problema es que el idioma está constituido por palabras con diferentes tipos de significados. Gramaticalmente, cada verbo tiene una versión en pasado, en futuro, en condicional, etc.; y los sustantivos pueden ser plurales, singulares, masculinos, femeninos. La lematización los reduce a su núcleo, se deshace de este problema. Se empieza con una primera reducción del material con la lematización, todas las variantes gramaticales se reducen a lo mismo. A los lingüistas les encanta hacer este tipo de paquete, y no hace falta que lo construya uno mismo como investigador, sino que se usa lo que hacen los lingüistas. Esto es sencillo para los idiomas latinos, pero no para el alemán, y creo que para el finlandés es aún más complejo. Porque el alemán es un idioma aglutinativo, una palabra tiene dos significados incorporados, aglutinados, pegados, y la computadora no puede identificar esas dos palabras contenidas en un único término. Así que si uno quiere usar una “palabra clave fuera de contexto”, un problema es qué tipo de idioma se está usando. Además, se hacen diccionarios para un propósito en particular o para un propósito en general. Por ejemplo, el Algoritmo David Liberman (ADL) es un diccionario español, y el principio guía son los conceptos psicoanalíticos. Un dicciona-



rio muy alabado es el diccionario de valores de Harvard, que está basado en conceptos de valores, morales o éticos. ¿Y qué se logra con este análisis de frecuencias y tabulaciones?. El diccionario tiene los conceptos y ustedes recorren el texto y consiguen distribuciones de esos conceptos. Entonces se encuentra la frecuencia relativa de cada concepto. Y en este marco se pueden hacer muchos tipos de comparaciones, porque se puede aplicar a diferentes textos. Por ejemplo, pueden estudiarse diferentes diarios y comparar la frecuencia de conceptos en dos de ellos. También pueden estudiarse diarios personales, por ejemplo los conceptos prevalecientes en cada persona. Entonces se puede hacer una tabulación cruzada, con una frecuencia doble en vez de simple. Después se realiza un análisis estadístico. Con posterioridad a esta época inicial, se desinfló el entusiasmo por este tipo de análisis, pero en este momento hay un revival, está volviendo a popularizarse porque existen muchos textos disponibles en computadora, y esto despertó el interés por cualquier tipo de análisis. El cuello de botella no aparece ya en el material (en los '60 imaginense que había que tipear primero los textos), sino en el análisis en sí. El problema que quitó entusiasmo a esta metodología es que el ejemplo siempre tiene que corresponder a un solo concepto. Si un ejemplo corresponde a concepto 2 no puede corresponder también a concepto 1. Esto va en contra de un principio lingüístico, que es que las palabras tienen múltiples significados. La computadora no identifica las palabras que tienen las mismas letras y en el mismo orden, y que se puede colocar bajo concepto 2 o concepto 1; hay que colocarlas en un solo lugar. Además, el contexto puede prestar ambigüedad a la palabra, por ejemplo, un verbo puede tener diferente significado. Además, está el uso retórico del lenguaje. Una metáfora es una palabra que está ligada a un concepto y uno adrede la coloca en la esfera de otro concepto.

Un primer paquete, el *General Enquirer*, es para uso macro y otro, el *Textpack*, es para PC. El Algoritmo David Liberman (ADL) está disponible para PC. El problema de esta serie de métodos es que va en contra de los principios lingüísticos. El progreso está surgiendo en los trabajos con el problema de la ambigüedad, mucha gente está trabajando en esto. También se ha aceptado que, si es posible clasificar el noventa por ciento de las palabras, esto es suficiente. El programa tipo "palabra clave fuera de contexto" pretende identificar palabras singulares, apilándolas para formar frecuencias, y al comparar estas frecuencias se pueden comparar los discursos o diferentes distribuciones de frecuencias, lo cual constituye una entrada al análisis muy fértil.

Creo que el SPSS ha introducido un nuevo módulo que se llama "texto inteligente" (*smart text*), que permite construir un diccionario, aplicado principalmente a las preguntas abiertas en las encuestas. David, ¿qué opinas de mis comentarios?.

David Maldavsky: La descripción que hiciste me parece correcta, los problemas metodológicos que planteas me parecen muy pertinentes, especialmente el de que cada palabra tiene un único valor, y no se la puede conectar con dos o tres conceptos. Eso



llevó a pensar que una palabra = 1, siempre vale 1 para la computadora. Nosotros hemos intentado resolver ese problema por otro camino. La cuestión de la ambigüedad y de la metáfora es una de las que más nos interesa dentro de nuestro terreno, el psicoanálisis. Nuestros avances apuntan en esa dirección, a resolver los problemas de la ambigüedad dando a cada palabra más de un valor, o sea la ubicamos en más de un concepto, con más de una significación erógena. Además, nuestro programa tiene dos alternativas de uso: por un lado podemos trabajar en forma automática y en ese caso es KWOC, pero también podemos emplearlo en forma interactiva. Entonces el que está trabajando con él tiene la posibilidad de decidir entre varias significaciones de la misma palabra. El programa le pregunta al usuario qué valor le da a un término entre varias alternativas que le propone, y el usuario toma la decisión considerando a cada palabra en su contexto.

Martín Bauer: Ya veo. Has hecho un desarrollo en términos de interactividad. Bien, ahora hablaré acerca de KWIC. Voy a hablarles de Alceste porque es el que conozco. Es un desarrollo francés, el estadístico que lo creó es un psicoanalista, el nombre del programa es de origen lacaniano. Este hombre está interesado en el problema de la asociación, cree que la esencia del significado es la asociación entre ejemplos semánticos. Sus primeros análisis fueron de poesía de los simbolistas del siglo XIX, donde lo onírico es significativo. ¿Qué hace el Alceste? Empieza con la lematización, se desprende del bagaje lingüístico, reduce la frecuencia de las palabras, quita las variantes gramaticales. Agregando un poco de dinero, entregan el instrumento con el programa de lematización para todos los idiomas latinos. Se inserta el texto, lo procesa y sale lematizado. Algunos colegas en un proyecto de biotecnología lo usaron para las entrevistas, y los alemanes construyeron, como líderes del proyecto, su propia lematización, y los finlandeses ni siquiera participaron en eso, se autoexcluyeron porque era imposible lematizar el finlandés. Así que este paquete de software también es una aparato para excluir. Después de la lematización, el vocabulario tiene otro aspecto: corta los extremos y mantiene las palabras de frecuencia media. Es en esta franja en la que uno trabaja. Estas palabras tendrán una distribución casi igual. La sugerencia lacaniana de que hay que buscar las asociaciones se traduce en un análisis estadístico de co-ocurrencia. Por ejemplo, tenemos las palabras "a", "b", y "c", el algoritmo chequea cuántas veces la palabra "a" está cerca de "b". La asociación es puramente espacial, por proximidad. En el texto el programa tiene una unidad de asociación, se puede especificar la cantidad de líneas que abarca este parámetro. Se le pregunta a la computadora si la palabra "a" está cerca de "b" en una línea, en dos, en tres, o en cuatro. Por supuesto, la probabilidad de co-ocurrencia es mayor cuanto más líneas se incluyen. Primero se analizan las ocurrencias nucleares y después se hace un análisis de concordancias comparando a aquellas. Algunas palabras pueden estar en el medio, lo cual significa que pueden ocurrir en cualquier lado. Lo interesante del texto es que se marca, se le da un nombre, se lo caracteriza, y el resultado es que se identifican los textos y las palabras que son más frecuentes en cada uno. Podemos



conjeturar que quizá el autor de Alceste analizó los textos de Nerval y encontró que en el primer período, en el segundo y en el tercero, predominaron diferentes simbolismos. La computadora da las agrupaciones y uno interpreta que cada agrupación se dio en un diferente período. Esto puede referirse a diferentes personas, diferentes tipos de textos, o diferentes tipos de diarios. Yo lo usé en este último tipo de investigación. Es una cuestión de suerte empírica que la computadora pueda agrupar claramente diferencias semánticas y que esto se advierta. Una vez mis colegas me mandaron algunos documentos y accidentalmente los metí en este software de análisis y conseguí cinco agrupaciones diferentes, estaba anonadado, era un cuadro divino. Si me hubiera fijado mejor, hubiera visto que la primera agrupación era texto en inglés, y las otras no. Eran cinco idiomas diferentes en total, y estaban bien agrupados, así que funcionó. Puede haber dos razones por las que obtuve un buen cuadro: primero, porque la estructura semántica del texto era así, y segundo, porque uno puede hacer malabarismos con esta orden en cuanto a las unidades de texto. Se puede testear la robustez del instrumento cambiando las unidades. Si se toma una línea, cinco, diez, veinte, y se obtienen diferentes cuadros, esto es una indicación de la robustez del análisis. Alceste tiene otro rasgo interesante y útil, que es la función de concordancia. Se puede usar la concordancia para una palabra, y la computadora dará todos los contextos en los que ésta aparece, o se la puede usar para el análisis de agrupaciones. El análisis de concordancia con los contextos diferentes de la palabra singular es un análisis muy sutil de los diferentes significados de un único término.

Les voy a hablar ahora de los últimos y más recientes paquetes CAQDAS (NU-DIST, ATLAS/ti, etc.). Tales paquetes permiten ofrecer opciones: o se puede hacer una descripción cualitativa del texto, o se puede usar la descripción y el análisis cuantitativo. El resultado es un clásico análisis de contenido con codificación, como siempre. Estos softwares tienen una función de indexación. Tenemos uno, dos, o tres textos, se identifica cierta parte de cada texto como correspondiendo a algo en particular. Lo interesante es que se trabaja con la codificación del texto, y se puede buscar, a partir del código otro texto con esa misma clasificación. Después se pueden crear jerarquías de temáticas. Por ejemplo, los textos pueden ser entrevistas, y uno de los temas puede ser familia (código 1, familia); después uno descubre que se dicen diferentes cosas en relación con la familia, y entonces esos ítems son código 1.1, código 1.2, y esto puede seguir hasta donde se necesite seguir. Tenemos así una función "memo", es una especie de diario personal sobre el proceso de la interpretación, el proceso de pensamiento acerca de lo que se está haciendo. Se puede abrir un memo en cualquier momento y decir: este código corresponde a este texto, o viceversa. ¿Por qué uno piensa que es así? En el curso de un año de trabajar con estos textos uno puede ver que hay una producción de textos del propio pensamiento acerca del trabajo, y esto es expresión del trabajo en progreso. En otras palabras, se puede documentar cómo el pensamiento de uno ha cambiado a través del proceso.



Estos paquetes trabajan con la relación entre código y texto. La utilidad de esto es que siempre se puede volver a encontrar el texto que se clasificó, se puede recuperar el texto donde se hizo esta clasificación. Entonces se pueden crear varios grupos de archivos para averiguar de qué habla la gente cuando habla de familia o de ese subtema dentro del tema familia. Y siempre se sabe de quién, de qué persona surge este texto. Así que, al profundizar la investigación de la indexación, uno puede descubrir qué tienen en común las personas que hablan de estos temas, si hay diferentes textos de la misma persona a través de un tiempo determinado, etc. Nunca se pierde esta conexión entre texto y código. Hablar de código implica que se pierde interés en la conexión con el texto original y sólo tiene uno interés en la frecuencia de ese código o índice. Así que al tener estos códigos, siempre existe la posibilidad de contar con un resultado de la frecuencia de estos códigos. Entonces se pueden comparar las frecuencias de códigos en dos personas, o en dos grupos de personas, o es posible quedarse con la representación gráfica del contenido. Se puede identificar que una persona en particular opera básicamente de una manera, y no de otra, se puede disponer de una representación gráfica de un material de texto. Esto es lo que hacen estos software. Un software permite identificar lo típico de un código. Por ejemplo, en dos textos que se refieren al mismo código uno puede tener una referencia típica y el otro, una marginal. Entonces uno le da un peso: el típico sería 1, y el otro 0,6. Entonces, cuando uno está luchando para escribir un informe, uno puede hacer que la computadora le busque sólo los códigos típicos o de peso.

Respecto de los tipos de programas, durante mucho tiempo NUDIST y ATLAS fueron grandes competidores. Un colega caracterizó al paquete NUDIST como un paquete protestante y al paquete ATLAS como católico. En el NUDIST no se puede hacer una referencia al código sin referirse al texto. ATLAS permite especular sobre cosas posibles en el mundo de los códigos, sobre cosas que no existen pero que son posibles, y se puede volver al texto a la búsqueda de esas cosas. La idea es que hay otras fuentes aparte de las Sagradas Escrituras para las descripciones, pero esto es sólo un chiste.

Estas son las tres filosofías de los tres tipos de paquetes de software, y me parece que es útil tratar de organizar toda la proliferación de programas que se ofrecen para ver en cuál de estas tres categorías los ubicamos. Es importante saber qué ofrecen estos paquetes en términos de análisis de texto, y la importancia que tienen es que aumentan la transparencia del análisis cualitativo. Esta es la contribución al proceso de investigación. Quisiera que ustedes me pregunten sobre lo que acabo de exponer.

Público 1: Pensaba en algo absurdo: la posibilidad de integrar la recepción del texto con el análisis del texto mismo, en analizar también el texto producido.

Martín Bauer: Habitualmente el investigador es el público, pero se podría tener un



texto y después un público que habla acerca de ese texto, lo cual se utilizaría como datos para hacer después el análisis. Creo que nunca se hizo un análisis así hasta ahora, pero es una idea interesante. Se han analizado las reacciones de un auditorio, pero no creo que se hayan usado estos análisis. Si se considera al texto como estímulo, y se clasifican las reacciones de las personas a ese texto haciendo las asociaciones, después puede hacerse un análisis de las agrupaciones para ver las características de las reacciones a ese texto.

Público 2: ¿Cómo se podría investigar la producción de un tema cuando no se nombra la palabra? Por ejemplo, un manipulador induce a que la otra persona haga o diga cosas sin decirlas.

Martín Bauer: Este caso requiere de otros datos aparte del texto. Siempre tomo el camino católico, me gusta saber lo que está ausente. Esto puede ser útil en este caso, es muy interesante lo que falta de un discurso.

Público 3: Su presentación me parece muy interesante y abre una perspectiva para pensar muchas cosas. Pensaba que un texto incluye palabras, silencios, pensamientos, afectos, apela al otro, produce un efecto en el otro. Un texto es riquísimo. Esto nos da una perspectiva para apoyarnos en algunos de estos puntos. Lo que no sé es en qué lugar está puesto el que escucha en esta investigación. Amí me surgieron dos recuerdos. Uno es la película *2001, Odisea del espacio*: la computadora fue programada con un secreto; el otro recuerdo es que pocos años atrás, en un viaje espacial que salió de Estados Unidos, el programa que se usaba había sido hecho en Inglaterra, con un sistema de medidas que no era el que estaban usando en Estados Unidos. Los científicos no se dieron cuenta de esto, lo cual dio lugar a un resultado totalmente absurdo.

Martín Bauer: Déjeme decirle primero una cosa. No consideren estos softwares como la solución. Hay cuestiones, rasgos de la comunicación que usted puede perder de vista completamente. Siempre empiezo hablando de la reducción y el aumento de la complejidad. La mayoría de estos paquetes que les describí tienen la idea de embudo, de reducir una gran cantidad de material en poco, pero la complejidad del lenguaje implica que una cosita puede significar cientos de cosas. Así que estas reducciones me parecen que son bastantes elementales. Yo no trabajo con el psicoanálisis, pero sé que hay un montón de gente que trabaja con estos parámetros para tratar de analizar el proceso psicoanalítico. Mergenthaler, por ejemplo, está trabajando con el elemento afectivo, pero tiene que usar otros enfoques, por ejemplo está usando sufijos, conectores, medidas automáticas, etc., para ver cómo cambian a través de un proceso analítico y crear indicadores de mejoría.

David Maldavsky: Quería hacer algunos comentarios. Se pueden combinar métodos. Por ejemplo, Lahlou recurrió al Alceste para estudiar la representación social de la co-



mida. Comparo una serie de respuestas a preguntas que les hizo a un gran número de personas sobre qué significa una buena comida. Los entrevistados tenían que dar cinco palabras como respuesta. Entonces comparó los resultados de esas respuestas con el Petit Robert, realizó agrupamientos, *clusters*. Lahlou tiene formación psicoanalítica, así que estableció algunas relaciones entre sus descripciones y la teoría de la libido. Sobre el resultado de la aplicación de su métodos, a mi vez, yo apliqué el mío. Respecto de la investigación sobre las relaciones entre uno que habla y otro que responde, deseo destacar que nosotros hicimos ese tipo de estudio, entre lo que dice el terapeuta y lo que dice el paciente. En psicoanálisis, Liberman afirma que existen respuestas más pertinentes del terapeuta, según cuál sea el estilo del paciente. Nosotros hicimos ese tipo de análisis, comparando las relaciones entre terapeuta y paciente. Por supuesto, existen muchas otras alternativas para la investigación, y me interesa que ustedes se sensibilicen respecto de ello. Las observaciones que hicieron acerca de las limitaciones del programa son atinadas; por ejemplo, la computadora no logra diferenciar entre la frase “quisiera no comer tanto, pero...” y la frase “pero no quisiera comer tanto”. Entonces, como lo planteaba Martín, es necesario combinar el análisis computacional con otros, por ejemplo un análisis de las frases y del relato. Aun así, el análisis computarizado tiene muchas utilidades, según el marco global en que se lo emplea.

Público 4: Dos preguntas: ¿Se pueden bajar alguno de estos tipos de programas por Internet? ¿Puede hacer algún comentario sobre la predictibilidad del desarrollo de este área, teniendo en cuenta lo que sucede hoy en día con la tecnología?

Martín Bauer: Si tengo que predecir el futuro, diría que el SPSS se va a tragar en diez años a los demás programas, ya que es una combinación de análisis, de factores, etc., Todos los tipos de análisis están incluidos (KWIC, KWOC, CAQDAS). La cuestión es quién lo va a hacer. Estábamos preocupados por NUDIST, que sale de Australia e hizo muy buen marketing. A veces ocurre que el software no es tan bueno, pero gana mercado por el marketing. Yo tengo que esforzarme para que se use el ATLAS, porque el NUDIST se hizo una especie de idioma o lenguaje y los usuarios no quieren aceptar otro. La producción de softwares es una industria casera por ahora. Mi interés no es vender ninguno, sino mantener una idea clara de lo que hace cada uno y elegir el más apropiado. Quisiera que el Algoritmo David Liberman (ADL) estuviera disponible para poder compararlo con otros instrumentos.

Público 5: En el último ejemplo que dieron sobre la obtención de racimos o *clusters*, ¿qué hallazgos obtuvieron, por ejemplo, en el caso de la comida? ¿Con qué asociaron los *clusters* en el nivel de otras variables culturales, demográficas, estilo de pensamiento?

Martín Bauer: El soft hace eso. Si usted marca el texto, por ejemplo, en el caso de la comida, Lahlou pudo identificar cuál grupo social está identificado con cuál racimo,



o cuál combinatoria de racimos. Se podría repetir el análisis: un estudiante mío lo hizo en Colombia, y se puede comparar la estructura. Se encuentra lo universal en este discurso sobre la comida y, además, se destacan variantes culturales. Lahlou se fijó en el diccionario, en el Petit Robert. Los diccionarios tienen la característica de que salen nuevas ediciones cada tanto, así que son depósitos de desarrollo cultural. No sé si lo va a hacer, pero podría volver a una edición antigua del diccionario, por ejemplo de 1890, y comparar con el diccionario actual; se pueden comparar dos análisis de racimo y ver dónde son diferentes. La metodología permite comparar diferentes rasgos, pero la decisión en cuanto a qué comparar está fuera de la metodología, es anterior, conceptual.

David Maldavsky: Creo que la pregunta tiene que ver con el aprovechamiento de los datos. Es diferente cuando uno trabaja con el Alceste, que es un sistema a-categorial, no tiene categorías previas y reúne lo que aparece, y cuando recurre a un diccionario. Los diccionarios necesitan de una discusión categorial previa acerca de los conceptos que uno cree importantes, con los cuales va a armar los conjuntos de palabras contenidos en la base de datos, y a partir de ahí va a estudiar el texto. Los diccionarios funcionan así, hay una serie de conceptos, y hay términos que representan a esos conceptos. Entonces uno se puede preguntar cómo se crean los conceptos, que son aquello que va a buscar en un texto determinado. Por ejemplo, Mergenthaler se apoya en ciertos argumentos para sostener que lo fundamental es detectar estados afectivos, abstracción, etc. Trata de detectar eso en un texto, y crea diccionarios para ello. Yo, que vengo del psicoanálisis, parto de otras preguntas, que definen a los conceptos con los cuales voy a armar las bases de datos. Yo me pregunto por las erogeneidades. Martín llama a esto el pansexualismo de mi teoría. Ese es mi sistema categorial, que tiene importancia cuando uno piensa en términos de un sistema hipotético deductivo. El Alceste, en cambio, no tiene ninguna hipótesis previa con la cual va a la búsqueda de los datos. Sin embargo, ambos instrumentos se pueden combinar. Se abren varias alternativas en la investigación concreta, y pueden alcanzarse resultados muy económicos en tiempo y esfuerzo.

Martín Bauer: Lo que muchas veces se siente como la mística de un software es en realidad la lenta progresión de encuentro de información o datos. Por ejemplo, se puede testear metodológicamente cuál es el software más eficiente, porque el tiempo es dinero en estas cosas. Esto se ha hecho con las encuestas, se pueden pasar al NUDIST y ver en mil preguntas cuál es el software que las procesa más rápidamente. Por ejemplo, teníamos una pregunta: ¿con qué asocia la palabra ingeniería genética? El resultado fue muy interesante, estudiamos con NUDIST y Alceste, y este último llegó a las mismas conclusiones que los contestadores o lectores. Si saben usar bien el software lo pueden hacer en dos o tres horas, pero con el NUDIST para este proceso se puede necesitar tres días. Por ejemplo, la concordancia es el trabajo de toda una vida, y en la computadora tarda tres segundos.