FL 33 11

Comenzamos ahora el tiempo dedicado a la revista de Filología. 30 minutos semanales dedicados a los alumnos de Filología de la Universidad de Adistancia y en general a todos aquellos que estén interesados en temas de lengua y literatura. Este espacio está elaborado por los profesores de los distintos departamentos que componen la Facultad de Filología y en él tratamos temas que interesan y pueden ayudar a todos los alumnos.

En el programa de hoy vamos a tratar de un tema de actualidad y que cada vez tiene más importancia. Se trata de los ordenadores, pero no sólo vamos a hablar de ordenadores, sino de la aplicación de estos a una ciencia concreta, la lingüística. Para introducirnos en el mundo de la informática y los ordenadores hemos contado con la presencia del profesor Sebastián Dormido, que es catedrático del Departamento de Informática y Automática en la Facultad de Matemáticas de la Universidad de Adistancia.

Son muchos los adjetivos que se han intentado asignar a estos años en que vivimos, la era atómica, la era espacial, pero ¿por qué no la era de la informática? El ordenador tal y como lo concebimos en la actualidad es un invento que no cuenta con 50 años, ni tan siquiera con 50 años de antigüedad, y sin embargo en tan corto espacio de tiempo está penetrando nuestra vida hasta tal punto que pocas cosas existen en nuestro entorno que no lleven detrás de sí un ordenador. Un ordenador prepara las nóminas con las que nos pagan, también un ordenador nos permite sacar los billetes de tren o de correo o de avión, otros ordenadores controlan en los bancos el estado de nuestras cuentas, la hacienda pública vigila nuestra actividad fiscal por medio de ordenadores, y así podríamos ir presentando mil ejemplos más. ¿Cómo es posible que solo en unas décadas se haya llegado a tan alto grado de implantación de los ordenadores en nuestra sociedad? La respuesta a esta pregunta tendríamos que buscarla en el espectacular progreso experimentado en el campo informático.

De la mano en muchos casos de la electrónica, la evolución experimentada por los ordenadores es de tal índole que son muchos los que prefieren asignar a tal periodo el nombre de revolución. Estamos pues inmersos en una auténtica revolución informática en la sociedad de nuestros días. Sin embargo, y pese a su uso generalizado, es poca la gente que conoce, aunque sea de una forma elemental, qué es y cómo funciona un ordenador, y aún más, cuáles son sus bases de funcionamiento, qué posibilidades tiene y cuál es el impacto que los ordenadores están produciendo en la sociedad y cuál es el futuro previsible que espera a los mismos.

En este sentido, las charlas sobre informática en relación con la lingüística hoy se inician tratan de acercar o de llevar aparentemente dos campos separados entre sí y poner en perspectiva cómo la utilización de los ordenadores puede llevar a una mejor comprensión en un campo como el de la lingüística en el que la introducción del ordenador va a cambiar sustancialmente sus métodos y sus técnicas de trabajo. Si bien la informática se ha introducido de una forma muy importante en ciertos aspectos de la vida académica, como es la pura investigación científica, hasta el extremo que ha causado en la misma esta verdadera revolución de la que

hemos hablado, existen disciplinas a las que en principio parecía o parecería no haber afectado la implantación de los ordenadores, y este es el caso de la investigación humanística o el campo de las artes. Evidentemente, la influencia que la informática ha tenido en estos campos es mucho menor que la experimentada en otros.

Sin embargo, también aquí los ordenadores están jugando cada día un papel más importante. Gran parte del tiempo que empleaba el investigador histórico estaba, por ejemplo, destinado a la búsqueda en bibliotecas y archivos de documentos y legajos, tiempo que ahora se está reduciendo considerablemente, gracias a los ordenadores instalados en estas instituciones, con las ayudas de bases de datos que permiten la recuperación y el acceso a dichos archivos. Pero no sólo la informática tiene un campo de aplicación en la investigación humanística en cuanto al ahorro de tiempo que pueda proporcionar el investigador, sino en aplicaciones muy concretas que gracias al ordenador se pueden llevar a cabo con mayor facilidad y exactitud.

Tomemos como ejemplo la labor de identificación de documentos originales. Para este tipo de tareas los grafólogos necesitaban comparar miles de similitudes estilísticas y grafológicas que habían de correlacionarse a los largos de texto que superaban las 100.000 palabras. Esta labor la realizan en la actualidad con mayor rapidez y seguridad ordenadores convenientemente preparados.

Así, por ejemplo, este tipo de técnicas se ha utilizado recientemente en Estados Unidos para determinar los autores de The Federalist Papers, conjunto de 77 ensayos políticos escritos a principios del siglo pasado y destinados a convencer a los ciudadanos de Nueva York para que ratificasen la constitución de los Estados Unidos. Con este espíritu, con esta idea de alcance, estimo que se inicia este coloquio, esta charla sobre informática y lingüística con la confianza de que lo que en ella se diga será del mayor interés para todos los que la escuchan. Después de escuchar al profesor Sebastián Dormido, que ha hecho un análisis a grandes rasgos de lo que es el mundo de la informática, vamos a pasar a la aplicación práctica de los ordenadores en lingüística.

José Manuel Fradejas Rueda, profesor encargado de curso de lingüística general en la UNED, se encargó de preparar este tema. Se me ha pedido que hable sobre lingüística y ordenadores, quizá no sea el más indicado para esta tarea, pues en España hay varios investigadores, tanto lingüistas como informáticos, más versados que yo en esta materia, pero como soy el profesor encargado de lingüística general de nuestra universidad y por bacho varios hechos fortuitos, trabajo con ordenadores, que no con computadoras, me veo en este, digamos, un tanto sorprendente trance de hablar en la radio. Lo primero que tendré que explicarles, aunque quizá muchos ya lo sepan, es qué es un ordenador.

Más o menos todos, en algún momento, nos hemos topado con una de estas fantásticas máquinas, por los más variados motivos, bien por medio del banco o bien por una ineludible sanción de tráfico, pero estos aparatos no sólo son esos monstruos llenos de botones, luces y siniestros zumbidos, esa es solamente la parte deslumbrante de ese mundo. Un ordenador

está constituido por dos sistemas muy diferenciados, cuyos nombres son artelocuentes, aunque, como en todas las técnicas modernas, sólo tengan nombre en inglés y son el hardware y software. El hardware, que traducido al español no quiere decir más que chatarra, es la parte deslumbrante y que a todos llama la atención, y que en más de una ocasión hemos maldecido.

Esto por sí sólo es incapaz de hacer cosa alguna, tan sólo consumir electricidad, y hoy casi ni eso. El software es lo opuesto a hard, que es lo duro, ya que soft quiere decir blando, suave, y es lo que hace que ese montón de tuercas y tornillos sirva para algo. Esto lo constituyen los llamados lenguajes de programación, que van desde el ininteligible código binario, que es lo que entiende la máquina, su sucesión de ceros y unos, cuyo significado es enciende uno, apaga cero, a los más sofisticados lenguajes de alto nivel, como el COBOL, FORTRAN, LISP, ALGOL, PASCAL y el hoy popularísimo BASIC.

Pero este mundo ha estado en cierta medida vetado a los estudios humanísticos, porque los primeros en usarlos fueron los llamados científicos, que necesitaban aparatos capaces de calcular a grandes velocidades. El tiempo en ellos se mide no en minutos, ni siquiera en segundos, sino en microsegundos y nanosegundos. Fuera de España, ese monopolio desapareció prontamente, y las viejas y maltratadas humanidades pudieron acceder a los llamados centros de cálculo.

Esto no quiere decir que en España los humanistas, y sobre todo los lingüistas, no tengan acceso a estas máquinas. En la actualidad son muchos los trabajos que se están realizando en los diversos centros de cálculo de nuestras universidades, y los libros y artículos que se han publicado sobre estas investigaciones. El origen del uso de los ordenadores en lingüística se remonta a los años 50, cuando surgió la necesidad de tener un método rápido y eficaz para traducir del inglés a otras lenguas naturales, manuales e información técnica.

Fuera de este ámbito restringido, la llamada traducción automática ha tenido escaso o nulo valor, ya que las máquinas son incapaces de pensar, interpretar, y así se cuenta la anécdota de que utilizando este procedimiento se tradujo del inglés al ruso y de nuevo al inglés una frase sencilla. La frase que sostuvo al final del proceso nada tenía que ver con la original. Pero no les voy a hablar de este problema, ni siquiera de la llamada lingüística computacional, sino de qué se hace en España con los ordenadores y en lingüística.

La primera premisa, y que no se ha de olvidar nunca, es que el ordenador no es nada más que una herramienta que nos puede librar del tedioso y nada productivo trabajo mecánico, con lo que podremos obtener mayores y mejores inventarios de datos que pueden ser actualizados y consultados con enorme rapidez. Tengan en cuenta que un pequeño micro ordenador, de los que anuncian en televisión o en la prensa, es capaz de leer 256.000 caracteres en poco más de un segundo, y es lento. Los trabajos que se están haciendo en España son de muy diversa índole.

Análisis morfosintácticos, de geografía lingüística, estadística lingüística, índices léxicos, etc. En definitiva, cualquier material lingüístico es susceptible de ser tratado con ordenador, y el único

requisito previo es saber qué se quiere hacer y por qué. De todas las investigaciones les voy a hablar de los índices léxicos, ya que son los trabajos en los que me he introducido.

Estos índices se realizan en muchos países, y en España el mayor proyecto se realiza conjuntamente con la Universidad de Wisconsin, y se trata del Dictionary of Old Spanish Language, y cuyo director español es el profesor Manuel Álvaro Ezquerra, de la Universidad de Málaga, pero que comenzó en la Complutense de Madrid. Manuel Álvaro Ezquerra es caderático de lengua española en la Universidad de Málaga, y es uno de los pocos investigadores españoles que está aplicando la informática en el estudio de la lengua española. Él mismo nos cuenta la experiencia de investigación que está llevando a cabo en este campo.

La aplicación de los ordenadores a la lingüística en España no ha tenido el desarrollo que hubiera sido conveniente y parece contradictorio, sobre todo porque la informatización de la lingüística requiere o es una nueva formulación de todo el positivismo anterior, y resulta chocante que una filología, como es tremendamente positiva la filología española, los ordenadores, que es un instrumental que desarrolla el positivismo, no se hayan aplicado a la lingüística. De todas maneras hay una serie de proyectos que son pioneros y que sí merece la pena tener en cuenta. En el nivel fonético se ha trabajado poco y lo poco que se ha hecho ha sido gracias a la tenaz labor de don Antonio Quilis, que es catedrático de la UNED y es la persona que más ha trabajado en fonética en España.

Quizás lo más espectacular que haya realizado ha sido en colaboración con la Escuela de Informática o de Telecomunicaciones de la Universidad Complutense y han llegado a desarrollar un simulador faringeo de manera que eso puede enseñar a hablar y a pronunciar correctamente a los a los sordos, enseñándoles por comparación de ondas sonoras cuando una articulación está bien hecha, cuando está mal o cuando se puede corregir. Por otra parte han diseñado también un lector magnético que transforma en voz lo que es escritura y eso es de una enorme utilidad sobre todo para los ciegos, porque siguiendo en la idea del positivismo la lingüística en un momento o en un mundo tan tremendamente pragmático como es el actual debe tener algunas aplicaciones de este tipo. En el nivel morfosintáctico en España apenas se ha hecho nada y no merece la pena ser señalado lo que se ha hecho y quizás los máximos desarrollos se han formulado a través del nivel léxico-semántico, quizás por su aparente simplicidad, ya que trabajando con palabras o considerando la palabra como una unidad de texto separada entre dos blancos de escritura es una delimitación muy fácil de hacer.

En este sentido se ha empezado a aplicar a textos literarios, en concreto yo empecé a trabajar en la Universidad Complutense sobre textos poéticos medievales y después cuando ya me trasladé a la Universidad de Málaga allí se siguió trabajando y no solamente sobre textos medievales, hemos trabajado sobre la obra de Berceo, el poema de Fernán González, el laberinto de Fortuna y otras cuantas obras anónimas. Se ha trabajado también algo sobre literatura de Siglos de Oro que recuerda así de memoria el Aragonés Jiménez Durrea y el Diablo Cojuelo y después ya en época de escritores más recientes que yo recuerdo de una forma inmediata está el Oficio de Tinieblas de Camilo José Cela. En realidad en un principio lo que se

pretende hacer son simplemente simples recuentos léxicos pero esto no tendría ningún sentido si no se le saca una aplicación o una utilidad práctica ulterior y esa aplicación práctica es ni más ni menos que el comentario de textos y la crítica literaria a partir de datos léxicos semánticos y estilísticos y son resultados realmente sorprendentes los que puede proporcionar un elemental recuento de palabras para demostrarnos cuál es la estructura interna del texto.

Últimamente sobre lo que más estamos trabajando es en la preparación de una nueva edición del Diccionario Vox, el Diccionario General Ilustrado de la Lengua Española. Para ello hemos recuperado toda la información de la imprenta, se trabaja en la impresión moderna en fotocomposición sobre cintas grabadas magnéticamente y hemos transformado esa impresión codificada para el ordenador. Nos ha costado un trabajo enorme descifrar todos los datos pero a partir de ahí ya vamos a poder hacer un nuevo diccionario que será el primer diccionario del español totalmente automatizado donde las ampliaciones, las reducciones, correcciones, etcétera, serán de una extrema facilidad y podremos ir adecuándolo a las necesidades diarias.

La utilidad práctica no solamente está en la utilidad pedagógica que tiene cualquier diccionario sino también todos los posibles desarrollos ulteriores que se puedan hacer desde posibilidades de comunicaciones telefónicas o a través de terminales instalados en cualquier punto de España o del mundo si se quiere, a través de redes de comunicación, llámense puntos de información cultural, llámense micro ordenadores personales pues uno podrá tener una información sobre las palabras y el significado de las palabras enorme pero si esos campos de aplicación los reducimos podríamos tener también diccionarios de tipo ortográfico para utilización en empresas y oficinas para correcciones de estilo, etcétera. Y por otro lado hemos trabajado también o estamos empezando a trabajar sobre diccionarios de tipo dialectal, estamos empezando a hacer pequeños vocabularios, digo pequeños por el tamaño que tienen ahora de cada uno de los dialectos del español, evidentemente sobre el que más hemos trabajado por nuestra situación geográfica es sobre el andaluz, las aplicaciones inmediatas ya son fácilmente imaginables sobre todo por lo que he dicho antes del diccionario box y en un futuro pues quizás podremos disponer de un amplio banco de datos léxicos no solamente de la lengua general sino de las variedades particulares de cada una de nuestras lecciones. Después de escuchar al profesor Manuel Álvarez Guerra que nos contó su experiencia en la investigación aplicada a la lingüística seguimos escuchando al profesor José Manuel Fradejas Rueda.

Esta investigación lo que trata es de crear o para ser una palabra del mundillo de la informática generar el diccionario del español medieval, éste se comenzó hace años manualmente pero tan pronto como los ordenadores estuvieron a disposición de los lingüistas se vio la posibilidad de utilizarlos ya que podían aligerar enormemente las tareas mecánicas que hubieran llevado a varios investigadores muchos años de trabajo tedioso y sujeto a múltiples errores. La forma de trabajar de este equipo requiere el concurso de lingüistas e informáticos, estos los informáticos son los que se preocupan de que las máquinas hagan lo que los investigadores necesitan ya que ellos plantean los problemas y qué es lo que desean obtener y los informáticos se ocupan de crear los programas adecuados que resolverán dichos problemas y darán los resultados

deseados. El Dictionary of Old Spanish Language sigue este proceso, un filólogo en cualquier parte del mundo transcribe un manuscrito siguiendo las normas establecidas por el equipo editor, todas ellas son inequívocas y se han de seguir al pie de la letra pues la más mínima variación puede resultar en una corrupción de los datos, una vez transcrito sobre un papel especial y utilizando un cabezal de máquina IBM especial que deja debajo de cada carácter, sea letra o no, un código de barras como los que hoy incorporan muchos productos y lo envía al centro de proceso, allí con un lector óptico se trasbasa el texto al ordenador que lo guardará en cintas magnéticas, también es posible hacer la transcripción directamente en un ordenador usando para ello la pantalla de rayos catódicos y el teclado y que irá grabando el texto o en tarjetas perforadas, sistema ya obsoleto, o en cintas o discos magnéticos, la ventaja de este método es que la corrección es mucho más rápida, una vez concluida la transcripción y grabación de los datos lo mandará al centro de proceso y los ordenadores lo podrán leer directamente sin necesidad de pasos intermedios como en el caso anterior, una vez que el texto está en los bancos de memoria de los ordenadores estos se encargarán de manipularlo para ofrecer los datos que los filólogos necesitan y así la máquina va leyendo el texto palabra por palabra fragmentándolo y asignando a cada elemento unas referencias de localización, cuando estos procesos han concluido nos encontramos con dos ficheros básicos, uno de textos y dos palabras, a partir del fichero de palabras, aunque este no es el nombre verdadero pero es útil para nuestra explicación, ya se puede realizar los más variados índices y cálculos, generalmente se hacen índices léxicos alfabetizados directamente, es decir, tal y como los encontramos en cualquier diccionario o inversos, o sea, empezando por la última letra de la palabra, ambos tipos de alfabetización tienen muy diversos usos a su vez, también se hacen recuentos y esto nos darán después las frecuencias de cada palabra, tanto absoluta, número de veces que aparece, como relativa, que suele expresarse en tantos por ciento y que nos habla de la aportación de cada palabra al conjunto del texto, todo este trabajo que un ordenador puede hacer en pocas horas podría emplear a muchos investigadores, como ya he dicho, durante muchas semanas y además estar sujeto a multitud de errores, pues en contra de lo que nos quieren hacer creer, los ordenadores no se equivocan nunca, son las personas que le dan la información los que los equivocan, por ejemplo, si yo estoy trabajando con números menores que uno, es decir, con decimales y me olvido de introducir la coma decimal, las máquinas, punto, el resultado final, que a lo mejor no tenía que pasar de 100, nos encontramos que es 10.000, la culpa no fue del ordenador, él admitió esa cifra porque yo se la introduje, no porque la máquina sumara mal, a partir del momento en que el ordenador nos ha dado los resultados, el lingüista entra de nuevo en acción, esta área única y exclusiva de él el interpretar y dar significado a los datos, no basta con publicar los listados que dio la máquina, estos no son más que un montón de papel con palabras y cifras que no quieren decir nada, ya son varios los textos y resultados publicados, tanto en Estados Unidos como en España, pero quizá el más llamativo sea el índice de concurrencias de la obra de Alfonso X el sabio, que debido a su gran volumen ha tenido que ser publicado en las en micro fichas, téngase en cuenta que en cada una de ellas pueden caber más de 100 folios mecanografiados, lo que les he contado durante los minutos precedentes no es más que una ínfima parte de las posibilidades que tiene el lingüista, sea de la lengua que sea, de emplear los ordenadores en sus investigaciones, y antes

de dar por finalizada esta pequeña charla quisiera decirles que no hay que tenerle miedo a estas máquinas, no son nada más que un montón de chatarra, pero eso sí, chatarra útil, y que no deben olvidar nunca que los ordenadores no son nada más que una herramienta que nos puede proporcionar más tiempo para la auténtica investigación.

Terminamos nuestro tiempo por hoy. El profesor José Manuel Fradejas Rueda se encargó de preparar y coordinar el espacio de hoy, ordenadores y lingüística. Agradecemos desde aquí la colaboración que nos prestaron el profesor Sebastián Dormido, catedrático de informática y automática en la universidad a distancia, y el profesor Manuel Álvarez Guerra, catedrático de lengua española en la universidad de Málaga.

Volvemos el próximo lunes a la hora habitual de este espacio, en el que trataremos de la épica medieval. El catedrático de literatura española de la UNED, José Fradejas Lebrero, será el encargado de desarrollar este tema. Hasta entonces, buenas noches.

Subtítulos realizados por la comunidad de Amara.org