

El habla como medio de acceso a la Sociedad de la Información

Por Joaquim Llisteri

Universidad Autónoma de Barcelona

Cuando pensamos en la Sociedad de la Información, inmediatamente asociamos esta nueva realidad a los ordenadores, y nos viene a la mente la imagen de una pantalla y un teclado que nos sirven, esencialmente, para enviar y recibir textos. No parece sorprendernos que un medio tan avanzado de transmisión de información y de interacción con otras personas requiera el uso de lo que, para nuestros propósitos, no dejan de ser instrumentos, aunque indudablemente sofisticados, de escritura.

Sin embargo, el modo natural de comunicación entre los seres humanos no es precisamente la lengua escrita, sino el habla. Aunque se trate de una cuestión polémica, se estima que el surgimiento de la lengua hablada pudo haberse producido hace unos 100.000 años [1], mientras que la escritura no empieza hasta hace aproximadamente 5.000 años. Aun prescindiendo de este argumento histórico, resulta difícil concebir en la actualidad que la escritura sea la manera más eficiente de que disponen dos personas para comunicarse. Un trabajo ya clásico de Chapanis, un experto en psicología laboral, mostró a mediados de los años 70 que el tiempo empleado por un equipo en resolver un problema se reducía a casi la mitad si los miembros del grupo podían hablarse, aunque fuera a distancia, en contraste con una situación en la que únicamente pudieran escribirse (Chapanis, 1975).

No deja de resultar pues llamativo que nuestra visión de la Sociedad de la Información, y fundamentalmente de la red Internet, a la que está inevitablemente ligada, gire a menudo alrededor de la escritura. Por citar algún dato concreto, los resultados de la 4ª encuesta de la Asociación para la Investigación de los Medios de Comunicación (AIMC) realizada entre usuarios de Internet en España (abril - junio de 2001) muestran que el 95,5% de los internautas mencionan el correo electrónico, un medio escrito, en respuesta a la pregunta "Durante el día de ayer ¿Qué servicios de Internet utilizó?". En el ámbito de las tecnologías lingüísticas [2], justamente una de las primeras que se ha incorporado de forma generalizada a Internet ha sido la traducción automática de textos, disponible ya incluso en algunos de los buscadores más populares (por ejemplo, Babel Fish en [Altavista](#) o Reverso en [Lycos](#)).

INTERACCIÓN ORAL CON SISTEMAS INFORMÁTICOS

Pese a esta aparente ubicuidad de la lengua escrita, desde hace tiempo numerosos expertos han reflexionado sobre la necesidad de acercar nuestro modo de interactuar con un sistema informático a los procedimientos naturales de comunicación a los que estamos habituados. Ha surgido por tal motivo un fuerte interés en los denominados "interfaces multimodales" (Mariani (Ed.), 1997), que incorporan la posibilidad de enviar y recibir información textual, gráfica o sonora, utilizando para ello un teclado convencional, una pantalla táctil - como la que se encuentra en muchos de los cajeros automáticos - o, como es lógico, el habla. El uso del habla en los casos en que la información visual no es estrictamente necesaria presenta, al menos, tres ventajas: por una parte, determinadas tareas pueden realizarse mediante un teléfono, sin necesidad de disponer de un ordenador; por otra, el envío o la recepción de

información pueden simultanearse con otras actividades; en tercer lugar, se puede prescindir del teclado en aparatos de pequeño tamaño como los asistentes personales digitales.

Algo, como acabamos de ver, tan común entre los usuarios de Internet como enviar o recibir correo electrónico se puede llevar a cabo mediante el habla con tecnologías lingüísticas existentes en la actualidad. Empresas españolas como Telefónica I+D han desarrollado conversores de texto a habla que se han incorporado a servicios como el [Correo Móvil de MoviStar](#), haciendo posible la lectura automática del correo electrónico a través del teléfono; por otro lado, existen sistemas comerciales de reconocimiento de habla que permiten dictar cualquier texto en español - [Via Voice](#) de IBM, [Free Speech](#) de Philips o [Naturally Speaking](#) de Dragon Systems - y tratarlo con los procesadores de texto más habituales.

La conversión de texto en habla - es decir, la lectura en voz alta de un texto realizada por un ordenador del modo más similar posible a como la haría una persona - y el reconocimiento del habla - la posibilidad de convertir automáticamente un enunciado hablado en un texto escrito - son las dos tecnologías básicas, conocidas como "tecnologías del habla" [3], necesarias para la interacción oral con los sistemas informáticos. En este trabajo nos centraremos en la primera de ellas, y remitimos al lector interesado en el reconocimiento a presentaciones generales como las de [Kurzweil](#) (1998).

LA CONVERSIÓN DE TEXTO EN HABLA

Aplicaciones de la conversión de texto en habla

Escuchar la lectura de un texto en lugar de consultarlo sobre el papel o en una pantalla puede ser sumamente útil en muchas ocasiones. Hemos mencionado ya la recepción del correo electrónico a través del teléfono, pero existen otras muchas aplicaciones de la conversión de texto en habla.

Los recientes portales de voz - en español, por ejemplo, el de [Terra](#) de Telefónica ([Fernández](#) et al., 2000) o [Ydilo](#) de People Communications ([Gómez, 2001](#)) - incorporan información que varía incluso en cuestión de horas, como la de los partes meteorológicos o del estado del tráfico; en este caso, una lectura automática permite ofrecer datos mucho más actualizados, ya que no requiere disponer de un locutor profesional dedicado a realizar constantemente una grabación.

Otros servicios se benefician también de una tecnología que permite recibir información hablada: horarios de los transportes públicos, información ciudadana - farmacias de guardia, horarios de museos -, consulta de carteleras de cine y teatro, datos de interés turístico sobre una determinada zona, y un largo etcétera. En todos estos casos, la información almacenada en una base de datos escrita llega en forma oral a la persona que desea conocerla mediante un teléfono y un sistema de conversión de texto en habla.

Las transacciones comerciales electrónicas son otra de las posibilidades que se han incorporado recientemente a la Sociedad de la Información. En ciertas ocasiones en las que la transacción no es excesivamente compleja - reserva y compra de billetes o de entradas a espectáculos, servicios de banca electrónica - es factible utilizar los denominados "sistemas de diálogo" o "sistemas conversacionales" ([Zue](#), 1999), que permiten llevar a cabo una interacción oral, más o menos natural, con el ordenador. Para ello es necesario disponer de un reconocedor de habla que recoja los enunciados del usuario, un conversor de texto a habla que proporcione las respuestas, un sistema de comprensión que interprete las peticiones del usuario y un gesto del diálogo que se encargue de organizar la "conversación".

Es muy importante destacar que la conversión de texto en habla es una herramienta indispensable para las personas privadas de visión, tanto en lo que se refiere a su integración en el mundo laboral como a su formación o a su vida cotidiana. La posibilidad de escuchar cualquier texto que previamente haya sido convertido a un formato electrónico - un documento escrito con un procesador de textos, un correo electrónico, una página web - evita depender de otras personas y proporciona también la intimidad necesaria para el estudio o para la lectura de textos de carácter personal. Los programas de reconocimiento óptico de caracteres (OCR, Optical Character Recognition) transforman la imagen de un texto escaneado en un documento electrónico que puede ser leído por un conversor, con lo que la combinación de ambos elementos - en sistemas como el de Lernout & Hauspie - [Kurzweil Educational Systems](#), por ejemplo - hace posible incluso la consulta de un texto impreso en papel.

Finalmente, las personas con dificultades en la expresión verbal - es bien conocido el caso del cosmólogo británico [Stephen Hawking](#) - pueden también hacer uso de sistemas de conversión de texto en habla integrados en pequeños ordenadores - una muestra de ello es el [Crespeaker](#) de Crestwood Communication Aids- para expresarse, incluso cuando existen severas limitaciones para utilizar un teclado.

Sistemas de conversión de texto en habla

No entraremos en esta presentación en los detalles de la estructura interna de un sistema de conversión de texto en habla. El lector interesado los encontrará en trabajos introductorios como Rodríguez (1997), [Olive](#) (1998), Dutoit (1999) o en publicaciones más especializadas como Dutoit (1997).

Mencionaremos, en cambio, que existen para el español sistemas de conversión de texto en habla desarrollados tanto en universidades como en empresas de nuestro país. Pueden escucharse, por ejemplo, las demostraciones accesibles en la web del [Grup de Tractament de la Parla de la Universitat Politècnica de Catalunya](#), del [Grupo de Tecnología del Habla de la Universidad Politécnica de Madrid](#), o del [Grupo de Tecnología de las Comunicaciones de la Universidad de Zaragoza](#). También otros centros de investigación vinculados a empresas han mostrado, en Estados Unidos y en Europa, su interés por la creación de sistemas de síntesis en español, incorporando en algunos casos variantes americanas, principalmente de México. Algunos ejemplos se encuentran en las páginas de [AT&T Labs](#), [Bell Labs](#) - Lucent Technologies, [Elan Informatique](#), [Infovox-Babel Technologies](#), [Lernout & Hauspie](#), o de [Loquendo](#). Se pueden oír igualmente ejemplos procedentes de centros académicos, que han desarrollado sistemas como [Festival](#) (CSLU, Center for Spoken Language Understanding, Oregon Graduate Institute) o [MBROLA](#) (TCTS Lab, Faculté Polytechnique de Mons).

Debe señalarse, por otra parte, la existencia de conversores de texto a habla para el catalán - se encuentran demostraciones en las páginas del [Grup de Tractament de la Parla de la Universitat Politècnica de Catalunya](#), y de la [Secció de Teoria del Senyal de la Universitat Ramon Llull](#) -, el gallego (Fernández, 1999) o el vasco (Hernández et al., 1994), así como el [sistema multilingüe](#) desarrollado en Telefónica I+D. Las personas interesadas en sistemas de conversión de texto en habla en otras lenguas pueden consultar el [Interactive Speech Synthesizer Comparison Site](#) preparado por el LDC (Linguistic Data Consortium de la Universidad de Pensilvania) conjuntamente con COCODSA (International Committee for the Co-ordination and Standardisation of Speech Databases and Assesment Techniques).

CONCLUSIÓN

Disponemos pues de una tecnología lingüística - la conversión de texto en habla - que hace posible acceder a través del habla a los contenidos de cualquier texto escrito que se encuentre en formato digital, y que es útil tanto para la recepción de información como para la realización de transacciones; es, además, una herramienta que abre inmensas perspectivas a personas con limitaciones visuales o

vocales. Esta tecnología está disponible en lengua española y en otras lenguas de España, y ha de servirnos para mejorar nuestro modo de interacción con los sistemas informáticos, permitiéndonos el acceso a la Sociedad de la Información mediante el procedimiento que nos es más natural a las personas: el habla.

REFERENCIAS

- AITCHISON, J. (1996) *The Seeds of Speech: Language Origin and Evolution*. Cambridge: Cambridge University Press.
- CHAPANIS, A. (1975) "Interactive Human Communication", *Scientific American* 232: 36-42.
- DUTOIT, T. (1997) *An Introduction to Text-to-Speech Synthesis*. Dordrecht: Kluwer Academic Publishers (Text, Speech and Language Technology, 3).
- DUTOIT, T. (1999) [A Short Introduction to Text-to-Speech Synthesis](#), TCTS Lab, Faculté Polytechnique de Mons.
- FERNÁNDEZ REI, E. (1999) "Tecnologías del habla y síntesis de voz en gallego", in GÓMEZ GUINOVART, J.- LORENZO SUÁREZ, A.- PÉREZ GUERRA, J.- ÁLVAREZ LUGRÍS, A. (Eds.) *Panorama de la investigación en lingüística informática*. RESLA, Revista Española de Lingüística Aplicada, Volumen monográfico. pp. 103-116.
- FERNÁNDEZ IGEA, S.- ORTEGA DEL RÍO, R.- SERRANO ARANDA, R. (2000) "[Portales de voz: Internet en el teléfono](#)", *Comunicaciones de Telefónica I+D* 19: 15-24.
- GÓMEZ MUÑOZ, C.F. (2001) "[Comercio electrónico y servicios avanzados de voz](#)", I Jornada EUROMAP 2001. Seminario Temático EUROMAP: El comercio electrónico y las tecnologías para el lenguaje humano. Ministerio de Ciencia y Tecnología, Madrid, 12 de junio de 2001.
- HERNÁEZ, I.- OLABE, J.C.- ETXEBERRIA, P.- ETXEBERRIA, B.- CUESTA, A. (1994) "Ahozka. Un sistema de conversión de texto a voz para el Euskara", *Boletín de la Sociedad Española para el Procesamiento del Lenguaje Natural* 14: 241-256.
- [Ingeniería lingüística](#). Cómo aprovechar la fuerza del lenguaje. Luxembourg: Language Engineering, Telematic Applications Programme. Versión española: Observatorio Español de Industrias de la Lengua, Instituto Cervantes.
- KURZWEIL, R. (1998) "[When Will HAL Understand What We Are Saying? Computer Speech Recognition and Understanding](#)", in STORK, D.G. (Ed.) *Hal's Legacy: 2001's Computer as Dream and Reality*. Cambridge, Mass.: The MIT Press.
- MARIANI, J. (Ed.) (1997) "Multimodality", in COLE, R.A.- MARIANI, J.- USZKOREIT, H.- ZAENEN, A.- ZUE, V. (Eds.) [Survey of the State of the Art in Human Language Technology](#). Cambridge: Cambridge University Press. pp. 329-370.
- MARTÍ, M.A.- LLISTERRI, J. (2001) "[L'enginyeria lingüística en la societat de la informació](#)", *Digit-HVM*, *Revista Digital d'Humanitats (Universitat Oberta de Catalunya)* 3.
- OLIVE, J.P. (1998) "[The Talking Computer: Text to Speech Synthesis](#)", in STORK, D.G. (Ed.) *Hal's Legacy: 2001's Computer as Dream and Reality*. Cambridge, Mass.: The MIT Press.
- QUARK (2000) "[El desarrollo de las tecnologías de la lengua](#)", *Quark. Ciencia, Medicina, Comunicación y Cultura* 19: 20-73.
- RODRÍGUEZ CRESPO, M.A. (1997) "Introducción a la conversión texto-voz", *Philologia Hispalensis* 11, 2: 177-192.
- RODRÍGUEZ CRESPO, M.A.- CORTÁZAR MÚGICA, I.- TAPIAS MERINO, D.- RELAÑO GIL, J. (2001) "[Estado del arte en tecnologías de voz](#)", *Comunicaciones de Telefónica I+D* 20: 117-136.
- ZUE, V. (1999) "[Talking with your computer](#)", *Scientific American*, August 1999. pp. 40-41.

[1] Las fechas que habitualmente se encuentran recogidas en los trabajos sobre el origen del lenguaje humano oscilan entre 250.000 y 50.000 años AC (Aitchison, 1996).

[2] Para una presentación general de las tecnologías lingüísticas véanse, por ejemplo, el documento [Ingeniería lingüística. Cómo aprovechar la fuerza del lenguaje](#), el volumen monográfico de Quark (2000) o [Martí - Llisteri](#) (2001).

[3] Presentaciones generales de las tecnologías del habla se encuentran, por ejemplo, en [Rodríguez et al.](#) (2001). Para una información más detallada se pueden consultar las páginas de la asignatura ["Tecnologías del habla"](#) preparadas por el autor de este trabajo.