

Llisterri, J. (2009). Las tecnologías del habla en las lenguas románicas ibéricas. *Studies in Hispanic and Lusophone Linguistics*, 2(1), 133-180. ISSN: 1939-0238

http://liceu.uab.cat/~joaquim/publicacions/Llisterri_09_Tecnologias_Habla_Lenguas_Romanicas_Ibericas.pdf

Las tecnologías del habla en las lenguas románicas ibéricas

Joaquim Llisterri
Universitat Autònoma de Barcelona

Abstract

This paper reviews the development of speech technologies for romance iberian languages (Catalan, Galician, Spanish and Portuguese) with a special emphasis on the phonetic knowledge involved in the three core areas in the field: text-to-speech synthesis, automatic speech recognition and spoken language systems. Several important technological advances have taken place during the last decades and there is a substantial base of active researchers in the areas of engineering and computer science. However, the trend towards the use of automatic learning techniques and corpus based models has seriously reduced the role of linguistics, favoring the reduction in time in the development of applications over basic research into the nature of the speech code.

1. Las tecnologías del habla

Si consideramos por un momento el modo cómo utilizamos un ordenador, nos percataremos de que buena parte de la interacción se realiza mediante una pantalla y un teclado, recurriendo a interfaces basadas en iconos o en textos. Este predominio del canal visual contrasta con la naturaleza vocal y auditiva del habla (Hockett 1960), nuestro principal medio de comunicación, con lo que el uso de sistemas informáticos requiere el aprendizaje de nuevas habilidades y una adaptación a procedimientos diferentes a los que empleamos en nuestra interacción con las personas.

Estudios ya clásicos como los de Chapanis (1975) pusieron de relieve la importancia de la comunicación oral frente a la escrita en la resolución de problemas en equipo. También son ilustrativos los datos proporcionados por Ainsworth (1988): mientras que una persona sin formación específica en mecanografía puede teclear entre 10 y 25 palabras por minuto, en el habla se alcanza una velocidad de 120-250 palabras en el mismo lapso de tiempo; por otra parte, el uso de teclados cada vez más reducidos, como los que se encuentran en los dispositivos móviles, aumenta todavía más la dificultad de la tarea. Estos argumentos, y otros que subrayan las ventajas del habla frente a la escritura, apoyan la necesidad de disponer de interfaces orales en el contexto de la interacción entre personas y ordenadores (HCI, *Human-Computer Interaction*), objetivo que, precisamente, se pretende alcanzar a través de las tecnologías del habla.

Por este motivo, las tecnologías del habla suelen describirse como todas aquellas que hacen posible la entrada y la salida vocal en un ordenador con el fin de facilitar su uso por parte de las personas (Laver 1987). La entrada se lleva a cabo mediante el reconocimiento automático del habla (§ 3), mientras que la salida se consigue a través de la síntesis (§ 2); los llamados sistemas de diálogo o sistemas conversacionales integran ambos aspectos, como veremos más adelante (§ 4), para que pueda darse una interacción lo más natural posible (Tapias 2002). Dado que todas estas operaciones implican procesos relacionados con el tratamiento lenguaje, las tecnologías del habla se suelen enmarcar en el contexto más amplio de las tecnologías lingüísticas, que incluyen tanto el procesamiento de la lengua hablada como de la escrita (Lavid 2005, Llisterri 2003, Martí 2003 y Ranchhod 2001).

Si bien las tecnologías del habla surgieron de la mano de la ingeniería de telecomunicación y se orientaron, en un principio, a resolver los problemas de la codificación y la transmisión de la señal del habla en el campo de la telefonía, los primeros intentos de síntesis y de reconocimiento pusieron ya de manifiesto la necesidad de contar con las aportaciones de la lingüística. Sin embargo, en las últimas décadas, los enfoques estadísticos en los que se realiza el aprendizaje automático a partir de un corpus de datos han tendido a reemplazar a las aproximaciones basadas en el conocimiento de la estructura de la lengua, por lo que el papel de los lingüistas se ha visto notablemente reducido, tal como se argumenta en las aportaciones recogidas en Barry & van Dommelen (2005). Aun así, recientemente se han propugnado enfoques híbridos que logren conjugar la contribución del saber lingüístico con las ventajas de los métodos estadísticos pues, como acertadamente señalaba Fant (2004, p.11): “Computing power can not substitute crucial knowledge”.

Por otro lado, para que las tecnologías del habla encuentren su utilidad real, tienen que estar a disposición de los usuarios en su propio idioma, lo que requiere, en muchos casos, una labor adicional de obtención de conocimientos para lenguas que, bien estudiadas desde el punto de vista de la lingüística, no se han abordado desde la perspectiva necesaria para el desarrollo de las tecnologías; en este sentido, Strik (2005, p.177) recuerda que, por la propia naturaleza de la disciplina, “phonetics does not provide ready-made quantitative models that can be plugged directly into a system”, por lo que es preciso reorientar la investigación de modo que, al tiempo que se logra un conocimiento científico básico, se disponga de datos cuantitativos que permitan la creación de modelos computacionales para las tecnologías del habla (Llisterri 2007b).

El objetivo de este trabajo es presentar, por una parte, la información lingüística y, especialmente, fonética, que requiere el desarrollo de los sistemas de síntesis (§ 2), de reconocimiento (§ 3) y de diálogo (§ 4) (Llisterri 2007b, Llisterri 2007c y Llisterri, Carbó, Machuca, Mota, Riera & Ríos 2003) y, por otra, ofrecer algunas indicaciones sobre los trabajos llevados a cabo en las lenguas románicas ibéricas: catalán (Llisterri 2002), español (Llisterri 2004 y Llisterri 2007a), gallego (García Mateo 2002 y García Mateo 2003) y portugués. Puesto que una contribución de esta naturaleza

no puede aspirar a ser exhaustiva, para una información más detallada sobre las actividades de los grupos de investigación en España remitimos a recopilaciones como el *Libro blanco de Tecnologías del Habla* (Rubio Ayuso & Hernández 2005) o a las actas de las jornadas de la Red Temática en Tecnologías del Habla¹ (Buera, Lleida, Miguel & Ortega 2006, Hernández 2008, López Soto & López-Cózar 2000, Rubio Ayuso 2002 y Sanchis 2004). Los lectores interesados en el portugués encontrarán información en parte de la bibliografía recogida en el proyecto *Linguateca*², en las actas de los congresos conocidos como PROPOR (*Computational Processing of the Portuguese Language*) y en las páginas en la web de la iniciativa *Tecnovoz (Tecnologia de Reconhecimento e Síntese de Voz)*³.

Asimismo, tanto en los eventos internacionales *Interspeech* (organizado bajo dos denominaciones diferentes hasta 1999: *Eurospeech, European Conference on Speech Communication and Technology* e *ICSLP, International Conference on Spoken Language Processing*), cuyas actas se recogen en los archivos en línea de ISCA (*International Speech Communication Association*)⁴, como en las revistas especializadas (*Computer Speech and Language, International Journal of Speech Technology, Speech Communication*, entre otras), pueden encontrarse contribuciones relacionadas con las lenguas que son objeto de este trabajo.

2. La conversión de texto en habla

Hemos señalado anteriormente que la salida vocal constituye uno de los requisitos para la interacción con los ordenadores mediante el habla. Las técnicas que permiten alcanzar este objetivo se agrupan bajo la denominación de síntesis del habla (Carlson & Granström 1997, Nusbaum & Shintel 2006 y Viana 2001) y, como puede suponerse, han experimentado a lo largo del tiempo una notable evolución, especialmente patente en la mejora de la calidad de las voces, que prácticamente han perdido ya buena parte de su anterior carácter artificial.

La principal aplicación de la síntesis es la posibilidad de transformar un texto escrito en su equivalente sonoro, lo que se consigue con los sistemas de conversión de texto en habla (TTS, *Text-to-Speech Synthesis*) (Bonafonte, Escudero & Riera 2006, Dutoit 1997 y Schroeter 2006). El proceso de conversión de texto en habla requiere una serie de etapas, en las que se suele distinguir entre las que se centran en el procesamiento de la señal sonora y las que realizan un tratamiento de la información lingüística. En lo que se refiere a este último aspecto, los principales módulos lingüísticos de un conversor se ocupan del procesamiento previo del texto, de la transcripción fonética automática y de la asignación de los valores relacionados con los parámetros prosódicos, tal como se describe brevemente a continuación y se presenta con más detalle en Llisterri, Carbó, Machuca, de la Mota, Riera & Ríos (2004).

Las operaciones implicadas en el tratamiento previo o normalización del texto de entrada -véase, por ejemplo, Rodríguez Crespo & Escalada (1990)- se relacionan con la necesidad de convertir en texto deletreado todos aquellos elementos que habitualmente no aparecen como tal: abreviaturas, números arábigos o romanos, símbolos como el del euro o del dólar, etc. Para ello es preciso solucionar, entre otros, problemas relacionados con los signos de puntuación, que deben recibir tratamientos distintos en función de su presencia en un texto o en expresiones numéricas. Es necesario también establecer reglas que conviertan los números en una forma aceptable en función del tipo de cantidad de la que se trate por lo que, en esta fase, puede ser relevante incorporar información morfológica para aquellos casos en los que determinadas expresiones numéricas deban concordar con el nombre al que acompañan.

La segunda etapa, la transcripción fonética automática, presenta una mayor complejidad lingüística, puesto que se trata de crear, a partir de la entrada normalizada, una representación fonética lo más cercana posible a la realización oral del texto. Ello implica, en primer lugar, definir el inventario de unidades fónicas -tanto si se trata de fonemas como de alófonos- de la lengua, decisión que tendrá repercusiones en el resto de la síntesis, pues determinará las unidades con las que trabajará el conversor. Como es bien sabido, no siempre existe un acuerdo entre los autores que han propuesto repertorios fonéticos y fonológicos de las lenguas; el lingüista que colabora en el desarrollo de una aplicación deberá, por tanto, tomar decisiones más orientadas por las consecuencias en el proceso de síntesis que por sus propias concepciones teóricas, principio que también se aplica en el caso de los modelos prosódicos, que trataremos más adelante.

En segundo lugar, deben establecerse las correspondencias entre las formas gráficas y su representación fonética, para lo cual es preciso adoptar criterios fonéticos sobre aspectos como la variedad geográfica y el estilo de habla que se desean en la salida del conversor. Cabe también considerar algunos casos como los extranjerismos, las siglas y los acrónimos (Monzo, Alías, Morán & Gonzalvo 2006) o la pronunciación de los nombres propios (San-Segundo, Montero, Córdoba & Gutiérrez Arriola 2000 y Trancoso, Viana, Silva, Marques & Oliveira 1994), que plantean problemas especiales muy a menudo poco tratados en los estudios descriptivos. Finalmente, en esta etapa se puede contemplar también la división silábica (Oliveira, Moutinho & Teixeira 2005) y, naturalmente, la asignación del acento léxico.

La transcripción fonética automática se ha estudiado detalladamente para el español peninsular en trabajos basados en conocimiento lingüístico, en los que se definen las reglas de asociación entre grafías y fonemas o alófonos (Bonaventura, Giuliani, Garrido & Ortín 1998, Enríquez & Casado 1991, Enríquez 1991, Llisterri & Mariño 1993, Moreno Sandoval, Torre, Curto & Torre 2006 y Ríos 1999), aunque desde el ámbito de la ingeniería se ha abordado también el diseño de este tipo de reglas (Castro, España, Marzal & Salvador 2001); más recientemente se ha recurrido a técnicas de aprendizaje automático que parten de enfoques estadísticos, en los que los conocimientos fonéticos juegan, como se ha dicho, un papel muy reducido (Monzo, Alías, Morán, et al. 2006 y San-Segundo, Montero, Córdoba, et al. 2000); contamos también con trabajos de orientación lingüística sobre la transcripción fonética automática del español de México (Cuétara 2004 y Uruga & Pineda 2002) y de Chile (Pérez & Armstrong 1998) y con una herramienta

interactiva de transcripción en español que puede probarse en López Morrás (2004).

El programa SEGRE, por su parte, realiza la transcripción fonética automática de textos en catalán tomando en consideración la variación dialectal (Pachès, Mota, Riera, Perea, Febrer, Estruch, Garrido, Machuca, Ríos, Llisterri, Esquerria, Hernando, Padrell & Nadeu 2000), mientras que en gallego se han desarrollado módulos de transcripción integrados en el conversor de texto en habla Cotovía (González González 2004) y sistemas más genéricos útiles para otras aplicaciones (Braga & Coelho 2006). Los trabajos sobre el portugués han considerado tanto las necesidades de la conversión de texto en habla (Barbosa, Pinto, Resende Jr., Gonçalves, Monserrat & Rosa 2003, Braga, Coelho & Resende Jr. 2007, Oliveira, Moutinho & Teixeira 2004 y Oliveira, Moutinho & Teixeira 2005) como las derivadas de la transcripción de corpus orales (Candeias & Perdigão 2008); también, en este caso, las últimas aproximaciones se han basado más en el aprendizaje a partir de corpus y en modelos estadísticos (Barros & Weiss 2006 y Paulo, Oliveira, Mendes, Figueira, Cassaca, Viana & Moniz 2008) que en las reglas concebidas por lingüistas propias de los primeros modelos (Albano & Moreira 1996).

La necesidad de disponer de un alfabeto fonético para usos informáticos ha llevado a realizar adaptaciones de algunos sistemas ya existentes para el tratamiento computacional del habla; SAMPA (*Speech Assessment Methods Phonetic Alphabet*), inicialmente concebido por Wells (2005) como parte de los estándares europeos para la representación fonológica de corpus orales, se ha adaptado al español peninsular (Llisterri & Mariño 1993), al español de América (Moreno & Mariño 1998), al catalán (Llisterri 1995), al gallego (Braga & Coelho 2006 y Losada 2004) y al portugués (Wells 1997); en el ámbito americano, MexBet (Cuétara 2004) constituye la adaptación de WorldBet (Hieronymus 1994) a las necesidades específicas de la transcripción de corpus en español de México.

El tercer módulo de un conversor de texto en habla en el que se realiza un tratamiento lingüístico es el que se encarga de asignar la información prosódica necesaria para una buena oralización del texto. Por su notable incidencia en la naturalidad de la síntesis éste es, precisamente, uno de los aspectos que más atención ha recibido en los últimos años. Los modelos prosódicos concebidos para la conversión de texto en habla se centran, en general, en los aspectos relacionados con la entonación: la delimitación de unidades prosódicas o *phrasing* (Rodríguez Crespo, Escalada, Armenta & Garrido 2008 y Viana, Oliveira & Mata 2003), las variaciones en duración segmental en español (Córdoba, Montero, Gutiérrez Arriola, Vallejo, Enríquez & Pardo 2002, Córdoba, Vallejo, Montero, Gutiérrez Arriola, López Carmona & Pardo 1999, Meza, Kirschning & Cervantes 2000 y Santos, Muñoz & Martínez Martín 1988), en catalán (Febrer, Padrell & Bonafonte 1998), en gallego (Fernández Salgado & Rodríguez Banga 1999) o en portugués (Teixeira & Freitas 2003) o la asignación de pausas (Adell, Bonafonte & Escudero 2007, Agüero & Bonafonte 2003, Barbosa 1997 y Puigvi, Jiménez & Fernández 1994), aunque en ocasiones contemplan también parámetros menos estudiados como la intensidad (Blecua & Acín 1995) y fenómenos globales como el ritmo (Barbosa 1999).

La diversidad de modelos prosódicos empleados hace inviable realizar aquí una revisión detallada, por lo que remitimos al lector a Hirschberg (2006) o a Llisterri, Machuca, de la Mota, Riera & Ríos (2003). Como es habitual en un campo interdisciplinar, encontramos aproximaciones que incorporan conocimiento lingüístico tanto en español (Garrido 1996 y Garrido, Ortín, Quazza, Salza & Mancini 2000) como en portugués (Freitas & Braga 2002), frente a las que adoptan un enfoque más computacional, sea en español (Escudero & Cardeñoso 2006, Escudero & Cardeñoso 2007, Escudero 2002, López Gonzalo & Hernández Gómez 1995 y López Gonzalo, Álvarez Cercadillo & Hernández Gómez 1994), en gallego (Campillo & Rodríguez Banga 2005, Campillo, van Santen & Rodríguez Banga 2006, Fernández Salgado & Rodríguez Banga 2000, Fernández Salgado & Rodríguez Banga 2000 y López Gonzalo, Villar & Hernández Gómez 2002) o en portugués (Teixeira 2004 y Teixeira, Freitas & Fujisaki 2005).

En relación con los aspectos prosódicos, cabe destacar la importancia que recientemente ha cobrado la síntesis del habla con emociones, también denominada síntesis del habla expresiva, tanto en español (Adell, Bonafonte & Escudero 2005, Esquerria 2006, Iriondo, Socoró & Alías 2007, Iriondo, Socoró, Formiga, Gonzalvo, Alías & Miralles 2006, Montero, Gutiérrez Arriola, Colás, Macías, Enríquez & Pardo 1999 y Montero, Gutiérrez Arriola, Palazuelos, Enríquez, Aguilera & Pardo 1998) como en catalán (Iriondo, Alías, Melenchón & Llorca 2004) o en portugués (Cabral & Oliveira 2006). Muchos de los trabajos suelen centrarse en las emociones conocidas como *The Big Six* -felicidad, tristeza, miedo, asco, enfado y sorpresa (Cornelius 2000)- pero, aun así, la definición de cada una de ellas no deja de ser problemática. Otra de las cuestiones que se plantean es la dificultad de obtener un corpus realista; por lo general, se ha recurrido a actores o a locutores profesionales -con el consiguiente peligro de la sobreactuación-, o se ha intentado, como en Iriondo (2008), establecer una relación entre emociones y enunciados publicitarios. Este interés por la expresividad en el campo de las tecnologías del habla no siempre ha ido acompañado por estudios fonéticos que detallen sus correlatos prosódicos, por lo que buena parte de la información disponible en las lenguas que estamos tratando procede de la aplicación de técnicas de aprendizaje y de clasificación automáticas a corpus tal vez no demasiado espontáneos.

Finalmente, no debe olvidarse que, en las últimas décadas, las tradicionales técnicas de síntesis por concatenación de unidades de pequeño tamaño como los difonemas, basadas en conjuntos de reglas y en un modelo de fuente y filtro para la parametrización acústica de las unidades -uno de los primeros sistemas de este tipo desarrollado para el español fue el que se describe en Santos (1985), mientras que, para el portugués, deben recordarse las primeras versiones de DIXI (Oliveira, Viana & Trancoso 1991)- se han visto desplazadas por la síntesis a partir de corpus o síntesis por selección de unidades. La síntesis se basa, en este caso, en la elección de las unidades óptimas, sea cual sea su longitud, procedentes de un corpus previamente grabado por un locutor profesional y adecuadamente segmentado (Torres & Gurlekian 2008) y etiquetado, generalmente de modo automático (Adell & Bonafonte 2004).

Aunque se necesitan también un módulo de transcripción fonética automática y un tratamiento específico de la entonación para los patrones melódicos que no están bien representados en el corpus, el papel que se otorga al

conocimiento lingüístico es mucho menor, puesto que el peso de la investigación se ha desplazado hacia los algoritmos que proporcionan la mejor selección de unidades y la concatenación óptima para lograr una síntesis más natural. Aun así, la elección del locutor, la supervisión de la grabación, el diseño del corpus, la preparación de modelos prosódicos cuando son necesarios y la evaluación del resultado final son tareas que los lingüistas podrían llevar a cabo adecuadamente.

También partiendo del entrenamiento automático a partir de un corpus se han desarrollado sistemas de conversión de texto en habla basados en Modelos Ocultos de Markov (HMM, *Hidden Markov Models*), una técnica de modelado habitualmente empleada en el reconocimiento automático del habla (Barra, Yamagishi, Montero, King, Lufti & Macías 2008, Barros, Maia, Tokuda, Resende Jr. & Freitas 2005, Bonafonte & Agüero 2008, Gonzalvo, Socoró, Iriondo, Monzo & Martínez Marroquín 2007 y Maia, Zen, Tokuda, Kitamura & Resende Jr. 2003); como puede suponerse, una vez más el papel del conocimiento lingüístico se ve notablemente reducido con respecto a los primeros sistemas de síntesis por reglas a los que antes aludíamos.

En esta misma línea, se ha explorado la posibilidad de crear rápidamente nuevas voces sin tener que invertir tiempo y recursos en la grabación de un hablante, transformando las características de la voz de un locutor ya existente para que sea percibido como otro diferente; estas investigaciones se centran, especialmente, en el procesamiento de la señal (Barra, Montero, Macías, Gutiérrez Arriola, Ferreiros & Pardo 2007, Erro & Moreno 2007 y Gutiérrez Arriola, Montero, Córdoba & Pardo 2003) sin tener en cuenta, al menos de un modo directo, aspectos fonéticos.

En conjunto, disponemos de sistemas de conversión de texto en habla para cada una de las lenguas que estamos considerando y, en algunos casos, con versiones para las variantes peninsulares y las americanas, tanto del español como del portugués. Por una parte, existen los sistemas comercializados por empresas del sector de las tecnologías del habla -muchas de ellas fuera del ámbito hispanohablante o lusófono- y, por otra, los conversores que se han desarrollado en universidades; a continuación, sin pretender ofrecer un listado completo, se mencionan algunos de ellos.

Entre los primeros, se cuentan, por ejemplo, los conversores de Acapela Group -en español, español norteamericano, portugués europeo y portugués brasileño-, AT&T Labs -en español de América-, Cepstral -en español de América-, CereVoice -en español y en catalán (Garrido, Bofías, Laplaza, Marquina, Aylett & Pidcock 2008)-, E-Speech -en español-, Fonix Speech -en español-, Loquendo -en español peninsular, español de América general, de Argentina, de Chile y de México, catalán, valenciano, gallego, portugués europeo y portugués brasileño (Llisterri, Machuca, Madrigal, Mancini, Massimino, Mota, Riera & Ríos 2004)-, NeoSpeech -en español-, Nuance, -en español peninsular, español de América, catalán, portugués europeo y portugués brasileño-, SoftVoice -en español-, SVOX -en español peninsular, español de América y portugués-, Telefónica I+D -en español peninsular, español americano general y español de Perú, catalán, gallego, portugués europeo y portugués brasileño (Armenta, Escalada, Garrido & Rodríguez Crespo 2003, Escalada, Armenta & Rodríguez Crespo 2008, Rodríguez Crespo, Escalada & Armenta 2008 y Rodríguez Crespo, Escalada & Torre 1998)-, Verbio -en español peninsular, de Argentina y de México, catalán, valenciano, gallego y portugués- y VoiceInteraction -en portugués europeo-. A todos ellos cabría añadir los múltiples productos, creados por otras empresas, a los que se incorporan las voces de algunos de los sistemas que acabamos de señalar.

Entre los desarrollados en las universidades, cabe mencionar, a modo de ejemplo, los sistemas de conversión de texto en habla creados por el centro TALP (Tecnologies i Aplicacions del Llenguatge i de la Parla) de la Universitat Politècnica de Catalunya -en español y en catalán (Bonafonte & Agüero 2008, Bonafonte, Esquerra, Febrer & Vallverdú 1997 y Bonafonte, Esquerra, Febrer, Fonollosa & Vallverdú 1998)-, el Área de Tecnologías de la Parla de la Universitat Ramon Llull -en español (Gonzalvo, Socoró, Iriondo, et al. 2007, Iriondo, Socoró & Alías 2007, Martí & Niñerola 1987 y Monzo, Formiga, Adell, Iriondo, Alías & Socoró 2008) y en catalán (Iriondo, Alías, Melenchón, et al. 2004)-, el Grupo de Teoría de la Señal de la Universidade de Vigo en colaboración con el Centro Ramón Piñeiro y el Departamento de Filología Galega de la Universidade de Santiago de Compostela -Cotovia, en español y en gallego (Fernández Rei & González González 1998, González González 2004, González González, Losada & Fernández Rei 1999, Rodríguez Banga, Campillo, Fernández Rei & Méndez 2002 y Rodríguez Banga, Méndez, Campillo, Iglesias & Docío 2008)-, el Grupo de Tecnología del Habla de la Universidad Politécnica de Madrid -en español (Montero, Gutiérrez Arriola, Colás, et al. 1999, Montero, Gutiérrez Arriola, Palazuelos, et al. 1998 y Santos, Olabe, Muñoz, et al. 1985)- o el grupo Aholab de la Universidad del País Vasco -AhoTTS, en español (Sainz, Hernández, Navas, Sánchez, Luengo, Saratxaga, Odriozola, Bilbao & Erro 2008)-.

En América, en lo que respecta al español, se han ocupado de la conversión de texto en habla, entre otros, el Grupo de Investigación en Ciencias Fonética de la Universidad de Los Andes (Mora & Rodríguez Hourcadette 2005, Rodríguez Hourcadette & Mora 2006 y Rodríguez Hourcadette, Mora & Cavé 2006) y el Laboratorio de Investigación en Tecnologías de Voz "Tlatoa" de la Universidad de las Américas (Flores, Vargas, Olivier, Kirschning & Cervantes 2001).

Por lo que se refiere al portugués, pueden señalarse, por ejemplo, los conversores de texto en habla creados por el Laboratório de Sistemas de Língua Falada del INESC-ID (Instituto de Engenharia de Sistemas e Computadores - Investigaçao e Desenvolvimento) -DIXI, en portugués europeo (Oliveira, Viana & Trancoso 1991 y Paulo, Oliveira, Mendes, et al. 2008)-, la Faculdade de Engenharia de la Universidade do Porto (Teixeira, Freitas, Gouveia, Olászy & Németh 1998), el Laboratório de Processamentos de Sinais de la Universidade Federal do Rio de Janeiro (Barros, Maia, Tokuda, et al. 2005, Silva, Resende Jr. & Netto 2001 y Torres, Seixas, Netto, Freitas & Brasil 2008) y por la Universidade Estadual de Campinas en colaboración con la Pontificia Universidade Católica de São Paulo -Aiuruê, en portugués brasileño (Barbosa, Violaro, Albano, Simões, Aquino, Madureira & Françoço 1999 y Violaro, Barbosa, Albano & Françoço 1996)-.

Finalmente, entre las herramientas de dominio público que permiten desarrollar un conversor, cabe destacar las adaptaciones de Festival (University of Edinburgh) y de FestVox (Carnegie Mellon University) al español y al catalán, así como los sistemas en español peninsular, de México y de Venezuela y en portugués europeo y brasileño derivados del proyecto MBROLA (Faculté Polytechnique de Mons).

3. El reconocimiento automático del habla

Un sistema de reconocimiento automático del habla (ASR, *Automatic Speech Recognition*) tiene como misión convertir la señal sonora del habla en una representación simbólica que, habitualmente, será un texto escrito (Juang & Rabiner 2006, Martins 2001 y Torres 2006). Para ello es necesario realizar, en primer lugar, una extracción de los parámetros acústicos relevantes (Mariño & Nadeu 2004) de modo que, a partir de las características acústicas de la señal, se puedan identificar las unidades en las que se basa el reconocedor y determinar, concatenando estas unidades con la ayuda de un diccionario de pronunciación y de un modelo de lenguaje, el enunciado que ha emitido el usuario.

Un reconocedor requiere un entrenamiento previo, que se realiza con un corpus de señal sonora segmentada y etiquetada para la creación de los modelos acústicos y con un corpus textual que se emplea para la creación de los modelos lingüísticos. Por modelo acústico se entiende la información relativa a los parámetros acústicos que necesitará el reconocedor para identificar una unidad -por ejemplo, un alófono, un difonema, un trifenema o una unidad mayor-, mientras que el modelo de lenguaje determina, esencialmente, la probabilidad de aparición de una palabra en función de las anteriores y las siguientes.

Por este motivo, la creación de corpus orales con un número cada vez mayor de hablantes y obtenidos en diversas condiciones ambientales está estrechamente ligada al desarrollo de los sistemas de reconocimiento automático del habla, lo que explica los esfuerzos dedicados a este ámbito (Llisterri, Machuca, Mota, Riera & Ríos 2005). Puesto que los recursos lingüísticos no son objeto de este trabajo, remitimos al lector a los catálogos en línea de ELDA (*Evaluations and Language Resources Distribution Agency*)⁵ en el contexto europeo y del LDC (*Linguistic Data Consortium*)⁶ en el americano para obtener más información sobre los materiales disponibles. Igualmente constituyen una referencia ineludible las actas de los congresos internacionales conocidos como LREC (*Language Resources and Evaluation*) organizados bianualmente desde 1998 por ELRA (*European Language Resources Association*)⁷.

El conocimiento fonético debería ocupar un lugar relevante en el desarrollo de este tipo de corpus en varios sentidos. En primer lugar, el corpus debe contener, en un número suficientemente representativo, los segmentos -fonemas y alófonos- de la lengua, así como sus posibles combinaciones, de modo que sea posible la creación de modelos acústicos adecuados; esto exige, al igual que en la conversión de texto en habla (§ 2), tomar decisiones sobre el inventario de unidades fónicas de la lengua y conocer las restricciones que operan sobre la distribución de las unidades en función de factores como la estructura silábica. Por otra parte, la variante geográfica, el estilo de habla o los acentos no nativos recogidos en el corpus incidirán en las prestaciones del reconocedor, por lo que es preciso considerar en profundidad los factores relacionados con la variación fonética (Benzeghiba, Mori, Deroo, Dupont, Erbes, Jouvét, Fissore, Laface, Mertins, Ris, Rose, Tyagi & Wellekens 2007).

Un sistema de reconocimiento incorpora también, como se ha indicado, un diccionario de pronunciación (*pronunciation lexicon*) en el que se recogen -mediante un alfabeto fonético apto para usos computacionales como SAMPA (§ 2)- la forma fonética o fonológica canónica, denominada, en ocasiones, *citation form*, con sus posibles variantes. Una de las tareas que suelen encargarse a los expertos en fonética es la revisión de estos diccionarios, ya que de ellos depende una parte importante del reconocimiento.

Pese al papel potencial del conocimiento lingüístico que acabamos de esbozar, el reconocimiento automático del habla se caracteriza por un predominio muy acusado de los modelos probabilísticos, entre los que son clásicos los ya mencionados Modelos Ocultos de Markov (Rabiner & Juang 2006) y, por tanto, por una presencia más bien escasa de expertos procedentes del mundo de la filología. Aun así, en algunos de los trabajos realizados desde la ingeniería se abordan cuestiones de interés fonético como el tratamiento de las variantes de pronunciación (Ferreiros, Macías, Pardo & Villarrubia 1998 y Macías, Ferreiros, Córdoba, Montero, Romeral & Pardo 2003) o la incorporación de los elementos prosódicos al proceso de reconocimiento (Bassi, Becerra Yoma & Loncomilla 2006, Escudero & Cardeñoso 2002, García & Tapias 2000 y Rubio Ayuso & Milone 2002, entre otros).

El reconocimiento automático del habla encuentra una de sus aplicaciones más populares en los programas de dictado automático, que facilitan la composición de cualquier tipo de texto, tanto de un documento redactado con un procesador de textos como de un correo electrónico o de un SMS, sin necesidad de recurrir a un teclado. Entre los programas comerciales que pueden utilizarse en un ordenador personal se encuentra *Naturally Speaking*, desarrollado por la antigua empresa Dragon Systems y propiedad, actualmente, de Nuance, del que sólo existe una versión en español. *Via Voice*, creado por IBM, incorporaba el español en sus primeras versiones, pese a que en la última de ellas, igualmente comercializada a través de Nuance, no se ofrece ni ésta ni ninguna de las demás lenguas que aquí consideramos.

En otras aplicaciones diseñados por empresas -por ejemplo, por IBM en español peninsular y americano y en portugués europeo; por Loquendo, en español peninsular, de Argentina, de Chile y de México, catalán, valenciano, gallego, portugués europeo y portugués brasileño; por Verbio, en once variedades de Español de América, en español peninsular, catalán, gallego, portugués europeo y portugués brasileño- no se ofrece el reconocedor como un producto independiente para el usuario final, sino que se integra lo que se denomina un motor de reconocimiento en un conjunto de programas que permiten automatizar una centralita telefónica o un servicio de información, de ventas o de atención

al cliente. Se han propuesto, también por su interés comercial, prototipos para automatizar la realización de encuestas telefónicas combinando el habla con una interfaz en la web (Toselli, Cubel & Sanchís 2006).

Por otra parte, el reconocimiento puede emplearse en aplicaciones relacionadas con el control de sistemas: para facilitar el uso de un ordenador personal (Callejas, Nouza, Cerva & López-Cózar 2007), en la navegación vocal a través de páginas web y en el manejo de electrodomésticos (Uria, Ortega, Torres, Miguel, Gujarrubia, Buera, Garmendia, Lleida, Aizpuru, Varona, Alonso & Saz 2006), de equipos de música (Fernández Martínez, Ferreiros, Sama, Montero & García Sánchez 2005) o de robots (Tamarit & Martínez Hinarejos 2006).

Al igual que en el caso de la conversión de texto en habla (§ 2), existen herramientas de dominio público como Sphinx (Carnegie Mellon University) o HTK *Speech Recognition Toolkit* (Cambridge University) que permiten desarrollar un reconocedor y que se han empleado con éxito en el ámbito hispanico -un ejemplo reciente para el español de México es Pérez Pavón (2006)- y lusófono.

Un campo que en los últimos años se está desarrollando con rapidez es el de la recuperación y extracción de información de grandes archivos sonoros. Puesto que en colecciones como las de una emisora de radio o de televisión no es factible realizar una transcripción manual de las grabaciones almacenadas, se emplean técnicas basadas en el reconocimiento automático del habla para transcribir automáticamente los programas -con las dificultades derivadas de la diversidad de estilos de habla y de los diferentes entornos en los que se han realizado las grabaciones-, localizar palabras clave o buscar fragmentos relacionados con un tema. Este tipo de aplicación podría ejemplificarse con el sistema Transcrigal, desarrollado para el español y el gallego (Diéguez, García Mateo & Cardenal 2005), o con los sistemas descritos en Amaral, Meinedo, Caseiro, Trancoso & Neto (2006) o en Meinedo & Neto (2003) para el portugués. En un terreno cercano al que estamos discutiendo, otro uso de las tecnologías de reconocimiento vendría dado por la posibilidad de consultar mediante el habla grandes bases de datos para obtener la información requerida, sea en forma oral (González Ferreras & Cardeñoso 2006) o escrita (Luque, Ferrés, Hernando, Mariño & Rodríguez 2006). También la transcripción automática de clases y conferencias ha sido objeto de atención en trabajos como los de Trancoso, Nunes, Neves, Viana, Moniz, Caseiro & Mata (2006) para el portugués.

Finalmente, en el contexto de los sistemas de diálogo (§ 4) ha surgido la necesidad de detectar la actitud o el estado cognitivo del usuario para poder, así, mejorar las prestaciones del servicio y, si es necesario, desviar la llamada a un operador humano de modo que los posibles problemas técnicos no provoquen la pérdida de clientes. Esto ha dado lugar a la investigación en el reconocimiento automático de emociones (Barra, Montero, Macías, D'Haro, San-Segundo & Córdoba 2006 y Nogueiras, Moreno, Bonafonte & Mariño 2001, entre otros), que se complementa con los trabajos realizados para lograr la generación de habla expresiva (§ 2).

3.1 El reconocimiento del hablante

Un ámbito estrechamente relacionado con el que nos ocupa en esta sección es el reconocimiento del hablante, en el que se suelen distinguir dos situaciones diferentes: la identificación (ASI, *Automatic Speaker Identification*), cuando se compara la voz de una persona desconocida con un conjunto de voces conocidas para tratar de determinar si ésta coincide con alguna de ellas y la verificación (ASV, *Automatic Speaker Verification*), en la que la voz de un hablante desconocido que dice ser un individuo determinado se compara con un grupo de voces entre las que se encuentra la de tal individuo (Furui 2006, Nolan 1983 y Nolan 1997).

Las tecnologías de reconocimiento del locutor, especialmente las de verificación, tienen un gran interés para los servicios telefónicos que exigen comprobar la identidad de un cliente, como es el caso de la banca telefónica, o en el campo, cada vez más relevante dadas sus implicaciones para la seguridad, de las técnicas biométricas de identificación de personas en entornos como aeropuertos o edificios de acceso restringido.

Por su parte, la identificación automática del locutor es pertinente en el ámbito judicial, cuando se trata de establecer la identidad de una persona de la que se dispone de una grabación, comparándola con una base de datos de voces previamente almacenadas. Una metáfora bien conocida es que la voz sería equivalente a una huella digital por lo que, en principio, puede constituir un elemento de gran importancia a la hora de identificar a una persona si su voz se encuentra registrada en una base de datos policial. También se emplea este tipo de técnicas para comparar las voces que aparecen en una grabación -una intervención telefónica ordenada por un juez, por ejemplo- con las de los sospechosos que han sido detenidos y determinar si corresponden a las mismas personas. La acústica judicial y la fonética judicial -o acústica forense y fonética forense- son especialidades que han adquirido un notable desarrollo en España, tal como se refleja en las actividades de la SEAF (Sociedad Española de Acústica Forense), a cuyas actas de congresos remitimos al lector para obtener una perspectiva más detallada (González González, Fernández Rei & González Rei 2007, Hernando 2003 y Ortega García 2000).

Es importante señalar que, pese a que no se trate de una tendencia mayoritaria, algunos trabajos realizados en las lenguas que estamos tratando consideran la importancia del conocimiento fonético en la identificación y la verificación automáticas del hablante, tanto desde una perspectiva metodológica (Cerdà, Farrús & Hernando 2005) como en lo que se refiere, especialmente, al uso de información relacionada con los elementos suprasegmentales (Escudero, Cardeñoso, Sánchez, Navas & Hernáez 2003, Farrús, Hernando & Ejarque 2007, Fernández Pozo, Fombella, Torre, López Gonzalo & Hernández Gómez 2006 y Montero, González Domínguez, Ramos, López Moreno, Torre & González Rodríguez 2006, entre otros).

3.2 La identificación automática de la lengua

El acceso cada vez más habitual a sistemas automáticos de atención telefónica y el creciente multilingüismo de nuestras sociedades han dado lugar a que se plantee la necesidad de identificar la lengua en la que se expresa un usuario, de modo que la interacción con el sistema no deba iniciarse con un menú para la selección del idioma. La tecnología que permite detectar la lengua empleada se conoce como identificación automática de la lengua (LID, *Automatic Language Identification*) y suele basarse tanto en la aparición de un determinado conjunto de alófonos como en el conocimiento estadístico de las restricciones fonotácticas, ya que existen combinaciones de alófonos que pueden presentar una alta frecuencia de aparición en una lengua dada y, en cambio, ser poco frecuentes en otra.

Una muestra de esta tecnología se encuentra, por ejemplo, en Adda-Decker, Fabien, Boula de Mareuil, Vasilescu, Lamel, Vaissière, Geoffrois & Liénard (2003), donde se presenta un experimento que incorpora el español y el portugués; por su parte, en aproximaciones como la propuesta por Córdoba, D'Haro, San-Segundo, Macías, Fernández & Plaza (2006) para el español y el inglés se pretende modelar la frecuencia de aparición de las secuencias de alófonos en varias lenguas, combinando esta información con un modelo de lenguaje en el que se ha especificado la probabilidad de que la frecuencia de aparición de una secuencia pueda encontrarse en una lengua dada; estas técnicas se utilizan conjuntamente con la información acústica procedente de la señal del habla.

Otra de las aplicaciones interesantes de la identificación automática de la lengua es la diferenciación no sólo entre lenguas, sino también entre variantes geográficas que aparecen en un mismo corpus; así, Rouas, Trancoso, Viana & Abreu (2008) utilizan datos acústicos, fonotácticos y prosódicos para diferenciar el portugués de otras lenguas y las variedades del portugués entre sí en noticiarios de televisión.

4. Los sistemas de diálogo

Si hasta ahora nos hemos referido de forma independiente a la salida y a la entrada vocales, éstas se conjugan en los sistemas de diálogo o sistemas conversacionales (SLS, *Spoken Language Systems*) (Llisterri & Machuca 2006 y López-Cózar & Araki 2005) para facilitar la interacción oral entre personas y máquinas, principalmente través del teléfono (Tapias & Hernández Gómez 2004), pero también mediante un dispositivo portátil o una interfaz en la web. Alcanzar tal objetivo requiere sistemas que, además del reconocimiento y de la síntesis, dispongan de procedimientos que extraigan la información necesaria de los enunciados del usuario, gestionen la interacción con el ordenador y generen una respuesta adecuada.

En un sistema de diálogo, una vez ha actuado el módulo de reconocimiento automático del habla (§ 3) para convertir la señal sonora en una representación textual, se realiza un proceso encaminado a determinar el contenido del enunciado; esto se lleva a cabo en un módulo centrado en la interpretación semántica, en el que se aplican las técnicas propias del ámbito conocido como comprensión del habla (SLU, *Spoken Language Understanding*) (Segarra 2006). No se trata, evidentemente, de una *comprensión* en el sentido humano del término, sino de una extracción de los datos relevantes en el dominio de la aplicación en la que se desarrolla el diálogo. Es preciso, también, contar con un módulo de gestión del diálogo (López-Cózar 2006) que se encargue de los aspectos relacionados con la interacción entre el usuario y el sistema, determinando la alternancia de turnos de palabra, recordando la historia de la conversación para poder interpretar los elementos anafóricos y, en conjunto, haciendo que la persona llegue al objetivo deseado -obtener una información o realizar una transacción- del modo más natural y eficaz posible.

Finalmente, un paso previo a la conversión de texto en habla que proporciona una respuesta oral es la propia generación del texto de esta respuesta; para ello se utilizan las tecnologías del generación del lenguaje natural (NLG, *Natural Language Generation*) (Lavid 2006) mediante las que, a partir de la representación abstracta de la información recogida en bases de conocimiento externas, se crea un enunciado gramaticalmente bien formado que, posteriormente, se enviará al módulo de conversión de texto en habla (§ 2).

Tras los primeros trabajos realizados a finales de los años noventa -entre los que se cuentan, por ejemplo, los de Álvarez, Crespo, Escalada, Tapias & López Soto (1996), de López-Cózar & Rubio (1997) o de Tapias, Álvarez & Cortázar (1997)-, el desarrollo de los sistemas de diálogo en España fue impulsado por proyectos llevados a cabo coordinadamente entre varios equipos de investigación universitarios, como BASURDE (Sistema de diálogo oral en dominios restringidos) (Bonafonte, Aibar, Castell, Lleida, Mariño, Sanchis & Torres 2000) y DIHANA (Sistema de diálogo para el acceso a la información mediante habla espontánea en diferentes entornos) (Benedí, Varona & Lleida 2004), en los que se sentaron las bases para la integración de la conversión de texto en habla, el reconocimiento automático del habla, la comprensión y la gestión del diálogo en el contexto de los servicios de información. Ello permitió los avances posteriores en ámbitos específicos y en nuevos contextos de aplicación en el marco de proyectos, también coordinados, como ALIADO (Tecnologías del habla y el lenguaje para un asistente personal) (Mariño & Rodríguez 2003), centrado en el empleo de tecnologías del habla multilingües, EDECÁN (Sistema de diálogo multidominio con adaptación al contexto acústico y de aplicación) (Lleida, Segarra, Torres & Macías 2006), en el que se investiga la adaptación al entorno, o como URBANO (Integración de robots autónomos en la sociedad mediante el uso de nuevas tecnologías) y ROBINT (Integración de comportamientos inteligentes en robots guía) (Lucas, Alcázar, Montero, Fernández Martínez, Barra, D'Haro, Ferreiros, Córdoba, Macías, San-Segundo & Pardo 2008), en los que se ha mostrado la incorporación de un sistema de diálogo y de un conversor de texto en habla con emociones a un robot que atiende a los visitantes de un museo.

Entre otros ámbitos de aplicación de los sistemas de diálogo que se han desarrollado en las lenguas de nuestro

interés pueden mencionarse, a modo de ejemplo, para el español, la automatización de centralitas y de servicios de información telefónica (Casacuberta, Martínez, Nevado & Vidal 2000, Córdoba, San-Segundo, Montero, Colás, Ferreiros, Macías & Pardo 2001 y Relaño, Villarrubia, Rodríguez Gancedo & Hernández Gómez 2003), la ayuda al control del tráfico aéreo (Sama, Fernández Martínez, Ferreiros, Macías, Córdoba, Montero, Colás, Campos & Pardo 2003) y el acceso mediante el habla a información sobre viajes en tren (López-Cózar & Granell 2004 y San-Segundo, Montero, Ferreiros, Macías & Pardo 2001) o a los contenidos de un periódico (González Ferreras & Cardeñoso 2005); en catalán se dispone de un sistema de información meteorológica por teléfono y por SMS que cuenta con una amplia base de usuarios (Hernando, Padrell & Rodríguez 2002 y Padrell & Hernando 2002) mientras que en gallego se ha diseñado, por ejemplo, un sistema bilingüe para la consulta por teléfono del correo electrónico (Rodríguez, García Mateo, Pardo & Darriba 2002); cabe también señalar, para el portugués, entre otros proyectos, el desarrollo de sistemas de diálogo de uso genérico (Neto, Mamede, Cassaca & Oliveira 2003) o para transacciones como la venta de entradas (Filipe & Mamede 2007).

4.1 La traducción automática del habla

Aunque al iniciar esta revisión situábamos las tecnologías del habla en el contexto de la interacción entre personas y sistemas informáticos, debe apuntarse también que algunas de sus aplicaciones facilitan la comunicación entre personas utilizando un ordenador como intermediario (CMC, *Computer-Mediated Communication*). Éste es el caso de la traducción automática del habla (SLT, *Spoken Language Translation*) (Casacuberta 2004, Waibel 2004 y Waibel, Bernardin & Wölfel 2007), cuyo objetivo es convertir un enunciado oral en la lengua de partida en su correspondiente traducción, también hablada, a la lengua de llegada, tanto si los interlocutores se encuentran en el mismo lugar como si se trata de una conversación telefónica.

Llevar a cabo la traducción del habla requiere, en primer lugar, un sistema de reconocimiento automático (§ 3) en la lengua de origen, a continuación un módulo que ejecute propiamente la traducción y, finalmente, un sistema de conversión de texto en habla (§ 2) en la lengua de destino. Podría pensarse que una vez el enunciado se ha transformado en un texto, sería factible aplicar las herramientas de traducción automática que ya existen para la lengua escrita; sin embargo, las diferencias en el proceso de planificación entre el discurso oral y el texto hacen difícil este planteamiento. Por ello, la investigación de los últimos años ha partido de corpus paralelos -es decir, con idéntico contenido en las dos lenguas- para entrenar sistemas estadísticos que realicen la traducción; como es fácil suponer, se trata nuevamente de un modo de trabajar muy alejado del conocimiento lingüístico que hasta no hace mucho tiempo había caracterizado algunas de las aproximaciones a la traducción automática de textos.

Entre los proyectos que han impulsado la traducción automática del habla cabe mencionar los pioneros EUTRANS (*Example-based Language Translation Systems*) y EUTRANS-I (*Example-based Understanding and Translation Systems*) (Pastor, Sanchis, Casacuberta & Vidal 2000) y, más recientemente, TC-STAR (*Technology and Corpora for Speech-to-Speech Translation*), que ha permitido crear recursos, herramientas y prototipos en catalán y en español, además de en otras lenguas; también los proyectos ALIADO (Tecnologías del habla y el lenguaje para un asistente personal) (Mariño & Rodríguez 2003), FAME (*Facilitating Agent for Multicultural Exchange*) (Arranz, Comelles & Farwell 2005), AVIVAVOZ (Tecnologías para la traducción de voz: reconocimiento, traducción estadística basada en corpus y síntesis) (Mariño 2006), que incorpora el gallego, y TECNOPARLA (Tecnologías de la parla en català) (Schulz, Costa-Jussà & Fonollosa 2008) han contribuido a que se avance notablemente en este campo. Otros ejemplos de sistemas de traducción automática del habla pueden encontrarse, por ejemplo, en Pérez, Torres, Guijarrubia & Casacuberta (2006), que describe un procedimiento de traducción de partes meteorológicas del español al inglés y al vasco, o en Sánchez, Martínez, Neto & Caseiro (2008), donde se aborda la traducción de noticias del español al portugués.

Es especialmente interesante mencionar la creación de prototipos para a la traducción del español hablado a la lengua de signos española; éste es el caso de un proyecto, orientado a la atención en los servicios públicos, realizado en colaboración entre el Grupo de Tecnología del Habla de la Universidad Politécnica de Madrid y la Fundación CNSE (Confederación Estatal de Personas Sordas) por la Supresión de las Barreras de Comunicación (Gallo, San-Segundo, Lucas, Barra, D'Haro & Fernández Martínez 2008a y Gallo, San-Segundo, Lucas, Barra, D'Haro & Fernández Martínez 2008b).

4.2 La multimodalidad

Debido al auge de la web y a la creciente sofisticación de los dispositivos móviles, una tendencia importante en las tecnologías del habla ha sido la incorporación de la multimodalidad, entendida como el uso simultáneo del habla y de la información visual, plasmada en un rostro o una figura animada, conocida como avatar, cuyos movimientos faciales o corporales están sincronizados con los enunciados que produce. Esta posibilidad ha dado lugar al surgimiento de los agentes conversacionales (ECA, *Embodied Conversational Agents*) o agentes animados interactivos, que constituyen un paso más en los sistemas de diálogo exclusivamente orales. Naturalmente, en su desarrollo intervienen equipos interdisciplinarios, que integran a especialistas en tratamiento digital de la imagen y a expertos en tecnologías del habla.

Una muestra de este tipo de trabajo la constituye Sam, un meteorólogo virtual que ofrece información del tiempo en catalán y en español mediante conversión de texto en habla (Alías, Iriando, Formiga, Gonzalvo, Monzo & Sevillano 2005), creado por la Universitat Ramon Lull y la Universitat Pompeu Fabra con la colaboración de la Corporació

Catalana de Ràdio i Televisió. También la empresa barcelonesa Umanify desarrolla asistentes virtuales para integrarlos en teléfonos móviles, combinando el reconocimiento automático del habla con la conversión de texto en habla, de modo que puede llevarse a cabo una interacción completamente oral con el asistente.

La utilidad de los sistemas de diálogo multimodales (MDS, *Multimodal Dialog Systems*) (López-Cózar & Araki 2005) podría encontrarse igualmente en los entornos educativos, como ayuda a estudiantes y profesores, tal como se plantea en el sistema DS-UCAT (López-Cózar, Callejas, Montoro & Haya 2006), en el uso interactivo de programas de diseño asistido por ordenador así como en la creación de un robot conversacional, como se ejemplifica para el español de México en el proyecto DIME (Diálogos inteligentes multimodales en español) (Pineda 2008 y Pineda, Massé, Meza, Salas, Schwarz, Uraga & Villaseñor 2002), o en el empleo de agentes conversacionales en la videoconferencia a través de teléfonos móviles (López Mencía, Hernández Trapote, Díaz Pardo, Torre, Hernández Gómez & López Gonzalo 2006).

5. Consideraciones finales

Dada la amplitud del campo, la diversidad de ámbitos de trabajo y el considerable número de equipos de investigación implicados, tanto en universidades como en empresas de Europa y de América, no es posible presentar aquí una panorámica razonablemente completa de las tecnologías del habla en catalán, español, gallego y portugués, las cuatro lenguas ibéricas que hemos pretendido abordar. Esta aproximación, necesariamente parcial, a algunos aspectos que hemos creído relevantes desde el punto de vista de la lingüística muestra que, aunque las tecnologías del habla no hayan alcanzado el mismo grado de desarrollo para todas las lenguas ibéricas ni, en conjunto, encontremos todavía los mismos productos y servicios que en inglés, los avances en los últimos años han sido muy importantes, gracias a la existencia de una comunidad investigadora activa y con una notable presencia internacional. Iniciativas como la mencionada Red Temática en Tecnologías del Habla, nacida el año 2000, o la creación en 2005 del SIG (*Special Interest Group*) sobre lenguas ibéricas⁸ en el seno de ISCA (*International Speech Communication Association*), que han llevado a la organización de varias reuniones científicas y a la publicación de un volumen monográfico de la revista *Speech Communication* (Trancoso, Becerra Yoma, Barbosa, San-Segundo & Paliwal 2008), permiten constatar la indudable vitalidad de las tecnologías del habla tanto en la Península Ibérica como en América.

Sin embargo, como hemos ido destacando a lo largo del trabajo, la tendencia a reemplazar el conocimiento lingüístico por los enfoques estadísticos se ha reforzado en los últimos años. Basta observar las publicaciones más recientes para constatar que casi todas ellas aparecen firmadas por autores vinculados a grupos de ingeniería o de informática, con una escasísima participación de lingüistas pese a que se traten de resolver problemas estrechamente relacionados con la estructura y el uso del lenguaje y del habla. Las causas de este desencuentro obedecen tanto a motivos científicos -o, en más de una ocasión, comerciales- como a las diferentes tradiciones académicas (Llisterri 2007b). Aunque puedan argumentarse y comprenderse las razones por las que una colaboración, que en algunos momentos ha sido estrecha y lo sigue siendo entre algunos equipos, no parece producirse con la frecuencia que sería deseable, la realidad es que las necesidades del desarrollo tecnológico no parecen impulsar, en general, la adquisición de conocimiento lingüístico, ya que el interés por crear rápidamente nuevos sistemas de la forma más automatizada posible no favorece el análisis lingüístico detallado ni requiere la participación de expertos de nuestro campo. Análogamente, la orientación muy marcada por las teorías vigentes de algunas de las investigaciones en el ámbito de la fonética tampoco facilita la obtención de modelos que puedan integrarse en un entorno computacional, incluso en áreas en las que la información lingüística podría ofrecer nuevas vías de soluciones a problemas que todavía no están completamente resueltos.

Puede pensarse, por tanto, que las acciones encaminadas a una mayor integración entre la fonética y las tecnologías del habla como las que proponían van Santen (2005) o Barry, van Dommelen & Koreman (2005) siguen manteniendo toda su vigencia, tanto en lo que se refiere a la formación interdisciplinar como a la creación y a la mejora de las tecnologías y de las aplicaciones. Como acertadamente explica Greenberg (2005, p.111) “Speech technology can proudly point to its apparent success with speech recognition and concatenative synthesis in defense of its machine-learning-centric approach. An indeed, imperfect science is capable of providing an effective foundation for technology —as long as the demands of the market are not exceedingly stringent or profound. However, as commercial expectations rise, immature science is unlikely to suffice as the empirical and theoretical foundation of future-generation technology”. Estas consideraciones deberían llevar a los que trabajamos en el ámbito de la fonética a una reflexión sobre la necesidad de estar presentes en un entorno que avanza rápidamente sin dejar, por ello, de ocuparnos del conocimiento del habla como código, que justamente Fant (2005, p.17) destacaba como el nexo de unión entre la fonética y las tecnologías del habla: “The challenge for future research is thus to force the speech code, i.e. to predict the articulatory, acoustic and perceptual manifestation of any utterance given the message transcript and the particular language, dialect, speaker and situational context. [...] In simple words, it is the ultimate aim of general phonetics and linguistics”. Así lograremos, tal vez, conjugar una ciencia madura con el progreso tecnológico.

6. Notas

1 Red Temática en Tecnologías del Habla: <http://lorien.die.upm.es/~lapiz/rtth/>

2 Linguateca, Centro de recursos para o processamento computacional da língua portuguesa: <http://www.linguateca.pt>

3 Tecnovoz, Tecnologia de Reconhecimento e Síntese de Voz: <http://www.tecnovoz.pt>

- 4 ISCA, International Speech Communication Association: <http://www.isca-speech.org>
 5 ELDA, Evaluations and Language Resources Distribution Agency: <http://www.elda.org>
 6 LDC, Linguistic Data Consortium: <http://www.ldc.upenn.edu>
 7 ELRA, European Language Resources Association: <http://www.elra.info>
 8 ISCA SIG on Iberian Languages: <http://www.laps.ufpa.br/il-sig/>

7. Referencias

- Adda-Decker, Martine, Antoine Fabien, Philippe Boula de Mareuil, Ioana Vasilescu, Lori Lamel, Jacqueline Vaissière, Edouard Geoffrois & Jean-Sylvain Liénard. 2003. Phonetic knowledge, phonotactics and perceptual validation for automatic language identification. In Maria Josep Solé, Daniel Recasens & Joaquín Romero (eds.), *ICPhS 2003. Proceedings of the 15th International Congress of Phonetic Sciences*, 747-750. Barcelona, Spain, 3-9 August 2003. Disponible en: <http://www.limsi.fr/Individu/mareuil/publi/PS021162.pdf>
- Adell, Jordi & Antonio Bonafonte. 2004. Towards phone segmentation for concatenative speech synthesis. In *Fifth ISCA Tutorial and Research Workshop on Speech Synthesis*, 139-144. Pittsburgh, PA, USA, June 14-16, 2004. Disponible en: http://gps-tsc.upc.es/veu/research/pubs/download/Ade_Tow_04.pdf
- Adell, Jordi, Antonio Bonafonte & David Escudero. 2005. Analysis of prosodic features: towards modelling of emotional and pragmatic attributes of speech. *Procesamiento del Lenguaje Natural* 35, 277-283. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/35/34.pdf>
- Adell, Jordi, Antonio Bonafonte & David Escudero. 2007. Statistical analysis of filled pauses' rhythm for disfluent speech synthesis. In *Sixth ISCA Tutorial and Research Workshop in Speech Synthesis*, 223-227. Bonn, Germany, August 22-24, 2007. Disponible en: http://gps-tsc.upc.es/veu/research/pubs/download/Ade_Sta_07.pdf
- Agüero, Pablo Daniel & Antonio Bonafonte. 2003. Phrase break prediction: a comparative study. *Procesamiento del Lenguaje Natural* 31, 107-114. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/31/31-Pag107.pdf>
- Ainsworth, William A. 1988. *Speech recognition by machine*. London: P. Peregrinus on behalf of the Institution of Electrical Engineers.
- Albano, Eleonora Cavalcante & Agnaldo Antonio Moreira. 1996. Archisegment-based letter-to-phone conversion for concatenative speech synthesis in Portuguese. In *ICLSP 1996. Proceedings of the 4th International Conference on Spoken Language Processing*, 1708-1711. Philadelphia, PA, USA, October, 3-6, 1996. Disponible en: <http://www.lafape.iel.unicamp.br/Publicacoes/ortofon.pdf>
- Alías, Francesc, Ignasi Iriondo, Lluís Formiga, Xavier Gonzalvo, Carlos Monzo & Xavier Sevillano. 2005. High quality Spanish restricted-domain TTS oriented to a weather forecast application. In *Interspeech 2005 - Eurospeech. Proceedings of the 9th European Conference on Speech Communication and Technology*, 2573-2576. Lisbon, Portugal, September 4-8, 2005. Disponible en: http://www.salle.url.edu/~falias/index_archivos/aliasES05.pdf
- Amaral, Rui, Hugo Meinedo, Diamantino Caseiro, Isabel Trancoso & João Paulo Neto. 2006. Automatic vs. manual topic segmentation and indexation in broadcast news. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), *IV Jornadas en Tecnología del Habla*, 123-128. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: <http://www.inesc-id.pt/pt/indicadores/Ficheiros/3577.pdf>
- Armenta, Ana, José Gregorio Escalada, Juan María Garrido & Miguel Ángel Rodríguez Crespo. 2003. Conversor texto a voz multilingüe de Telefónica I+D. *Procesamiento del Lenguaje Natural* 31, 331-332. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/31/31-Pag331.pdf>
- Arranz, Victoria, Elisabet Comelles & David Farwell. 2005. Sistema de traducción oral para el castellano, catalán e inglés. *Procesamiento del Lenguaje Natural* 35, 343-350. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/35/42.pdf>
- Álvarez Cercadillo, Jorge, Carlos Crespo, José Gregorio Escalada, Daniel Tapias & María Teresa López Soto. 1996. Modelado de lenguaje y gestor de diálogo en un sistema de operador asistido por voz. *Procesamiento del Lenguaje Natural* 19, 179-188. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/19/19-Todo.pdf>
- Barbosa, Filipe, Guilherme Pinto, Fernando Gil V Resende Jr., Carlos Alexandre Gonçalves, Ruth Monserrat & Maria Carlota Rosa. 2003. Grapheme-Phone transcription algorithm for a Brazilian Portuguese TTS. In Nuno J Mamede, Jorge Baptista, Isabel Trancoso & Maria das Graças Volpe Nunes (eds.), *PROPOR 2003. Computational Processing of the Portuguese Language. Sixth International Workshop*, Faro, Portugal, June 26-27, 2003. *Proceedings*, 23-30. Heidelberg: Springer.
- Barbosa, Plínio Almeida. 1997. A model of segment (and pause) duration generation for Brazilian Portuguese text-to-speech synthesis. In *Eurospeech 1997. Proceedings of the 5th European Conference on Speech Communication and Technology*, 2655-2658. Rhodes, Greece, September 22-25, 1997. Disponible en: <http://www.lafape.iel.unicamp.br/Publicacoes/eurosp1997.pdf>
- Barbosa, Plínio Almeida. 1999. Revelar a estrutura rítmica de uma língua construindo máquinas falantes: pela integração entre ciência e tecnologia de fala. In Esther Mirian Scarpa (ed.), *Estudos de Prosódia*, 21-52. Campinas: Editoria da Unicamp. Disponible en: <http://www.unicamp.br/iel/site/docentes/plinio/EstudosProsodia.pdf>
- Barbosa, Plínio Almeida, Fábio Violaro, Eleonora Cavalcante Albano, Flávio Olmos Simões, Patrícia Aquino, Sandra Madureira & Edson França. 1999. Aiuruetê: A high-quality concatenative text-to-speech system for Brazilian Portuguese with demisyllabic analysis-based units and a hierarchical model of rhythm production. In *Eurospeech 1999. Proceedings of the 6th European Conference on Speech Communication and Technology*, 2059-2062.

- Budapest, Hungary, September 5-9, 1999. Disponible en: <http://www.lafape.iel.unicamp.br/Publicacoes/eurosp1999.pdf>
- Barra, Roberto, J Yamagishi, Juan Manuel Montero, S King, S Lufti & Javier Macías. 2008. Generación de una voz sintética en castellano basada en HSMM para la evaluación Albayzín 2008: Conversión texto a voz. In Inmaculada Hernández (ed.), V Jornadas en Tecnología del Habla, 115-118. Bilbao: Universidad del País Vasco - Red Temática en Tecnologías del Habla. Disponible en: <http://www-gth.die.upm.es/research/documentation/AG-068Gen-08.pdf>
- Barra, Roberto, Juan Manuel Montero, Javier Macías, Juana María Gutiérrez Arriola, Javier Ferreiros & José Manuel Pardo. 2007. On the limitations of voice conversion techniques in emotion identification tasks. In Interspeech 2007. Proceedings of the 8th Annual Conference of the International Speech Communication Association, 2233-2236. Antwerp, Belgium, August 27-31, 2007. Disponible en: <http://www-gth.die.upm.es/research/documentation/AI-104Ont-07.pdf>
- Barra, Roberto, Juan Manuel Montero, Javier Macías, Luis Fernando D'Haro, Rubén San-Segundo & Ricardo Córdoba. 2006. Prosodic and segmental rubrics in emotion identification. In ICASSP 2006. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 1085-1088. Toulouse, France, 14-19 May 2006. Disponible en: <http://www-gth.die.upm.es/research/documentation/AG-39Pro-06.pdf>
- Barros, Maria João & Christian Weiss. 2006. Maximum Entropy motivated grapheme-to-phoneme, stress and syllable boundary prediction for Portuguese text-to-speech. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), IV Jornadas en Tecnología del Habla, 177-182. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Barros, Maria João, Ranniery Maia, Keiichi Tokuda, Fernando Gil V Resende Jr. & Diamantino Rui da Silva Freitas. 2005. HMM-based European Portuguese TTS system. In Interspeech 2005 - Eurospeech. Proceedings of the 9th European Conference on Speech Communication and Technology, 2581-2584. Lisbon, Portugal, September 4-8, 2005.
- Barry, William J & Wim A van Dommelen (eds.). 2005. The Integration of Phonetic Knowledge in Speech Technology. Dordrecht: Springer.
- Barry, William J, Wim A van Dommelen & Jacques Koreman. 2005. Phonetic knowledge in speech technology - and phonetic knowledge from speech technology?. In William J Barry & Wim A van Dommelen (eds.), The Integration of Phonetic Knowledge in Speech Technology, 1-12. Dordrecht: Springer. Disponible en: <http://www.coli.uni-saarland.de/~koreman/Publications/2005/Eurospeech2001.pdf>
- Bassi, Alejandro, Néstor Becerra Yoma & Patricio Loncomilla. 2006. Estimating tonal prosodic discontinuities in Spanish using HMM. Speech Communication 48,(9), 1112-1125. Disponible en: <http://dx.doi.org/10.1016/j.specom.2006.03.006>
- Benedí, José Miguel, Amparo Varona & Eduardo Lleida. 2004. DIHANA: Sistema de diálogo para el acceso a la información mediante habla espontánea en diferentes entornos. In Emilio Sanchis (ed.), III Jornadas en Tecnología del Habla, 141-146. València: Universitat Politècnica de València - Red Temática en Tecnologías del Habla. Disponible en: <http://www.dihana.upv.es/publicaciones/documento44.pdf>
- Benzeghiba, Mohamed Faouzi, Renato de Mori, Olivier Deroo, Stéphane Dupont, Teodora Erbes, Denis Jovet, Luciano Fissore, Pietro Laface, Alfred Mertins, Christophe Ris, Richard Rose, Vivek Tyagi & Christian Wellekens. 2007. Automatic speech recognition and speech variability: A review. Speech Communication 49,(10-11), 763-786. Disponible en: <http://dx.doi.org/10.1016/j.specom.2007.02.006>
- Blecua, Beatriz & Vanessa Acín. 1995. Propuesta de un modelo de intensidad vocálica del castellano y el catalán aplicable a un sistema de conversión de texto a habla. Procesamiento del Lenguaje Natural 17, 257-271. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/17/17-Pag257.pdf>
- Bonafonte, Antonio & Pablo Daniel Agüero. 2008. The UPC TTS system description. In Inmaculada Hernández (ed.), V Jornadas en Tecnología del Habla, 130-133. Bilbao: Universidad del País Vasco - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/V/pdfs/actas.pdf>
- Bonafonte, Antonio, David Escudero & Montserrat Riera. 2006. La conversión de texto en habla. In Joaquim Llisterri & María Jesús Machuca (eds.), Los sistemas de diálogo, 177-208. Bellaterra - Soria: Universitat Autònoma de Barcelona - Fundació Duques de Soria.
- Bonafonte, Antonio, Ignasi Esquerra, Albert Febrer & Francesc Vallverdú. 1997. A bilingual tex-to-speech system in Spanish and Catalan. In Eurospeech 1997. Proceedings of the 5th European Conference on Speech Communication and Technology, 2455-2458. Rhodes, Greece, September 22-25, 1997.
- Bonafonte, Antonio, Ignasi Esquerra, Albert Febrer, José A R Fonollosa & Francesc Vallverdú. 1998. The UPC text-to-speech system for Spanish and Catalan. In ICLSP 1998. Proceedings of the 5th International Conference on Spoken Language Processing, Sidney Convention Centre, Sidney, Australia, 30 November - 4 December, 1998. Disponible en: <http://gps-tsc.upc.es/veu/research/pubs/download/Bon98c.pdf>
- Bonafonte, Antonio, Pablo Aibar, Núria Castell, Eduardo Lleida, José B Mariño, Emilio Sanchis & María Inés Torres. 2000. Desarrollo de un sistema de diálogo oral en dominios restringidos. In María Teresa López Soto & Ramón López-Cózar (eds.), I Jornadas en Tecnología del Habla, Sevilla: Universidad de Sevilla - Universidad de Granada - Red Temática en Tecnologías del Habla. Disponible en: http://gps-tsc.upc.es/veu/basurde/download/Bon00a_sevilla.pdf
- Bonaventura, Patrizia, Fabio Giuliani, Juan María Garrido & Isabel Ortín. 1998. Grapheme-to-phoneme transcription rules for Spanish, with application to automatic speech recognition and synthesis. In CVIR' 98. Workshop on

- Content Visualization and Intermedia Representations (COLING-ACL' 98. 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics), 33-39. Montreal, Quebec, Canada. Disponible en: <http://acl.ldc.upenn.edu/W/W98/W98-0804.pdf>
- Braga, Daniela & Luís Coelho. 2006. Letter-to-sound conversion for Galician TTS systems. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), IV Jornadas en Tecnología del Habla, 171-176. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Braga, Daniela, Luís Coelho & Fernando Gil V Resende Jr.. 2007. Homograph ambiguity resolution in front-end design for Portuguese TTS systems. In Interspeech 2007. Proceedings of the 8th Annual Conference of the International Speech Communication Association, 1761-1764. Antwerp, Belgium, August 27-31, 2007.
- Buera, Luis, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.). 2006. IV Jornadas en Tecnología del Habla. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Cabral, João P & Luís Caldas de Oliveira. 2006. EmoVoice: A system to generate emotions in speech. In Interspeech 2006 - ICSLP. Proceedings of the 9th International Conference on Spoken Language Processing, 1798-1801. Pittsburgh, PA, USA, September 17-21, 2006. Disponible en: http://homepages.inf.ed.ac.uk/jscabral/artigos/jpc_interspeech_2006.pdf
- Callejas, Zoraida, Jan Nouza, Petr Cerva & Ramón López-Cózar. 2007. MyVoice goes Spanish. Cross-lingual adaptation of a voice controlled PC tool for handicapped people. *Procesamiento del Lenguaje Natural* 39, 227-278. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/39/34.pdf>
- Campillo, Francisco & Eduardo Rodríguez Banga. 2005. Evaluación del modelado acústico y prosódico del sistema de conversión texto-voz Cotovía. *Procesamiento del Lenguaje Natural* 35, 5-12. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/35/01.pdf>
- Campillo, Francisco, Jan P H van Santen & Eduardo Rodríguez Banga. 2006. A model for the f0 reset in corpus-based intonation approaches. In Interspeech 2006 - ICSLP. Proceedings of the 9th International Conference on Spoken Language Processing, 2362-2365. Pittsburgh, PA, USA, September 17-21, 2006.
- Candeias, Sara & Fernando Perdigão. 2008. Conversor de grafemas para fones baseado em regras para português. In Luís Costa, Nuno Cardoso & Diana Santos (eds.), *Linguateca: 10 anos / Actas do encontro na Curia, Linguateca*. Disponible en: <http://www.linguateca.pt/Linguateca10anos/ResumosAlargados/CandeiasPerdigaoL10.pdf>
- Carlson, Rolf & Björn Granström. 1997. Speech synthesis. In William J Hardcastle & John Laver (eds.), *The Handbook of Phonetic Sciences*, 768-788. Oxford: Blackwell.
- Casacuberta, Francisco. 2004. Traducción automática del habla. In Maria Antònia Martí & Joaquim Llisterra (eds.), *Tecnologías del texto y del habla*, 121-144. Barcelona - Soria: Edicions de la Universitat de Barcelona - Fundación Duques de Soria.
- Casacuberta, Francisco, Carlos Martínez, Francisco Nevado & Enrique Vidal. 2000. Implementación de una centralita automática dirigida por voz. In María Teresa López Soto & Ramón López-Cózar (eds.), *I Jornadas en Tecnología del Habla*, Sevilla: Universidad de Sevilla - Universidad de Granada - Red Temática en Tecnologías del Habla.
- Castro, María José, Salvador España, Andrés Marzal & Ismael Salvador. 2001. Transcriptor ortográfico-fonético para el castellano. *Procesamiento del Lenguaje Natural* 27, 241-246. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/27/27-articulo28.pdf>
- Cerdà, Ramon, Mireia Farrús & Javier Hernando. 2005. Hacia una sinergia metodológica en la identificación de locutores. In *Filología y lingüística. Estudios ofrecidos a Antonio Quilis*, 1515-1528. Madrid: Consejo Superior de Investigaciones Científicas - Universidad Nacional de Educación a Distancia - Universidad de Valladolid.
- Chapanis, Alphonse. 1975. Interactive human communication. *Scientific American* 232, 36-42. Disponible en: <http://web.media.mit.edu/~geek/class/ChapanisSciAm.pdf>
- Cornelius, Randolph R. 2000. Theoretical approaches to emotion. In *Speech and Emotion*, ISCA Tutorial and Research Worksh, 3-10. Newcastle, Northern Ireland, UK, September 5-7, 2000.
- Córdoba, Ricardo, José Ángel Vallejo, Juan Manuel Montero, Juana María Gutiérrez Arriola, Miguel Ángel López Carmona & José Manuel Pardo. 1999. Automatic modeling of duration in a Spanish text-to-speech system using neural networks. In *Eurospeech 1999. Proceedings of the 6th European Conference on Speech Communication and Technology*, 1619-1622. Budapest, Hungary, September 5-9, 1999. Disponible en: <http://www-gth.die.upm.es/research/documentation/AI-52Aut-99.pdf>
- Córdoba, Ricardo, Juan Manuel Montero, Juana María Gutiérrez Arriola, José Ángel Vallejo, Emilia Enríquez & José Manuel Pardo. 2002. Selection of the most significant parameters for duration modelling in a Spanish text-to-speech system using neural networks. *Computer Speech and Language* 16,(2), 183-203.
- Córdoba, Ricardo, Luis Fernando D'Haro, Rubén San-Segundo, Javier Macías, Fernando Fernández Martínez & Juan Carlos Plaza. 2006. A multiple-Gaussian classifier for Language Identification using acoustic information and PPRM scores. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), *IV Jornadas en Tecnología del Habla*, 45-48. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://www-gth.die.upm.es/~macias/doc/pubs/rthabla06/4jth_117.pdf
- Córdoba, Ricardo, Rubén San-Segundo, Juan Manuel Montero, José Colás, Javier Ferreiros, Javier Macías & José Manuel Pardo. 2001. An interactive directory assistance service for Spanish with large-vocabulary recognition. In *Eurospeech 2001 Scandinavia. Proceedings of the 7th European Conference on Speech Communication and*

- Technology, 2nd Interspeech Event, 1279-1282. Aalborg, Denmark, September 3-7, 2001. Disponible en: <http://www-gth.die.upm.es/~macias/doc/pubs/eurosp01/rick-idas-system.pdf>
- Cuéstara, Javier O. 2004. Fonética de la ciudad de México. Aportaciones desde las tecnologías del habla. Tesis para obtener el título de Maestro en Lingüística Hispánica. Maestría en Lingüística Hispánica, Posgrado en Lingüística, Universidad Nacional Autónoma de México. Disponible en: http://leibniz.iimas.unam.mx/~luis/DIME/publicaciones/tesis/Cuetara_Tesis_MLH-UNAM.pdf
- Diéguez, Javier, Carmen García Mateo & Antonio Cardenal. 2005. Comparación de modelos de lenguaje en tareas de transcripción automática de noticiarios televisivos. *Procesamiento del Lenguaje Natural* 35, 269-276. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/35/33.pdf>
- Dutoit, Thierry. 1997. *An introduction to text-to-speech synthesis*. Dordrecht: Kluwer.
- Enríquez, Emilia & Celia Casado. 1991. Hacia un algoritmo para la conversión automática de fonema en grafema en español. *Anuario de Lingüística Hispánica* 7, 151-204.
- Enríquez, Emilia. 1991. El problema de las ambigüedades fonéticas y su tratamiento automático. *Boletín de la Real Academia Española* LXXI,(XXLII), 157-183.
- Erro, Daniel & Asunción Moreno. 2007. Frame-alignment method for cross-lingual voice conversion. In *Interspeech 2007. Proceedings of the 8th Annual Conference of the International Speech Communication Association*, 1969-1972. Antwerp, Belgium, August 27-31, 2007. Disponible en: http://gps-tsc.upc.es/veu/research/pubs/download/err_fra_07.pdf
- Escalada, José Gregorio, Ana Armenta & Miguel Ángel Rodríguez Crespo. 2008. Descripción del sistema II de Telefónica I+D presentado a la evaluación ALbayzín' 08 para CTV. In Inmaculada Hernández (ed.), *V Jornadas en Tecnología del Habla*, 107-109. Bilbao: Universidad del País Vasco - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/V/pdfs/actas.pdf>
- Escudero, David & Valentín Cardeñoso. 2002. Una experiencia de reconocimiento automático de tipos de unidades melódicas a partir de su perfil de entonación. In Jesús Díaz García (ed.), *Actas del II Congreso de Fonética Experimental*, 161-166. Sevilla: Laboratorio de Fonética, Facultad de Filología, Universidad de Sevilla. Disponible en: <http://WWW.infor.uva.es/~descuder/investig/pdfs/cfeII.pdf>
- Escudero, David & Valentín Cardeñoso. 2006. Visualization of prosodic knowledge using corpus driven MEMOInt intonation modelling. In Petr Sojka, Ivan Kopeček & Karel Pala (eds.), *TDS 2006. 9th International Conference on Text, Speech and Dialog*, 645-652. Berlin - Heidelberg: Springer. Disponible en: <http://www.infor.uva.es/~descuder/investig/pdfs/tsd2006.pdf>
- Escudero, David & Valentín Cardeñoso. 2007. Applying data mining techniques to corpus based prosodic modeling. *Speech Communication* 49,(3), 213-229. Disponible en: <http://dx.doi.org/10.1016/j.specom.2007.01.008>
- Escudero, David. 2002. Modelado estadístico de entonación con funciones de Bézier: Aplicaciones a la conversión texto-voz en español. Tesis doctoral. Departamento de Informática, Universidad de Valladolid. Disponible en: <http://www.infor.uva.es/~descuder/investig/tesis/>
- Escudero, David, Valentín Cardeñoso, Juan María Sánchez, Eva Navas & Inmaculada Hernández. 2003. Uso de entonación en reconocimiento automático de locutor: Resultados preliminares. In Javier Hernando (ed.), *SEAF 2003. Actas del II Congreso de la Sociedad Española de Acústica Forense*, 167-174. Barcelona: Universitat Politècnica de Catalunya - Sociedad Española de Acústica Forense. Disponible en: <http://www.infor.uva.es/~descuder/investig/pdfs/SEAF2003.pdf>
- Esquerra, Ignasi. 2006. Síntesis de habla emocional por selección de unidades. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), *IV Jornadas en Tecnología del Habla*, 161-165. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Fant, Gunnar. 2004. More than half a century in phonetics and speech research. In *Speech Acoustics and Phonetics: Selected Writings*, 1-14. Dordrecht: Kluwer. Disponible en: <http://www.speech.kth.se/~gunnar/halfcentury.pdf>
- Fant, Gunnar. 2005. Historical notes. *Speech, Music and Hearing - Quarterly Progress and Status Report* 47, 9-19.
- Farrús, Mireia, Javier Hernando & Pascual Ejarque. 2007. Jitter and shimmer measurements for speaker recognition. In *Interspeech 2007. Proceedings of the 8th Annual Conference of the International Speech Communication Association*, 778-781. Antwerp, Belgium, August 27-31, 2007. Disponible en: http://gps-tsc.upc.es/veu/research/pubs/download/Far_Jit_07.pdf
- Febrer, Albert, Jaume Padrell & Antonio Bonafonte. 1998. Modeling phone duration: Application to Catalan TTS. In *Third ESCA/COCOSDA Workshop on Speech Synthesis*, 43-46. Jenolan Caves House, Blue Mountains, Australia, November 26-29, 1998. Disponible en: <http://gps-tsc.upc.es/veu/research/pubs/download/Feb98c.ps.gz>
- Fernández Martínez, Fernando, Javier Ferreiros, Valentín Sama, Juan Manuel Montero & Rafael García Sánchez. 2005. Demostración de una interfaz vocal para el control de un sistema de alta fidelidad. *Procesamiento del Lenguaje Natural* 35, 451-452. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/35/60.pdf>
- Fernández Pozo, Rubén, Carlos Fombella, Doroteo Torre, Eduardo López Gonzalo & Luis A Hernández Gómez. 2006. Estudio del uso de información prosódica en reconocimiento de locutor en ámbito forense. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), *IV Jornadas en Tecnología del Habla*, 343-348. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Fernández Rei, Elisa & Manuel González González. 1998. Un sintetizador de voz para el gallego. In Gilles Luquet

- (ed.), *Travaux de linguistique hispanique*, 65-76. Paris: Presses de la Sorbonne Nouvelle.
- Fernández Salgado, Xavier & Eduardo Rodríguez Banga. 1999. Segmental duration modelling in a text-to-speech system for the Galician language. In *Eurospeech 1999. Proceedings of the 6th European Conference on Speech Communication and Technology*, 1635-1638. Budapest, Hungary, September 5-9, 1999.
- Fernández Salgado, Xavier & Eduardo Rodríguez Banga. 2000a. A hierarchical intonation model for synthesising F0 contours in Galician language. In *ICSLP - Interspeech 2000. Proceedings of the 7th International Conference on Spoken Language Processing*, 625-628. Beijing, China, October 16-20, 2000.
- Fernández Salgado, Xavier & Eduardo Rodríguez Banga. 2000b. Proposición de un marco adecuado para el estudio de contornos de F0 para síntesis de voz. *Procesamiento del Lenguaje Natural* 24, 175-182. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/26/fernandez-salgado.pdf>
- Ferreiros, Javier, Javier Macías, José Manuel Pardo & Luis Villarrubia. 1998. Introducing multiple pronunciations in Spanish speech recognition systems. In *Modeling Pronunciation Variation for Automatic Speech Recognition*, 29-34. Rolduc, The Netherlands, May 4-6, 1998. Disponible en: <http://www-gth.die.upm.es/research/documentation/AI-43Int-98.pdf>
- Filipe, Porfirio P & Nuno J Mamede. 2007. Adaptação de um sistema de diálogo ao domínio bilhética. In *TIL 2007. 5th Workshop in Information and Human Language Technology*, Rio de Janeiro, Brazil, June 2007. Disponible en: <http://www.inesc-id.pt/pt/indicadores/Ficheiros/3984.pdf>
- Flores, Leonardo, Alcira Vargas, Alejandra Olivier, Ingrid Kirschning & Ofelia Cervantes. 2001. Síntesis en español mexicano con el método de selección de unidades de longitud variable. In *ENC 2001. Encuentro Nacional de Computación*, Disponible en: http://ict.udlap.mx/people/ingrid/ingrid/articulo_33.pdf
- Freitas, Diamantino Rui da Silva & Daniela Braga. 2002. Towards an intonation module for a Portuguese TTS system. In *ICSLP - Interspeech 2002. Proceedings of the 7th International Conference on Spoken Language Processing*, 161-164. Denver, Colorado, USA, September 16-20, 2002.
- Furui, Sadaoki. 2006. Automatic speaker recognition and verification. In Keith Brown (ed.), *Encyclopedia of Language & Linguistics*, 619-627. Amsterdam: Elsevier. Disponible en: <http://dx.doi.org/10.1016/B0-08-044854-2/00919-6>
- Gallo, Beatriz, Rubén San-Segundo, Juan Manuel Lucas, Roberto Barra, Luis Fernando D'Haro & Fernando Fernández Martínez. 2008a. Aplicación de métodos estadísticos para la traducción de voz a Lengua de Signos. *Procesamiento del Lenguaje Natural* 41, 251-258. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/41/sec9-art2.pdf>
- Gallo, Beatriz, Rubén San-Segundo, Juan Manuel Lucas, Roberto Barra, Luis Fernando D'Haro & Fernando Fernández Martínez. 2008b. Sistema de traducción estadística de voz a lengua de signos para personas sordas. In *Telecom I+D*, Bilbao. Disponible en: http://lorien.die.upm.es/~lapiz/voz2lse/papers/Traducci%C3%B3n_TELECOMI+D_4p_v2.pdf
- García Mateo, Carmen. 2002. Recursos e actividades necesarias para desenvolver tecnoloxía da fala en galego. In María Xesús Bugarín, José Cajide, Agustín Dosil, Xavier Ferreiro, Manuel González González & Miguel Anxo Santos (eds.), *Actas da VIII Conferencia Internacional de Linguas Minoritarias. Políticas lingüísticas e educativas na Europa Comunitaria*, 151-156. Santiago de Compostela: Xunta de Galicia.
- García Mateo, Carmen. 2003. Tecnologías del habla y lenguas minoritarias. *Procesamiento del Lenguaje Natural* 31, 381-384. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/31/31-Pag381.pdf>
- García, Carlos & Daniel Tapias. 2000. La frecuencia fundamental de la voz y sus efectos en reconocimiento de habla continua. *Procesamiento del Lenguaje Natural* 26, 163-167. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/26/garcia.pdf>
- Garrido, Juan María. 1996. Modelling Spanish intonation for text-to-speech applications. Tesis doctoral. Departament de Filologia Espanyola, Universitat Autònoma de Barcelona. Disponible en: <http://www.tdx.cat/TDX-0428108-155145>
- Garrido, Juan María, Eva Bofías, Yesika Laplaza, Montserrat Marquina, Matthew Aylett & Chris Pidcock. 2008. The CereVoice speech synthesiser. In Inmaculada Hernández (ed.), *V Jornadas en Tecnología del Habla*, 126-129. Bilbao: Universidad del País Vasco - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/V/pdfs/actas.pdf>
- Garrido, Juan María, Isabel Ortín, Silvia Quazza, Pier Luigi Salza & Franca Mancini. 2000. Desarrollo de un módulo de asignación de parámetros prosódicos para la versión en español del sistema de conversión texto-habla ACTOR®. *Procesamiento del Lenguaje Natural* 26, 183-190. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/26/garrido-alminana.pdf>
- Gonzalvo, Xavier, Joan Claudi Socoró, Ignasi Iriondo, Carlos Monzo & Elisa Martínez Marroquín. 2007. Linguistic and mixed excitation improvements on a HMM-based speech synthesis for Castilian Spanish. In *Sixth ISCA Tutorial and Research Workshop in Speech Synthesis*, 362-367. Bonn, Germany, August 22-24, 2007.
- González Ferreras, César & Valentín Cardeñoso. 2005. Development and evaluation of a spoken dialog system to access a newspaper website. In *Interspeech 2005 - Eurospeech. Proceedings of the 9th European Conference on Speech Communication and Technology*, 857-860. Lisbon, Portugal, September 4-8, 2005.
- González Ferreras, César & Valentín Cardeñoso. 2006. Experiments in speech driven information retrieval for Spanish language. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), *IV Jornadas en Tecnología del Habla*, 149-153. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- González González, Manuel. 2004. A síntese de voz en lingua galega: o proxecto Cotovía. *Revista Galega do Ensino*

- González González, Manuel, Elisa Fernández Rei & Begoña González Rei (eds.). 2007. SEAF 2005. Actas do III Congreso da Sociedade Española de Acústica Forense. Santiago de Compostela: Xunta de Galicia.
- González González, Manuel, Ruth Losada & Elisa Fernández Rei. 1999. O galego e as tecnoloxías da fala: o caso do sintetizador de voz. In Actas do V Congreso Internacional de Estudos Galegos, 703-716. Trier: Edicións do Castro - Galicien-Zentrum der Universität Trier.
- Greenberg, Steven. 2005. From here to utility - Melding phonetic insight with speech technology. In William J Barry & Wim A van Dommelen (eds.), *The Integration of Phonetic Knowledge in Speech Technology*, 107-132. Dordrecht: Springer. Disponible en: http://www.icsi.berkeley.edu/~steveng/PDF/Phonetic_Insight.pdf
- Gutiérrez Arriola, Juana María, Juan Manuel Montero, Ricardo Córdoba & José Manuel Pardo. 2003. ¿Podemos imitar la voz de una persona? Técnicas de conversión de hablante. In Javier Hernando (ed.), SEAF 2003. Actas del II Congreso de la Sociedad Española de Acústica Forense, 35-46. Barcelona: Universitat Politècnica de Catalunya - Sociedad Española de Acústica Forense. Disponible en: <http://www-gth.die.upm.es/research/documentation/AN-39Pod-03.pdf>
- Hernando, Javier (ed.). 2003. SEAF 2003. Actas del II Congreso de la Sociedad Española de Acústica Forense. Barcelona: Universitat Politècnica de Catalunya - Sociedad Española de Acústica Forense.
- Hernando, Javier, Jaume Padrell & Horacio Rodríguez. 2002. Sistema de información meteorológica automática por teléfono ATTEMPS. *Procesamiento del Lenguaje Natural* 29, 311-312. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/29/29-Pag311.pdf>
- Hernández, Inmaculada (ed.). 2008. V Jornadas en Tecnología del Habla. Bilbao: Universidad del País Vasco - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/V/pdfs/actas.pdf>
- Hieronymus, James L. 1994. ASCII phonetic symbols for the world's languages: Worldbet. Disponible en: <http://www.ling.ohio-state.edu/~edwards/WorldBet/worldbet.pdf>
- Hirschberg, Julia. 2006. Speech synthesis: Prosody. In *Encyclopedia of Language & Linguistics*, 49-55. Amsterdam: Elsevier. Disponible en: <http://dx.doi.org/10.1016/B0-08-044854-2/00914-7>
- Hockett, Charles F. 1960. The origin of speech. *Scientific American* 203, 89-96.
- Iriondo, Ignasi. 2008. Producción de un corpus oral y modelado prosódico para la síntesis del habla expresiva. Tesis doctoral. Departament de Comunicacions i Teoria del Senyal, Escola Tècnica Superior d'Enginyeria Electrònica i Informàtica La Salle, Universitat Ramon Llull.
- Iriondo, Ignasi, Francesc Alías, Javier Melenchón & María Ángeles Llorca. 2004. Modeling and synthesizing emotional speech for Catalan text-to-speech synthesis. In ADS 2004. Affective Dialogue Systems. Tutorial and Research Workshop, Kloster Irsee, Germany, June 14-16, 2004. Proceedings, 197-208. Heidelberg: Springer.
- Iriondo, Ignasi, Joan Claudi Socoró & Francesc Alías. 2007. Prosody modelling of Spanish for expressive speech synthesis. In ICASSP 2007. Proceedings of the 32nd IEEE International Conference on Acoustics, Speech and Signal Processing, 821-824. Honolulu, HI, USA.
- Iriondo, Ignasi, Joan Claudi Socoró, Lluís Formiga, Xavier Gonzalvo, Francesc Alías & Pere Miralles. 2006. Modelado y estimación de la prosodia mediante razonamiento basado en casos. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), IV Jornadas en Tecnología del Habla, 183-188. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Juang, Biing-Hwang & Lawrence R Rabiner. 2006. Automatic speech recognition: History. In Keith Brown (ed.), *Encyclopedia of Language & Linguistics*, 806-819. Amsterdam: Elsevier. Disponible en: <http://dx.doi.org/10.1016/B0-08-044854-2/00906-8>
- Laver, John. 1987. New horizons in European speech technology. Report on the ESPRIT Workshop on Speech Technology. Aarhus, Denmark, 21-22 May 1987.. In Proceedings and Following Workshop on Speech Technology. Status on Technology and Programmes Worldwide, Jutland Telephone, Denmark, May 20, 1987.
- Lavid, Julia. 2005. Lenguaje y nuevas tecnologías. Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI. Madrid: Cátedra.
- Lavid, Julia. 2006. La generación del lenguaje en los sistemas de diálogo. In Joaquim Llisterra & María Jesús Machuca (eds.), *Los sistemas de diálogo*, 153-176. Bellaterra - Soria: Universitat Autònoma de Barcelona - Fundació Duques de Soria.
- Lleida, Eduardo, Encarna Segarra, María Inés Torres & Javier Macías. 2006. EDECÁN: sistEma de Diálogo multidominio con adaptación al contExto aCústico y de AplicacióN. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), IV Jornadas en Tecnología del Habla, 294-296. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Llisterra, Joaquim & María Jesús Machuca (eds.). 2006. *Los sistemas de diálogo*. Bellaterra - Soria: Universitat Autònoma de Barcelona - Fundació Duques de Soria.
- Llisterra, Joaquim. 1993. Spanish adaptation of SAMPA and automatic phonetic transcription. SAM-A/UPC/001/v1. ESPRIT project 6819 (SAM-A Speech Technology Assessment in Multilingual Applications). Disponible en: http://liceu.uab.cat/~joaquim/publicacions/SAMPA_Spanish_93.pdf
- Llisterra, Joaquim. 1995. A proposal for Catalan SAMPA. Bellaterra: Departament de Filologia Espanyola, Universitat

- Autònoma de Barcelona. Disponible en: http://liceu.uab.cat/~joaquim/language_resources/SAMPA_Catalan.html
- Llisterri, Joaquim. 2002. Las tecnologías del habla en lengua catalana. In María Xesús Bugarín, José Cajide, Agustín Dosil, Xavier Ferreiro, Manuel González González & Miguel Anxo Santos (eds.), *Actas da VIII Conferencia Internacional de Linguas Minoritarias. Políticas lingüísticas e educativas na Europa Comunitaria*, 163-169. Santiago de Compostela: Xunta de Galicia. Disponible en: http://liceu.uab.cat/~joaquim/publicacions/Llisterri_02_TecnolHabla_Catalan.pdf
- Llisterri, Joaquim. 2003. Lingüística y tecnologías del lenguaje. *Lynx. Panorámica de estudios lingüísticos* 2, 9-71. Disponible en: http://liceu.uab.cat/~joaquim/publicacions/Llisterri_03_Linguistica_Tecnologias_Lenguaje.pdf
- Llisterri, Joaquim. 2004. Las tecnologías del habla para el español. In Reyes Sequera (ed.), *Ciencia, tecnología y lengua española: la terminología científica en español*, 123-141. Madrid: Fundación Española para la Ciencia y la Tecnología. Disponible en: http://liceu.uab.cat/~joaquim/publicacions/Llisterri_04_Tecnologias_Habla_Espanol.pdf
- Llisterri, Joaquim. 2007a. El español y las nuevas tecnologías. In Manel Lacorte (ed.), *Lingüística aplicada del español*, 483-520. Madrid: Arco/Libros.
- Llisterri, Joaquim. 2007b. El papel de la fonética en las tecnologías del habla. In Manuel González González, Elisa Fernández Rei & Begoña González Rei (eds.), *Actas do III Congreso Internacional de Fonética Experimental*, 23-37. Santiago de Compostela: Xunta de Galicia. Disponible en: http://liceu.uab.cat/~joaquim/publicacions/Llisterri_05_Fonetica_Tecnologias_Habla.pdf
- Llisterri, Joaquim. 2007c. Màquines que parlen i que escolten: el paper de la fonètica en el desenvolupament de les tecnologies de la parla. In Josefina Carrera & Clàudia Pons (eds.), *Aplicacions de la fonètica. Catorzè Col·loqui Lingüístic de la Universitat de Barcelona*, 139-169. Barcelona: Promociones y Publicaciones Universitarias. Disponible en: http://liceu.uab.cat/~joaquim/publicacions/Llisterri_07_Fonetica_Tecnologies_Parla.pdf
- Llisterri, Joaquim, Carme Carbó, María Jesús Machuca, Carme de la Mota, Montserrat Riera & Antonio Ríos. 2003. El papel de la lingüística en el desarrollo de las tecnologías del habla. In Miguel Casas & Carmen Varo (eds.), *VII Jornadas de Lingüística*, 137-191. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz. Disponible en: http://liceu.uab.cat/~joaquim/publicacions/Llisterri_Carbo_Machuca_Mota_Riera_Rios_03_Linguistica_Tecnologias_Habla.pdf
- Llisterri, Joaquim, Carme Carbó, María Jesús Machuca, Carme de la Mota, Montserrat Riera & Antonio Ríos. 2004. La conversión de texto en habla: aspectos lingüísticos. In Maria Antònia Martí & Joaquim Llisterri (eds.), *Tecnologías del texto y del habla*, 145-186. Barcelona: Edicions de la Universitat de Barcelona - Fundación Duques de Soria. Disponible en: http://liceu.uab.cat/~joaquim/publicacions/Llisterri_Carbo_Machuca_Mota_Riera_Rios_04_Conversion_Texto_Habla.pdf
- Llisterri, Joaquim, María Jesús Machuca, Carme de la Mota, Montserrat Riera & Antonio Ríos. 2003. Entonación y tecnologías del habla. In Pilar Prieto (ed.), *Teorías de la entonación*, 209-243. Barcelona: Ariel. Disponible en: http://liceu.uab.es/~joaquim/publicacions/Llisterri_Machuca_Mota_Riera_Rios_03_Entonacion_Tecnologias_Habla.pdf
- Llisterri, Joaquim, María Jesús Machuca, Carme de la Mota, Montserrat Riera & Antonio Ríos. 2005. Corpus orales para el desarrollo de las tecnologías del habla en español. *Oralia. Análisis del discurso oral* 8, 289-325. Disponible en: http://liceu.uab.cat/~joaquim/publicacions/Llisterri_Machuca_Mota_Riera_Rios_05_Corpus_Orales_Tecnologias_Habla_Espanol.pdf
- Llisterri, Joaquim, María Jesús Machuca, Natalia Madrigal, Franca Mancini, Paolo Massimino, Carme de la Mota, Montserrat Riera & Antonio Ríos. 2004. Aspectos lingüísticos en el diseño de un conversor de texto en habla en castellano y en catalán: El sistema LoquendoTTS®. In VI Congreso de Lingüística General, 521-522. Santiago de Compostela: Disponible en: http://liceu.uab.cat/~joaquim/publicacions/Llisterri_et_al_04_Conversor_Texto_Habla_Castellano_Catalan_Loquendo.pdf
- Losada, Ruth. 2004. Unha adaptación do SAMPA para a lingua galega. In Rosario Álvarez Blanco, Francisco Fernández Rei & Antón Santamarina (eds.), *A Lingua Galega. Historia e actualidade. Actas do I Congreso Internacional*, 615-625. Santiago de Compostela: Instituto da Lingua Galega - Consello da Cultura Galega. Disponible en: http://consellodacultura.org/mediateca/pubs.pdf/galego_historia_2.pdf/
- López Gonzalo, Eduardo & Luis A Hernández Gómez. 1995. Automatic data-driven prosodic modeling for text to speech. In *Eurospeech 1995. Proceedings of the 4th European Conference on Speech Communication and Technology*, 585-588. Madrid, Spain, September 18-21, 1995.
- López Gonzalo, Eduardo, Jorge Álvarez Cercadillo & Luis A Hernández Gómez. 1994. Metodología para el modelado prosódico de un sistema de conversión de texto en habla en castellano. *Procesamiento del Lenguaje Natural* 15, Disponible en: <http://www.sepln.org/revistaSEPLN/revista/15/todo.pdf>
- López Gonzalo, Eduardo, Juan María Villar & Luis A Hernández Gómez. 2002. Automatic prosody modelling of Galician and its applications to Spanish. In Erik Keller, Gérard Bailly, Alex Monaghan, Jacques Terken & Mark Huckvale (eds.), *Improvements in speech synthesis. Cost 258: The naturalness of synthetic speech*, 218-227. Chichester: John Wiley & Sons.
- López Mencía, Beatriz, Álvaro Hernández Trapote, David Díaz Pardo, Doroteo Torre, Luis A Hernández Gómez &

- Eduardo López Gonzalo. 2006. A good gesture: Exploring nonverbal communication for robust SLDSs. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), IV Jornadas en Tecnología del Habla, 39-44. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- López Morrás, Xavier. 2004. Transcriptor fonético automático del español. Disponible en: <http://www.aucel.com/pln/transbase.html>
- López Soto, María Teresa & Ramón López-Cózar (eds.). 2000. I Jornadas en Tecnología del Habla. Sevilla: Universidad de Sevilla - Universidad de Granada - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/I/ACTAS.zip>
- López-Cózar, Ramón & Antonio J Rubio Ayuso. 1997. SAPLEN: Un sistema de diálogo en lenguaje natural para una aplicación comercial. *Procesamiento del Lenguaje Natural* 20, 65-81. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/20/20-Pag65.pdf>
- López-Cózar, Ramón & Masahiro Araki. 2005. Spoken, Multilingual and Multimodal Dialogue Systems: Development and Assessment. Chichester: John Wiley & Sons.
- López-Cózar, Ramón & Ramón Granell. 2004. Sistema de diálogo basado en VoiceXML para proporcionar información de viajes en tren. *Procesamiento del Lenguaje Natural* 33, 171-178. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/33/33-Pag171.pdf>
- López-Cózar, Ramón. 2006. Análisis y gestión del diálogo. In Joaquim Llisterrí & María Jesús Machuca (eds.), Los sistemas de diálogo, 119-152. Bellaterra - Soria: Universitat Autònoma de Barcelona - Fundació Duques de Soria.
- López-Cózar, Ramón, Zoraida Callejas, Germán Montoro & Pablo Haya. 2006. DS-UCAT: Sistema de diálogo multimodal y multilingüe para un entorno educativo. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), IV Jornadas en Tecnología del Habla, 135-140. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Lucas, Juan Manuel, Rosario Alcázar, Juan Manuel Montero, Fernando Fernández Martínez, Roberto Barra, Luis Fernando D'Haro, Javier Ferreiros, Ricardo Córdoba, Javier Macías, Rubén San-Segundo & José Manuel Pardo. 2008. Desarrollo de un robot-guía con integración de un sistema de diálogo y expresión de emociones: Proyecto ROBINT. *Procesamiento del Lenguaje Natural* 40, 51-58. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/40/09p12.pdf>
- Luque, Jordi, Daniel Ferrés, Javier Hernando, José B Mariño & Horacio Rodríguez. 2006. GeoVAQA: A voice activated geographical question answering system. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), IV Jornadas en Tecnología del Habla, 309-314. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Macías, Javier, Javier Ferreiros, Ricardo Córdoba, Juan Manuel Montero, José David Romeral & José Manuel Pardo. 2003. Estrategias de generación y reducción de variantes de pronunciación en sistemas de reconocimiento automático de habla: consideraciones arquitecturales. *Procesamiento del Lenguaje Natural* 31, 91-98. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/31/31-Pag91.pdf>
- Maia, Ranniery, Heiga Zen, Keiichi Tokuda, Tadashi Kitamura & Fernando Gil V Resende Jr.. 2003. Towards the development of a Brazilian Portuguese text-to-speech system based on HMM. In Eurospeech 2003. Proceedings of the 8th European Conference on Speech Communication and Technology, 2645-2648. Geneva, Switzerland, September 1-4, 2003. Disponible en: <http://www.sp.nitech.ac.jp/~zen/publications/maia-euro03.pdf>
- Mariño, José B & Climent Nadeu. 2004. La representación de la voz para el reconocimiento del habla. In Maria Antònia Martí & Joaquim Llisterrí (eds.), Tecnologías del texto y del habla, 187-224. Barcelona: Edicions de la Universitat de Barcelona - Fundació Duques de Soria.
- Mariño, José B & Horacio Rodríguez. 2003. Proyecto ALIADO: Tecnologías del habla y el lenguaje para un asistente personal. *Procesamiento del Lenguaje Natural* 31, 305-306. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/31/31-Pag305.pdf>
- Mariño, José B. 2006. AVIVAVOZ: Tecnologías para la traducción de voz. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), IV Jornadas en Tecnología del Habla, 285-290. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Martins, Fernando. 2001. Bases linguísticas em sistemas de reconhecimento de fala. In Elisabete Marques Ranchhod (ed.), Tratamento das línguas por computador. Uma introdução à linguística computacional e suas aplicações, 195-228. Lisboa: Caminho.
- Martí, Josep & Daniel Niñerola. 1987. SINCAS: un conversor texto-voz en castellano. *Procesamiento del Lenguaje Natural* 5, 111-122. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/5/5-Pag111.pdf>
- Martí, Maria Antònia (ed.). 2003. Tecnologías del lenguaje. Barcelona: Editorial UOC.
- Meinedo, Hugo & João Paulo Neto. 2003. Automatic speech annotation and transcription in a Broadcast News task. In MSDR 2003. ISCA Workshop on Multilingual Spoken Document Retrieval, 95-100. Hong Kong, April 4-5, 2003. Disponible en: <http://www.12f.inesc-id.pt/documents/papers/2003MeinedoA.pdf>
- Meza, Horacio, Ingrid Kirschning & Ofelia Cervantes. 2000. Estimation of duration models for phonemes in Mexican speech synthesis. In ICSLP 2000. Proceedings of the 6th International Conference on Spoken Language Processing, 685-688. Beijing, China, October 16-20, 2000. Disponible en: <http://ict.udlap.mx/people/ingrid/ingrid/ICSLP2000.pdf>
- Montero, Alberto, Javier González Domínguez, Daniel Ramos, Ignacio López Moreno, Doroteo Torre & Joaquín

- González Rodríguez. 2006. On the use of high-level information in speaker and language recognition. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), IV Jornadas en Tecnología del Habla, 355-360. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Montero, Juan Manuel, Juana María Gutiérrez Arriola, José Colás, Javier Macías, Emilia Enríquez & José Manuel Pardo. 1999. Development of an emotional speech synthesiser in Spanish. In Eurospeech 1999. Proceedings of the 6th European Conference on Speech Communication and Technology, 2099-2102. Budapest, Hungary, September 5-9, 1999. Disponible en: <http://www-gth.die.upm.es/research/documentation/AI-53Dev-99.pdf>
- Montero, Juan Manuel, Juana María Gutiérrez Arriola, Sira Palazuelos, Emilia Enríquez, Santiago Aguilera & José Manuel Pardo. 1998. Emotional speech synthesis: from speech database to TTS. In ICSLP 1998. Proceedings of the 5th International Conference on Spoken Language Processing, 923-926. Sidney Convention Centre, Sidney, Australia, 30 November - 4 December, 1998. Disponible en: <http://www-gth.die.upm.es/research/documentation/AI-45Emo-98.pdf>
- Monzo, Carlos, Francesc Alías, Antonio Morán & Xavier Gonzalvo. 2006. Transcripción fonética de acrónimos en castellano utilizando el algoritmo C4.5. *Procesamiento del Lenguaje Natural* 37, 275-282. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/37/34.pdf>
- Monzo, Carlos, Lluís Formiga, Jordi Adell, Ignasi Iriondo, Francesc Alías & Joan Claudi Socoró. 2008. Adaptación del CTH-URL para la competición Albayzín 2008. In Inmaculada Hernández (ed.), V Jornadas en Tecnología del Habla, 87-90. Bilbao: Universidad del País Vasco - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/V/pdfs/actas.pdf>
- Mora, Elsa & Manuel Rodríguez Hourcadette. 2005. Síntesis de habla en español venezolano: usos prácticos y de investigación. *Revista Investigación* 12, 26-29. Disponible en: <http://www.saber.ula.ve/bitstream/123456789/21405/1/avance-001.pdf>
- Moreno Sandoval, Antonio, Doroteo Torre, Natalia Curto & Raúl de la Torre. 2006. Inventario de frecuencias fonémicas y silábicas del castellano espontáneo y escrito. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), IV Jornadas en Tecnología del Habla, 77-81. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Moreno, Asunción & José B Mariño. 1998. Spanish dialects: phonetic transcription. In ICSLP 1998. Proceedings of the 5th International Conference on Spoken Language Processing, Sidney Convention Centre, Sidney, Australia, 30 November - 4 December, 1998. Disponible en: <http://www.shlrc.mq.edu.au/proceedings/icslp98/PDF/AUTHOR/SL980598.PDF>
- Neto, João Paulo, Nuno J Mamede, Renato Cassaca & Luís Caldas de Oliveira. 2003. The development of a multi-purpose spoken dialogue system. In Eurospeech 2003. Proceedings of the 8th European Conference on Speech Communication and Technology, 1933-1936. Geneva, Switzerland, September 1-4, 2003. Disponible en: <http://www.l2f.inesc-id.pt/documents/papers/2003NetoB.pdf>
- Nogueiras, Albino, Asunción Moreno, Antonio Bonafonte & José B Mariño. 2001. Speech emotion recognition using Hidden Markov Models. In Eurospeech 2001 Scandinavia. Proceedings of the 7th European Conference on Speech Communication and Technology, 2nd Interspeech Event, 2679-2682. Aalborg, Denmark, September 3-7, 2001. Disponible en: http://gps-tsc.upc.es/veu/research/pubs/download/nog_emo_01.pdf
- Nolan, Francis. 1983. *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.
- Nolan, Francis. 1997. Speaker recognition and forensic phonetics. In William J Hardcastle & John Laver (eds.), *The Handbook of Phonetic Sciences*, 744-767. Oxford: Blackwell.
- Nusbaum, Howard C & Hadas Shintel. 2006. Speech Synthesis. In Keith Brown (ed.), *Encyclopedia of Language & Linguistics*, 19-31. Amsterdam: Elsevier. Disponible en: <http://dx.doi.org/10.1016/B0-08-044854-2/00913-5>
- Oliveira, Catarina, Lourdes Castro Moutinho & António Teixeira. 2004. Un novo sistema de conversão grafema-fone para PE baseado em transdutores. In *Actas do III Congresso Internacional de Fonética e Fonologia*, Maranhão, Brasil.
- Oliveira, Catarina, Lourdes Castro Moutinho & António Teixeira. 2005. On European Portuguese automatic syllabification. In *Interspeech 2005 - Eurospeech*. Proceedings of the 9th European Conference on Speech Communication and Technology, 2933-2936. Lisbon, Portugal, September 4-8, 2005. Disponible en: <http://www.ii.ua.pt/cidlc/gcl/files/publicacoes/IS052277.pdf>
- Oliveira, Luís Caldas de, Maria do Céu Viana & Isabel Trancoso. 1991. DIXI - Portuguese text-to-speech system. In Eurospeech 1991. Proceedings of the 2nd European Conference on Speech Communication and Technology, 1239-1242. Genova, Italy, September 24-26, 1991. Disponible en: <http://www.inesc-id.pt/pt/indicadores/Ficheiros/3195.pdf>
- Ortega García, Javier (ed.). 2000. SEAF 2000. Actas del I Congreso de la Sociedad Española de Acústica Forense. Madrid: Escuela Universitaria de Ingenieros Técnicos de Telecomunicación - Sociedad Española de Acústica Forense.
- Pachès, Pau, Carme de la Mota, Montserrat Riera, Maria Pilar Perea, Albert Febrer, Mònica Estruch, Juan María Garrido, María Jesús Machuca, Antonio Ríos, Joaquim Llisterri, Ignasi Esquerra, Javier Hernando, Jaume Padrell & Climent Nadeu. 2000. Segre: An automatic tool for grapheme-to-allophone transcription in Catalan. In Donncha

- Ó Cróinín (ed.), *Proceedings of the Workshop on Developing Language Resources for Minority Languages: Reusability and Strategic Priorities* (LREC 2000. 2nd International Conference on Language Resources and Evaluation), 52-61. Athens, Greece, May 30 - June 2, 2000. Disponible en: http://liceu.uab.es/~joaquim/publicacions/Paches_et_al_00_SEGRE_Phonetic_Transcription_Catalan.pdf
- Padrell, Jaume & Javier Hernando. 2002. ACIMET: Access to meteorological information by telephone. In *ICSLP - Interspeech 2002. Proceedings of the 7th International Conference on Spoken Language Processing*, Denver, Colorado, USA, September 16-20, 2002. Disponible en: http://gps-tsc.upc.es/veu/research/pubs/download/Pad_Aci_02.pdf
- Pastor, Moisés, Alberto Sanchís, Francisco Casacuberta & Enrique Vidal. 2000. *EuTrans. Prototipo de traducción automática de voz a voz*. In María Teresa López Soto & Ramón López-Cózar (eds.), *I Jornadas en Tecnología del Habla*, Sevilla: Universidad de Sevilla - Universidad de Granada - Red Temática en Tecnologías del Habla. Disponible en: <http://prhlt.iti.es/papers/2000/Pastor00a.pdf>
- Paulo, Sérgio, Luís Caldas de Oliveira, Carlos Mendes, Luís Figueira, Renato Cassaca, Maria do Céu Viana & Helena Moniz. 2008. DIXI - A generic text-to-speech system for European Portuguese. In *PROPOR 2008. Computational Processing of the Portuguese Language. Eighth International Conference*, Aveiro, Portugal, September 8-10, 2008, *Proceedings*, 91-100. Heidelberg: Springer. Disponible en: <http://www.inesc-id.pt/pt/indicadores/Ficheiros/5009.pdf>
- Pérez Pavón, Elia Patricia. 2006. *Construcción de un reconocedor de voz utilizando Sphinx y el corpus DIMEx100*. Tesis para la obtención del título de Ingeniera en Computación. Facultad de Ingeniería, Universidad Nacional Autónoma de México. Disponible en: <http://leibniz.iimas.unam.mx/~luis/DIME/publicaciones/tesis/Tesis-Paty.pdf>
- Pérez, Alicia, María Inés Torres, Víctor Gujjarrubia & Francisco Casacuberta. 2006. *Basque-Spanish-English machine translation using finite state transducers*. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), *IV Jornadas en Tecnología del Habla*, 49-53. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Pérez, Hernán Emilio & Thomas Armstrong. 1998. *Diseño e implementación de un transcriptor fonético automático de textos generales del español*. *Onomazein* 3, 315-324. Disponible en: <http://www2.udec.cl/~hvaldivi/articulos/Perez1998b.pdf>
- Pineda, Luis Alberto. 2008. *El proyecto DIME y el robot conversacional Golem: una experiencia multidisciplinaria entre la computación y la lingüística*. Departamento de Ciencias de la Computación, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México. Disponible en: <http://leibniz.iimas.unam.mx/~luis/DIME/publicaciones/papers/DIME-Golem.pdf>
- Pineda, Luis Alberto, Antonio Massé, Iván Meza, Miguel Salas, Eric Schwarz, Esmeralda Uraga & Luis Villaseñor. 2002. *The DIME project*. In Carlos A Coello, Álvaro de Albornoz, Luis E Sucar & Osvaldo C Battistutti (eds.), *MICAI 2002. Advances in Artificial Intelligence. Proceedings of the Second Mexican International Conference on Artificial Intelligence*, 166-175. Dordrecht: Springer. Disponible en: <http://leibniz.iimas.unam.mx/~luis/DIME/publicaciones/papers/dime-micai.pdf>
- Puigvi, David, David Jiménez & Josep Maria Fernández. 1994. *Parametrización de las pausas ortográficas en castellano. Aplicación a un conversor de texto en habla*. *Procesamiento del Lenguaje Natural* 15, Disponible en: http://liceu.uab.cat/publicacions/Puigvi_Jimenez_Fernandez_94_Pausas_Sintesis_Castellano.pdf
- Rabiner, Lawrence R & Bing-Hwang Juang. 2006. *Speech recognition: Statistical methods*. In Keith Brown (ed.), *Encyclopedia of Language & Linguistics*, 1-18. Amsterdam: Elsevier. Disponible en: <http://dx.doi.org/10.1016/B0-08-044854-2/00907-X>
- Ranchhod, Elisabete Marques (ed.). 2001. *Tratamento das línguas por computador. Uma introdução à linguística computacional e suas aplicações*. Lisboa: Caminho.
- Relaño, José, Luis Villarrubia, Mari Carmen Rodríguez Gancedo & Luis A Hernández Gómez. 2003. *AGORA. Multilingual multiplatform architecture for the development of natural language voice services*. *Procesamiento del Lenguaje Natural* 31, 325-326. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/31/31-Pag325.pdf>
- Ríos, Antonio. 1999. *La transcripción fonética automática del Diccionario Electrónico de Formas Simples Flexivas del Español: un estudio fonológico en el léxico*. *Estudios de Lingüística Española* 4, Disponible en: <http://elies.rediris.es/elies4/>
- Rodríguez Banga, Eduardo, Francisco Campillo, Elisa Fernández Rei & Francisco Méndez. 2002. *Sistema de conversión texto-voz en lengua gallega basado en la selección combinada de unidades acústicas y prosódicas. Procesamiento del Lenguaje Natural* 29, 153-158. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/29/29-Pag153.pdf>
- Rodríguez Banga, Eduardo, Francisco Méndez, Francisco Campillo, Gonzalo Iglesias & Laura Docío. 2008. *Descripción del sintetizador de voz Cotovia para la evaluación Albayzín TTS 2008*. In Inmaculada Hernáez (ed.), *V Jornadas en Tecnología del Habla*, 100-103. Bilbao: Universidad del País Vasco - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/V/pdfs/actas.pdf>
- Rodríguez Crespo, Miguel Ángel & José Gregorio Escalada. 1990. *Text analysis system with automatic letter to allophone conversion for a Spanish text to speech synthesizer*. In *The ESCA Workshop on Speech Synthesis*, 105-108. Autrans, France, September 25-28, 1990.
- Rodríguez Crespo, Miguel Ángel, José Gregorio Escalada & Ana Armenta. 2008. *Descripción del sistema I de Telefónica I+D presentado a la evaluación Albayzín' 08 para CTV*. In Inmaculada Hernáez (ed.), *V Jornadas en*

- Tecnología del Habla, 104-106. Bilbao: Universidad del País Vasco - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/V/pdfs/actas.pdf>
- Rodríguez Crespo, Miguel Ángel, José Gregorio Escalada & Doroteo Torre. 1998. Conversor texto-voz multilingüe para el español, catalán, gallego y euskera. *Procesamiento del Lenguaje Natural* 23, 16-23. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/23/23-Pag16.pdf>
- Rodríguez Crespo, Miguel Ángel, José Gregorio Escalada, Ana Armenta & Juan María Garrido. 2008. Nuevo módulo de análisis prosódico del conversor texto-voz multilingüe de Telefónica I+D. In Inmaculada Hernández (ed.), *V Jornadas en Tecnología del Habla*, 157-160. Bilbao: Universidad del País Vasco - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/V/pdfs/actas.pdf>
- Rodríguez Hourcadette, Manuel & Elsa Mora. 2006. Conversor texto a voz en el dialecto venezolano por medio de la concatenación de difonos. *Ciencia e Ingeniería* 27,(2), 79-87. Disponible en: <http://oai.saber.ula.ve/db/ssaber/Edocs/papers/humanidades/mora-elsa/conversor-texto.pdf>
- Rodríguez Hourcadette, Manuel, Elsa Mora & Christian Cavé. 2006. Síntesis de voz en el dialecto venezolano por medio de la concatenación de difonos. *Ciencia e Ingeniería* 27,(1), 17-24. Disponible en: <http://www.saber.ula.ve/bitstream/123456789/16453/1/dialecto-venezolano.pdf>
- Rodríguez, Leandro, Carmen García Mateo, Santiago Pardo & Víctor Darriba. 2002. Un sistema de diálogo para la consulta de correo electrónico en lenguaje natural. *Procesamiento del Lenguaje Natural* 29, 181-187. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/29/29-Pag181.pdf>
- Rouas, Jean-Luc, Isabel Trancoso, Maria do Céu Viana & Mónica Abreu. 2008. Language and variety verification on broadcast news for Portuguese. *Speech Communication* 50,(11-12), 965-979. Disponible en: <http://www.inesc-id.pt/pt/indicadores/Ficheiros/4871.pdf>
- Rubio Ayuso, Antonio J & Diego H Milone. 2002. Información prosódica y acentual para el reconocimiento automático del habla. In Jesús Díaz García (ed.), *Actas del II Congreso de Fonética Experimental*, 56-77. Sevilla: Laboratorio de Fonética, Facultad de Filología, Universidad de Sevilla.
- Rubio Ayuso, Antonio J & Inmaculada Hernández (eds.). 2005. Libro blanco de Tecnologías del Habla. Granada: Universidad de Granada - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/docs/LibroBlancoTecnologiasDelHabla.pdf>
- Rubio Ayuso, Antonio J (ed.). 2002. II Jornadas en Tecnología del Habla. Granada: Universidad de Granada - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/II/index.htm>
- Sainz, Iñaki, Inmaculada Hernández, Eva Navas, Jon Sánchez, Iker Luengo, Ibon Saratxaga, Igor Odriozola, Eneritz de Bilbao & Daniel Erro. 2008. Descripción del conversor de texto a voz AHOTTS presentado a la evaluación Albayzín TTS 2008. In Inmaculada Hernández (ed.), *V Jornadas en Tecnología del Habla*, 96-99. Bilbao: Universidad del País Vasco - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/V/pdfs/actas.pdf>
- Sama, Valentín, Fernando Fernández Martínez, Javier Ferreiros, Javier Macías, Ricardo Córdoba, Juan Manuel Montero, José Colás, Eduardo Campos & José Manuel Pardo. 2003. Sistema de comprensión de comunicaciones habladas para el control de tráfico aéreo del proyecto INVOCA. *Procesamiento del Lenguaje Natural* 31, Disponible en: <http://www.sepln.org/revistaSEPLN/revista/31/31-Pag313.pdf>
- San-Segundo, Rubén, Juan Manuel Montero, Javier Ferreiros, Javier Macías & José Manuel Pardo. 2001. Sistema de información ferroviaria por teléfono: propuesta de una metodología de diseño de gestores de diálogo. In *Actas del Segundo Taller Internacional de Procesamiento Computacional del Español y Tecnologías del Lenguaje*, 241-245. Jaén, 14-15 de septiembre de 2001. Disponible en: <http://www-gth.die.upm.es/~macias/doc/pubs/slpl01/MethodologiaDialogov3.pdf>
- San-Segundo, Rubén, Juan Manuel Montero, Ricardo Córdoba & Juana María Gutiérrez Arriola. 2000. Stress assignment in Spanish proper names. In *ICSLP 2000. Proceedings of the 6th International Conference on Spoken Language Processing*, 346-349. Beijing, China, October 16-20, 2000. Disponible en: <http://www-gth.die.upm.es/research/documentation/AG-06Str-00.pdf>
- Sanchis, Emilio (ed.). 2004. III Jornadas en Tecnología del Habla. València: Universitat Politècnica de València - Red Temática en Tecnologías del Habla. Disponible en: <http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/III/actas3JTH.pdf>
- Santen, Jan P H van. 2005. Phonetic knowledge in text-to-speech synthesis. In William J Barry & Wim A van Dommelen (eds.), *The Integration of Phonetic Knowledge in Speech Technology*, 149-166. Dordrecht: Springer.
- Santos, Andrés, Juan Carlos Olabe, Elías Muñoz, Carlos López Barrio, Antonio Quilis & Miguel Martínez Martín. 1985. Sistema de conversión de texto a voz en español el tiempo real. *Procesamiento del Lenguaje Natural* 3, 21-28. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/3/3-articulo-3.pdf>
- Santos, Andrés, Pablo Muñoz & Miguel Martínez Martín. 1988. Diseño y evaluación de reglas de duración en la conversión de texto a voz. *Procesamiento del Lenguaje Natural* 6, 69-92.
- Sánchez Martínez, Raquel, João Paulo Neto & Diamantino Caseiro. 2008. Statistical machine translation of broadcast news from Spanish to Portuguese. In *PROPOR 2008. Computational Processing of the Portuguese Language. Eighth International Conference*, Aveiro, Portugal, September 8-10, 2008, *Proceedings*, 112-121. Heidelberg: Springer. Disponible en: <http://www.inesc-id.pt/pt/indicadores/Ficheiros/5008.pdf>
- Schroeter, Juergen. 2006. Text to-speech (TTS) synthesis. In *The Electrical Engineering Handbook*, 16.1-16-13. Boca Baton, FL: CRC Press. Disponible en: http://www.research.att.com/~ttsweb/tts/papers/2005_EEHandbook/tts.pdf

- Schulz, Henrik, Marta R Costa-Jussà & José A R Fonollosa. 2008. TECNOPARLA - Speech technologies for Catalan and its application to Speech-to-speech Translation. *Procesamiento del Lenguaje Natural* 41, 319-320. Disponible en: <http://www.sepln.org/revistaSEPLN/revista/41/proy5.pdf>
- Segarra, Encarna. 2006. La interpretación semántica. In Joaquim Llisterra & María Jesús Machuca (eds.), *Los sistemas de diálogo*, 99-118. Bellaterra - Soria: Universitat Autònoma de Barcelona - Fundación Duques de Soria.
- Silva, Solimar, Fernando Gil V Resende Jr. & Sergio L Netto. 2001. A text-to-speech system for the Brazilian Portuguese based on syllabic units. In *Proceedings of the Second IEEE South-American Workshop on Circuits and Systems*, Disponible en: <http://www.lps.ufjf.br/profs/sergioln/papers/BC08.pdf>
- Strik, Helmer. 2005. Is phonetic knowledge of any use for speech technology?. In William J Barry & Wim A van Dommelen (eds.), *The Integration of Phonetic Knowledge in Speech Technology*, 167-180. Dordrecht: Springer. Disponible en: <http://lands.let.kun.nl/literature/strik.2005.1.pdf>
- Tamarit, Vicent & Carlos D Martínez Hinarejos. 2006. Control de un robot "RoboSapien" mediante voz. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), *IV Jornadas en Tecnología del Habla*, 303-308. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Tapias, Daniel & Luis A Hernández Gómez. 2004. Los sistemas de diálogo en los servicios telefónicos: evolución y consideraciones de diseño. In María Antònia Martí & Joaquim Llisterra (eds.), *Tecnologías del texto y del habla*, 225-253. Barcelona - Soria: Edicions de la Universitat de Barcelona - Fundación Duques de Soria.
- Tapias, Daniel. 2002. Interfaces de voz con lenguaje natural. In María Antònia Martí & Joaquim Llisterra (eds.), *Tratamiento del lenguaje natural. Tecnología de la lengua oral y escrita*, 189-207. Barcelona - Soria: Edicions de la Universitat de Barcelona - Fundación Duques de Soria.
- Tapias, Daniel, Jorge Álvarez Cercadillo & Ismael Cortázar. 1997. ATOS: un sistema de control automático del teléfono mediante computador. *Philologia Hispalensis* 11,(2), 151-161.
- Teixeira, João Paulo Ramos & Diamantino Rui da Silva Freitas. 2003. Evaluation of a segmental durations model for TTS. In Nuno J Mamede, Jorge Baptista, Isabel Trancoso & Maria das Graças Volpe Nunes (eds.), *PROPOR 2003. Computational Processing of the Portuguese Language. Sixth International Workshop*, Faro, Portugal, June 26-27, 2003. *Proceedings*, 40-48. Heidelberg: Springer. Disponible en: <http://www.ipb.pt/~joaopt/publicacoes/artigos/Evaluation%20of%20a%20Segmental%20-%20PROPOR%2003.pdf>
- Teixeira, João Paulo Ramos. 2004. A prosody model to TTS. Tese de Doutoramento. Departamento de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia, Universidade do Porto. Disponible en: http://www.ipb.pt/~joaopt/publicacoes/Phd/PhD_JPT_complete.pdf
- Teixeira, João Paulo Ramos, Diamantino Rui da Silva Freitas & Hiroya Fujisaki. 2005. Evaluation of a system for F0 contour prediction for European Portuguese. In *Interspeech 2005 - Eurospeech. Proceedings of the 9th European Conference on Speech Communication and Technology*, 3249-3251. Lisbon, Portugal, September 4-8, 2005. Disponible en: <http://www.ipb.pt/~joaopt/publicacoes/artigos/F0%20Eurospeech%2005%20ver4.pdf>
- Teixeira, João Paulo Ramos, Diamantino Rui da Silva Freitas, Paulo Duarte Ferreira Gouveia, Gábor Olászy & Géza Németh. 1998. MULTIVOX - Conversor texto fala para português. In *II Encontro Para o Processamento Computacional da Língua Portuguesa Escrita e Falada*, Porto Alegre, Brasil, Novembro de 1998. Disponible en: http://www.ipb.pt/~joaopt/publicacoes/artigos/MULTIVOX%20Conversor%20_%20PROPOR%2098.pdf
- Torres, Humberto M & Jorge A Gurlekian. 2008. Acoustic speech unit segmentation for concatenative synthesis. *Computer Speech and Language* 22,(2), 196-206.
- Torres, María Inés. 2006. El reconocimiento del habla. In Joaquim Llisterra & María Jesús Machuca (eds.), *Los sistemas de diálogo*, 81-98. Bellaterra - Soria: Universitat Autònoma de Barcelona - Fundación Duques de Soria.
- Torres, Rodrigo C, José M de Seixas, Sergio L Netto, Diamantino Rui da Silva Freitas & Eduardo F Brasil. 2008. Portable implementation of a text-to-speech system for Portuguese. In *Proceedings of the European Signal Processing Conference*, Disponible en: <http://www.lps.ufjf.br/profs/sergioln/papers/IC43.pdf>
- Toselli, Alejandro H, Elsa Cubel & Alberto Sanchís. 2006. Prototipo de sistema de automatización de encuestas telefónicas. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), *IV Jornadas en Tecnología del Habla*, 316-320. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: http://lorien.die.upm.es/~lapiz/rtth/JORNADAS/IV/4jth_cdrom.html
- Trancoso, Isabel, Maria do Céu Viana, Fernando M Silva, Gonçalo C Marques & Luís Caldas de Oliveira. 1994. Rule-based vs neural-network based approaches to letter-to-phone conversion for Portuguese common and proper names. In *ICSLP 1994. Proceedings of the 3rd International Conference on Spoken Language Processing*, 1767-1770. Yokohama, Japan, September 18-22, 1994.
- Trancoso, Isabel, Néstor Becerra Yoma, Plínio Almeida Barbosa, Rubén San-Segundo & Kuldeep Paliwal (eds.). 2008. Iberian Languages. Special issue of *Speech Communication* 50(11-12).
- Trancoso, Isabel, Ricardo Nunes, Luís Neves, Maria do Céu Viana, Helena Moniz, Diamantino Caseiro & Ana Isabel Mata. 2006. Recognition of classroom lectures in European Portuguese. In *Interspeech 2006 - ICSLP. Proceedings of the 9th International Conference on Spoken Language Processing*, 281-284. Pittsburgh, PA, USA, September 17-21, 2006. Disponible en: <http://www.inesc-id.pt/pt/indicadores/Ficheiros/3410.pdf>
- Uraga, Esmeralda & Luis Alberto Pineda. 2002. Automatic generation of pronunciation lexicons for Spanish. In Alexander Gelbukh (ed.), *CICLing 2002. Computational Linguistics and Intelligent Text Processing. Proceedings of the Third International Conference*, 330-338. Heidelberg: Springer. Disponible en:

<http://springerlink.metapress.com/content/crl4dnrde4jkhlp4/fulltext.pdf>

- Uria, A, Alfonso Ortega, María Inés Torres, Antonio Miguel, Víctor Gujarrubia, Luis Buera, J Garmendia, Eduardo Lleida, O Aizpuru, Amparo Varona, E Alonso & Óscar Saz. 2006. A virtual butler controlled by speech. In Luis Buera, Eduardo Lleida, Antonio Miguel & Alfonso Ortega (eds.), IV Jornadas en Tecnología del Habla, 297-301. Zaragoza: Universidad de Zaragoza - Red Temática en Tecnologías del Habla. Disponible en: <http://diec.unizar.es/intranet/articulos/uploads/A%20virtual%20butler%20controlled%20by%20speech.PDF>
- Viana, Maria do Céu. 2001. Síntese de fala. In Elisabete Marques Ranchhod (ed.), Tratamento das línguas por computador. Uma introdução à linguística computacional e suas aplicações, 133-193. Lisboa: Caminho.
- Viana, Maria do Céu, Luís Caldas de Oliveira & Ana Isabel Mata. 2003. Prosodic phrasing: Machine and human evaluation. *International Journal of Speech Technology* 6, 83-94. Disponible en: <http://www.inesc-id.pt/pt/indicadores/Ficheiros/930.pdf>
- Violaro, Fábio, Plínio Almeida Barbosa, Eleonora Cavalcante Albano & Edson Françaço. 1996. Um conversor texto-fala para o português brasileiro com processamento lingüístico de alta qualidade. In Anais do VII Simpósio Brasileiro de Microondas e Optoeletrônica, XV Simpósio Brasileiro de Telecomunicações, 361-366. Disponible en: <http://www.lafape.iel.unicamp.br/Publicacoes/telemo1996.pdf>
- Waibel, Alex. 2004. Speech translation: Past, present and future. In *ICSLP - Interspeech 2004. Proceedings of the 8th International Conference on Spoken Language Processing*, 353-356. Jeju Island, Korea, October 4-8, 2004.
- Waibel, Alex, Keni Bernardin & Matthias Wölfel. 2007. Computer-supported human-human multilingual communication. In *Interspeech 2007. Proceedings of the 8th Annual Conference of the International Speech Communication Association*, 14-21. Antwerp, Belgium, August 27-31, 2007.
- Wells, John. 1997. SAMPA for Portuguese. London: Division of Psychology and Language Sciences, University College London. Disponible en: <http://www.phon.ucl.ac.uk/home/sampa/portug.htm>
- Wells, John. 2005. SAMPA Computer Readable Phonetic Alphabet. London: Division of Psychology and Language Sciences, University College London. Disponible en: <http://www.phon.ucl.ac.uk/home/sampa/index.html>