

Llisterri, J. (2007). Màquines que parlen i que escolten: el paper de la fonètica en el desenvolupament de les tecnologies de la parla. In J. Carrera & C. Pons (Eds.), *Aplicacions de la fonètica. Catorzè Col·loqui Lingüístic de la Universitat de Barcelona* (pp. 139-169). Barcelona: Promociones y Publicaciones Universitarias.

http://liceu.uab.cat/~joaquim/publicacions/Llisterri_07_Fonetica_Tecnologies_Parla.pdf

Màquines que parlen i que escolten: el paper de la fonètica en el desenvolupament de les tecnologies de la parla

Joaquim Llisterra
(Universitat Autònoma de Barcelona)

1 Fonètica i tecnologies de la parla

Les relacions entre la fonètica i les tecnologies de la parla han passat per diversos moments al llarg de la història de les dues disciplines i s'han considerat des de perspectives molt diverses. Mentre que alguns investigadors prou rellevants han defensat la necessitat d'una aproximació basada en el coneixement de la parla com a sistema de codificació de la informació lingüística que es transmet entre l'emissor i el receptor, d'altres han destacat, en canvi, que és possible arribar a desenvolupar productes amb bones prestacions sense gaires coneixements sobre el llenguatge.

Un exemple clar del primer punt de vista —el que defensa aquest treball— és Gunnar Fant, potser el principal investigador del segle XX en l'àmbit que ens ocupa, que el 1983 es preguntava:

«Can we leave it to the computer to learn about speech or shall we insist on developing our own insights in the many dimensions of the speech code?» (Fant 1983: 17)

Reafirmit-se en la seva posició de sempre, Fant ens recordava l'any 2004 que:

«Computing power can not substitute crucial knowledge» (Fant 2004: 11)

L'altre extrem es troba en les ja mítiques paraules atribuïdes a Fred Jelinek, enginyer d'IBM i un dels especialistes més rellevants d'aleshores

en reconeixement automàtic de la parla: «Every time I fire a linguist, my performance goes up», expressió paradigmàtica de l'època en què els mètodes estadístics es començaven a imposar per sobre dels fonamentats en el coneixement.

En l'actualitat trobem, afortunadament, posicions menys extremes que la representada per Jelinek, però no deixa de ser cert que després d'un període de col·laboració molt estreta entre fonetistes i tecnòlegs en la dècada dels cinquanta i a principis dels seixanta,¹

«the linguistic approach soon lost terrain, in recognition applications at least, to (nonlinguistically oriented) engineers who were less concerned with formal linguistic insights, treating the signal as a pattern just like any other, and this with outstanding success». (Barry *et al.* 2005: 1).

El mateix podria dir-se, *mutatis mutandis*, de la síntesi de la parla: del concepte inicial de síntesi per regles que incorporaven, gairebé de forma manual, informació lingüística recollida per experts,² s'ha arribat a l'anomenada síntesi per selecció d'unitats o síntesi basada en corpus,³ que empra, com n'indica el nom, un corpus de realitzacions orals obtingudes d'un únic locutor, del qual s'extreuen automàticament els fragments més adequats al context de l'enunciat que cal sintetitzar, intentant manipular el mínim possible l'enregistrament original.

¹ Un bon exemple d'això són els estudis sobre els indicis perceptius de la parla mitjançant el *Pattern Playback* duts a terme als Laboratoris Haskins per un equip multidisciplinari format per l'enginyer Franklin S. Cooper, el fonetista Pierre Delattre i el psicòleg Alvin Liberman, entre d'altres. Veg., com a mostra dels resultats d'aquest enfocament, Cooper *et al.* (1951) i Delattre *et al.* (1955).

² En aquest terreny, es poden destacar novament els treballs de l'equip dels Laboratoris Haskins (Liberman *et al.* 1959) i, com a sistema completament desenvolupat, el MITalk (Allen *et al.* 1987).

³ Vegeu, per a una presentació general, Möbius (2000) i Campbell (2005). Com a mostra de sistemes desenvolupats amb aquests mètodes es pot consultar, per exemple, Quazza *et al.* (2001).

L'ús cada vegada més estès d'algoritmes que aprenen automàticament a partir d'una quantitat important de dades, combinat amb el fet disposar, almenys per a certes llengües, de grans corpus orals, ha tingut com a conseqüència que, tant en síntesi com en reconeixement, la participació dels lingüistes en els equips que es dediquen a les tecnologies de la parla hagi perdut bona part de la seva rellevància inicial. Si a la mateixa tecnologia s'afegeixen altres obstacles⁴ de naturalesa acadèmica i els que es deriven del que s'ha anomenat la separació entre dues «cultures» —és a dir, la humanística i la tecnològica— s'arriba a la paradoxa que els qui posen al mercat les «màquines que parlen i que escolten» no sempre tenen en compte els coneixements, acumulats des de fa molts anys pels especialistes, sobre com parlem i escoltem les persones.

Per aquest motiu, en aquesta contribució es presenten alguns exemples de sistemes desenvolupats per al català i es mira de fer palesa la necessitat de disposar d'informació lingüística —i, especialment, fonètica— acurada, per tal d'incorporar-la a les tecnologies de la parla.⁵

2 La conversió de text en parla

Un conversor de text en parla és un sistema que transforma qualsevol text, escrit en ortografia convencional, en el seu equivalent sonor;⁶ dit d'una altra manera, és un programa informàtic que ha de dur a terme la mateixa tasca que un bon lector, amb la limitació de no poder comprendre el text tal com ho faria una persona.

En el disseny d'un conversor s'incorporen components informàtics i d'altres propis del processament digital del senyal —que no es tractaran aquí— i mòduls lingüístics, en els quals se centrarà aquest article. Des d'aquesta perspectiva, es poden distingir cinc mòduls bàsics: tractament

⁴ Per a una discussió d'aquests obstacles, veg. Llisterra (2005).

⁵ Un tractament més detallat es pot trobar a Llisterra *et al.* (2003, 2004).

⁶ Sobre la conversió de text en parla, podeu consultar, entre d'altres, Dutoit (1997), Llisterra (2001), Bonafonte *et al.* (2006) i Schroeter (2006).

previ del text, anàlisi lingüística, transcripció fonètica automàtica, assignació de la prosòdia i selecció i concatenació de les unitats de síntesi. En la figura 1 es presenta un esquema, necessàriament simplificat, de les etapes en la conversió de text en parla que es relacionen amb la informació lingüística.

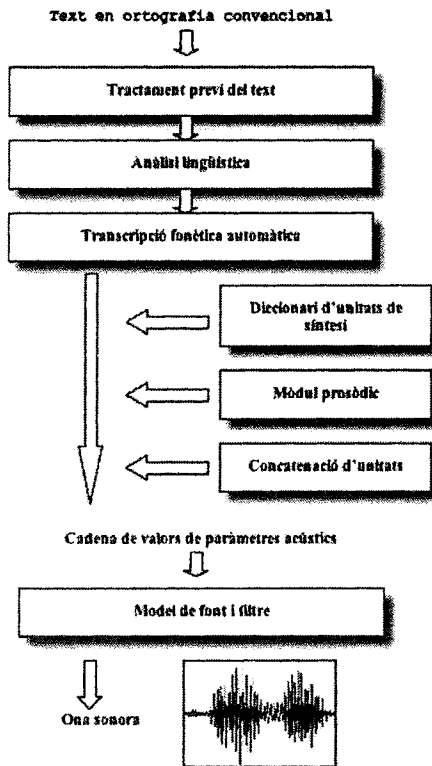


FIGURA 1. Principals mòduls lingüístics i etapes en la conversió de text en parla.

En els apartats següents es repassaran alguns dels problemes de tipus lingüístic que poden sorgir en el procés de conversió d'un text en una manifestació oral. L'article se centra en exemples concrets; s'ha seleccionat, en aquest sentit, una notícia d'actualitat i s'ha sintetitzat

mitjançant tres conversors de text a parla en català que, com a demostració, ofereixen aquesta possibilitat a la xarxa:⁷ el sistema desenvolupat per la secció de Teoria del Senyal del Departament de Comunicacions i Teoria del Senyal d'Enginyeria i Arquitectura La Salle a la Universitat Ramon Llull⁸ i els que comercialitzen l'empresa barcelonina Verbio Technologies⁹ i la torinesa Loquendo.¹⁰ Per completar aquest inventari de sistemes en català, caldria afegir el conversor de Telefónica I+D, que no té demostració en línia,¹¹ i el creat pel Grup de Tractament de la Parla del Departament de Teoria del Senyal i Comunicacions de la Universitat Politècnica de Catalunya.¹² En aquest darrer cas, problemes tècnics en el servidor no han permès provar la demostració amb el nostre text. Cal advertir que les pàgines que segueixen són purament un exercici il·lustratiu en el qual no es pretén, de cap manera, avaluar sistemàticament els conversors emprats ni qüestionar-ne la qualitat.

⁷ Els exemples recollits en aquest treball es poden escoltar a:

http://liceu.uab.es/~joaquin/speech_technology/CLUB14_06/CLUB14_06.html

⁸ <http://cepheus.salleurl.edu/www/ttsweb.htm>. Una presentació del sistema es troba a Gaus i Iriondo (2000) i Alías i Iriondo (2002).

⁹ <http://www.verbio.com/>. De fet, Verbio és l'empresa anomenada abans ATLAS (*Applied Language and Speech Technologies*), originalment un «spin-off» del Grup de Tractament de la Parla del Departament de Teoria del Senyal i Comunicacions de la Universitat Politècnica de Catalunya.

¹⁰ http://www.loquendo.com/en/demos/demo_tts.htm. El tractament de la informació lingüística en aquest sistema es descriu a Bonaventura *et al.* (1998), Garrido *et al.* (2000) i Llisteri *et al.* (2004). Una presentació més general es troba a Quazza *et al.* (2001).

¹¹ Vegeu Cortázar *et al.* (2002) i Villarrubia *et al.* (2003).

¹² <http://gps-tsc.upc.es/veu/soft/demos/tts.php3>. El conversor es descriu a Bonafonte *et al.* (1998).

2.1 El tractament previ del text

Es llegirà, en primer lloc, la notícia que s'ha triat tal com es va publicar al diari digital *VilaWeb* el dissabte 11 de novembre del 2006:¹³

«Unes 1.500 persones es manifesten contra el Barcelona Meeting Point

Diversos col·lectius que lluiten contra l'especulació immobiliària s'han manifestat aquest migdia a Barcelona per a protestar contra la celebració del Barcelona Meeting Point. La concentració, que ha aplegat unes 1.500 persones, ha sortit de la plaça dels Països Catalans darrere una pancarta amb el lema 'Es riuen de nosaltres a la cara!' i ha anat fins al recinte firal de Montjuïc. Allà, davant un important cordó policial, s'ha dissolt. Abans de dissoldre's, els manifestants han representat una partida de Monopoly entre un jove i un especulador, que es disputaven un pis de protecció oficial de dotze metres quadrats. Aquesta no és pas la primera acció de protesta contra la celebració del Barcelona Meeting Point. Dimarts una cinquantena de manifestants es van concentrar a la plaça de Sant Jaume».

Un dels primers problemes que troba un conversor en el processament previ del text és la manera de tractar de les quantitats quan estan expressades amb números. Això implica, entre altres coses, reconèixer que el punt que separa els milers no té la mateixa funció que el que apareix al final de cada frase. També es podrien trobar en un text altres expressions numèriques com ara dates, hores, distàncies o preus i, per tant, cal definir un conjunt de regles que, en aquesta etapa, defineixin quina és l'equivalència, en la forma ortogràfica completa, de cadascuna

¹³ La selecció d'aquest text obeeix al fet que a les pàgines de Verbio es troba el servei «Portal de notícies», que realitza la lectura en temps real, en català, en espanyol (peninsular i de Mèxic) i en portuguès, de notícies publicades a la xarxa. La font de notícies en català és VilaWeb. S'hi pot accedir directament a: http://www.verbio.com/webverbio/html/demos_news.php

d'aquestes expressions. En cas contrari, es produeixen lectures com la del conversor de La Salle: «u cinc-cents» per «1.500».

Aquest mòdul també s'ha d'ocupar dels signes de puntuació. En el text de *VilaWeb* el lema de la pancarta apareix entre cometes simples i, en el sistema de Verbio, per exemple, això produeix una realització sense pauses. Un error tipogràfic del text, la manca d'espai entre «dissolt» i «Abans», fa que el conversor de Loquendo llegeixi «dissolt punt abans».

Una qüestió també sovint difícil de resoldre és la de les sigles i els acrònims; per al seu tractament, com en el cas de les abreviatures, se solen construir taules que associen cadascun d'aquests elements amb la forma ortogràfica expandida o, si convé, amb la transcripció fonètica.

2.2 L'anàlisi lingüística

Un cop s'han superat tots els entrebancs en el text d'entrada, cal dur a terme una anàlisi lingüística que, idealment, hauria d'abraçar des del nivell morfològic fins al pragmàtic. El més habitual, però, és que aquesta anàlisi es redueixi a la morfologia i, en algun cas, a l'estructura dels constituents majors. Les dificultats que apareixen si no es disposa d'informació d'aquesta mena queden ben exemplificades en la lectura de «1.500 persones» com a «mil cinc-cents persones», que es troba en el conversor de Loquendo.

La solució d'aquests problemes no la dona —és evident— el coneixement fonètic. En aquest punt cal incorporar eines de processament del llenguatge natural com són els categoritzadors, que fan explícites les categories gramaticals dels morfemes i la categoria lèxica del mot, i els analitzadors sintàctics automàtics, que defineixen l'estructura dels constituents.¹⁴

¹⁴ Un exemple del funcionament d'aquestes eines en català es pot veure a les pàgines del Centre de Lingüística i Computació de la Universitat de Barcelona: <http://clic.fil.ub.es/>

2.3 La transcripció fonètica automàtica

La finalitat del mòdul de transcripció fonètica automàtica és convertir el text ortogràfic d'entrada en una representació fonètica, ja més propera a la forma oral. Les dificultats d'aquesta tasca depenen, en bona part, de la transparència o de l'opacitat de l'ortografia de cada llengua.

En el cas del català, en el marc del Centre de Referència en Enginyeria Lingüística de la Generalitat de Catalunya (CREL), es va desenvolupar una eina de transcripció, anomenada *Segre*, capaç de generar formes fonètiques en quatre variants dialectals (central, nord-occidental, balear i valencià). Tal com s'exposa a Pachès *et al.* (2000), en català es troben sovint problemes en la realització oberta o tancada de les vocals ortografiades com a *e* i *o*, en la reducció vocàlica quan intervenen restriccions lèxiques, morfològiques o contextuals, en les consonants geminades, en el tractament de la realització fricativa o africada de consonants a final de mot, o en la sonorització o ensordiment també de consonants finals. A l'eina *Segre* es tenen en compte, per tal d'assolir una transcripció acurada, les assimilacions de sonoritat, de mode i de lloc d'articulació. Amb les regles dissenyades per un equip de lingüistes de la Universitat Autònoma de Barcelona i de la Universitat de Barcelona i programades per enginyers de la Universitat Politècnica de Catalunya, s'arriba a obtenir un 99,1% de transcripcions correctes de paraules aïllades i un 99,39% en textos.

S'observaran ara alguns exemples dels transcriptors que s'estan comentant. En el sistema de La Salle es troben alguns casos en què no es fa la reducció vocàlica: «celebració del» se sent com a *[sələβrəsi'o ðel] i «Plaça dels Països» com a *[ˈplasə ðels paˈizus]. També a Verbio TTS s'aprecien algunes mostres de problemes amb la reducció vocàlica: «entre», realitzat com a *[ˈɛ̃trɛ] per la veu masculina i com a *[ˈɛ̃trə] per la femenina, i «protesta» llegit *[pruˈtɛstɛ] en veu masculina. En la paraula «policia» apareixen novament diferències entre les dues veus:

*[polisi'əl] (veu femenina) i [pulisi'əl] (veu masculina). El sistema de Loquendo ofereix la forma *[polisi'əl], mentre que en el de La Salle es pot sentir la versió correcta [pulisi'əl]. Un altre mot conflictiu per a un dels sistemes és «migdia», llegit *[mig'diə] (La Salle) o *[midʒ'ðiə]? (Verbio i Loquendo). Finalment, fixem-nos en la lectura de «concentrar a la» del conversor de La Salle, sense elisió de la *r* ortogràfica final: *[kunsən'trarələ].

El cas de «celebració» que s'acaba de comentar palesa que una altra de les tasques purament lingüístiques implicades en la conversió de text en parla és la definició d'un model d'estàndard oral per al sistema que s'estigui desenvolupant. En el conversor de La Salle s'opta, si ens fixem en [sələβresi'o], per seguir els criteris de l'IEC (1990) pel que fa a la divisió sil·làbica, i es deixa de banda una pronúncia pròpia del registre informal com hauria estat [sələβrə'sjo], que és justament la que es troba a Verbio i a Loquendo. En la figura 2 es pot observar un altre cas equivalent en el mot «especulació». Remarquem també que «aquest» en els tres sistemes es realitza com a [ə'ket] en el fragment «aquest migdia», de manera que es té en compte que el mot següent comença en consonant i, per tant, s'elideix la *s* (Vallverdú 1986).

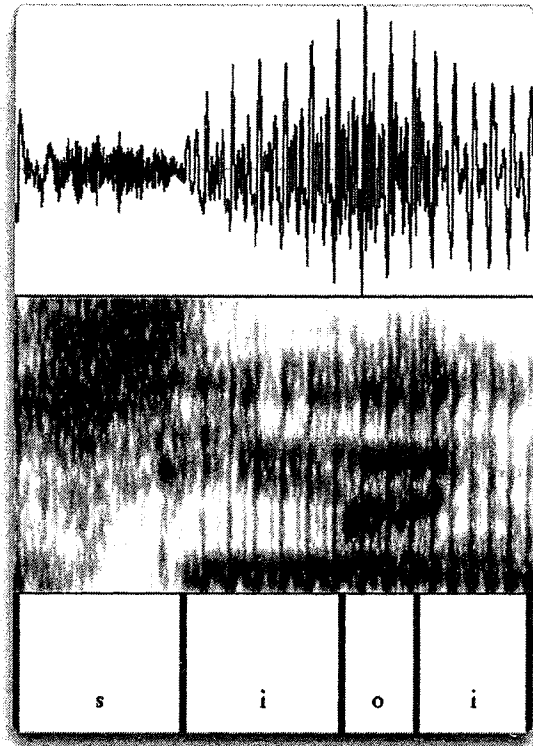


FIGURA 2. Forma d'ona, espectrograma i segmentació del fragment [si'oi] (a «especulació immobiliària») llegit pel conversor de La Salle.

Aquestes mostres il·lustren la dificultat de la transcripció fonètica automàtica que, si ja és elevada en determinades paraules catalanes, es torna molt més difícil quan apareix un mot provinent d'una altra llengua. Centrem-nos en «Meeting Point», transformat en [mæ'tim'pojŋ] (La Salle) o en [mæ'tiŋpu'in] (Verbio i Loquendo); mentre que en el primer cas s'aplica la reducció vocàlica en el context *ee*, s'elideix la *g* final, s'aplica la regla d'assimilació de lloc d'articulació a *n* i es realitza la diftongació de *oi*, considerada una síl·laba accentuada, en el segon actuen unes regles completament diferents que reflecteixen la consideració del context *ee* en la reducció vocàlica, l'assimilació de *n* al lloc d'articulació de l'oclusiva

velar sonora següent, que es manté, i que assignen l'accent a la síl·laba final sense reduir la *o*.

Possiblement cap de les dues versions és prou encertada, però sense una anàlisi prèvia del text que marqui els mots als quals no s'han d'aplicar les regles de transcripció del català, els resultats responen al funcionament regular del sistema. Tot i així, s'arriben a detectar inconsistències: en el sistema de Verbio, per exemple, «Monopoly» es converteix en *[ˈmonopuli] en la veu d'home i en [munuˈpɔli]? en la de dona, com succeïa a «policia».

Una manera de resoldre aquest inconvenient és elaborar llistes de paraules estrangeres amb la corresponent transcripció fonètica, cosa que no deixa de plantejar noves dificultats, per dos motius: en primer lloc, en un text periodístic és pràcticament impossible de preveure un lèxic que, per altra banda, va canviant segons els esdeveniments, sobretot pel que fa als noms propis; en segon lloc, caldria prendre decisions ortològiques sobre la pronúncia dels noms estrangers, la font dels quals hauria de ser, ara per ara, els llibres d'estil dels mitjans de comunicació orals, que no sempre cobreixen tota la casuística.

2.4 L'assignació de la prosòdia

Un cop el text d'entrada al conversor ha estat processat, analitzat en la mesura en què ho permeten les eines disponibles i fonèticament transcrit, arriba el moment d'incorporar la informació sobre els elements prosòdics.

Es farà referència, per començar, a les pauses, entre les quals cal distingir les que ja vénen marcades pels signes de puntuació, que plantegen poques dificultats, i les que no ho estan, però que es trobarien en una bona lectura. En aquest darrer cas, cal recórrer a informació sobre la categoria lèxica —de manera que, per exemple, no s'insereixi una pausa entre un article i un nom o un nom i un adjectiu— i sobre l'estructura sintàctica si es disposa d'un analitzador adient. S'observa, en la figura 3, la realització del fragment «acció de protesta» pel conversor

de Verbio, amb una pausa d'uns 90 ms entre «acció» i «de», que probablement no realitzaria un lector humà.

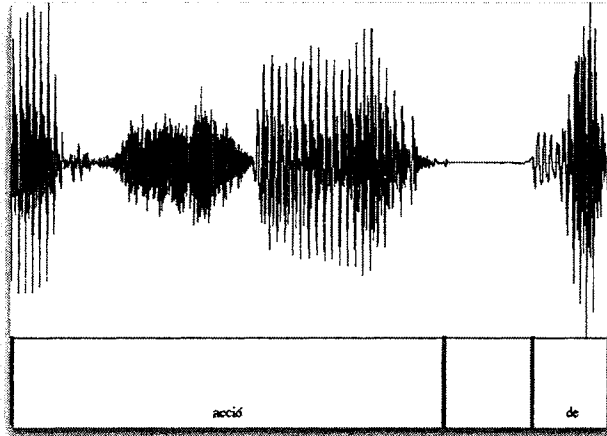


FIGURA 3. Forma d'ona i segmentació del fragment «acció de» en la realització del conversor de Verbio.

Novament al sistema de Verbio es troba una pausa que no és del tot natural: a la figura 4 s'aprecia que en l'enunciat «contra la celebració» apareix una pausa d'uns 48 ms després de «contra».

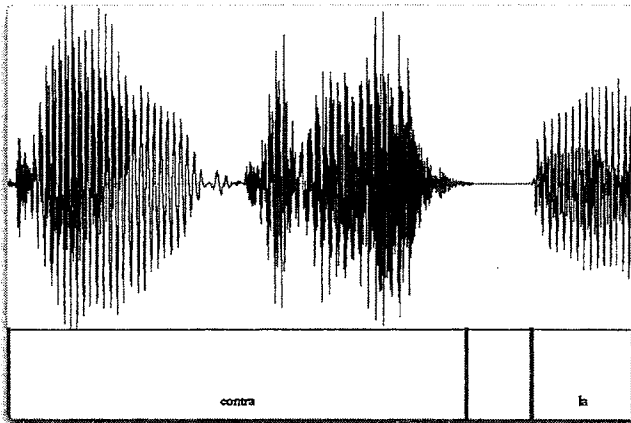


FIGURA 4. Forma d'ona i segmentació del fragment «contra la» en la realització del conversor de Verbio.

Un segon mòdul del component prosòdic d'un conversor de text en parla és el que s'encarrega de determinar el perfil de la corba melòdica dels enunciats o, en termes acústics, l'evolució en el temps de la freqüència fonamental (F_0). Juntament amb les pauses, aquest és, probablement, un dels factors que més contribueix a la naturalitat de la síntesi i a la bona comprensió dels enunciats. Per aquest motiu, el tractament de la melodia —i el dels elements suprasegmentals en general— ha estat, en els darrers anys, un dels temes centrals de la recerca en tecnologies de la parla.¹⁵

En els conversors que s'estan fent servir per il·lustrar aquest treball no es deixen de trobar alguns problemes pel que fa a la melodia. Un exemple el forneix Verbio TTS, que, en la realització de «a la Plaça de Sant Jaume.», en comptes del final descendent propi de les enunciatives que assenyala, també, final de l'enunciat, presenta una pujada que es pot observar en la figura 5 i que, en el moment de sentir-lo, fa la impressió que la frase hagi de continuar.

¹⁵ Per a una revisió d'aquests aspectes, veg. Llisterra *et al.* (2003) i Batliner i Möbius (2005). Un estat de la qüestió específicament orientat a la conversió de text en parla es pot llegir a Monaghan (2002 a, b).

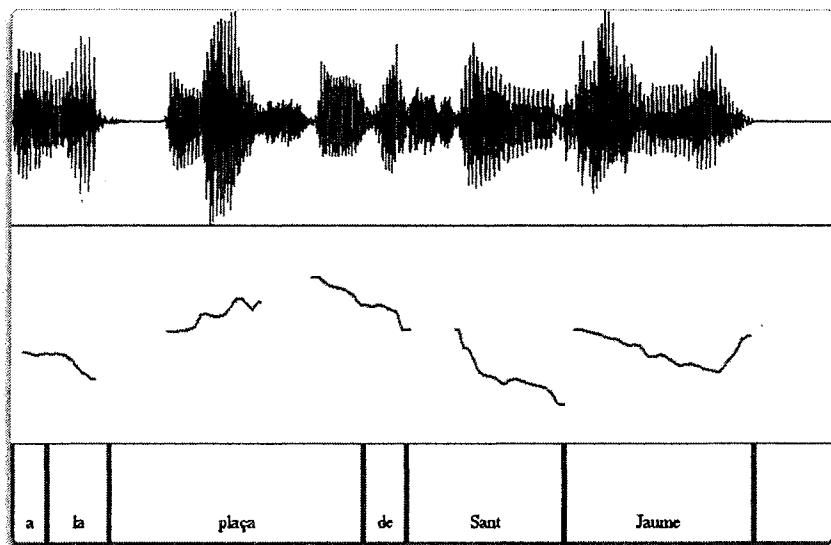


FIGURA 5. *Forma d'ona, corba melòdica i segmentació del fragment «a la Plaça de Sant Jaume.» en el conversor de Verbio.*

Quan es fa servir una tècnica de síntesi basada en corpus cal tenir cura, en la selecció d'unitats, de buscar-les en contextos que permetin evitar realitzacions com la que es visualitza en la figura 6: és fàcil adonar-se que, en la veu femenina de Verbio, existeix una discontinuïtat melòdica important entre «manifesten» i «contra».

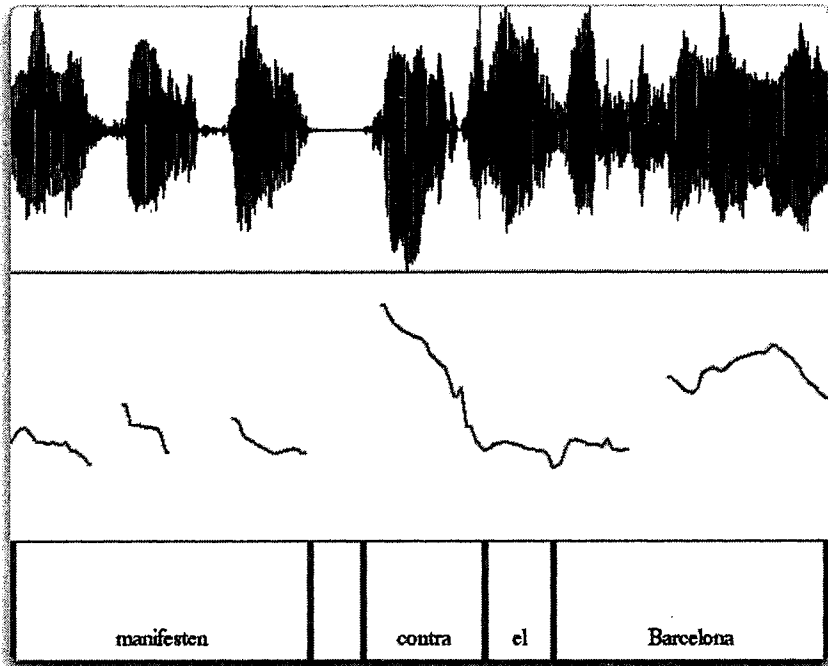


FIGURA 6. Forma d'ona, corba melòdica i segmentació del fragment «manifesten contra el Barcelona» en el conversor de Verbio.

Els reinicis de la corba melòdica poden obeir a una pausa i, si aquesta no s'ha situat en un punt adient, arriben a sorgir fenòmens com el que s'il·lustra en la figura 7: observeu que la trajectòria de la corba és descendent a «migdia» i a «contra» però que, en canvi, comença en un punt més alt respecte al valor d' F_0 que trobem en acabar aquests dos mots després de les pauses que precedeixen «a Barcelona» i «la celebració». Per contra, en la figura 8 s'aprecia que, com que no es fan pauses, la corba melòdica manté un perfil diferent.

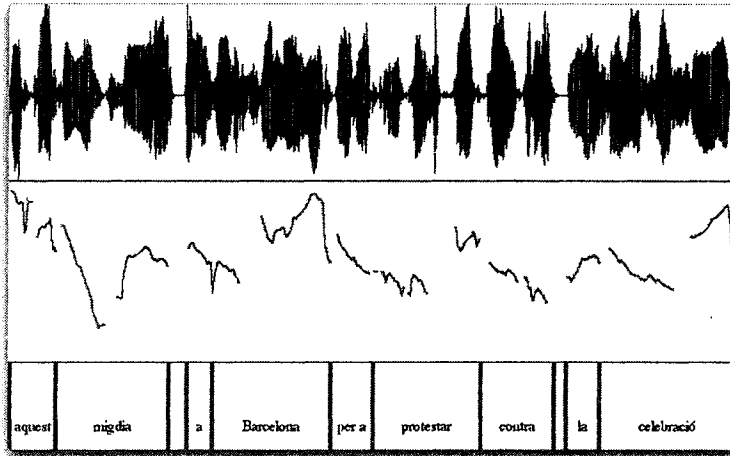


FIGURA 7. Forma d'ona, corba melòdica i segmentació del fragment «aquest migdia a Barcelona per a protestar contra la celebració» en el convertor de Verbio.

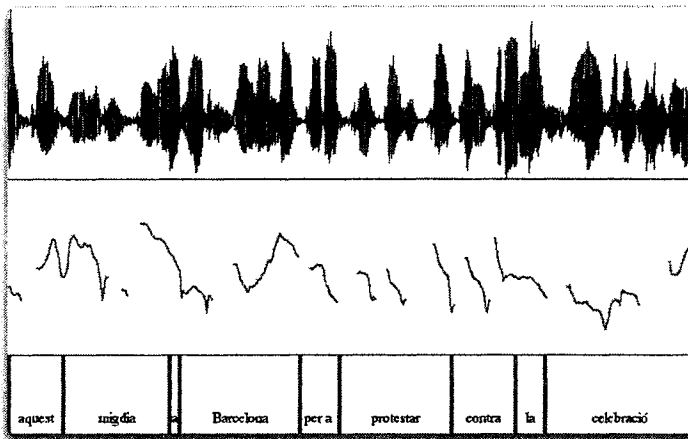


FIGURA 8. Forma d'ona, corba melòdica i segmentació del fragment «aquest migdia a Barcelona per a protestar contra la celebració» en el convertor de Loquendo.

Altres aspectes que s'han d'ajustar en el mòdul prosòdic són la durada de cadascun dels sons, especialment si es treballa a partir de la concatenació d'unitats curtes com ara el difonema,¹⁶ i el perfil de la corba d'intensitat.

2.5 La selecció d'unitats i les unitats de síntesi

Com es mostra en la figura 1, un conversor de text a parla ha de disposar d'un inventari d'unitats que, concatenades, formaran els enunciats que cal llegir. Tradicionalment aquestes unitats havien estat els difonemes (veg. la nota 16), combinats amb trifonemes o quadrifonemes per tractar, per exemple, els grups consonàntics formats per una oclusiva i una líquida, que s'envoltaven de vocals per facilitar la concatenació per una zona acústicament estable del senyal. Treballar amb aquestes unitats implica una tasca lingüística prèvia, que és la definició de l'inventari de fonemes i d'al·lòfons que es tindran en compte en la síntesi.

En aquest sentit, és interessant estudiar la taula de sons del català emprats en el conversor de La Salle¹⁷, que inclou l'africada palatal sorda però no la sonora —recordem la pronúncia *[mig'diə] a la qual es feia referència en el § 2.1—, que descriu la [β] com una aproximant palatal sonora i considera fricatiu labiodental el so [θ] (incorporat, igual que [x], per fer possible la síntesi de noms castellans). Si afegim que s'agrupa sota la denominació de «fonemes» les tres aproximants [β ð γ] i la nasal velar, no sembla que hi hagi motius per qüestionar que la definició de l'inventari de segments a partir dels quals es dissenyen les unitats de síntesi és una activitat que requereix disposar de la participació de lingüistes.

¹⁶ Un difonema es defineix habitualment com el segment que va des del principi de la segona meitat del primer so que el forma fins al final de la primera meitat del segon so. Així, el difonema [ma] començaria al centre de la [m] i acabaria al centre de la [a]. D'aquesta manera s'eviten, en la síntesi, les concatenacions per les parts menys estables del senyal sonor que es troben, en general, a la transició entre consonants i vocals.

¹⁷ Es pot consultar a: <http://cepheus.salleurl.edu/www/fonemes.htm>

Un cop s'ha fixat el conjunt de fonemes i al·lòfons, cal determinar-ne les possibilitats combinatòries per tal d'arribar al conjunt de difonemes i altres unitats majors que s'hauran de concatenar per confegir els enunciats. No tenir en compte algunes de les combinacions pot portar a lectures com [pə'riðə] per «partida», tal com es troba en el conversor de La Salle (Figura 9).

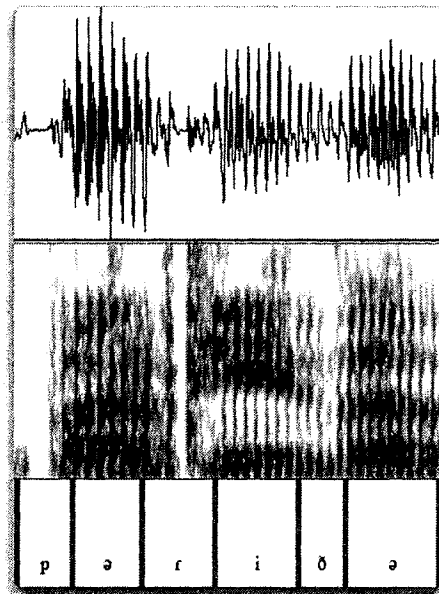


FIGURA 9. Forma d'ona, espectrograma i segmentació del mot «partida» en el conversor de La Salle.

El sistema basats en corpus, tal com s'ha exposat, busquen en els enregistraments d'un locutor el fragment més semblant al text que s'ha de sintetitzar. Això implica que moltes vegades es treballa amb combinacions més llargues que els difonemes i, per aquest motiu, la qualitat sol ser més elevada, però també cal valorar el grau d'adequació prosòdica i, si convé, corregir la corba melòdica per tal de no introduir

discontinuitats excessives —com les que s’observaven en la figura 6— ni perturbacions en el patró accentual.

3 Els sistemes de diàleg

Un conversor de text a parla pot tenir moltes aplicacions, però potser la més freqüent en la vida quotidiana és com a part d’un sistema de diàleg. Els sistemes de diàleg o sistemes conversacionals¹⁸ inclouen, a més d’un conversor, un sistema de reconeixement automàtic de la parla que transforma en text allò que ha dit un usuari; un mòdul d’interpretació semàntica, que n’extreu la informació rellevant en el context d’una determinada aplicació; un mòdul de generació de respostes, que crea un enunciat coherent en funció del que s’ha interpretat que volia l’usuari, i, finalment, un conversor de text en parla encarregat de proporcionar una versió sonora d’aquest enunciat; l’operació de tots aquests mòduls és controlada per un gestor del diàleg, responsable que es mantingui una interacció adequada (*fig. 10*).

¹⁸ Els sistemes de diàleg es descriuen, per exemple, a Tapias (2002), Dahl (2004), McTear (2004), Llisterrí i Machuca (2006) i López-Cózar i Araki (2005).

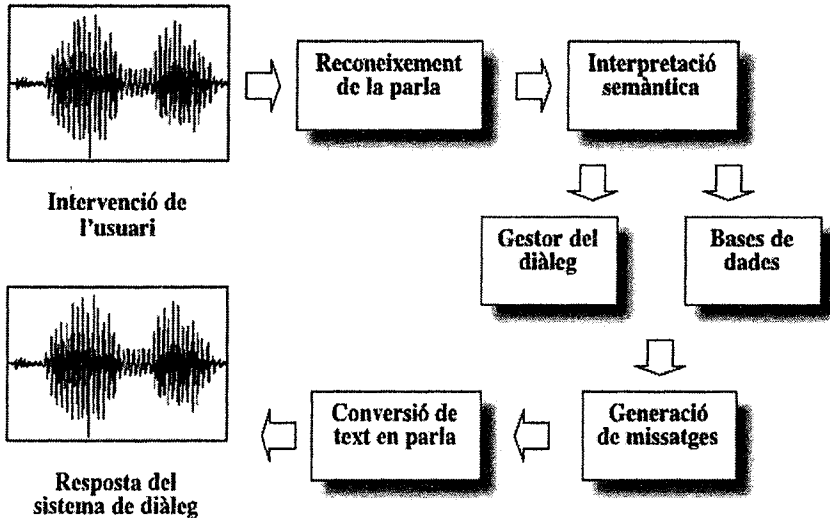


FIGURA 10. Principals mòduls lingüístics d'un sistema de diàleg.

Un exemple de sistema conversacional en català, desenvolupat per un equip multidisciplinari de lingüistes del Departament de Filologia Espanyola de la Universitat Autònoma de Barcelona i del Departament de Lingüística de la Universitat de Barcelona, enginyers de telecomunicació de la Universitat Politècnica de Catalunya i empreses especialitzades en missatgeria electrònica i en tecnologies lingüístiques, és el sistema de diàleg que actualment empra el Servei Meteorològic de Catalunya per tal d'oferir informació telefònica sobre el temps; a més d'integrar la síntesi i el reconeixement, permet programar alertes i rebre la informació pertinent en forma de trucada o de missatge curt al telèfon mòbil.¹⁹ L'empresa Verbio Technologies també ofereix, entre els seus productes en català, el convertidor al qual ja s'ha fet referència (§ 2) i un sistema de

¹⁹ Es pot trobar més informació a les dues adreces següents:

<http://gps-tsc.upc.es/veu/attempts/>

http://www.meteocat.com/pagines/inf_productes/txt.pdf. El número de telèfon del servei és el 807012012.

reconeixement automàtic de la parla que, juntament amb altres eines com Dialog@, fan possible la interacció a distància amb sistemes informàtics i, per tant, automatitzar diversos serveis d'atenció al client.²⁰

Els coneixements fonètics d'un sistema de diàleg es localitzen, principalment, en el mòdul de conversió de text en parla i en el de reconeixement. Ja s'ha presentat el primer i, pel que fa al segon, cal recordar que els reconeixadors es construeixen mitjançant l'aprenentatge automàtic a partir d'un corpus de dades que sigui prou representatiu, motiu pel qual la intervenció del lingüista és sovint mínima, per no dir inexistent, tal com s'avançava en el § 1.

Tot i així, igualment cal definir les unitats que utilitzarà el reconeixedor —tal com es veia en el § 2.5, és el que s'havia de fer per a la síntesi— i considerar acuradament el disseny del corpus amb el qual s'entrenarà el sistema, ja que no només es tracta de recollir dades d'un nombre molt elevat de locutors,²¹ sinó que s'han de tenir en compte criteris lingüístics com la variació geogràfica —per això s'ha de recórrer als coneixements dels dialectòlegs— i la diversitat d'estils de parla. D'aquests corpus s'extreuen també les variants de realització que es poden trobar en un mot, se sistematitzen i es confegeixen diccionaris de pronúncia que ajudaran, ja en una situació d'ús real, al bon funcionament del reconeixedor.

Si s'observen els mòduls que aporten informació lingüística en un sistema de reconeixement de la parla (*fig. 11*), es constata fàcilment que trobem determinats aspectes, entre els quals hi ha els que s'acaben de descriure, que requereixen un coneixement especialitzat.

²⁰ Les instruccions per accedir a una demostració del sistema de reconeixement es troben a http://www.verbio.com/webverbio/html/demos_basic.php

²¹ Per exemple, el corpus de *SpeechDat* per al català, desenvolupat pel Grup de Tractament de la Parla de la Universitat Politècnica de Catalunya amb la col·laboració del Grup de Fonètica del Departament de Filologia Espanyola de la Universitat Autònoma de Barcelona en el marc del CREL, conté elocucions de 1.005 locutors. En altres projectes més recents es preveu recollir, per telèfon, dades de 4.000 locutors (*Generació de recursos...*).

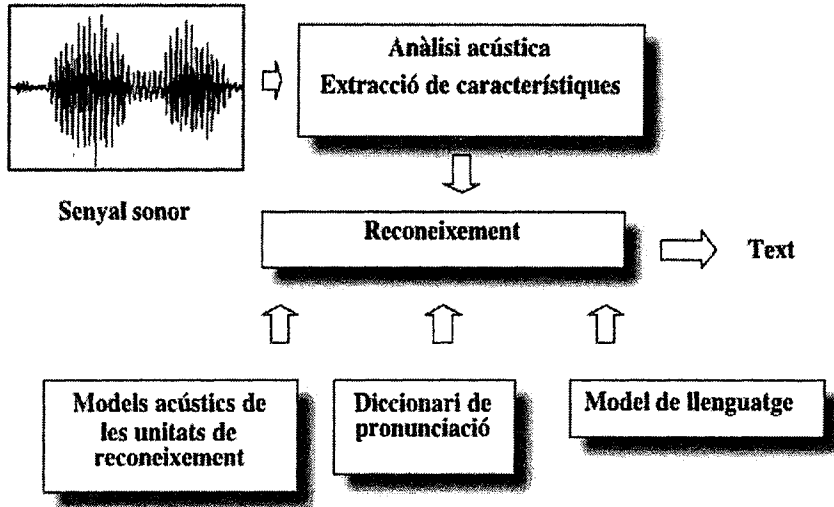


FIGURA 11. Principals mòduls lingüístics d'un sistema de reconeixement automàtic de la parla.

De fet, dit d'una manera molt simplificada, es podria considerar que un reconeixedor és un programa que compara el senyal d'entrada produït pel parlant amb els models de les unitats acústiques que ha extret del corpus d'entrenament, amb el model de llenguatge (o «gramàtica»), que consisteix en probabilitats d'aparició de seqüències també apreses del corpus, i amb el diccionari de pronúncia, per tal d'arribar, combinant aquestes dades, a una hipòtesi sobre el text que pot correspondre a l'enunciat d'entrada²².

En els mòduls dels sistemes de diàleg relacionats amb la comprensió i amb la generació de respostes (fig. 10) s'empren tècniques pròpies del processament del llenguatge natural, que ja són fora de l'abast d'aquest treball. Finalment, el disseny de les estratègies de diàleg —com es

²² Sobre els sistemes de reconeixement automàtic de la parla vegeu, entre d'altres, Deroo (1999), Mariño i Nadeu (2004) i Torres (2006).

pregunta, com es confirma que s'ha interpretat correctament el contingut d'allò que ha dit l'usuari, com se solucionen els errors que puguin sorgir, com s'obre i es tanca el diàleg, etc.— no hauria de deixar de banda tot el que s'ha après sobre la interacció entre persones en els estudis centrats en l'anàlisi de la conversa; d'aquesta manera, una altra disciplina, la pragmàtica, s'implica en el camp de les tecnologies de la parla.

4 Reflexió final

En començar aquest treball, s'esmentava que en l'àmbit de les tecnologies de la parla sovint es desenvolupen sistemes que no tenen en compte el coneixement fonètic o, en general, lingüístic, acumulat després de molts anys de recerca bàsica sobre els mecanismes de la comunicació parlada. S'ha intentat donar alguns exemples de la complexitat del procés de conversió de text a parla, i s'ha fet referència, també, al reconeixement i al diàleg per posar en evidència que, com encertadament assenyala Steven Greenberg,

«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». (Greenberg 2005: 111).

El mateix autor, investigador a l'International Computer Science Institute de la Universitat de Califòrnia a Berkeley, també adverteix que:

«[...] an emphasis on machine-learning algorithms and training regimes often comes at the expense of genuine insight into the nature

of spoken language and not infrequently violates the precepts of the hypotetico-deductive method». (Greenberg 2005: 111).

Precisament per aquest motiu s'ha pretès destacar que la col·laboració entre experts de disciplines diferents és imprescindible si es vol investigar seriosament i assolir un nivell de qualitat alt. Això és necessari no només per motius pràctics lligats a les demandes del mercat, sinó, sobretot, per fer ciència madura, reprenent la terminologia de Greenberg. Crec que, al cap i a la fi, un lingüista i un enginyer treballant plegats són, com diria Mario Benedetti,²³ «mucho más que dos».

Referències bibliogràfiques

- ALÍAS, Francesc; IRIONDO, Ignasi (2002): «La evolución de la síntesis del habla en Ingeniería La Salle». A: RUBIO, A. (ed.) (2002): *Actas de las II Jornadas en Tecnologías del Habla*. Granada: Universidad de Granada, Departamento de Electrónica y Tecnología de Computadores.
<http://lorien.die.upm.es/~lapiz/rth/JORNADAS/II/articulos/34.pdf>
- ALLEN, Jonathan; HUNNICUTT, M. Sharon; KLATT, Dennis H. (en col·laboració amb Robert C. ARMSTRONG i David PISONI) (1987): *From Text to Speech: The MITalk System*. Cambridge: Cambridge University Press.
- BARRY, William J.; VAN DOMMELEN, Wim A.; KOREMAN, Jacques (2005): «Phonetic knowledge in speech technology - and phonetic knowledge from speech technology?». A: BARRY, W. J.; VAN DOMMELEN, W. A. (eds.) (2005): *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pàg. 1-12.
<http://www.coli.uni-saarland.de/~koreman/Publications/2005/Eurospeech2001.pdf>

²³ «Te quiero». Benedetti (1984).

- BATLINER, Anton; MÖBIUS, Bernd (2005): «Prosodic models, automatic speech understanding, and speech synthesis: Towards the common ground?». A: BARRY, W. J.; VAN DOMMELEN, W. A. (eds.) (2005): *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pàg. 21-44.
- BENEDETTI, MARIO (1984) *Poemas de otros*. Madrid: Visor.
- BONAFONTE, Antonio; ESCUDERO, David; RIERA, Montserrat (2006): «La conversión de texto en habla». A: LLISTERRI, J.; MACHUCA, M. J. (eds.) (2006): *Los sistemas de diálogo*. Bellaterra: Universitat Autònoma de Barcelona - Fundación Duques de Soria. pàg. 177-208.
- BONAFONTE, Antonio; ESQUERRA, Ignasi; FEBRER, Albert; FONOLLOSA, José A. R.; VALLVERDÚ, Francesc (1998): «The UPC Text-to-Speech System for Spanish and Catalan». A: *ICSLP 1998. Proceedings of the 5th International Conference on Spoken Language Processing*. Sidney (Austràlia), desembre del 1998. Vol. 5. pàg. 1667-1670.
<http://gps-tsc.upc.es/veu/research/pubs/download/Bon98c.pdf>
- BONAVENTURA, Patrizia; GIULIANI, Fabio; GARRIDO, Juan M.; ORTÍN, Isabel (1998): «Grapheme-to-phoneme transcription rules for Spanish, with application to automatic speech recognition and synthesis». A: *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)*. Montreal (Canadà), agost del 1998. pàg. 33-39.
<http://acl.ldc.upenn.edu/W/W98/W98-0804.pdf>
- CAMPBELL, N. (2005): «Developments in corpus-based speech synthesis: Approaching natural conversational speech». *IEICE Transactions on Information and Systems* E88-D, 3. pàg. 376-383.
- COOPER, Franklin S.; LIBERMAN, Alvin M.; BORST, John M. (1951): «The Interconversion of Audible and Visible Patterns as a Basis for Research in the Perception of Speech». *Proceedings of the National*

- Academy of Sciences* 37. pàgs. 318- 325. A: FLANAGAN, J. L.; RABINER, L. R. (eds.) (1973): *Speech Synthesis*. Stroudsburg: Dowden, Hutchinson & Ross Inc. pàg. 59-66. A: FRY, D. B. (ed.) (1976): *Acoustic Phonetics. A Book of Basic Readings*. Cambridge: Cambridge University Press. pàg. 204-207.
- CORTÁZAR, Ismael; RODRÍGUEZ Miguel Á.; GARRIDO Juan M.; CAMINERO Francisco J.; BERNAT, Jesús; RELAÑO José; GARIJO, Francisco J; HERNÁNDEZ, Luis A. (2002): «Últimos desarrollos en tecnologías del voz y del lenguaje». *Comunicaciones de Telefónica I+D* 24. pàg. 25-64.
http://www.tid.es/documentos/revista_comunicaciones_i%2Bd/numero24.pdf
- DAHL, Deborah (ed.) (2004): *Practical Spoken Dialog Systems*. Dordrecht: Kluwer.
- DELATTRE, Pierre C.; LIBERMAN, Alvin M.; COOPER, Franklin S. (1955): «Acoustic Loci and Transitional Cues for Consonants». *Journal of the Acoustical Society of America* 27. pàg. 769-773. A: LEHISTE, I. (ed.) (1967): *Readings in Acoustic Phonetics*. Cambridge: The MIT Press. pàg. 283-287. A: FRY, D. B. (ed.) (1976): *Acoustic Phonetics. A Book of Basic Readings*. Cambridge: Cambridge University Press. pàg. 273-283.
- DEROO, Olivier (1999): *A Short Introduction to Speech Recognition*. TCTS Lab. Faculté Polytechnique de Mons.
<http://tcts.fpms.ac.be/asr/intro.php>
- DUTOIT, Thierry (1997): *An Introduction to Text-to-Speech Synthesis*. Dordrecht: Kluwer.
- FANT, Gunnar (1983): «Phonetics and Speech Technology». *Speech Transmission Laboratory - Quarterly Progress and Status Report* 2-3. pàg. 20-35. A: VAN DEN BROECKE, M. P. R.; COHEN, A. (eds.) (1984): *Proceedings of the 10th International Congress of Phonetic Sciences*. Dordrecht: Foris. pàg. 13-24.

- http://www.speech.kth.se/qpsr/pdf/1983/1983_24_2-3_020-035.pdf
FANT, Gunnar (2004) «More than half a century in phonetics and speech research». A: FANT, G. (2004): *Speech Acoustics and Phonetics*. Dordrecht: Kluwer. pàg. 1-14.
<http://www.speech.kth.se/~gunnar/halfcentury.pdf>
- GARRIDO, Juan M.; ORTÍN, Isabel; QUAZZA, Silvia; SALZA, Pier.Luigi; MANCINI, Franca (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. pàg. 183-190.
<http://www.sepln.org/revistaSEPLN/revista/26/garrido-alminana.pdf>
Generació de recursos lingüístics per al desenvolupament de tecnologies de la parla en català. Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya - Universitat Politècnica de Catalunya. Pàg. 1-14.
<http://www10.gencat.net/dursi/generados/catala/departament/recurs/doc/tecnologies.pdf>
- GREENBERG, Steven (2005): «From here to utility». A: BARRY, W. J.; VAN DOMMELEN, W. A. (eds.) (2005): *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pàg. 107-132.
http://www.icsi.berkeley.edu/%7Esteveng/PDF/Phonetic_Insight.pdf
- GUAUS, Roger; IRIONDO, Ignasi (2000): «Diphone based Unit Selection for Catalan Text-to-Speech Synthesis». A: *TSD 2000. Workshop on Text, Speech and Dialogue*. Brno (República Txeca), setembre del 2000.
<http://serpens.salleurl.edu/intranet/pdf/241.pdf>
- INSTITUT D'ESTUDIS CATALANS. SECCIÓ FILOLÒGICA (1990): *Proposta per a un estàndard oral de la llengua catalana I. Fonètica*. Barcelona: Institut d'Estudis Catalans. [3a edició revisada, 1998].
<http://www.iecat.net/institucio/seccions/Filologica/pdf/Proposta%20estandard%201.pdf>

- LIBERMAN, Alvin M.; INGEMANN, Frances; LISKER, Leigh; DELATTRE, Pierre; COOPER, Franklin S. (1959): «Minimal Rules for Synthesizing Speech». *Journal of the Acoustical Society of America* 31. pàg. 1490-1499. A: FLANAGAN, J. L.; RABINER, L. R. (eds.) (1973): *Speech Synthesis*. Stroudsburg: Dowden, Hutchinson & Ross Inc. pàgs. 351-367. A: FRY, D. B. (ed.) (1976): *Acoustic Phonetics. A Book of Basic Readings*. Cambridge: Cambridge University Press. Pàg. 445-466. A: LEHISTE, I. (ed.) (1967): *Readings in Acoustic Phonetics*. Cambridge: The MIT Press. pàg. 333-342.
- LLISTERRI, Joaquim (2001): «La conversión de texto en habla». *Quark. Ciencia, Medicina, Comunicación y Cultura* 21. pàg. 79-89.
- LLISTERRI, J. (2005) «El papel de la fonética en las tecnologías del habla». A: *Actas del III Congreso de Fonética Experimental*. Santiago de Compostela (Espanya), octubre del 2005. (en premsa).
http://liceu.uab.es/~joaquim/publicacions/Llisterri_05_Fonetica_Tecnologias_Habla.pdf
http://liceu.uab.es/~joaquim/publicacions/CTH_Quark_01.pdf
- LLISTERRI, Joaquim; CARBÓ, Carme; MACHUCA, María J.; DE LA MOTA, Carme; RIERA, Montserrat; RÍOS, Antonio (2003): «El papel de la lingüística en el desarrollo de las tecnologías del habla». A: CASAS, M. (dir.); VARO, C. (ed.) (2005): *VII Jornadas de Lingüística*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz. pàg. 137-191.
http://liceu.uab.es/~joaquim/publicacions/Llisterri_Carbo_Machuca_Mota_Riera_Rios_03_Linguistica_Tecnologias_Habla.pdf
- LLISTERRI, Joaquim; CARBÓ, Carme; MACHUCA, María J.; DE LA MOTA, Carme; RIERA, Montserrat; RÍOS, Antonio (2004): «La conversión de texto en habla: aspectos lingüísticos». A: MARTÍ, M. A.; LLISTERRI, J. (eds.) (2004): *Tecnologías del texto y del habla*. Barcelona. Edicions de la Universitat de Barcelona – Fundación Duques de Soria. pàg. 145-186.

http://liceu.uab.es/~joaquim/publicacions/Llisterri_Carbo_Machuca_Mota_Riera_Rios_04_Conversion_Texto_Habla.pdf

LLISTERRI, Joaquim; MACHUCA, María J. (eds.) (2006): *Los sistemas de diálogo*. Bellaterra: Universitat Autònoma de Barcelona - Fundación Duques de Soria.

LLISTERRI, Joaquim; MACHUCA, María J.; DE LA MOTA, Carme; RIERA, Montserrat; RÍOS, Antonio (2003): «Entonación y tecnologías del habla». A: PRIETO, P. (ed.) (2003): *Teorías de la entonación*. Barcelona: Ariel. pàg. 209-243.

http://liceu.uab.es/~joaquim/publicacions/Llisterri_Machuca_Mota_Riera_Rios_03_Entonacion_Tecnologias_Habla.pdf

LLISTERRI, Joaquim; MACHUCA, María J.; MADRIGAL, Natalia; MANCINI, Franca; MASSIMINO, Paolo; DE LA MOTA, Carme; RIERA, Montserrat; RÍOS, Antonio (2004): «Aspectos lingüísticos en el diseño de un conversor de texto en habla en castellano y en catalán: El sistema LoquendoTTS®». A: *VI Congreso de Lingüística General*. Santiago de Compostela (Espanya), maig del 2004.

http://liceu.uab.es/~joaquim/publicacions/Llisterri_et_al_04_Conversor_Texto_Habla_Castellano_Catalan_Loquendo.pdf

LÓPEZ-CÓZAR, Ramón; ARAKI, Masahiro (2005): *Spoken, Multilingual and Multimodal Dialogue Systems: Development and Assessment*. Chichester: John Wiley & Sons.

MARIÑO, José B.; NADEU, Climent (2004): «La representación de la voz para el reconocimiento del habla». A: MARTÍ, M. A.; LLISTERRI, J. (eds.) (2004) *Tecnologías del texto y del habla*. Barcelona. Edicions de la Universitat de Barcelona – Fundación Duques de Soria. pàg. 187-224.

MCTEAR, Michael F. (2004): *Spoken Dialogue Technology: Toward the Conversational User Interface*. Londres: Springer-Verlag.

- MÖBIUS, Bernd (2000): «Corpus-based speech synthesis: methods and challenges». *AIMS, Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (Universität Stuttgart)* 6, 4. pàg. 87-116.
<http://www.ims.uni-stuttgart.de/~moebius/papers/unitsel.pdf>
- MONAGHAN, Alex (2002a): «State-of-the-art summary of European synthetic prosody R&D». A: KELLER, E.; BAILLY, G.; MONAGHAN, A.; TERKEN, J.; HUCKVALE, M. (eds.) (2002): *Improvements in Speech Synthesis. Cost 258: The Naturalness of Synthetic Speech*. Chichester: John Wiley & Sons. pàg. 93-103.
- MONAGHAN, Alex (2002b): «Prosody in synthetic speech: Problems, solutions and challenges». A: KELLER, E.; BAILLY, G.; MONAGHAN, A.; TERKEN, J.; HUCKVALE, M. (eds.) (2002): *Improvements in Speech Synthesis. Cost 258: The Naturalness of Synthetic Speech*. Chichester: John Wiley & Sons. pàg. 89-92.
- PACHÈS, Pau; DE LA MOTA, Carme; RIERA, Montserrat; PEREA, M. Pilar; FEBRER, Albert; ESTRUCH, Mònica; GARRIDO, Juan M.; MACHUCA, María J.; RÍOS, Antonio; LLISTERRI, Joaquim; ESQUERRA, Ignasi; HERNANDO, Javier; PADRELL, Jaume; NADEU, Climent (2000): «Segre: An automatic tool for grapheme-to-allophone transcription in Catalan». A: Ó CRÓINÍN, D. (ed.) (2000): *Proceedings of the Workshop on Developing Language Resources for Minority Languages: Reusability and Strategic Priorities (LREC-2000 Second International Conference on Language Resources and Evaluation)*. Atenes (Grècia), maig del 2000. pàg. 52-61.
http://liceu.uab.es/~joaquim/publicacions/Paches_et_al_00_SEGRE_Phonetic_Transcription_Catalan.pdf
- QUAZZA, Silvia; DONETTI, Laura; MOISA, Loreta; SALZA, Pier-Luigi (2001): «ACTOR: A multilingual unit-selection speech synthesis system». A: *Proceedings of the 4th ISCA Tutorial and Research Workshop on Speech Synthesis*. Pertshire (Regne Unit), setembre del 2001.

- http://www.loquendo.com/en/brochure/art_TTS_2001.pdf
- SCHROETER, Juergen (2006): «Text-to-Speech (TTS) Synthesis». A: DORF, R. C. (ed.) (2006): *The Electrical Engineering Handbook*. Boca Baton: CRC Press. pàg. 16-1 - 16-13. [3a edició].
http://www.research.att.com/~ttsweb/tts/papers/2005_EEHandbook/tts.pdf
- TAPIAS, Daniel (2002): «Interfaces de voz con lenguaje natural». A: MARTÍ, M. A.; LLISTERRI, J. (eds.) (2002): *Tratamiento del lenguaje natural. Tecnología de la lengua oral y escrita*. Barcelona: Edicions Universitat de Barcelona - Fundación Duques de Soria. pàg. 189-207.
- TORRES, M. Inés (2006): «El reconocimiento del habla». A: LLISTERRI, J.; MACHUCA, M. J. (eds.) (2006): *Los sistemas de diálogo*. Bellaterra: Universitat Autònoma de Barcelona - Fundación Duques de Soria. pàg. 81-98.
- VALLVERDÚ, Francesc (1986): *Elocució i ortologia catalanes*. Barcelona: Junc.
- VILLARRUBIA, Luis; RODRÍGUEZ Miguel Á.; RELAÑO, José; GARIJO, Francisco J.; BERNAT, Jesús; HERNÁNDEZ, Luis A.; SAN SEGUNDO, Rubén; TAPIAS, Daniel; MARÍA, Luis A. (2003): «Tecnología del habla para aplicaciones multilingües, multiservicio y multiplataforma». *Comunicaciones de Telefónica I+D* 30. pàg. 47-78.
http://www.tid.es/documentos/revista_comunicaciones_i%2Bd/numero30.pdf