

Llisterri, J. (2009). Les tecnologies de la parla. *Llengua, Societat i Comunicació. Revista de Sociolingüística de la Universitat de Barcelona*, 7, 11-19.

[http://liceu.uab.cat/~joaquim/publicacions/  
Llisterri\\_09\\_Tecnologies\\_Parla.pdf](http://liceu.uab.cat/~joaquim/publicacions/Llisterri_09_Tecnologies_Parla.pdf)

## LES TECNOLOGIES DE LA PARLA

Joaquim Llisteri

Departament de Filologia Espanyola  
Universitat Autònoma de Barcelona

### 1. QUÈ SÓN LES TECNOLOGIES DE LA PARLA?

Al cap del dia sentim moltes veus, la majoria de les quals, de les persones amb qui parlem. Però, de tant en tant, en el paisatge sonor apareixen veus sorgides de la megafonia dels transports públics, de les màquines que fem servir per comprar tabac i bitllets de tren, del GPS del taxi que agafem, o veus que escoltem a l'altra banda de la línia telefònica quan truquem al banc, a un servei d'atenció al client o a una companyia aèria. En aquestes darreres situacions, més d'un cop ens trobem conversant amb una màquina que ens fa preguntes i que, amb una mica de sort, ens arriba a donar els horaris que buscàvem o ens confirma que acabem de reservar una entrada per a un espectacle.

Algunes d'aquestes veus són, en realitat, de persones de carn i ossos que han enregistrat prèviament els missatges que escoltem, de la mateixa manera que nosaltres gravem una salutació en el contestador automàtic del mòbil. D'altres, en canvi, han estat generades artificialment per un ordinador i representen el resultat del treball d'un equip de recerca dedicat al que s'anomenen, globalment, *tecnologies de la parla*. Els investigadors que treballen en aquest camp fan possible que arribem a mantenir un diàleg reeixit amb un sistema informàtic o que, quan conduïm, no ens calgui separar les mans del volant per marcar, amb la veu, un número de telèfon. Arribar a disposar d'aquestes tecnologies, que ara considerem relativament naturals i quotidianes, ha

requerit moltes dècades de treball d'enginyers de telecomunicació, d'informàtics i, vull remarcar-ho especialment, també de lingüistes, que han estudiat de quina manera es poden reproduir artificialment els mecanismes de producció i de percepció de la parla.

De fet, l'interès per les màquines parlants no és gens nou. Ja en el segle XVIII, Wolfgang von Kempelen —també conegut per inventar un autòmat que jugava a escacs i que va resultar ser un frau— va idear un sistema que, imitant els nostres articuladors, permetia reproduir alguns sons aïllats i, pel que sembla a partir de les reproduccions que se n'han fet, fins i tot algunes paraules i frases curtes. Durant les primeres dècades del segle XX, amb l'aparició de l'electrònica, es van crear sistemes com el *Voder* de Homer Dudley, presentat en l'Exposició Universal de Nova York el 1939, al qual van seguir molts altres intents, precursors dels sistemes que coneixem avui (Traunmüller 2000). Els salts quantitatiu i qualitatiu es van produir amb el naixement de la informàtica, que fa possible el tractament digital dels senyals sonors i que és la base de les tecnologies actuals.

Entenem per *tecnologies de la parla* un àmbit de recerca i de desenvolupament que té per objectiu facilitar la interacció entre les persones i els ordinadors mitjançant l'ús de la parla, el nostre sistema de comunicació més natural. En aquest sentit, les tecnologies de la parla contribueixen a augmentar les possibilitats de les interfícies tradicionals, alhora que ens ajuden a fer d'una manera

més senzilla i més ràpida tota una sèrie de tasques que habitualment exigeixen la lectura o l'escriptura. Cal disposar, doncs, com a mínim, de dues tecnologies bàsiques: per una banda, la síntesi de la parla, que fa possible la sortida vocal i, per una altra, el reconeixement de la parla, que permet fer servir la veu per a moltes de les operacions que requeririen un teclat o un ratolí.

L'aplicació més coneguda de la síntesi de la parla és l'anomenada *conversió de text en parla*, ja que serveix per transformar qualsevol text escrit en l'equivalent de la seva lectura en veu alta (Bonafonte, Escudero i Riera 2006). S'empra, sobretot, en el context dels serveis telefònics quan la informació que cal fer arribar al destinatari canvia molt sovint o és del tot imprevisible —pensem en les notícies d'actualitat, la informació sobre el temps o la lectura del correu electrònic—, quan no és factible fer servir una pantalla —aquesta és la situació, per exemple, de les persones amb limitacions de la capacitat visual— o en totes aquelles situacions en què és més pràctic escoltar que llegir.

En canvi, podríem dir que el reconeixement de la parla fa, en certa manera, la funció inversa a la de la síntesi, ja que proporciona un mitjà d'entrada d'informació a un sistema informàtic sense necessitat de tenir ocupades les mans i la vista (Mariño i Nadeu 2004). Per això troba les aplicacions més comunes en el dictat automàtic de texts —que poden anar d'un article sencer a un SMS—, en el comandament vocal de programes informàtics o en accions com ara marcar amb la veu un número de telèfon o cercar un contacte en una agenda.

A vegades, la interacció no s'ha de fer només en una direcció, sinó que cal arribar a un intercanvi per fer una transacció que implica preguntes i respostes, com ara comprar un bitllet o concertar una visita. En aquestes situacions es recorre als sistemes de

diàleg, que incorporen la síntesi, el reconeixement i, el que és més important, la possibilitat d'una «conversa» entre la persona i l'ordinador (López-Cózar 2006). En les versions més avançades, poden incorporar també el rostre del que es coneix com un *assistent virtual* o un *assistent digital interactiu*, per tal que augmenti la sensació d'estar duent a terme una interacció real; una mostra d'aquestes tecnologies la podem trobar, per exemple, en els productes que ofereix l'empresa barcelonina Umanify.<sup>1</sup>

Resumint, la síntesi, el reconeixement i el diàleg constitueixen els tres pilars fonamentals de les tecnologies de la parla i són la base de les aplicacions que presentaré en aquest treball. El seu desenvolupament exigeix, com ja he assenyalat, la col·laboració entre experts de diferents disciplines, entre les quals s'inclou la lingüística i, molt especialment, la fonètica, ja que és la que s'ocupa de manera específica del vessant sonor de la llengua (Llisterrí 2007). De fet, es considera que les tecnologies de la parla formen part d'un domini més ampli, el de les tecnologies del llenguatge o tecnologies lingüístiques, enteses com «la aplicación de los conocimientos sobre la lengua al desarrollo de sistemas informáticos para que puedan comprender, interpretar y generar el lenguaje humano en todas sus formas» (LingLink, sd). Una d'aquestes formes, l'oral, concerneix les tecnologies de la parla.

Disposar de tecnologies de la parla adequades a les diverses situacions que he anat descrivint implica tenir a l'abast un conjunt de recursos i de coneixements lingüístics que no sempre es troben per a totes les llengües. Això explica que en anglès es puguin utilitzar moltes aplicacions i, en canvi, en llengües com el català, el gallec, el basc o, fins i tot, el castellà, aquestes

<sup>1</sup> <http://www.umanify.com>.

A <http://youtube.com/user/umanify> es presenten vídeos de demostració.

aplicacions tinguin una presència molt més reduïda, malgrat l'existència d'un nombre relativament elevat de grups de recerca de qualitat i els esforços de les diferents administracions (Llisterra 2009). Només a tall d'exemple esmentaré que, un cop retirat del mercat el sistema *FreeSpeech* de Philips,<sup>2</sup> presentat el 1999, no sembla que es pugui trobar, fins on sé, cap producte comercial que permeti el dictat automàtic en català; és també simptomàtic que la versió 10 del *ViaVoice* d'IBM,<sup>3</sup> un altre sistema de dictat, ara comercialitzat per Nuance, no incorpori el castellà, que es trobava en versions anteriors.

## 2. LES TECNOLOGIES DE LA PARLA A LA VIDA QUOTIDIANA

Si comencem pensant en les aplicacions de la conversió de text en parla, ens adonarem que troben un lloc natural en la lectura de texts que habitualment mirariem en una pantalla. Per exemple, al final de 2008 Acapela Group va anunciar una versió del conversor especialment preparada per a l'iPhone,<sup>4</sup> mentre que Ivona adaptava *Expressivo* per escoltar els missatges que van arribant a la finestra de xat de l'Skype.<sup>5</sup> De fet, ja fa temps que algunes companyies telefòniques ofereixen l'opció d'escoltar el correu electrònic en comptes de llegir-lo, una possibilitat molt còmoda si tenim en

compte que cada cop més tendim a consultar el correu en dispositius com els mòbils o els organitzadors personals, que solen tenir pantalles més aviat petites. Des d'un punt de vista lingüístic, aquesta mena d'aplicacions requereixen un estudi previ de les estratègies que es fan servir per redactar correus electrònics, SMS i missatges de xat, ja que sovint s'empren recursos específics que allunyen aquests texts de l'ortografia convencional.

En altres contextos no es necessita disposar d'un sistema complet de conversió de text en parla i es treballa amb plantilles de missatges prèviament enregistrats que s'acaben d'omplir amb la informació necessària, extreta de llistes de paraules o d'expressions també gravades per un locutor professional. És el cas, per exemple, dels anuncis dels transports públics que, si s'escolten amb atenció, mostren molt sovint discontinuïtats melòdiques en elements com ara els noms de les estacions, que han estat triats d'una llista i inserits en un enunciat que conté «buits» en els punts on cal afegir la informació pertinent. Podríem dir que aquí ens trobem davant d'una manca de sensibilitat lingüística, ja que les empreses que dissenyen aquests sistemes no solen disposar de l'assessorament d'experts en fonètica. També s'empra aquest tipus de missatges en els sistemes de navegació per GPS incorporats als vehicles, en les informacions sobre l'entorn proporcionades pels sistemes que governen les anomenades *llars intel·ligents* o en certs electrodomèstics que incorporen una sortida vocal.

El reconeixement automàtic de la parla, en la versió més senzilla, la que tracta paraules o expressions curtes aïllades, és la tecnologia pròpia de les aplicacions pensades per controlar tota mena de sistemes, des d'un equip de música fins a un ordinador. Programes com *Dragon Naturally Speaking* de

<sup>2</sup> «Philips i la UPC desenvolupen la versió catalana del programa de reconeixement de la parla humana i conversió a text, *FreeSpeech*», Sala de Premsa, Universitat Politècnica de Catalunya, 30 de setembre de 1999. [http://www.upc.edu/saladeprensa/aldia/mes-noticies/noticia\\_702?set\\_language=ca](http://www.upc.edu/saladeprensa/aldia/mes-noticies/noticia_702?set_language=ca)

<sup>3</sup> <http://www.nuance.com/viavoice>

<sup>4</sup> <http://www.acapela-group.com/acapela-for-iphone-26-speech-solutions.html>.  
Podeu veure'n diverses demostracions en vídeo a <http://www.youtube.com/acapelagroup>.

<sup>5</sup> [http://www.expressivo.com/about\\_expressivo/integration.html](http://www.expressivo.com/about_expressivo/integration.html). Se'n troben vídeos de demostració a: <http://www.youtube.com/user/IvonaTTS>.

Nuance<sup>6</sup> i d'altres de semblants ofereixen la possibilitat de fer servir un explorador amb comandes vocals; els sistemes operatius mateixos —Windows Vista n'és un exemple—<sup>7</sup> porten ja incorporada aquesta tecnologia. El reconeixement també és útil quan no és gaire pràctic utilitzar un teclat, tal com succeeix en els mòbils i els organitzadors personals. Marcar números de telèfon amb la veu és una possibilitat que existeix ja fa temps, però altres aplicacions són més innovadores, com ara la que proposa Google Mobile, que permet fer cerques per la xarxa mitjançant la pronunciació en veu alta de les paraules clau que introduiríem en la finestra de cerca.<sup>8</sup>

Quan cal crear un text sencer en format digital se sol recórrer al dictat automàtic. En aquest cas, el sistema de reconeixement ha de ser capaç de processar més enllà de les paraules aïllades i, per tant, necessita més sofisticació tecnològica, però també més coneixements lingüístics, tant en el nivell fonètic acústic com en el gramatical. El dictat automàtic facilita, en primer lloc, l'ús d'aparells amb teclats poc eficients per escriure texts. L'empresa Nuance va organitzar una competició entre la persona que té el rècord de velocitat en escriptura d'SMS i el sistema de dictat automàtic *Dragon Naturally Speaking* adaptat per a telèfons mòbils i, com era d'esperar, la màquina va superar clarament l'ésser humà.<sup>9</sup> En el terreny de les xarxes socials també sembla que el reconeixement està guanyant terreny: per exemple, QuickTate ha proposat un servei, conegut com a *Tweetcall*, que permet

actualitzar el Twitter fent una trucada telefònica.<sup>10</sup>

A més dels productes que reconeixen la parla, també hi ha programes especialitzats a verificar la identitat d'una persona a partir de la veu (Furui 2006); més endavant faré referència als usos en l'àmbit de la justícia, però com que la veu es pot considerar, en certa manera, un equivalent de l'empremta digital, s'han desenvolupat el que anomenaríem *panyes vocals* per autoritzar l'accés a un edifici, a un ordinador o, fins i tot, per obrir una maleta, com és el cas del *NanoLock* de l'empresa Sensory.<sup>11</sup> Aquestes tècniques, que formen part dels sistemes d'identificació biomètrica, es fonamenten en comparacions estadístiques entre la veu que cal verificar i una base de dades d'elocucions de la persona o del grup que estan autoritzats. El problema que això suposa és evident, ja que cal buscar aquells elements acústics individuals en la veu d'una persona que no variïn al llarg del temps ni segons circumstàncies diverses, que no es trobin en la veu de ningú més i que no es puguin imitar amb facilitat. Actualment aquests sistemes desperten molt d'interès comercial atès que, més enllà de l'entorn domèstic o empresarial, s'està plantejant que serveixin per controlar de manera ràpida la identitat en aeroports i en altres llocs on es formen cues fàcilment. En els serveis telefònics també facilitaríen la identificació de la persona únicament per la veu, eliminant l'inconvenient de memoritzar una contrasenya; un producte que ofereix aquesta possibilitat és *Kivox*, comercialitzat per l'empresa espanyola Agnitio.<sup>12</sup>

Explicava al principi que hi ha serveis que

<sup>6</sup> <http://spain.nuance.com/naturallyspeaking/>

<sup>7</sup> <http://www.microsoft.com/enable/products/windowsvista/speech.aspx>

<sup>8</sup> <http://www.google.com/mobile/apple/app.html>

<sup>9</sup> Es pot veure un vídeo de la competició, per exemple, a <http://www.youtube.com/watch?v=-L4Jk6GDud0>.

<sup>10</sup> <http://tweetcall.com/>

<sup>11</sup> <http://www.sensoryinc.com/>.

Una demostració en vídeo es troba a <http://www.youtube.com/watch?v=e79EyIOSe8>.

<sup>12</sup> <http://www.agnitio.es/espanol/kivox.php>

requereixen una interacció, combinant el reconeixement i la síntesi en un sistema de diàleg. Això succeeix quan hem de demanar una informació, com en el cas de la consulta al *Temps per Telèfon* del Servei Meteorològic de Catalunya, que permet obtenir diverses dades meteorològiques i programar avisos dels canvis de temps.<sup>13</sup> Els avantatges dels serveis automàtics d'informació que fan servir sistemes avançats de reconeixement de la parla és que ens estalvien passar per una sèrie de menús (el clàssic «premi la tecla 1 si vol..., premi 2 si vol...») i que, gràcies a la conversió de text en parla, poden oferir dades molt actualitzades sense haver d'esperar que un locutor professional les enregistri. També fan possible disposar d'un servei constant, de 24 hores al dia durant 365 dies a l'any, sense les despeses de personal que implicaria mantenir un equip d'operadors humans. Els àmbits més habituals d'aquests serveis, a més de la informació sobre el temps, són la informació ciutadana, l'adreçada al turisme —amb l'avantatge que suposa disposar de tecnologies multilingües— i tota la que cal posar al dia amb certa regularitat, com les cartelleres de cinema o les farmàcies de guàrdia.

La interacció és igualment necessària per fer transaccions. En aquest sentit, els sistemes de diàleg són especialment útils en el comerç electrònic —venda d'entrades d'espectacles, reserves de bitllets, etc.—, en la banca electrònica i en els serveis d'atenció al client o al públic en general —per concertar una cita prèvia, per exemple—. Empreses del país com Verbio,<sup>14</sup> NaturalVox<sup>15</sup> o Grupo Voice,<sup>16</sup> entre d'altres, ofereixen solucions als sistemes de

diàleg, han desenvolupat productes basats en les tecnologies de la parla per a oferir aquests serveis.

Finalment, una dimensió que arribarà a tenir una importància cabdal en la vida quotidiana en un món cada cop més globalitzat és la del multilingüisme. Ja disposem de sistemes de traducció automàtica de texts escrits, però s'estan desenvolupant també tecnologies que tradueixin la llengua parlada, tant en el context d'una conversa telefònica com en situacions que, ara per ara, necessiten la presència d'un intèrpret (Waibel, Bernardin i Wölfel 2007). L'especificitat de la llengua oral fa que no s'obtinguin bons resultats aplicant un reconeixedor per transformar la parla en text i enviant el resultat a un sistema convencional de traducció automàtica. Cal desenvolupar traductors especialment pensats per al discurs oral, i aquesta és una tasca que s'ha dut a terme en projectes com TC-STAR<sup>17</sup> o AVIVAVOZ,<sup>18</sup> que incorporen el català i el castellà. De tota manera, cal destacar que es tracta d'una tasca molt complexa i que, com succeeix en la traducció automàtica de texts, el bon funcionament de les aplicacions depèn, en part, de la limitació del domini i del grau de coneixement lingüístic implicat.

### 3. TECNOLOGIES DE LA PARLA EN ÀMBITS ESPECIALITZATS

Si en l'apartat anterior he intentat presentar algunes aplicacions de les tecnologies de la parla que poden incidir directament en la vida quotidiana, cal no oblidar que en entorns professionals determinats aquestes tecnologies també tenen aplicacions. Un exemple molt clar és el dictat automàtic, necessari quan es fa una tasca que ocupa les mans o la vista i,

<sup>13</sup> [http://www.meteo.cat/mediamb\\_xemec/servmet/pagines/inf\\_productes/txt.pdf](http://www.meteo.cat/mediamb_xemec/servmet/pagines/inf_productes/txt.pdf)

<sup>14</sup> <http://www.verbio.com/>

<sup>15</sup> <http://www.natvox.es/>

<sup>16</sup> <http://www.grupovoice.com/>

<sup>17</sup> <http://www.tc-star.org/>

<sup>18</sup> <http://www.avivavoz.es/>

simultàniament, cal prendre notes o redactar un informe. Per això, sistemes com el *Dragon Naturally Speaking* (DNS) es presenten en versions especials per a metges, ja que sovint han de dictar un text mentre miren una radiografia o fan qualsevol altra activitat. Hi ha, igualment, una versió del DNS orientada a l'àmbit jurídic, ja que es tracta d'un camp que exigeix la preparació de molts documents.

En realitat, qualsevol feina que impliqui passar una gran quantitat de temps escrivint texts —la traducció, per exemple— sembla que podria fer-se d'una manera més ràpida emprant un sistema de dictat automàtic, sempre que l'usuari, sobretot al començament, tingui cura d'anar corregint-ne els errors i d'adaptar-lo a idiosincràcies i necessitats. El fet que sistemes operatius com Windows Vista portin ja incorporada la tecnologia per dictar correus electrònics o texts en Word i que empreses com MacSpeech<sup>19</sup> ofereixin productes equivalents per a Mac OS X és una prova de l'interès comercial d'aquesta aplicació de les tecnologies de la parla.

Les tecnologies basades en el reconeixement de la parla s'han incorporat també en el món de l'ensenyament de llengües assistit per ordinador. Programes com el *Tell Me More* d'Auralog<sup>20</sup> avaluen la pronúncia a partir de la comparació automàtica entre un model i l'enregistrament de l'estudiant: marquen, en una representació acústica de la producció de l'alumnat, els punts en els quals l'estudiant s'ha desviat més del model. Val a dir, però, que ara per ara no poden substituir encara el professorat: tot i assenyalar els errors, aquests programes no ofereixen estratègies per corregir-los més enllà d'algunes indicacions generals sobre com es pronuncia

un so determinat, les quals van sovint acompanyades d'una representació, estàtica o animada, de la configuració dels articuladors.

Per acabar aquest apartat, val la pena desfer alguns mites sobre les gairebé miraculoses identificacions de persones a partir de les veus que veiem en sèries com CSI. És cert que hi ha tecnologies d'identificació automàtica del locutor (Furui 2006), però per utilitzar-les cal disposar d'una base de dades específicament creada recollint veus de delinqüents, de manera que es pugui calcular la probabilitat estadística que la veu del sospitós coincideixi amb alguna de les que s'han emmagatzemat prèviament. També s'han desenvolupat sistemes automàtics —*BatVox* de l'empresa Agnitio<sup>21</sup> n'és un exemple— que ajuden els experts en fonètica judicial a comparar la veu d'un sospitós amb la d'un detingut per descobrir si es tracta de la mateixa persona. Els resultats d'aquestes operacions, però, se solen expressar, com ja hem dit, en forma de probabilitats i gairebé mai de la manera tan taxativa com se'ns mostra a la televisió (Rose 2002). Entre les dificultats principals, cal destacar la poca qualitat dels enregistraments obtinguts per les forces de seguretat -per ordre d'un jutge, habitualment mitjançant una intervenció dels telèfons dels sospitosos- i les diferents circumstàncies en què es recullen les mostres de veu quan la persona ja està detinguda i, com és natural, presenta un comportament lingüístic ben diferent del que s'observa quan no sap que està sotmès a vigilància i que se'n graven les trucades.

#### 4. TECNOLOGIES DE LA PARLA PER LA INCLUSIÓ

No voldria acabar aquest treball sense posar de relleu la importància de les tecnologies de la parla per als usuaris que

<sup>19</sup> <http://www.macspeech.com/>

<sup>20</sup> <http://www.tellmemore.com/>

<sup>21</sup> <http://www.agnitio.es/espanol/batvox.php>

presenten alguna necessitat especial, entre els quals cal incloure les persones grans, a més dels col·lectius que es consideren habitualment (Aguilera, Godino, Palazuelos i Martín 2001).

La sortida vocal que proporciona la síntesi i, especialment, la conversió de text en parla, fa que les persones amb visió limitada o inexistent puguin fer servir aparells tan comuns com un rellotge, una bàscula o un termòmetre i que tinguin accés a la informació escrita en la pantalla de qualsevol dispositiu electrònic. L'autonomia i la privacitat que proporciona un conversor de text en parla, combinat, si cal, amb un sistema de reconeixement òptic de caràcters per consultar documents impresos, faciliten la integració en el món laboral i obren noves vies de comunicació en una societat en la qual predomina l'escriptura. Una altra de les aplicacions en aquest àmbit és ben coneguda gràcies a la popularitat de personalitats com Stephen Hawking, que es comunica amb un sistema de conversió de text en parla i un programa dissenyat per l'empresa Words+ per confegir fàcilment els texts.<sup>22</sup>

El reconeixement de la parla, en canvi, és una tecnologia especialment rellevant per a les persones sordes o amb dificultats d'audició, ja que permet transformar qualsevol missatge parlat en text. És igualment útil quan falta la visió, ja que facilita l'ús de mòbils o d'organitzadors personals sense haver de fer servir un teclat. Tampoc no es pot menystenir la importància del control vocal de l'entorn domèstic per a les persones amb mobilitat reduïda.

Dos dels camps orientats a la comunitat sorda que actualment aglutinen molts esforços pel que fa a la recerca són la generació de subtítols, per una banda, i la traducció automàtica de les llengües de signes, per una altra. En el primer cas,

s'apliquen les tècniques de reconeixement automàtic de la parla, amb la dificultat afegida que en cinema o televisió és necessari tractar veus, accents i registres molt diversos i separar la parla de la música o del soroll de fons, tot això en temps real. Per evitar aquests problemes, en ocasions es recorre a una tècnica semiautomàtica, en la qual una persona que veu i escolta l'emissió dicta els subtítols a un sistema de reconeixement de la parla (Aliprandi i Verruso 2006), amb l'ajuda de programes com, per exemple, *ProTitle* de l'empresa Ninsight.<sup>23</sup>

Per a la traducció de les llengües de signes (LS) es combinen tecnologies de reconeixement de la parla i de processament digital d'imatges, ja que s'ha de transformar un senyal sonor en un conjunt de moviments coordinats, si traduïm a l'LS, o convertir moviments significatius en ones sonores, si passem de l'LS a la llengua parlada; a més a més, s'ha de recordar que les llengües de signes constitueixen llengües naturals amb gramàtica pròpia, la qual cosa afegeix un problema de traducció d'estructures lingüístiques. Malgrat les dificultats, les tecnologies actuals permeten ja el desenvolupament de prototips que fan possible, per exemple, la traducció del castellà a la llengua de signes espanyola per ajudar a fer gestions administratives.<sup>24</sup>

## 5. CONSIDERACIONS FINALS

En aquest treball he pretès il·lustrar el paper cada cop més rellevant que les tecnologies de la parla estan assolint tant en la vida quotidiana com en determinats entorns professionals, sense deixar de banda la utilitat que tenen per a les persones amb necessitats especials. Val a dir, però, que la

<sup>22</sup> <http://www.words-plus.com/>

<sup>23</sup> <http://www.ninsight.fr/FR/index.php?page=Produits&f=3>

<sup>24</sup> <http://www.traduccionvozlse.es/>

disponibilitat d'aquestes tecnologies en totes les llengües es veu sovint limitada per qüestions purament relacionades amb el mercat, tot i la vitalitat i la qualitat dels grups de recerca que hi treballen. En alguns moments, els lingüistes i, especialment, els fonetistes, han tingut un paper molt determinant en el desenvolupament de les tecnologies de la parla, encara que en els darrers anys els mètodes basats en l'aprenentatge automàtic a partir de grans corpus de dades han fet que el coneixement lingüístic perdi protagonisme en favor de la comercialització ràpida d'aplicacions centrades en un domini específic. Tot i així, descobrir de quina manera codifiquem una informació lingüística molt complexa en forma d'ones sonores —l'única realitat tangible de la llengua— i conèixer a fons els mecanismes implicats en la producció i la percepció de la parla per dissenyar tecnologies més avançades continua sent, com en els temps de von Kempelen, un repte intel·lectual fascinant.

## 6. REFERÈNCIES BIBLIOGRÀFIQUES

- AGUILERA, S., GODINO, J. I., PALAZUELOS, S. E. i MARTÍN, J. L. (2001). «Aplicaciones sociales de las tecnologías de la lengua». *Quark. Ciencia, Medicina, Comunicación y Cultura*, 21, 90-94. <http://www.prbb.org/quark/21/021090.htm>.
- ALIPRANDI, C. i VERRUSO, F. (2006). «Tecnologie del linguaggio naturale e sottotitolazione multilingue diretta. Lo stato dell'arte in Italia e l'esperienza dei campionati intersteno». *Intralinea. Special Issue on Respeaking*. [http://www.intralinea.it/specials/respeaking/eng\\_more.php?id=453\\_0\\_41\\_0\\_M](http://www.intralinea.it/specials/respeaking/eng_more.php?id=453_0_41_0_M).
- BONAFONTE, A., ESCUDERO, D. i RIERA, M. (2006). «La conversión de texto en habla». LLISTERRI, J. i MACHUCA, M. J. (eds.). *Los sistemas de diálogo*. Bellaterra-Soria: Universitat
- Autònoma de Barcelona-Fundación Duques de Soria, 177-208.
- FURUI, S. (2006). «Automatic speaker recognition and verification». BROWN, K. (ed.). *Encyclopedia of Language & Linguistics*. Amsterdam: Elsevier, 619-27. <http://dx.doi.org/10.1016/B0-08-044854-2/00919-6>.
- LINGLINK (sd). *Ingeniería lingüística. Cómo aprovechar la fuerza del lenguaje* (Folleto preparado por Anite Systems para el proyecto LingLink en nombre de los participantes del sector Ingeniería Lingüística del Programa de Aplicaciones Telemáticas y adaptado al español por el Observatorio Español de Industrias de la Lengua del Instituto Cervantes). Luxemburg: Anite Systems.
- LLISTERRI, J. (2007). «Màquines que parlen i que escolten: el paper de la fonètica en el desenvolupament de les tecnologies de la parla». Carrera, J. i Pons, C. (eds.), *Aplicacions de la fonètica. Catorzè Col·loqui Lingüístic de la Universitat de Barcelona*. Barcelona: Promociones y Publicaciones Universitarias, 139-69. [http://liceu.uab.cat/~joaquin/publicacions/LListerri\\_07\\_Fonetica\\_Tecnologies\\_Parla.pdf](http://liceu.uab.cat/~joaquin/publicacions/LListerri_07_Fonetica_Tecnologies_Parla.pdf).
- LLISTERRI, J. (2009). «Las tecnologías del habla en las lenguas románicas ibéricas». *Studies in Hispanic and Lusophone Linguistics*, 2, 1, 133-180.
- LÓPEZ-CÓZAR, R. (2006). «Análisis y gestión del diálogo». LLISTERRI, J. i MACHUCA, M. J. (eds.). *Los sistemas de diálogo*. Bellaterra-Soria: Universitat Autònoma de Barcelona-Fundación Duques de Soria, 119-52.
- MARIÑO, J. B. i NADEU, C. (2004). «La representación de la voz para el reconocimiento del habla». MARTÍ, M. A. i LLISTERRI, J. (eds.). *Tecnologías del texto y del habla*. Barcelona: Edicions de la Universitat de Barcelona-Fundación Duques de Soria, 187-224.
- ROSE, P. (2002). *Forensic speaker identification*.

Londres: Taylor & Francis.

TRAUNMÜLLER, H. (2000). «Wolfgang von Kempelen's speaking machine and its successors». *Wolfgang von Kempelen on the Web*. Estocolm: Stockholm University. <http://www.ling.su.se/staff/hartmut/kemplane.htm>.

WAIBEL, A., BERNARDIN, K. i WÖLFEL, M. (2007). «Computer-Supported human-human multilingual communication». *Interspeech 2007. Proceedings of the 8<sup>th</sup> Annual Conference of the International Speech Communication Association*. Antwerp (Bèlgica): 27-31 d'agost, 2007, 14-21. [http://www.isca-speech.org/archive/interspeech\\_2007/i07\\_0014.html](http://www.isca-speech.org/archive/interspeech_2007/i07_0014.html).