

NOTAS

LA COMPUTADORA EN LAS INVESTIGACIONES HUMANÍSTICAS

En las humanidades hay dos posibles enfoques de trabajo: el sintético y el analítico. Al sintetizar un punto de vista filosófico, al llegar a una conclusión política, o al hacer una aportación a la literatura, el investigador selecciona los elementos intelectuales que están dentro de su campo de visión, impone sobre ellos algún nuevo esquema de organización, quizá ligando otros elementos a ellos por primera vez, y hace surgir así una totalidad nueva. Pero esta actividad intelectual está indeleblemente entrelazada con el segundo enfoque, el analítico. Antes de hacer una síntesis, hay que aislar los elementos verdaderamente importantes, examinar sus propiedades, y averiguar que lo que se sostiene acerca de ellos concuerda con la realidad. Antes de hablar, conviene entender bien de qué se habla.

Tanto en el análisis como en la síntesis, es el investigador quien decide cuáles son los objetos de su interés, y qué relaciones entre ellos quiere imponer o sugerir. Las consecuencias de su decisión pueden tomar simultáneamente dos trayectorias. En un aspecto, sobresale la personalidad del innovador; tanto en los esfuerzos artísticos como en los académicos, la visión de la persona encauza una expresión verbal o plástica cuyo valor depende en parte de su individualidad, y en parte de la disciplina a que ha sometido su plan normativo. En cambio, hay un aspecto en que la expresión de la decisión creadora (lo que en las ciencias suele llamarse hipótesis, pero que a fin de cuentas no es diferente de la visión del artista) obliga al investigador a reunir una colección de entidades, u organizar un sistema racional, o probar varias posibilidades lógicas de organización e interpretación. El crítico literario, por ejemplo, no debe aseverar que el efecto estilístico de la prosa de determinado escritor se debe a cierta característica rítmica de sus oraciones periódicas, sin haber examinado en detalle la obra de ese autor; tampoco debe el historiador llamar revolución de las masas a lo que es en realidad un simple cambio de guardia, por no haberse enterado de la evidencia.

EL TRATAMIENTO DE DATOS

En este sentido resulta indudable el valor que para las humanidades tiene el uso de los medios técnicos modernos adecuados para el tratamiento de la información. Las cuestiones de los objetos conceptuales que se manejan y de la forma en que han de ser manejados, quedan, como antes, bajo la responsabilidad del individuo que hace uso de ellos. Es sólo en lo que se refiere a su organización y manipulación racional, y a la facilidad con que el investigador tiene acceso a grandes cantidades de ellos, en lo que consiste la ventaja que hoy se nos presenta. Una vez que el investigador decida con qué datos quiere trabajar y qué quiere hacer con ellos, existen ya sistemas mecánicos más rápidos y exactos que los esclavos griegos de la antigüedad o que las secretarías de nuestra época, capaces de hacer una organización y presentación lógica de lo que se necesita.

Así considerados, los equipos electrónicos para el tratamiento de datos no pueden ya considerarse como una incursión del materialismo en tierra sagrada, sino como un método de realizar los trabajos que todos tenemos que hacer, sin necesidad de perder tiempo en actividades rutinarias y en repeticiones mecánicas. Para el humanista, la computadora es un almacén de información, no sólo en forma de un simple catálogo unilineal, sino más bien en una forma compleja pero que es de fácil acceso por medio de muchos criterios y combinaciones de criterios lógicos, y que tiene la capacidad adicional de poder, en el momento en que se necesita, localizar un dato o un grupo de datos, formar nuevas agrupaciones, aplicar pruebas y tomar decisiones sobre varias características, realizando condicionalmente las acciones consecuentes.

En la lingüística, por ejemplo, desde hace años se ha reconocido que la concordancia es una forma de presentación de datos muy útil para facilitar el examen del uso de las formas de un idioma con relación a su contexto. Sin embargo, como la compilación manual de concordancias es un trabajo muy laborioso —una concordancia paralela de los textos del Nuevo Testamento en griego e inglés consumió unas treinta mil horas de labor humana— la producción de concordancias se ha limitado a obras conocidas de la literatura y la teología: el *Poema de Mío Cid*, las obras de Shakespeare, la *Biblia*, y Santo Tomás de Aquino, entre otras.

En nuestra época, resulta absurdo gastar miles de horas en hacer manualmente una concordancia, por loable que sea su contenido

o su uso. Hoy se puede tomar la cinta de papel que instruye al equipo automático encargado de seleccionar y disponer los tipos de imprenta, y a base de ella (sin volver a copiar el texto) en cinco o veinte horas —de acuerdo con la clase de máquina computadora que se usa— sale una concordancia más completa y más exacta que una hecha a mano. Y, sin menospreciar el trabajo de los que no tenían esa ayuda mecánica, cabe observar que el valor de una concordancia, como el de muchas otras presentaciones de materiales informativos, no consiste en la actividad de producirla, sino en el uso a que se somete después de terminada.

Como ya es posible hacer concordancias no sólo de obras literarias bien conocidas, sino de cualquier texto lingüístico, la concordancia se ha convertido en un instrumento útil para todo tipo de investigación lingüística. Tiene importancia para la descripción de sistemas gramaticales, la estilística, la comparación de usos variantes (por ejemplo, entre un dialecto regional o social y otro, o entre un género literario y otro), el estudio de colocaciones léxicas y semánticas, y la lexicografía. En el Instituto Lingüístico de Verano, donde los investigadores están examinando la variedad de manifestaciones del lenguaje humano en 370 idiomas anteriormente desconocidos, hemos sometido diversas colecciones de narraciones y de literatura oral tradicional al tratamiento descrito, con el resultado de que el trabajo analítico, así como la elaboración de materias educativas y la traducción de libros de alto valor en aquellos idiomas, ha experimentado un gran adelanto.¹

La lingüística, por supuesto, no es la única ciencia humanística en que hay necesidad de hacer presentaciones especiales de cantidades elevadas de datos. Sin embargo, ilustra la manera en que las técnicas modernas se prestan para que el investigador no tenga que preocuparse por su manejo y organización, y pueda así concentrar sus esfuerzos en lo que debe concentrarlos: en la interpretación de los datos.

LA REDUCCIÓN DE LA INFORMACIÓN

Un aspecto de la investigación que es característico de las ciencias físicas y del comportamiento —aunque tiene lugar también en los estudios humanísticos— es la reducción de la información.

¹ JOSEPH E. CRIMES, reseña del libro de LAMB Y GOULD, *Concordances from computers*, en *International Journal of American Linguistics*, 31 (1965), pp. 178-181.

Tomemos el caso del historiador que, al analizar el ambiente cultural de nuestros días, menciona la preponderancia de la juventud en la sociedad actual. Podríamos preguntarnos si, al decir eso, basa su juicio en la observación superficial de las jóvenes que desfilan ante su mesa favorita de la cafetería, o en el hecho de que la noche antes de escribir tal cosa no pudiera dormir por el ruido de un *rock and roll* que le golpeaba los oídos, o si en realidad ha investigado el caso. Aunque todos reaccionáramos como él en las dos circunstancias primeras, como historiador le prestamos fe únicamente si se basa en algo más concreto. Su juicio ha de representar la reducción de mucha información censal, hecha de acuerdo con las normas del muestreo y la estimación numérica.

Cuando se trata de una reducción de información como ésta, que se nos ofrece en forma numérica, no es difícil advertir las ventajas de las técnicas de la computación. Es obvio que en la manipulación de números, la máquina comete menos errores que una persona, y que lo hace en una milésima parte del tiempo. Pero no toda la información que queremos reducir aparece siempre en forma numérica. Hay casos en que la información verbal permite una reducción a forma numérica; pero hay también información que puede reducirse sin perder su forma no numérica.

Mucho de lo que consideramos estilística se basa en la investigación de patrones de distribución de vocablos, de sílabas, de sonidos, con relación entre sí y con relación a agrupaciones lingüísticas mayores. Cuando podemos distinguir estas agrupaciones mayores (oraciones o líneas, por ejemplo) y las unidades cuya distribución nos interesa, es trabajo idóneo para una computadora el examinar vastas cantidades de textos para precisar cuáles son las distribuciones de los elementos y qué variación tienen. Muchas veces esto se hace en forma numérica; por ejemplo, en términos de frecuencias, o de la separación en el contexto entre una forma y otra, o de la posición del elemento dentro de agrupaciones. Los promedios y otras cifras relacionadas con las distribuciones constituyen una reducción a un compás pequeño de cantidades de información originalmente no numérica.²

La reducción de información no numérica a una forma también no numérica se ilustra con los diversos tipos de redacción que se aplican a los textos lingüísticos. El problema de la redacción es

² Un ejemplo del uso de la reducción a números de grandes cantidades de textos lingüísticos se ve en GUSTAV HERDAN, "Vocabulary statistics and phonology: a parallel", *Language*, 37 (1961), pp. 247-255.

el de transformar un documento en otro de sentido equivalente, bajo restricciones específicas de estilo, de idioma, del número de palabras en el producto, y otras. Si esto se hace en más o menos la misma escala y dentro del mismo idioma, se conoce como perifrasis o redacción ordinaria; si se hace en la misma escala (casi siempre un poco aumentada, por razones que no discuto ahora) y en otro idioma, es una traducción. Si reducimos la escala del producto, el resultado es una sinopsis, que debe contener toda la información esencial del original, pero que suele representar una reducción por un factor de diez o más en el número de palabras. Se practican reducciones más radicales en el caso de la catalogación bibliotecaria, en que el contenido de una obra tiene que ser representado por un número reducido de temas; hay otras formas de reducción de información que operan en la misma escala, pero que representan análisis más tallados de la estructura semántica.

La traducción de un texto de un idioma al mismo idioma —la perifrasis— se ha logrado hacer hasta cierto punto, principalmente con relación a la descomposición de preguntas y a la incorporación de los elementos contenidos en ellas, junto con elementos nuevos, a respuestas. La traducción entre distintos idiomas no ha dado semejantes resultados, porque las correspondencias sintácticas y semánticas entre dos idiomas son tan complejas, que los análisis suficientemente detallados para permitir la traducción están todavía en la etapa tentativa. Sin embargo, tanto por la perfección progresiva de los métodos ya en uso, como por la investigación de nuevos procesos de traducción, no debe desecharse la posibilidad de éxito en esta forma de la traducción. Conviene mencionar también el sistema de la traducción ayudada por máquinas; es decir, la presentación por medio de un sistema mecánico al traductor humano de toda la información sintáctica y semántica que ha sido almacenada bajo cada forma, para evitarle la pérdida de tiempo en búsquedas de información acerca de las formas del texto.

La producción automática de sinopsis se ha intentado también, ya por métodos como el de Luhn (seleccionando oraciones que destacan el peso de ciertas palabras importantes), ya por el de Harris (analizando la sintaxis de las oraciones, y tachando las partes dependientes), ya por otros métodos que toman en cuenta factores como la posición y presentación gráfica de varios elementos. La calidad de las sinopsis es mejor que la de las traducciones, pero aún queda mucho por hacer. La catalogación semántica de la in-

formación contenida en un texto será tratada como un aspecto de la recuperación de la información.³

LA SIMULACIÓN

La computadora es un mecanismo que obedece y ejecuta instrucciones. Por eso, si queremos examinar el efecto de una operación sobre las entidades que nos interesan, basta con especificar qué es la operación, qué forma tienen los datos, y cuáles son las contingencias que pueden modificar el tratamiento que hemos definido. De esta manera podemos formarnos una idea de las posibles consecuencias de la aplicación de diversos tipos de intervención, con una complejidad casi sin límite.

Por ejemplo, el cuento de hadas es un género literario mundialmente difundido y de arquitectura muy sencilla. Muchas investigaciones realizadas sobre el tema se han limitado a hacer la catalogación de los elementos del cuento; pero Propp sugiere que, además, se debe hablar de una estructura global del cuento, que se manifiesta en cada uno de ellos, y que es lo que permite el desarrollo de nuevos cuentos hechos según el mismo plan. He simulado la estructura que define Propp en una computadora, por medio de un proceso que selecciona elementos al azar y los coloca en su secuencia debida, con el resultado de que salen argumentos para cuentos jamás contados, pero que se reconocen como tales.

En gramática, la simulación hace posible la evaluación de una descripción gramatical. Con un esquema muy semejante al de la simulación folklórica, se trazan trayectos al azar entre proposiciones que describen los esquemas sintácticos (*patrones*) y elementos de un idioma. Si el resultado de este proceso es una serie de oraciones o textos que los hablantes del idioma califican como aceptables, la descripción fundamental tiene que ser buena. Así pues, la adecuación de un análisis gramatical se puede caracterizar como una función de la proporción de oraciones inaceptables entre los resultados de la simulación de una gramática.⁴

³L. B. DOYLE, "Expanding the editing function in language data processing", *Communications of the ACM*, 8 (1965), 238-243.

⁴VLADIMIR PROPP, *Morphology of the folktale* (Publication 10 of the Research Center in Anthropology, Folklore, and Linguistics), Bloomington, 1958. Traducido del ruso por Laurence Scott.—El programa que sintetiza cuentos lo preparé para la IBM 1620 en la Universidad de Oklahoma y el Instituto Politécnico Nacional. Varios programas para generar oraciones basadas en una gramática han sido producidos en el Instituto Tecnológico de Massachusetts.

Las máquinas computadoras se han empleado para simular procesos del aprendizaje en la psicología, sistemas económicos, percepciones claras y equivocadas, estrategias militares y de juegos, patrones diagnósticos en la medicina y otras cosas. La principal ventaja de la simulación lógica es que se puede obtener una idea aproximada (en el caso de la simulación, nunca exacta por causa de la limitación de factores que se practica) en muy poco tiempo, sin el desgaste de materias quizás irremplazables. Además, la simulación permite examinar una variedad de posibles procesos, con todas sus consecuencias, en número que quizá nunca podría ser considerado en la observación directa de la experiencia.

LA RECUPERACIÓN DE INFORMACIÓN

Imaginemos por un momento una biblioteca completamente mecanizada (cosa inalcanzable por ahora, pero que puede ayudarnos a entender la cuestión); de ella podíamos solicitar dos clases de información: datos y citas. Por ejemplo, si nos interesa saber cuántos elefantes tenía Aníbal al pasar los Alpes, lo que pedimos es un dato. En cambio, si queremos saber si alguien ha formado una opinión sobre la similitud entre los problemas logísticos que impulsaron a Aníbal a cruzar los Alpes y a Pizarro a llegar al Cuzco, lo que exigimos es una lista de citas bibliográficas, con o sin los textos o partes de textos correspondientes.

La diferencia principal entre los dos tipos de acceso a la información es que cuando pedimos un dato, la bibliotecaria mecánica debe poder contestar "no sé"; mientras que cuando pedimos citas bibliográficas, debe actuar a base de probabilidades: "estoy casi segura de que esto te interesa", o "me parece esto algo alejado del tema, pero conviene examinarlo". Además, la información almacenada con el fin de ofrecer datos, tiene que ser suficiente para distinguir, a base de preguntas, entre el objeto deseado y el resto del mundo; la información asociada con una cita, en cambio, se considera desde el punto de vista de la probabilidad de que la obra citada tenga algo que ver con el tema que se pide. Se sigue un proceso lógico para localizar respuestas correspondientes a las preguntas relativas a datos concretos, pero se usa un proceso aritmético en la búsqueda de citas, otorgando a cada documento de la colección una cifra que es indicio de la asociación existente entre el documento y el pedido de información.⁵

⁵ El acceso a datos se explica en R. F. SIMMONS, "Answering English questions by computer: a survey", *Communications of the ACM*, 8 (1965), 53-70.

La clasificación semántica que sirve para la recuperación de la información puede considerarse como una extensión en muchas dimensiones lógicas del mismo sistema de clasificación bibliotecaria que se usa hoy. En la biblioteca ordinaria se almacenan no solamente libros, sino diversas informaciones acerca de cada libro: su autor, su título, varios temas, y (en la colocación material) su relación con otras obras de contenido semejante. Aunque hay un solo orden, correspondiente a la disposición que los libros guardan en las repisas de la biblioteca, la clasificación lógica manejada por la computadora permite el uso simultáneo de varios tipos de ordenamiento relativo a diversos principios clasificatorios, y no está limitada al principio único que determina cuáles libros serán almacenados en la misma zona. Sin embargo, el resultado de una búsqueda lógica en el catálogo mecanizado puede expresarse como un número de clasificación que localiza la obra seleccionada con precisión en términos de la disposición material de la biblioteca.

Para permitir el acceso a la información bibliográfica con relación a los pedidos que se hagan, el contenido de una obra tiene que ser reducido al mínimo suficiente para permitir su identificación. En algunos sistemas el contenido se caracteriza por una serie de palabras y frases que se llaman descriptores, semejantes a los temas que usa el bibliotecario, pero más flexibles y organizados de acuerdo con varios esquemas independientes. En otros sistemas de recuperación de información, la obra se conceptualiza como un punto en un espacio semántico ordenado por dimensiones múltiples. Las operaciones de búsqueda se hacen sobre un vector que caracteriza el contenido de la obra de acuerdo con las escalas que corresponden a cada dimensión.

Algunos sistemas de caracterización del contenido de una obra dependen de que el autor o un lector profesional asigne a la obra una lista de descriptores, o la califique de acuerdo con algún esquema normativo. Son de mucho más interés, tanto práctico como teórico, los sistemas que aprovechan las peculiaridades internas del documento para elaborar su descripción característica. Por ejemplo, si se toma la frecuencia de ciertos términos técnicos como índice del contenido del documento, el vector resultante de frecuencias puede ser comparado con otro vector de la misma estructura que caracteriza otro documento, o que caracteriza el pedido de información que se ha hecho. Hay medidas de asociación relativa que pueden calcularse para cada par de vectores, que repre-

sentan la semejanza del contenido de uno y otro. El proceso de recuperación de la información entonces es así: El pedido de información se transforma en un vector análogo a los vectores característicos de los documentos de la colección, de tal modo que la importancia relativa de cada término representado en el vector queda indicada por un número. Se calcula el coeficiente de asociación entre el vector de pedido y los vectores característicos de la colección, y los documentos que muestren la más alta asociación con el contenido del pedido, serán los que el sistema sugiere al usuario para su consideración.

Este sencillo principio, según el cual es posible expresar numéricamente la semejanza de contenido entre un pedido de información y un documento, puede aun perfeccionarse. Si en lugar de emplear las frecuencias brutas de palabras seleccionadas, colocamos primero a cada palabra dentro de una categoría de un diccionario ideológico, y contamos la frecuencia con que se hace referencia a cada una de esas categorías, se aumenta la probabilidad de que el contenido de los documentos seleccionados por el sistema concuerde con las necesidades de la persona que hace uso del mismo. Además, pueden agregarse restricciones de tipo sintáctico a la relación semántica, para evitar, por ejemplo, que un pedido que hace referencia a "los antepasados españoles" sea tratado en los mismos términos que uno que menciona "el español de los antepasados".⁶

Cuando los sistemas automáticos de recuperación de la información sean de uso general, se conseguirá liberar a las personas interesadas en las encuestas académicas de la pérdida de tiempo que ahora implica la búsqueda de fuentes bibliográficas. Este tiempo podrá entonces invertirse en la lectura y la evaluación crítica de ellas. Como en otras áreas, la mecanización entrega las tareas puramente rutinarias a un equipo especializado en hacer trabajos rutinarios, reservando las tareas críticas y creadoras a las personas especializadas en ellas.

Otro resultado de los esfuerzos hechos para diseñar sistemas comprensivos de recuperación de información ha sido el de mostrar la necesidad de aclarar los conceptos actuales de la semántica. Es difícil, por ejemplo, precisar cuáles son los criterios que nos permiten decir que dos oraciones no difieren en cuanto a su sentido. Pero el desarrollo de sistemas mecanizados para el manejo de la

⁶ GERARD SALTON, "Data manipulation and programming problems in automatic information retrieval", *Communications of the ACM*, 9 (1966), 210-214.

información, ha dado un impulso a estas investigaciones, que puede ser de provecho para la filosofía en general.

PROBLEMAS

Los sistemas actuales de tratamiento de datos, aunque están altamente desarrollados en algunos aspectos, en otros sufren limitaciones que, creo, debo mencionar, no solamente para manifestar mi discrepancia con la idea popular de que las máquinas computadoras nos han hecho avanzar mil años, sino también para mostrar la posible dirección de su desarrollo en los próximos años.

La gran cantidad de datos que se usan en las investigaciones humanísticas requiere una preparación previa para que puedan ser asimilados por la máquina. Normalmente se perforan en tarjetas, a razón de hasta ochenta caracteres por tarjeta, proceso muy parecido al de pasar en máquina de escribir un manuscrito. Se usan también cintas perforadas de papel o de plástico, como, por ejemplo, cintas de teletipo. Todo este esfuerzo de preparación tiende a limitar la cantidad de datos que se someten al tratamiento mecánico. Sin embargo, de dos maneras se está buscando la solución del problema de la preparación de datos. Primero, muchos libros se preparan ya para la imprenta por medio de una máquina de teclas que produce una cinta especial de papel. Una vez corregida, esta cinta alimenta a otra máquina que automáticamente selecciona y dispone los tipos de imprenta. La misma cinta puede aprovecharse para la presentación de datos a una computadora. Existe ya un prototipo de lectora de cinta "monotipo", que transforma la información usada por la imprenta en una forma que la computadora acepta. Segundo, funcionan ya varios equipos ópticos que leen caracteres escritos a máquina o impresos. Aunque todavía no aceptan mucha variedad tipográfica, y aunque son muy sensibles a factores externos (como suciedad en el papel), dentro de poco se podrá meter directamente en la memoria de una computadora la información contenida en un manuscrito o en un libro.

Otra dificultad que proviene de la cantidad de datos que nuestras investigaciones nos obligan a manejar, es la limitación de la memoria de las computadoras actuales. Tanto la información que manipula la computadora como las instrucciones que especifican el tratamiento, tienen que estar presentes dentro de la memoria de la máquina en el momento en que se toman en consideración. La memoria las maneja en unidades lógicas de uno a ocho caracteres, según la máquina. La computadora de más capacidad que

actualmente está instalada en el país contiene sólo 32,768 palabras o unidades de información en su memoria; la capacidad de otros equipos va de la mitad a la décimosexta parte de esa cifra. Ni siquiera para los trabajos puramente numéricos es siempre suficiente; así, resulta necesario ampliarlo mediante unidades de memoria secundaria —como cintas magnéticas, discos magnéticos, o tambores magnéticos— en las que la información que no se necesita se graba para ser leída después. Los equipos de memoria secundaria difieren en cuanto a la rapidez de acceso a la información contenida en ellos. La que está dentro de la memoria principal se selecciona en un tiempo que se mide en millonésimos de segundo, mientras que la que está en la memoria secundaria tiene que localizarse y traerse a la memoria primaria antes de que sea asequible para la operación de las instrucciones. En los mejores (y más costosos) equipos secundarios, la información puede ser introducida en la memoria en unos milésimos de segundo; pero si se trata de información que está en el otro extremo del medio kilómetro de cinta magnética, a menudo tarda más de un minuto en aparecer; o sea, millones de veces más que la información que ya está dentro de la memoria. Considerando el enorme volumen de datos que suelen manejarse, y atendiendo a las partes del proceso que tienen que esperar mientras se obtiene la información, esto puede resultar económicamente desastroso. Resulta, pues, necesario perfeccionar los discos y tambores de almacenaje secundario, para poder seleccionar una unidad de información sin pérdidas costosas de tiempo. Además, con la fabricación en masa de elementos miniaturas de memoria, pronto será posible producir memorias primarias realmente gigantescas, en las que la mayor parte de la información que se necesita podrá estar presente al instante.

Para hacer uso de un equipo de computación electrónica ha sido necesario, en el pasado, o que el investigador estudiara detalladamente la programación de máquinas, o que se pusiera en comunicación con un experto en computación, cuyos antecedentes intelectuales diferían mucho, por lo general, de los del propio investigador. En ambos casos, el investigador tenía que emprender algo muy semejante al aprendizaje de un idioma extranjero: el de la máquina, o el del maquinista. Como el segundo es más fácil que el primero, aunque cualitativamente no sea distinto, el computólogo ha sido investido de cierto aspecto sacerdotal de intermediario entre el lego y la omnisciencia. Conviene aclarar una cosa y prometer otra para hacer menos aguda la diferencia entre los que

tienen acceso al uso de las computadoras y los que quieren tenerlo, aunque no entiendan el lenguaje de los misterios. La aclaración es ésta: el proceso de instruir a una computadora no es tan difícil de aprender como muchos creen. Requiere cierto sentido lógico y un apego a los detalles de un proceso, más bien que una preparación especializada en matemáticas. La promesa es ésta: hay un número creciente de lenguajes diseñados para facilitar la comunicación entre el hombre y la máquina, aunque el investigador no sea experto en máquinas. El intermediario entre el usuario y la computadora se coloca cada vez más dentro del equipo y no fuera de él. Aunque todavía no es posible instruir una computadora en lenguaje natural, como por ejemplo en español, los lenguajes formales especiales que están disponibles ya no son tan extraños a la forma de las lenguas humanas como antes. Además, ahora se pone mucho énfasis en los sistemas conversacionales: en el momento en que la computadora encuentra una dificultad para interpretar las instrucciones que se le presentan, esos sistemas pueden pedir una aclaración a la persona que les instruye y, por medio de un diálogo, van formando la definición exacta del proceso que es necesario para alcanzar el fin que procura el investigador.

Es la necesidad de definiciones exactas de los procesos de tratamiento lo que, en muchos casos, se ha considerado como obstáculo para el uso de equipos electromecánicos en las humanidades. Nadie diría que la precisión sea una desventaja en la investigación; sin embargo, muchas encuestas empiezan en un estado de imprecisión necesaria, y a veces nunca se aclaran totalmente. En estos casos, tanto los lenguajes que permiten la conversación y modificación del proceso como la simulación de posibles resultados, ofrecen amplias perspectivas. Además, el uso de criterios probabilísticos y asociativos, semejantes a los que se están explorando en la recuperación de documentos, permite que el investigador apunte a un blanco amplio, y no necesariamente a un punto fijo.

Por lo que respecta al sentimiento que para muchos colegas rige su actuación como humanistas (que la materia de su interés no es básicamente susceptible de definición precisa o de tratamiento formal), sin discutir ahora la cuestión, sugiero que los límites de lo indefinible nunca deben imponerse mientras no se haga un esfuerzo para saber hasta qué punto tiene validez la definición exacta.

JOSEPH E. GRIMES

Instituto Lingüístico de Verano, México.