



**Universitat Autònoma  
de Barcelona**

# **TECNOLOGIES DE LA PARLA: CREACIÓ D'UN PUNT D'INFORMACIÓ**

Memòria del projecte  
d'Enginyeria Tècnica en  
Informàtica de Sistemes

realitzat per

*Xavier Closa Serra*

i dirigit per

*Montserrat Meneses Benítez*

**Escola d'Enginyeria**

*Sabadell, setembre de 2012*

La sotasignat, **Montserrat Meneses Benítez**,  
professora de l'Escola d'Enginyeria de la UAB,

**CERTIFICA:**

Que el treball al que correspon la present memòria  
ha estat realitzat sota la seva direcció per

**Xavier Closa Serra**

I per a que consti firma la present.

Sabadell, **setembre** de **2012**

-----  
Signat: **Montserrat Meneses Benítez**

## **Resum**

En aquesta memòria hi ha reflectit tot el procés de creació d'un sistema de diàleg persona-màquina basat en la idea de muntar un punt d'informació turístic que transmeti dades parlades al turista quan aquest les hi demani. S'hi trobarà una breu introducció a aquests sistemes, s'explicaran les parts que s'utilitzen a l'hora de construir-ne un, se'n veurà el funcionament, les eines fetes servir durant la creació del nostre projecte, les proves realitzades per comprovar-ne el correcte funcionament i per acabar també es comentarà el resultat de l'estudi de viabilitat dut a terme.

## **Resumen**

En esta memoria está reflejado todo el proceso de creación de un sistema de diálogo persona-máquina basado en la idea de montar un punto de información turístico que transmita datos hablados al turista cuando éste se los pida. Se encontrará una breve introducción a estos sistemas, se explicarán las partes que se utilizan a la hora de construir uno, se verá el funcionamiento, las herramientas utilizadas durante la creación de nuestro proyecto, las pruebas realizadas para comprobar su correcto funcionamiento y para terminar también se comentará el resultado del estudio de viabilidad llevado a cabo.

## **Abstract**

In this report there is reflected throughout the process of creating a human-machine dialogue system based on the idea of setting up a tourist information point to convey spoken information to the tourist when he asks for them. Furthermore, there is a brief introduction to these systems, it explains the parts that are used to construct one of this and his operation, the tools used during the creation of our project, the tests done to verify the correct operation and finally also discuss the results of the feasibility study done.

# Índex

|                                                                                             |           |
|---------------------------------------------------------------------------------------------|-----------|
| <b>Índex</b>                                                                                | <b>4</b>  |
| <b>1 Introducció</b>                                                                        | <b>7</b>  |
| 1.1 Objectius . . . . .                                                                     | 8         |
| 1.2 Motivació . . . . .                                                                     | 9         |
| 1.3 Estructura de la memòria . . . . .                                                      | 10        |
| <b>2 Reconeixement automàtic de la parla</b>                                                | <b>12</b> |
| 2.1 Introducció . . . . .                                                                   | 12        |
| 2.2 Models acústics . . . . .                                                               | 16        |
| 2.2.1 Parametrització . . . . .                                                             | 17        |
| 2.2.2 Inventari d'unitats . . . . .                                                         | 17        |
| 2.2.3 Selecció de la tecnologia . . . . .                                                   | 18        |
| 2.3 Model de llenguatge . . . . .                                                           | 18        |
| 2.4 Eines corresponents a la part de ASR que seran utilitzades en aquest projecte . . . . . | 20        |
| <b>3 Conversió de text a veu</b>                                                            | <b>22</b> |
| 3.1 Introducció . . . . .                                                                   | 22        |
| 3.2 Anàlisi del text i transcripció fonètica . . . . .                                      | 23        |
| 3.2.1 Estructura del text . . . . .                                                         | 23        |
| 3.2.2 Normalització del text . . . . .                                                      | 25        |
| 3.2.3 Transcripció fonètica . . . . .                                                       | 27        |
| 3.3 Modelat prosòdic . . . . .                                                              | 28        |
| 3.3.1 Corbes melòdiques . . . . .                                                           | 29        |
| 3.3.2 Duració . . . . .                                                                     | 31        |
| 3.3.3 Intensitat . . . . .                                                                  | 33        |
| 3.3.4 Pausas . . . . .                                                                      | 33        |
| 3.4 Síntesi de veu . . . . .                                                                | 35        |
| 3.4.1 Síntesi per formants . . . . .                                                        | 36        |
| 3.4.2 Síntesi articulatòria . . . . .                                                       | 37        |

|          |                                                                                         |           |
|----------|-----------------------------------------------------------------------------------------|-----------|
| 3.4.3    | Síntesi per concatenació . . . . .                                                      | 37        |
| 3.5      | La conversió de text a veu dins els sistemes de diàleg . . . . .                        | 40        |
| 3.6      | Eines corresponents a la part de TTS que seran utilitzades en aquest projecte . . . . . | 46        |
| <b>4</b> | <b>Estudi de viabilitat</b>                                                             | <b>48</b> |
| 4.1      | Introducció . . . . .                                                                   | 48        |
| 4.1.1    | Tipologia i paraules clau . . . . .                                                     | 48        |
| 4.1.2    | Descripció . . . . .                                                                    | 48        |
| 4.1.3    | Objectius del projecte . . . . .                                                        | 49        |
| 4.1.4    | Definicions, acrònims i abreviacions . . . . .                                          | 50        |
| 4.1.5    | Parts interessades . . . . .                                                            | 51        |
| 4.2      | Estudi de la situació actual . . . . .                                                  | 53        |
| 4.2.1    | Context . . . . .                                                                       | 53        |
| 4.2.2    | Lògica del sistema . . . . .                                                            | 53        |
| 4.2.3    | Descripció física . . . . .                                                             | 54        |
| 4.2.4    | Usuaris i/o personal del sistema . . . . .                                              | 54        |
| 4.2.5    | Diagnòstic del sistema . . . . .                                                        | 54        |
| 4.2.6    | Normatives i legislació . . . . .                                                       | 55        |
| 4.3      | Requisits del sistema . . . . .                                                         | 55        |
| 4.3.1    | Requisits funcionals . . . . .                                                          | 55        |
| 4.3.2    | Requisits no funcionals . . . . .                                                       | 56        |
| 4.3.3    | Restriccions del sistema . . . . .                                                      | 56        |
| 4.3.4    | Catalogació i prioritització dels requisits . . . . .                                   | 57        |
| 4.4      | Alternatives i selecció de la solució . . . . .                                         | 57        |
| 4.4.1    | Alternativa 1 . . . . .                                                                 | 57        |
| 4.4.2    | Alternativa 2 . . . . .                                                                 | 59        |
| 4.4.3    | Solució proposada . . . . .                                                             | 59        |
| 4.5      | Work Breakdown Structure (WBS) . . . . .                                                | 60        |
| 4.5.1    | Fases i activitats del projecte . . . . .                                               | 60        |
| 4.5.2    | Diagrama WBS . . . . .                                                                  | 60        |
| 4.5.3    | Milestones . . . . .                                                                    | 61        |
| 4.6      | Recursos del projecte . . . . .                                                         | 62        |
| 4.6.1    | Recursos del projecte . . . . .                                                         | 62        |
| 4.6.2    | Calendari dels recursos . . . . .                                                       | 62        |
| 4.7      | Calendari del projecte . . . . .                                                        | 63        |
| 4.7.1    | Dependències . . . . .                                                                  | 63        |
| 4.7.2    | Quadre de tasques del projecte . . . . .                                                | 64        |
| 4.7.3    | Calendari temporal . . . . .                                                            | 65        |
| 4.8      | Avaluació de riscos . . . . .                                                           | 65        |
| 4.8.1    | Llista de riscos . . . . .                                                              | 65        |

|          |                                                       |            |
|----------|-------------------------------------------------------|------------|
| 4.8.2    | Catalogació de riscos . . . . .                       | 66         |
| 4.8.3    | Pla de contingència . . . . .                         | 66         |
| 4.9      | Pressupost . . . . .                                  | 67         |
| 4.9.1    | Estimació cost de personal . . . . .                  | 67         |
| 4.9.2    | Estimació cost de recursos . . . . .                  | 67         |
| 4.9.3    | Estimació costos indirectes . . . . .                 | 67         |
| 4.9.4    | Resum i anàlisi cost-benefici . . . . .               | 67         |
| 4.10     | Conclusions finals del estudi de viabilitat . . . . . | 69         |
| <b>5</b> | <b>Disseny de l'aplicació</b>                         | <b>71</b>  |
| 5.1      | Diàleg . . . . .                                      | 71         |
| 5.2      | Reconeixement automàtic de la parla . . . . .         | 81         |
| 5.3      | Síntesi de veu . . . . .                              | 88         |
| <b>6</b> | <b>Proves</b>                                         | <b>92</b>  |
| <b>7</b> | <b>Conclusions</b>                                    | <b>98</b>  |
| 7.1      | Conclusions del projecte . . . . .                    | 98         |
| 7.1.1    | Millores futures . . . . .                            | 99         |
| 7.2      | Conclusions personals . . . . .                       | 100        |
|          | <b>Bibliografia</b>                                   | <b>101</b> |

# Capítol 1

## Introducció

Des de sempre, el llenguatge oral ha sigut, i continua sent, el principal vehicle de comunicació entre éssers humans gràcies al fet que es tracta d'un mètode natural, intuïtiu i eficient. Tenint en compte això i veient com avança de ràpid la tecnologia i com n'és de present a les nostres vides, fa temps que enginyers i lingüistes van desenvolupant estudis i projectes que tenen com a finalitat aconseguir que els usuaris puguin interaccionar amb les màquines tot utilitzant la veu, com si d'un ésser humà es tractés, per així fer més fàcil la utilització d'aquestes ja que, per tothom és sabut, és més senzill i ràpid xarrar que no pas escriure.

Per dur a terme aquesta comunicació bidireccional persona-màquina necessitem, bàsicament, dos tipus de tecnologia:

- Reconeixement automàtic de la parla o ASR (*Automatic Speech Recognition*).
- Conversió de text a veu o TTS (*Text-To-Speech*).

Com a breu introducció a aquests dos tipus de tecnologia, podem afirmar que el reconeixement de la parla és el què fa possible que les màquines escoltin i executin les nostres ordres sense necessitat d'haver-les-hi de passar per teclat o per qualsevol altre dispositiu d'entrada. Per citar un conegut exemple d'aquesta tecnologia, seria la què permetria a Alí Babà entrar a la seva cova tot pronunciant el famós "Obre't Sèsam".

Per altra banda, la conversió de text a veu, també anomenada síntesi de veu, permet la comunicació en sentit contrari gràcies al fet que dóna a les màquines la possibilitat de llegir-nos, de la manera més similar a com ho faria una persona, qualsevol text i fer-nos arribar així informació de forma oral amb total flexibilitat. Aquí, com a exemple, podem citar la tecnologia que porten els autobusos urbans de la ciutat de Barcelona i que ens informa, a través de l'àudio que surt per megafonia, de quina és la propera parada.

És important destacar els avantatges que aporta la tecnologia TTS vers

el fet de gravar els missatges de veu prèviament. Si ens fixem en l'exemple anterior podem veure que, emprant un locutor, es poden gravar totes les parades de la línia i llavors utilitzar aquestes gravacions quan toqui. Però, què passarà si es modifica el recorregut de la línia? I si s'afegeixen noves parades? I si es canvia el nom d'alguna d'aquestes? Bé, doncs la solució passaria per tornar a contractar el locutor i que gravi de nou tot el què s'ha modificat. Això, lògicament, suposa un cost addicional, tant econòmic com de temps, bastant important, ja que hi ha moltes línies d'autobús a la ciutat comtal que poden sofrir modificacions i, a més, l'acció de gravar pot comportar haver-se d'esperar un cert temps (disponibilitat locutor, disponibilitat estudi de gravació...).

La solució, doncs, passa per utilitzar la síntesi de veu. Només cal tenir una llista escrita amb el nom de les parades i, utilitzant aquesta tecnologia, automàticament s'obtidran les gravacions desitjades per utilitzar en qualsevol moment a l'autobús. I si hi ha alguna modificació com les comentades abans? Molt fàcil, només cal modificar la llista de parades i tornar-la a sintetitzar. Assumpte resolt.

Aquí hem vist un petit exemple de com les tecnologies de la parla poden ajudar a millorar, en diferents aspectes, situacions quotidianes de la vida moderna i, encara que no ens ho sembli, com en són de presents en el dia a dia de moltes persones.

## 1.1 Objectius

Aquest projecte té com a principal objectiu crear una aplicació que utilitzi conjuntament les tecnologies de la parla, tant el ASR com el TTS. La idea sobre la qual està basada aquesta aplicació és la d'aconseguir un punt d'informació turístic on els usuaris que arribin de visita al municipi vagin demanant tot el què desitgin conèixer sobre aquest fent servir la seva veu i que la màquina els respongui, mitjançant veu i text mostrat per pantalla, tot el què van preguntant, de manera que entre usuari i màquina s'estableixi un diàleg. L'aplicació estarà dividida en diferents categories com ara informació d'interès general, serveis i equipaments, llocs d'interès... per facilitar així la comprensió al turista.

Bàsicament el paper que tindrà el ASR dins l'aplicació serà el de reconèixer cada paraula o frase que pronunciï l'usuari i processar-la per, segons sigui el què s'hagi dit, fer una acció o una altra. Aquestes accions van des d'activar l'aplicació fins a moure's per diferents subcategories d'informació. L'ús del TTS és per convertir a veu el text que conté la informació desitjada i que s'ha de fer arribar a l'usuari.



El fet de realitzar aquest projecte permetrà que ens endinsem, no gaire profundament, en el funcionament de les tecnologies de la veu i que, en conseqüència, en surti una memòria on hi quedi plasmada tant la informació sobre ASR i TTS que es vagi recopilant a mesura que avanci el projecte, com la planificació i disseny de l'aplicació.

Cal tenir en compte que hi ha una sèrie de factors entorn a aquest projecte que faran que tingui un resultat moderat si el comparem amb el que podria tenir, per exemple, el creat per una empresa especialitzada. Alguns d'aquests factors poden ser: recursos tècnics i humans limitats (projecte creat per un únic estudiant que utilitza aplicacions lliures instal·lades a l'ordinador domèstic), poc temps per dur-lo a terme, manca d'experiència amb aquest tipus de tecnologies... Més endavant es veuran més en profunditat aquestes limitacions i les conseqüències directes que tenen sobre el projecte.

## 1.2 Motivació

Principalment, la raó per la qual vaig decidir escollir aquest projecte d'entre una llarga llista de propostes que tenia a davant, va ser la curiositat que em va picar per descobrir aquestes noves, per a mi, tecnologies de la parla. La veritat és que gairebé mai les hi havia prestat atenció: mai m'havia parat a pensar en com funciona el reconeixement de veu que molts cotxes moderns porten incorporat, o com ho fan per tenir gravades totes les parades de les moltes línies de tren del país. Simplement ho veia o ho sentia.

Com que m'anava a endinsar en un món en el qual la majoria de coses eren noves per a mi, van sorgir certs dubtes dins del meu cap, fet que va propiciar que consultés el tutor del projecte i, quan aquest em va respondre que no calia que em preocupés, que es podia començar el projecte amb coneixements nuls sobre ASR i TTS i que ell em posaria a l'abast la informació més bàsica per a poder tirar-lo endavant, tots aquests dubtes es van anar esvaint i la motivació d'adquirir nous coneixements va fer que m'hi apuntés de seguida.

Des del meu punt de vista, aprendre el funcionament, encara que sigui a un nivell molt bàsic, d'aquestes tecnologies sempre aporta més del que resta. Per ficar un parell d'exemples, comentar que et pot ajudar a trobar feina o, si més no, et permet crear una aplicació pròpia per ordinador o mòbil que faci quelcom interessant i funcional, o bé, que simplement deixi bocabadades a les persones a qui els hi ensenyis.

Ara que estan tant de moda els reconeixadors de la parla, només cal observar la quantitat de cotxes que hi ha al mercat que en porten de sèrie, tots els telèfons d'última generació que ho utilitzen com a entrada de comandes..., és fàcil adonar-se'n que la gran majoria són en anglès i només

una petita minoria funcionen en castellà. Dels que funcionen en la meua llengua materna, el català, ja ni en parlem. Per això, un cop vaig tenir clar que faria una aplicació que utilitzés ASR i TTS, no em va fer falta pensar gens la llengua sobre la qual funcionaria: la llengua que més utilitzo i que sovint queda al marge de les noves tecnologies. A més, gràcies al fet d'utilitzar el català i tenint en compte que no hi ha gaire oferta d'aplicacions creades en aquesta llengua, pot sorgir alguna oportunitat de fer negoci. De moment però, cal crear l'aplicació del projecte i més endavant ja es veurà si cal prosseguir i millorar-la o bé deixar-la tal i com acabi.

## 1.3 Estructura de la memòria

Aquesta memòria està estructurada en 7 capítols (sense contar la bibliografia, ja que aquesta al ser obligatòria en totes les memòries no s'ha considerat com a tal), el primer dels quals és la Introducció. Aquí dins hi trobarem una breu presentació sobre el tema del qual tractarà el projecte que es crearà, els objectius que s'han fixat abans de començar-lo i les motivacions que ens van dur a escollir aquest i no un altre projecte.

Seguidament veurem, en dos capítols separats, les tecnologies de la parla. En el primer d'aquests dos capítols (el número 2 en el global de la memòria) es tractarà el reconeixement automàtic de la parla: què és, la base del seu funcionament, una explicació sobre els seus components (models acústics i models de llenguatge), quines eines relacionades amb el ATK seran utilitzades durant la creació del projecte...

La síntesi de veu serà la protagonista del capítol 3. Es veuran, bastant en profunditat, les tres etapes per les quals s'ha de passar per convertir text a veu: anàlisi del text, generació prosòdica i generació de la veu. Cada una d'aquestes etapes, com és lògic, té diferents parts, les quals també seran explicades en aquest capítol, ficant algun exemple per tal de facilitar-ne la comprensió. També en aquest tercer capítol es mostrarà com s'utilitza la conversió de text a veu dins els sistemes de diàleg. Ja per acabar s'explicaran quines eines relacionades amb el TTS seran utilitzades durant la creació del projecte.

El quart capítol és l'estudi de viabilitat que es va realitzar abans de començar el projecte en sí per tal d'establir si era viable crear-lo o no. Aquí hi trobarem tots els apartats típics d'un estudi de viabilitat, des dels objectius establerts al calendari, passant per les parts interessades, l'estudi de la situació actual, els requisits que ha de tenir l'aplicació i els recursos que s'hi destinaran.

El capítol número 5 segurament és el més important de tots els que hi

ha en aquesta memòria, ja que explica com s'ha pensat i gestionat el diàleg que l'usuari tindrà amb la màquina: s'hi trobarà quins temes d'informació poden ser tractats dins l'aplicació, com s'ha de demanar aquesta informació, com s'activa la màquina... També tracta sobre com s'han implementat i combinat les dues tecnologies de la parla explicades als capítols 2 i 3 a l'hora de crear aquest sistema de diàleg: com s'han configurat i/o modificat les eines utilitzades, els passos seguits, algun exemple de com han quedat els fitxers...

En el penúltim capítol hi han explicades les proves realitzades pels creadors del projecte per tal de garantir el correcte funcionament de l'aplicació i evitar així que quan la màquina estigui en funcionament surtin a la llum.

Finalment, l'últim capítol extrau les conclusions i sensacions, tant a nivell de projecte com a nivell personal, a les quals s'ha arribat un cop acabat el projecte. També hi ha un apartat on s'hi trobaran les possibles millores a aplicar en versions futures d'aquest projecte i que ara no s'han pogut dur a terme per diferents raons (temps, falta de gent...).

## Capítol 2

# Reconeixement automàtic de la parla

### 2.1 Introducció

Tal i com s'ha vist en el primer capítol, perquè l'ordinador pugui entendre el què l'usuari diu és necessari un reconeixedor de la parla, una eina computacional capaç de processar un senyal de veu, ja sigui una gravació o un àudio que entra des d'un perifèric, i reconèixer la informació que conté. Per poder dur a terme aquest procés cal que hi intervinguin disciplines com ara la fisiologia, l'acústica, el processament de senyal, la intel·ligència artificial, la ciència computacional i l'estadística.

Si s'observa la figura 2.1 es pot veure com el sistema transcriu el senyal de veu que li arriba, s'assumeix que aquest senyal acústic està format per un o més símbols, en seqüències d'unitats lingüístiques,  $W = \omega_1, \omega_2, \dots, \omega_{|\Omega|}$ , generalment paraules, que pertanyen a un vocabulari finit  $\Sigma = \{\omega_j\}$ ,  $j = 1, \dots, |\Sigma|$ .

Cada frase pronunciada pel locutor passa un procés de parametrització i arriba al reconeixedor com una seqüència de mesures (observacions) acústiques  $O = o_1, o_2, \dots, o_{|O|}$ , on  $o_t$  és un vector observat al temps  $t$ , i aquest té la feina de decodificar i proporcionar la seqüència d'unitats lingüístiques  $W$  amb major probabilitat de ser la que ha estat pronunciada per l'usuari quan s'ha observat la seqüència acústica  $O$ . És a dir, es tracta de resoldre el següent problema:

$$\hat{W} = \operatorname{argmax}_W P(W|O) \quad (2.1)$$

L'equació 2.1 representa la cerca de la seqüència de paraules  $\hat{W}$  que maximitza  $P(W|O)$  entre totes les seqüències  $W$  possibles. La probabilitat

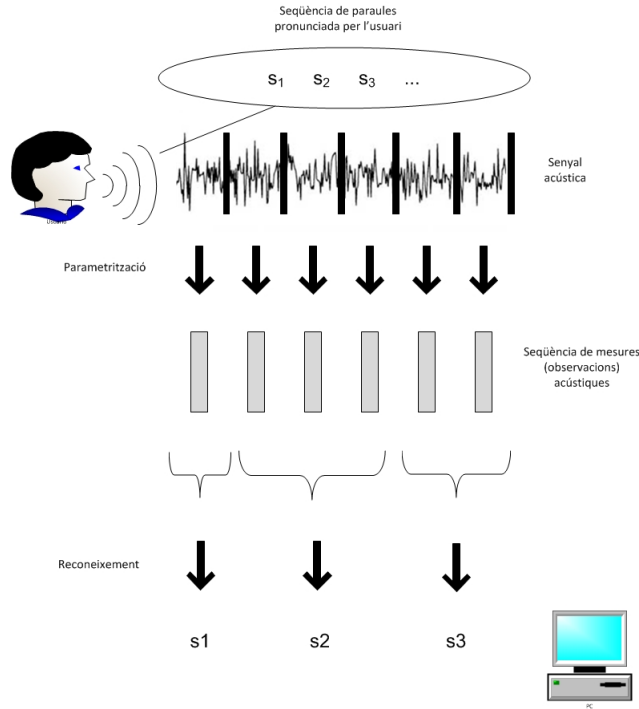


Figura 2.1: Funcionament d'un reconeixedor de la parla

que es busca, però, no és directament calculable, així que és necessari aplicar el Teorema de Bayes sobre aquesta equació:

$$\operatorname{argmax}_W P(W|O) = \operatorname{argmax}_W \left( \frac{P(O|W)P(W)}{P(O)} \right) \quad (2.2)$$

$P(O)$  és la probabilitat de la seqüència de mesures acústiques, probabilitat difícil de conèixer. Ara bé, sabem que s'examina la mateixa seqüència d'observació  $O$  per a cada ellocució en potència, així que  $P(O)$  sempre tindrà el mateix valor. Aquest fet permet simplificar, sense que això alteri el resultat, l'equació 2.2 de la següent manera:

$$\hat{W} = \operatorname{argmax}_W P(O|W)P(W) \quad (2.3)$$

L'equació 2.3 és la base dels sistemes de reconeixement de la parla actuals i té conseqüències importants. Primerament, estableix la separació dels dos principals components del sistema: el model de llenguatge, per calcular la probabilitat  $P(W)$  i el model acústic, per construir el lèxic de l'aplicació. Tots dos models representen fonts de coneixement diferents i, en principi, independents, que intervenen en el procés del reconeixement de la parla: per una banda tenim el coneixement de totes les possibles paraules (sintaxi) que

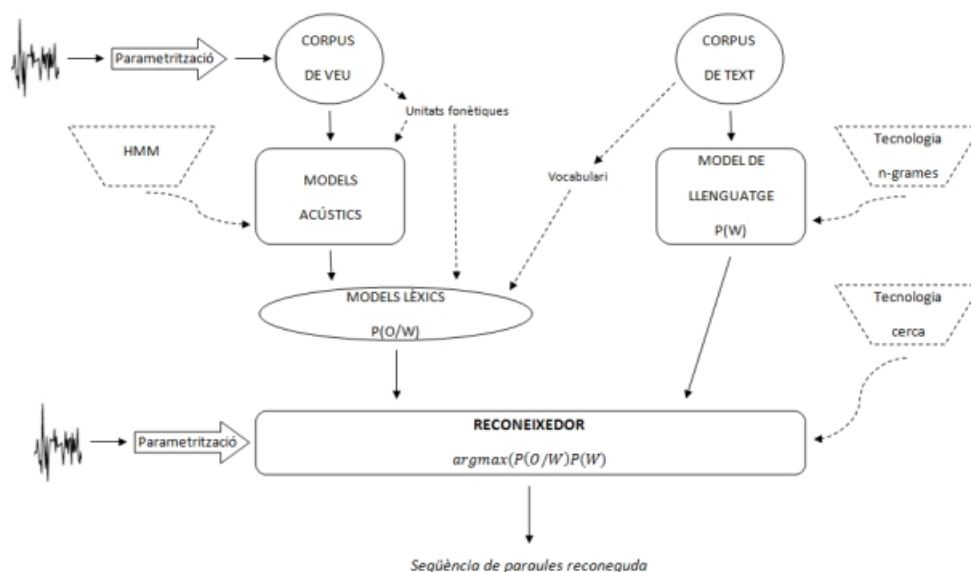


Figura 2.2: Esquema d'un sistema de reconeixement automàtic de la parla

s'han reconèixer i, per l'altra, el coneixement dels sons (acústic i fonètic) de la llengua. La segona conseqüència que deriva de l'equació 2.3 és que el procés de reconeixement, cerca i obtenció de la millor seqüència de paraules  $\hat{W}$  es realitza sobre el producte entre  $P(O|W)$  i  $P(W)$ , on el primer paràmetre és la probabilitat de la seqüència d'observacions per a una seqüència de paraules  $W$  donada i el segon és la probabilitat de que la seqüència  $W$  hagi estat pronunciada.

La figura 2.2 mostra l'esquema d'un sistema de reconeixement automàtic de la parla. En ell es representen dues fases: la d'entrenament, tant per a l'obtenció del model de llenguatge com per a l'obtenció del model acústic i la pròpia fase de reconeixement.

Els models acústics, que normalment consisteixen en models ocults de Markov (HMM, de l'anglès *Hidden Markov Models*), s'entrenen a partir de corpus orals que representen la variació fonètica de la llengua. Si concatenem cada model acústic amb la cadena de fonemes que forma la seva corresponent paraula, obtenim els models de les paraules o models lèxics, que conformen el vocabulari de l'aplicació. El model de llenguatge representa les combinacions de paraules possibles i més freqüents d'aquesta aplicació.

Un model ocult de Markov (HMM) és una màquina d'estats finita que es pot entendre com un generador de seqüències d'observacions acústiques (els vectors de paràmetres). En cada instant de temps canvien d'estat, i

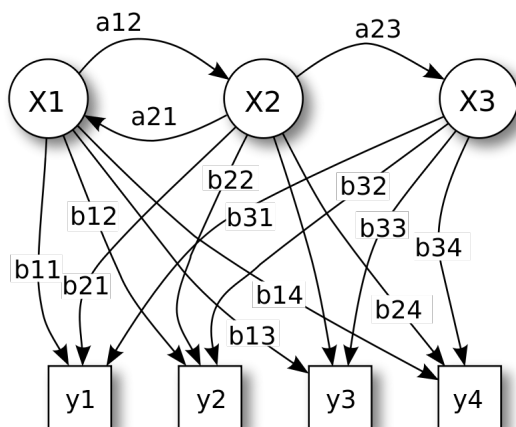


Figura 2.3: *Exemple de transició d'estats en un model ocult de Markov*

en la transició (que pot ser un bucle al mateix estat) generen un vector de paràmetres amb una certa probabilitat. Aquests vectors seran les observacions acústiques que serviran per obtenir els models acústics (HMM) de cada fonema en la fase d'entrenament.

En la imatge 2.3 es pot veure un exemple de HMM, a on les  $x$  representen els estats ocults, les  $y$  les observacions acústiques, les  $a$  les probabilitats de transició i les  $b$  les probabilitats de sortida.

Per acabar, comentar que quan el reconeixedor de la parla està inclòs dins d'un sistema de diàleg i aquest últim està dirigit al públic general, sovint inexpert en aquestes tecnologies, cal tenir present que aquests parlaran amb total llibertat davant la màquina, com si parlessin amb un altre ésser humà. Això vol dir que existeixen certs factors que influeixen directament en la manera de funcionar del reconeixedor, com ara la parla contínua -l'usuari enllaça una paraula rere una altra i així fins a completar tot el què vol dir, sense interrupció alguna-, l'origen del locutor -encara que el sistema estigui pensat per funcionar en una sola llengua per fer-lo senzill, hi pot haver moltes variants d'aquesta llengua, deixant de banda que dues persones, encara que siguin de la mateixa regió, poden pronunciar una paraula de maneres completament diferents-, la parla espontània -l'usuari, ja sigui perquè està nerviós, per ser la primera vegada que utilitza una aplicació d'aquest tipus o per qualsevol altre motiu, pot ser que s'equivoqui i torni a començar, que es quedi en blanc, que repeteixi paraules...- o un entorn advers -el soroll ambiental és un element a tenir en compte, ja que es cola pel micròfon com

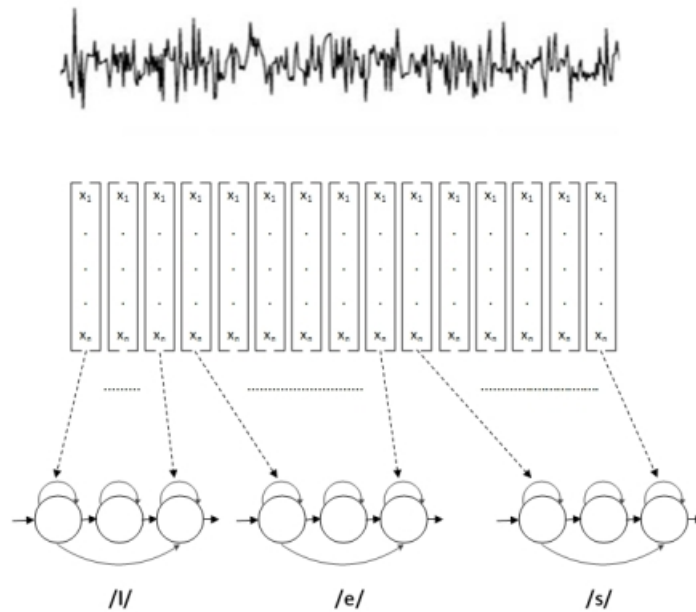


Figura 2.4: Obtenció dels vectors de paràmetres que representen

si es tractés d'una paraula pronunciada pel locutor-.

## 2.2 Models acústics

Els models acústics intenten representar els sons de la llengua. Són independents de l'aplicació, encara que de vegades s'adapten a la mateixa, i s'entrenen a partir de corpus orals *fonèticament equilibrats*, com bé s'ha comentat fa un moment. Aquests corpus es dissenyen de manera que els fonemes i els al·lòfons apareixen amb la mateixa distribució de freqüències amb la qual ho fan a la llengua (equilibri fonètic) i en número suficient per assegurar que cada fonema estigui representat en tots els seus contextos possibles (amb tots els seus al·lòfons) un número mínim de vegades.

Per construir models acústics és necessari definir el conjunt de sons que utilitzarà el reconeixedor (selecció d'unitats), per a cada un dels quals s'ha d'obtenir un model (selecció del tipus de model) a partir d'una representació adequada del senyal (parametrització). La figura 2.4 representa aquest procés.



### 2.2.1 Parametrització

Aquesta fase es coneix com el pre-procés del senyal. La forma d'ona és una representació de l'evolució de l'amplitud del senyal amb el temps. Amb les freqüències de mostreig habituals (8 Khz per a senyals telefòniques i 16 Khz per a la resta) el número de punts necessaris per emmagatzemar el senyal és molt elevat. Per altra part, les característiques més importants per la percepció i diferenciació dels sons es troben en la distribució de freqüències del senyal. No obstant, és difícil apreciar i obtenir aquestes característiques en la forma d'ona a pesar de la gran quantitat d'informació (número de punts) necessària per a la seva representació. Per tant, l'objectiu de la parametrització és buscar un domini de representació en el qual no es necessiti emmagatzemar molta informació i en el qual es ressaltin les diferències entre sons. El espectrograma és una opció de parametrització que consisteix en el càlcul d'un vector que representa la distribució de freqüències (transformada de Fourier) d'una petita trama de senyal, generalment de 25,6 ms. Aquest càlcul es va repetint, habitualment cada 10 ms, per garantir que les trames de senyal s'encavalquin (per exemple, la paraula diccionari podria quedar partida en les següents trames: dic - icci - cio - ona - nari); aquest procediment permet estudiar l'evolució temporal del l'anàlisi espectral. Amb aquest tipus de parametrització es redueixen 256 punts a un vector d'entre 12 i 20 components corresponents a un banc de 12 a 20 filtres. Per tant es tracta d'una representació més compacta, que necessita menys dades que la representació directa de la forma d'ona. Aquests filtres es distribueixen seguint una escala logarítmica -escala Mel-, que simula l'escala de percepció de freqüències de l'oïda humana.

En els sistemes de reconeixement de veu actuals s'utilitzen de manera generalitzada els *coeficients cepstrals (MFCC)* com a vectors de paràmetres. Es calculen obtenint la antitransformada del logaritme dels paràmetres del banc de filtres (transformada de Fourier). S'utilitzen, per cada trama, 10 ó 12 coeficients, és a dir, un vector de 10 ó 12 components. S'ha demostrat que constitueixen la millor representació per a sistemes de reconeixement de veu que operen en ambients sense soroll. Per completar la representació s'afegeixen l'energia i els paràmetres dinàmics representats per les primeres i segones derivades dels *coeficients cepstrals*.

### 2.2.2 Inventari d'unitats

La selecció de l'inventari d'unitats sublèxiques defineix l'inventari de models acústics que s'han d'obtenir. Entenem per unitat sublèxica la unitat més petita que es modela en un sistema de reconeixement automàtic de veu.

Pot tractar-se d'unitats pròximes a la síl·laba, al fonema, al al·lòfon o inclús a porcions de fonemes. Els models de les paraules es construiran a partir de concatenacions dels models d'aquestes unitats. Encara que existeixen altres alternatives, generalment es defineix un inventari amb un número d'unitats similar al dels fonemes de la llengua com a conjunt bàsic d'unitats independents del context. A continuació es consideren, per a cada una d'aquestes unitats fonètiques, els contextos precedents i següents més freqüents, considerant aquells que produeixen majors variacions en la seva pronunciació. D'aquesta manera es defineix un conjunt de trifenemes que incrementa el número d'unitats amb les quals es representa cada fonema bàsic. L'inventari i número de trifenemes dependrà del corpus i de la tecnologia utilitzada per a la seva obtenció.

### 2.2.3 Selecció de la tecnologia

Els sistemes basats en regles (*Knowledge Based Systems*) es van utilitzar ja fa algunes dècades per desenvolupar sistemes de descodificació acústic - fonètica. No obstant, actualment s'utilitzen predominantment els sistemes d'interferència inductiva, que obtenen models probabilístics a partir de grans corpus d'entrenament. D'entre les diferents possibilitats d'aquest tipus de models, inferència gramatical, xarxes neuronals... s'ha demostrat que els models ocults de Markov són els més potents per modelar les unitats sublèxiques amb un cost computacional controlat. Per representar el senyal de veu s'utilitzen models d'esquerra a dreta, és a dir, que només contenen transicions cap a davant, ja que el model també representa l'evolució temporal dels vectors de paràmetres. Els models acústics solen ser de tres estats: un estat intermedi representa la part estacionària del fonema i els altres dos estats representen la unió amb el fonema anterior i posterior.

## 2.3 Model de llenguatge

El model de llenguatge és l'encarregat de definir les restriccions a les possibles combinacions de paraules; intenta crear un model de les relacions que s'estableixen entre les paraules per formar estructures sintàctiques i/o semàntiques. És independent dels fonemes i dels al·lòfons de la llengua. En la majoria de casos es dissenyen models de llenguatge de manera que recullen les particularitats de l'aplicació per a la qual es desenvolupa el reconeixedor, cosa que vol dir que se centra en les possibles combinacions de paraules rellevants per a l'aplicació en concret, no té en compte les restriccions sintàctiques i les combinacions de la llengua en general.

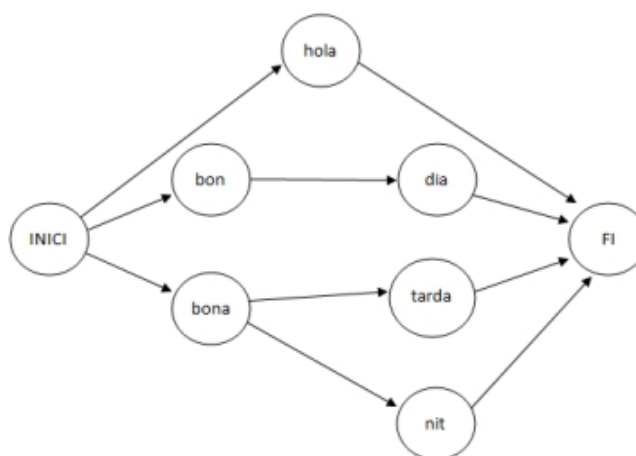


Figura 2.5: Relacions/combinacions entre paraules per construir frases

En la figura 2.5 podem observar un graf que representa, a un nivell molt bàsic, algunes de les possibles expressions que pot utilitzar l'usuari per aconseguir comunicar-se amb l'aplicació desenvolupada en aquest projecte.

En l'aproximació inductiva, l'obtenció dels models de llenguatge es realitza a partir de la informació continguda en mostres d'entrenament (corpus de text en la figura 2.2). La idea bàsica rau en el fet que quan el llenguatge presenta regularitats, és possible obtenir un model a partir de l'estudi d'un subconjunt finit d'enunciats en els quals es manifestin aquestes regularitats. Perquè la distribució de probabilitat sigui adequada, és fonamental contar amb una gran quantitat de dades. En aquesta aproximació cal distingir fonamentalment entre dos tipus de models: estadístics o gramaticals.

Els primers assignen *a priori* les probabilitats a les seqüències de paraules a partir del número de vegades que apareix una determinada paraula darrera la o les paraules que la precedeixen. Dins l'àrea del processament del llenguatge natural, les primeres tècniques utilitzades van ser les estimacions de les probabilitats basades en *n-grames*. Els models estadístics de *n-grames* calculen la probabilitat d'aparició de cada paraula condicionada a les *n-1* paraules que la precedeixen. Tenint en compte que les restriccions del llenguatge es poden modelar millor mitjançant una aproximació sintàctica, també és possible utilitzar formalismes gramaticals per desenvolupar els models de llenguatge, com per exemple gramàtiques de context lliure o gramàtiques regulars, cas en el qual es treballa amb autòmats d'estats finits estocàstics.

Un problema important associat a la generació dels models de llenguatge

és la gran quantitat de mostres necessàries per a l'aprenentatge. Aquest problema és especialment crític i molt habitual en el desenvolupament de sistemes de diàleg en els que l'usuari ha de poder utilitzar la parla espontània. Per intentar limitar el problema existeixen varies alternatives. Una d'elles consisteix en utilitzar models de llenguatge basats en categories de paraules. En els primers algorismes es treballava amb bigrames i s'establia la limitació de que cada paraula només podia estar inclosa en una classe. Actualment es tendeix a escollir correctament les classes de paraules de manera automàtica. Podem trobar un mètode automàtic de divisió del conjunt de paraules basat en el comportament sintàctic de la paraula dins la frase. Altres mètodes basen aquesta divisió automàtica en optimitzar alguna mesura de la qualitat del model de llenguatge sobre el conjunt d'entrenament. En aquest cas, l'obtenció automàtica de categories està guiada per les dades. Com a mesura d'avaluació de la qualitat dels models se sol utilitzar la perplexitat, que es defineix com la probabilitat normalitzada que el model de llenguatge assigna a una seqüència de paraules. És un valor que ens dóna una idea del *factor por de ramificació* del corpus. La perplexitat serà màxima quan totes les paraules són equiprobables i el seu valor sigui igual al invers de la grandària del vocabulari.

## 2.4 Eines corresponents a la part de ASR que seran utilitzades en aquest projecte

Per muntar la part corresponent al reconeixement automàtic de la parla dins de la nostra aplicació s'utilitzarà ATK (Application Toolkit for HTK), una API dissenyada per facilitar la creació d'aplicacions experimentals de HTK. Tot seguit s'explicarà, amb més detall, què és cada cosa comentada en el paràgraf anterior. HTK (Hidden Markov Model Toolkit) és un conjunt d'eines portables per a la creació i manipulació de models ocults de Markov. Aquesta aplicació és útil sobretot en el camp del reconeixement automàtic de la parla, tot i que en nombroses ocasions s'utilitza en altres terrenys, com ara en la síntesi de veu, en el reconeixement de caràcters o bé en la seqüenciació de l'ADN. Veiem doncs, que HTK és usat en centenars de llocs arreu del món.

Es compon d'un conjunt de mòduls biblioteca i eines disponibles en codi font pel llenguatge de programació C. Aquestes eines proporcionen una infraestructura prou sofisticada com per dur a terme l'anàlisi de la parla, l'entrenament dels models ocults de Markov i proves i anàlisi dels resultats obtinguts.

Originàriament va ser desenvolupat al *Machine Intelligence Laboratory* (conegut també com a *Speech Vision and Robotics Group*) de la universitat de Cambridge, dins del departament d'enginyeria, a on era utilitzat per crear llargs sistemes de reconeixement de vocabulari. Més tard va canviar de mans quan *Entropic Research Laboratory Inc.* en va adquirir els drets per poder-hi comerciar.

ATK consisteix en una capa programada en C++ que es troba un nivell per sobre de les llibreries HTK estàndard. Això permet que nous reconeixadors siguin construïts utilitzant versions personalitzades de l'eina HTK. Suporta gramàtiques d'estats finits que poden ser carregades des del disc i creades o editades sobre la marxa en temps d'execució. També és compatible amb trigrames de models de llenguatge estadístics. Una bona característica és que és multi-thread, cosa que vol dir que permet múltiples entrades per part de l'usuari (comandes de veu, clics de ratolí...) que es combinen en un únic flux de dades. És capaç de retornar els resultats del reconeixement paraula per paraula, tal i com són reconegudes per tal de reduir la latència al mínim, permetent així construir sistemes d'alta capacitat de resposta.

# Capítol 3

## Conversió de text a veu

### 3.1 Introducció

Generar parla es pot aconseguir mitjançant diversos mecanismes en funció de la complexitat de la comunicació i dels recursos dels quals es disposi. Avui en dia trobem sistemes de diàleg molt útils però que només generen un número limitat de missatges, com pot ser el cas del sistema que trobem quan truquem per demanar hora per a portar el cotxe a passar la ITV. Els missatges de contestació poden ser missatges gravats anteriorment o bé es poden formar mitjançant la concatenació de gravacions de paraules o frases soles. Concatenar gravacions no només serveix per a aplicacions amb respostes limitades, sinó per a totes, ara bé, com es pot comprendre la feina de gravació (temps, locutors...) i l'espai que ocupen aquestes gravacions seran importants i un factor a tenir en compte. Una cosa bona però, és que la qualitat de la gravació, i per tant la resposta que rep l'usuari, és excel·lent. Si això s'aplica en portals d'Internet que ofereixen, a través de veu, les últimes notícies actualitzades cada X temps, seria inviable, ja que hi hauria d'haver uns quants locutors preparats cada moment per enregistrar la informació, no es podrien actualitzar al moment per la demora a l'hora de fer la gravació... Així doncs aquí seria necessari utilitzar un sistema automàtic de generació de missatges parlats, lloc on entra en escena la tecnologia de conversió de text en parla, que com s'ha explicat en el primer capítol, són sistemes capaços de produir parla a partir d'una representació escrita de qualsevol text en una determinada llengua. Tot i que no es poden comparar les dues qualitats, aquests sistemes han anat evolucionant de tal manera que ara representen una bona alternativa a les gravacions.

Un sistema de conversió de text a parla ha de passar per tres etapes mostrades a la figura 3.1: anàlisi del text, generació prosòdica i generació de

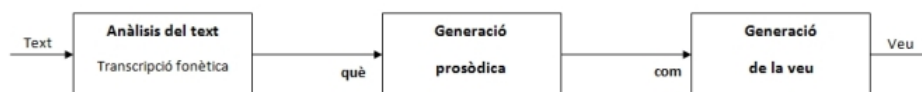


Figura 3.1: Etapes d'un sistema de conversió de text a veu

la veu. En la primera etapa es determina què és el que l'usuari haurà de sentir, és a dir, els sons que s'han de pronunciar. Aquí s'estudia el text, s'interpreten formes d'expressió no explícites (abreviatures, xifres...) i es determina els sons associats a cada paraula. En la segona etapa s'estableix com s'ha de pronunciar cada so considerat individualment i dins de la cadena fònica; cada so (vocals, consonants...) pot pronunciar-se de moltes formes, a diferents velocitats, amb tons greus o aguts, amb intensitat fluixa o forta... La velocitat d'elocució, el to o la intensitat constitueixen alguns dels trets denominats prosòdics, emprats també per transmetre informació sobre el significat del missatge o sobre certes característiques del locutor. Ja per acabar, en l'últim procés s'ha de produir un senyal sonor que correspongui als sons determinats a la primera etapa, la d'anàlisi del text, i a les característiques fixades en l'etapa de generació prosòdica.

## 3.2 Anàlisi del text i transcripció fonètica

### 3.2.1 Estructura del text

El primer que s'ha de fer si es vol estudiar un text és analitzar-ne la seva estructura. Segons sigui el tipus de text, pot presentar títols, capítols, paràgrafs, taules, figures... Tot i que poden ser importants a l'hora de llegir i entendre el text, aquests elements han rebut poca atenció. Tot i que les respostes que sol donar un convertidor de text a veu emmarcat dins d'un sistema de diàleg no són relativament llargues, hi pot haver casos en què es requereixi la lectura de textos de longituds variables, com per exemple quan ho utilitzem en les aplicacions d'accés per a persones cegues. Per tractar aquest problema actualment s'utilitzen documents estructurats, generalment emprant marques XML, i es dissenyen els sistemes de conversió de text a veu de manera que es tinguin en compte. Existeixen varis sistemes de marques dissenyats expressament per a la conversió del text a veu. Per destacar-ne un, citar el SSML (*Speech Synthesis Markup Language*), definit pel W3C (*World Wide Web Consortium*) i que constitueix el resultat d'un procés d'estandardització i consens partint de diverses propostes anteriors (Walker,

|                                                |                                                                                                                                                                                                                                                   |
|------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Marques generals</b>                        |                                                                                                                                                                                                                                                   |
| <lang>                                         | Idioma en el qual ha de ser llegit un text                                                                                                                                                                                                        |
| <b>Marques sobre l'estructura del text</b>     |                                                                                                                                                                                                                                                   |
| <paragraph>                                    | Permet indicar els paràgrafs del text                                                                                                                                                                                                             |
| <sentence>                                     | Permet indicar les oracions del text                                                                                                                                                                                                              |
| <b>Marques sobre la normalització del text</b> |                                                                                                                                                                                                                                                   |
| <say-as>                                       | Indica com s'ha d'interpretar una part del text (cardinal, data, dígit, telèfon...), el format de la informació (d/m/a: dia/mes/any) i el nivell de detall en la lectura (si només s'ha de dictar, si s'han de llegir els signes de puntuació...) |
| <b>Marques sobre la transcripció fonètica</b>  |                                                                                                                                                                                                                                                   |
| <phoneme>                                      | Indica la transcripció fonètica per a una determinada paraula. Permet indicar l'alfabet fonètic en el qual s'expressa                                                                                                                             |
| <lexicon>                                      | Indica quin diccionari de transcripcions fonètiques s'ha d'utilitzar                                                                                                                                                                              |
| <b>Marques sobre la prosòdia</b>               |                                                                                                                                                                                                                                                   |
| <emphasis>                                     | Indica el nivell d'èmfasi d'un text                                                                                                                                                                                                               |
| <break>                                        | Indica una pausa i permet assenyalar la seva duració de manera qualitativa o quantitativa                                                                                                                                                         |
| <prosody>                                      | Permet especificar alguns paràmetres prosòdics: el nivell base de la corba melòdica, el seu rang, el seu contorn, la duració d'una part del text, la intensitat...                                                                                |
| <b>Marques sobre la generació de veu</b>       |                                                                                                                                                                                                                                                   |
| <voice>                                        | Permet especificar característiques del locutor: edat, sexe...                                                                                                                                                                                    |
| <audio>                                        | Permet incorporar dins la síntesi de la parla un fitxer de so, que pot contenir veu, música...                                                                                                                                                    |

Figura 3.2: Marques SSML proposades pel consorci W3C

Burnett, Hunt 2002...). Aquest llenguatge sembla ser que serà adoptat per l'estàndard d'especificacions de diàlegs persona-màquina VoiceXML i per les tecnologies impulsades per Microsoft (SALT, .NET, ASP...). A la taula 3.2 es pot observar un resum d'aquestes marques explicades.

Destacar que, mitjançant les citades marques, es pot indicar l'estructura del text, influir en diferents funcionalitats del sistema com per exemple en la normalització, la pronunciació i la prosòdia. Si el convertidor troba un text que no disposa d'aquestes marques ha de ser capaç de poder-lo interpretar igualment i llegir-lo, per això ha d'incorporar un cert grau d'anàlisi per tractar les estructures bàsiques del text. Mitjançant heurístiques senzilles es pot determinar els diferents paràgrafs del document, entonant així una correcta pronunciació. Dins els paràgrafs, com és normal, hi ha frases, que cal distingir mitjançant els signes de puntuació (comes, punts, signes d'exclamació, signes d'interrogació...) tot i que moltes vegades el seu ús és ambigu (per exemple el cas del punt, que tant pot indicar que s'ha acabat



una frase com que la paraula que té al davant és una abreviació). Malgrat tot, cal tenir en compte que els sistemes per convertir text en veu es dissenyen intentant que funcionin el millor possible en la major part de les situacions.

### 3.2.2 Normalització del text

Segurament, mentre s'estigui fent l'estudi del text per determinar-ne l'estructura, es trobaran amb freqüència elements *no normalitzats*, és a dir, cadenes de caràcters la correcta pronunciació de les quals no es dedueix usant un diccionari fonètic, sinó que s'ha d'escollir una de les possibles pronunciacions que tenen segons el context on es troben. Per citar algun exemple podem observar la següent frase:

*Hi havia 200 persones que van arribar a bord de 200 vehicles.*

Com es pot observar els números solen ser elements no normalitzats. En aquest cas la primera pronunciació és *dues-centes* i la segona *dos-cents*. Això depèn del què acompanyi a aquest número. Tot i que molts números s'han de llegir com a cardinals, no es considera natural expressar, per exemple, el número de compta o el DNI com una quantitat amb milions o bilions; no obstant, és la única opció quan es fa referència a una quantitat explícita. Així doncs, la coma i el punt s'utilitzen per marcar la separació entre la part entera i la part decimal o per agrupar les xifres de tres en tres (mil, milió...). Si un número va associat a una marca o a un producte en concret pot ser que tingui una manera de pronunciar-se concreta, com per exemple els cotxes de la marca *Peugeot*: 205 (dos-cents cinc), 306 (tres-cents sis) o bé els antics processadors *Intel*: 286 (dos vuit sis), 386 (tres vuit sis)...

Els números romans també poden presentar dificultats a l'hora de ser pronunciats, especialment quan es tracta de números d'un sol caràcter: *I trobada gegantera del municipi*. La lletra *I*, a part de representar un nombre romà, també és una conjunció en la nostra llengua i s'ha de tenir present quan es tracta d'un cas o d'un altre.

Les hores i les dates són representades per números, tenint una forma especial de pronunciació, al igual que els codis postal, les direccions IP, les matrícules, formules matemàtiques i un llarg etcètera.

Aquests exemples posen de manifest que els mateixos caràcters (dígit) s'han de llegir de manera diferent atenent no només al format, sinó també al context lingüístic.

Les abreviatures són una altra font d'ambigüitat. En català podem trobar tot tipus de paraules abreviades, des de les que serveixen per tractar a la gent, com ara *Sr.* (senyor), *Sra.* (senyora) o *Dr.* (doctor) a les que representen

unitats de mesura, com per exemple *kg* (quilogram) o *mm* (mil·límetre), passant per les que ens assenyalen llocs: *C/* (carrer). Cal que el convertidor, encara que vegi l'abreviatura, pronunciï bé les paraules, les digui pel seu nom, sinó pot passar que l'usuari no arribi a entendre bé el missatge. Per fer-ho cal mirar el context en el qual es troben aquests mots, al igual que passa amb els números. Les paraules abreviades que s'envien en els missatges de text són un complex cas a part.

Les sigles i acrònims són abreviatures que es formen mitjançant lletres, normalment les inicials, de noms propis. *ONU*, *UPC*, *UAB* són exemples de sigles que podem trobar al vocabulari humà. Moltes sigles d'aquestes amb el pas del temps s'adapten a l'estructura fonètica pròpia de la llengua i els usuaris d'aquesta perden consciència del seu origen i l'acrònim passa a considerar-se paraula: així ha passat, per exemple amb la paraula *OVNI* (Objecte Volador No Identificat). Les sigles, en alguns casos, s'expandeixen i es pronuncien les paraules completes que representen; en altres no s'expandeixen sinó que, si és possible, es llegeixen com una paraula (*ONU*) i, si no, es lletregen (*UAB*). La regla general és que quant més habituat està l'usuari a pronunciar la sigla, major és la tendència a utilitzar-ne la versió curta. Com en els altres casos, el context lingüístic és fonamental per a una correcta interpretació, ja que un mateix acrònim es pot utilitzar per a diferents referents.

Com s'ha mencionat, la correcte normalització del text requereix un ampli context lingüístic que en alguns casos inclou informació semàntica. No obstant, la majoria dels sistemes utilitzen regles basades en els propis caràcters o en un context molt limitat, en ocasions adaptades al domini en el qual s'utilitzarà el convertidor. En aquests casos, el criteri que se segueix s'inspira en la prudència: es preferible una lectura literal que realitzar una mala interpretació. Per exemple: llegir *JJO* com *jota jota o o* no és la millor opció. Ara bé, més desconcertant seria llegir *Codi A20 JJO B3* com *codi a vint jocs olímpics b tres*. En aquest exemple, en el que la primera opció és molt més probable de que passi que no pas la segona, es pot acceptar el risc d'utilitzar l'expressió *jocs olímpics*. No obstant, en altres casos no tant evidents és preferible adoptar una posició més conservadora.

La normalització del text se sol dur a terme en dues fases. En la primera es classifiquen les cadenes de caràcters en paraules, abreviatures, signes de puntuació, dates, hores... En la segona, sabent ja el tipus d'informació que representa, s'expandeixen. En ambdues etapes se solen utilitzar regles basades en el coneixement. Per exemple, es poden escriure regles que contemplin la majoria dels possibles formats en els quals pot estar escrita una data:

*17/VI/2012*

17.06.2012

17-06-2012

La regla de classificació identifica també els components de l'expressió, en aquest cas el dia, el mes i l'any. Amb aquesta informació, les regles d'expansió són relativament senzilles: per exemple, transformar el VI a juny, sabent que es tracta d'un mes, requereix únicament una llista en la qual hi hagi associats els dotze primers números romans al mes al qual representen. El mateix amb l'expansió d'abreviatures o sigles, tot i que en aquest cas la llista pot ser molt i molt gran.

### 3.2.3 Transcripció fonètica

La transcripció fonètica consisteix en determinar la correcta pronunciació de cada paraula, expressant-la mitjançant algun alfabet fonètic, que aquí a Europa el que s'ha anat utilitzant més en els darrers anys és el SAM Phonetic Alphabet (SAMPA). En la conversió de text a veu es requereix, a part de la transcripció fonètica pròpiament dita, determinar la posició del accent i, si és possible, la divisió de la paraula en síl·labes.

Aquesta tasca, que segons en quina llengua es vulgui utilitzar en el convertidor pot ser molt complicada, és, a primera vista, senzilla pel català, ja que la relació entre sons i grafies en la pronunciació estàndard es pot sistematitzar mitjançant un nombre reduït de regles. Com també passa en altres idiomes, el català té paraules estrangeres que poden trobar-se, amb un percentatge força significatiu, als textos que volem que siguin llegits pel convertidor. Aquestes paraules no pertanyents a l'idioma no acostumen a regir-se per les regles habituals de pronunciació d'aquesta llengua. En general, la transcripció fonètica es basa en l'aplicació d'un diccionari o bé d'unes regles en el cas de que la paraula no es trobi al diccionari. En llengües com l'anglès, en les que la correspondència entre grafia i so és irregular, els diccionaris contenen desenes de mils de paraules i les regles s'apliquen a les paraules desconegudes, en canvi en llengües com el castellà o català, que són més regulars, el diccionari només conté unes certes paraules, les que corresponen a les excepcions de les regles de pronunciació. Casos, com es pot observar, totalment oposats, ja que el primer cas tibia de diccionari i el segon de les regles. Una de les principals causes del perquè això és així és que en la llengua anglesa pots trobar sistemes que utilitzen més de mil regles...

Com és lògic, escriure a mà aquesta gran quantitat de regles és inviable, així que s'utilitzen mètodes com ara utilitzar un classificador per decidir, per a cada grafia, quin so li correspon segons el context on es troba. Com que hi ha vegades que les grafies no tenen una manifestació fonètica -o dígrafs que es

corresponguin amb un únic so-, es considera la possibilitat de que a una grafia no li correspongui cap so. Per fer això es pot utilitzar la tècnica coneguda amb el nom d'arbres de decisió. Un altre mètode proposat consisteix en buscar, dins al diccionari, paraules el més semblants possible a la que no hi és i adaptar-ne la seva pronunciació, tenint en compte les diferències entre la paraula buscada i la trobada. Així doncs, el coneixement de la pronunciació d'una paraula que no podria obtenir-se mitjançant les regles regulars de conversió de grafia a so, com per exemple *byte*, pot utilitzar-se per derivar la pronunciació d'altres de similars: *bytes*, *megabytes*.... Tots aquests mètodes exposats poden aplicar-se tant a la transcripció fonètica com a la accentuació i a la separació de síl·labes.

Els diccionaris solen estar estructurats partint de formes, de manera que s'inclouen les paraules tal i com apareixen dins el text, a més a més de les variacions més freqüents en funció del seu gènere, número, persona o sexe.

Tal i com s'ha explicat, la transcripció fonètica es basa únicament en la forma ortogràfica. No obstant això, aquest procediment no pot distingir la pronunciació dels homògrafs, és a dir, paraules que s'escriuen igual però que es pronuncien diferent. Com es veurà més endavant, l'informació sobre la categoria lèxica de cada paraula és necessària també per derivar la prosòdia, fet pel qual els diccionaris incorporen aquest tipus d'informació, coneguda amb el nom de *part of speech* (POS). Junt a cada paraula habitualment s'hi poden trobar les seves possibles categories lèxiques. Aquesta solució no arregla el problema dels homògrafs, però sí que contribueix notablement a la disminució dels errors.

### 3.3 Modelat prosòdic

La representació dels sons que s'obté una vegada realitzada la transcripció fonètica automàtica no és suficient, ja que un mateix conjunt de sons pot pronunciar-se de diferents maneres, modificant el què es coneix com a trets prosòdics: entonació, accent, intensitat, to, duració, velocitat d'elocució, pauses... Aquests trets són determinants per aconseguir una major naturalitat en aquests sistemes de conversió i, a més a més, poden aportar un canvi de significat o indicar una actitud o una intenció concretes per part del locutor. En resum, la pronunciació d'una mateixa cadena fònica pot presentar resultats diferents segons els trets prosòdics:

*La Joana juga a futbol.*  
*La Joana juga a futbol?*

Podem constatar doncs, la gran capacitat expressiva que la prosòdia aporta al llenguatge oral, encara que no resulta fàcil que un convertidor de text a veu gaudeixi d'aquesta riquesa expressiva, ja que això depèn de la informació de tipus sintàctic, semàntic i pragmàtic que es pot deduir del text, així com de les característiques dialectals, socioculturals i personals del vocal.

Així doncs, l'objectiu d'aquesta part del procés de conversió de text a veu és, per una part, determinar, a partir del text, l'informació prosòdica que s'ha de representar i, per altra banda, convertir aquesta informació en un senyal sonor, és a dir, en paràmetres físics concrets i quantitatius.

Un dels trets prosòdics sobre el qual s'hi ha dedicat més estudi a l'hora d'evolucionar aquesta tecnologia és l'entonació, és a dir, l'efecte perceptiu que produeixen a l'oient les variacions de la freqüència fonamental (to), del temps (duració) i de l'amplitud (intensitat) durant la conversa. Les pauses també tenen una gran importància en la naturalitat de la parla. La seva aparició i duració poden venir determinades per l'estructura sintàctica dels enunciats i per l'organització jeràrquica dels constituents que els componen.

### 3.3.1 Corbes melòdiques

Les variacions melòdiques proporcionen informació lingüística, sociolingüística i expressiva. Des del punt de vista lingüístic, l'entonació compleix la funció *distintiva*, ja que és capaç de diferenciar la modalitat oracional d'enunciats el contingut dels quals és idèntic:

*Abans d'ahir vas anar a córrer.*

*Abans d'ahir vas anar a córrer?*

Una altra funció que també compleix és la *demarcativa*, que pot representar un valor distintiu al delimitar, tot utilitzant pauses i/o inflexions de la freqüència fonamental, unitats sintàctiques més o menys complexes, és a dir, grups d'entonació:

*Les noies que van de compres són a Barcelona.*

*Les noies, que van de compres, són a Barcelona.*

Mirat des del punt de vista sociolingüístic, aporta informació sobre el locutor (edat, sexe...) i sobre el grup dialectal, social o cultural al qual pertany. Finalment, el vocal pot recórrer a l'entonació per expressar intencions i actituds:

*Aniràs a la platja?* Amb to d'incrèdilitat  
*Aniràs a la platja?* Amb to de súplica

Generalment, en la conversió de text a veu es tenen en compte aquells factors que deriven fàcilment del text, per això no se solen modelar determinats fenòmens, com per exemple el focus, ja que la seva determinació a partir del text no és una tasca trivial.

En concret, els fenòmens que s'acostumen a modelar, encara que no sempre d'una forma explícita, en la síntesi de la parla són:

- La declinació progressiva de la freqüència fonamental que té lloc al llarg de l'enunciat.
- La modalitat oracional (enunciativa, exclamativa i interrogativa).
- Les fronteres prosòdiques entre els enunciats, estiguin o no marcades usant una pausa.
- Les síl·labes accentuades de cada paraula.

Per això és necessari tenir en compte informació sobre:

- El tipus d'oració.
- La posició del grup d'entonació dins de l'enunciat (principi, mig o fi).
- La duració del grup d'entonació, generalment contant síl·labes.
- Els límits sintàctics que delimiten els grups d'entonació.
- La posició en la qual apareixen les síl·labes tòniques.
- El tipus de veu que es vol generar: masculina, femenina...

Durant els últims anys s'han desenvolupat un gran nombre de sistemes per modelar l'entonació i una de les maneres de classificar-los és dividir-los en fonètics i fonològics. Mentre que els primers es centren en l'anàlisi de l'entonació a partir dels seus correlats físics, els segons parteixen de l'estructura de la llengua per establir un inventari de categories abstractes. En qualsevol cas, atès que durant la conversió es requereixen uns valors numèrics per generar la corba melòdica, els models fonològics s'han de complementar amb un procediment per derivar els contorns melòdics. De la mateixa manera, els models fonètics han d'incloure un procediment per assignar els diferents patrons en funció de la informació textual, la única de la que parteixen aquests sistemes de conversió que estem explicant.

Veurem un parell d'exemples de cada classificació. Començant pels models fonètics, citar el model del IPO (1990) o el de Garrido (1996). Per altra banda, de models fonològics podem trobar el mètric i auto-segmental (1980, 1996) o el model de Mertens (1997).

Una altra manera de classificar els sistemes de modelatge de l'entonació és la que divideix aquests models en jeràrquics i lineals, segons si les corbes melòdiques es generen a partir de la superposició de patrons de diferents nivells o bé s'obtenen mitjançant la concatenació de les unitats en un sol nivell.

Un exemple de classificació jeràrquica és la proposada per Fujisaki (1991). Es tracta d'una aproximació de base fisiològica que modela els contorns de freqüència fonamental a partir de dos components: un de frase, que genera un contorn global d'entonació i un altre d'accent, que genera els contorns locals, és a dir, els pics sobre les síl·labes accentuades.

Un model que trobem classificat com a lineal és el Tilt (2000).

### 3.3.2 Duració

La duració dels segments de la parla és fonamental per dotar d'una major naturalitat la síntesi. També pot ser útil per assenyalar el límit d'una frontera prosòdica. Per exemple, un allargament significatiu de la duració dels sons d'una síl·laba final de paraula pot ser una estratègia del locutor per marcar la separació entre dos constituents prosòdics sense haver de realitzar una pausa. La duració també aporta informació sobre l'estat d'ànim de la persona que parla: exaltat, tranquil...

Partit d'aquest objectiu, els sistemes de conversió de text a veu generen la duració dels segments en funció de diversos factors:

→ El tipus de segment. Cada un dels sons de la parla té una duració intrínseca, condicionada per l'esforç necessari per realitzar el gest articulatori que requereix la seva pronunciació.

→ La posició del segment dins del grup fònic. És freqüent un increment de la duració dels sons que es troben en posició anterior a una pausa.

→ La tonicitat de la síl·laba. La duració de les vocals augmenta considerablement quan es troben situades en una síl·laba tònica.

→ L'estructura de la síl·laba. La duració vocàlica acostuma a ser superior quan la vocal es troba en una síl·laba oberta que quan apareix en una síl·laba travada.

→ La posició del segment dins la síl·laba. La duració de les consonants varia si estan en posició d'atac o de coda sil·làbica o bé si formen part d'un atac simple o compost.

→ El context fonètic immediat. La duració d'un so pot veure's influïda pel punt, el mode i la sonoritat del segment següent i/o precedent.

Tenint en compte aquests factors, s'han desenvolupat diferents models de duració que es poden dividir en dos grups: els basats fonamentalment en regles i els que es recolzen en gran mesura en algorismes d'exploració de dades, encara que aquests últims poden utilitzar coneixements lingüístics tant per seleccionar les característiques rellevants com per formular el propi model.

En els models basats en regles es sol partir d'una duració intrínseca per a cada segment, obtinguda del anàlisi d'un corpus de parla natural, que es modifica en funció de diferents factors com ara els citats anteriorment.

La disponibilitat de corpus cada vegada més grans permet utilitzar mètodes d'exploració de dades, com els arbres de classificació i regressió (CART) que estimen la duració dels segments realitzant separacions binàries de les dades per minimitzar la varietat de les duracions en els dos subconjunts resultants. Així, pot utilitzar-se un arbre de regressió que aprengui a estimar la duració en funció dels paràmetres lingüístics que es considerin, a priori, rellevants: la identitat del segment, l'estructura de la síl·laba que conté el segment, la posició del segment dins la síl·laba, la posició de la síl·laba dins del grup fònic... Un problema que presenta aquest procediment és que cada vegada que l'algoritme decideix que certs factors són importants -per exemple la identitat del so- considera les dades de cada partició de forma independent i no utilitza la informació referent al efecte que té un factor sobre l'altre. Un exemple del problema comentat és que no es podrà utilitzar informació sobre l'efecte que té l'accent damunt una determinada vocal en posició anterior a una pausa utilitzant el coneixement de l'efecte que l'accent produeix sobre aquesta vocal en general o de l'efecte que produeix damunt el conjunt de les vocals d'abans d'una pausa.

Pot ser avantatjós definir models basats en el coneixement, encara que siguin molt fluïxos, i estimar els paràmetres del model, especialment quan les dades de les quals es disposa són limitades. Un d'aquests models, proposat per van Santen (1998) és el conegut com a *Suma de productes*. En aquest model, la duració d'un so en un context ve donat per la suma de varis sumatoris, cadascun dels quals surten de la contribució multiplicativa de varis factors. Escollint adequadament els factors que hi intervenen, s'imposa al model el què l'autor denomina *invariància direccional*, que es refereix a que alguns factors afecten en el mateix sentit (en aquest cas incrementen) la duració de tots els sons. Campbell (1992) també proposa un model, en aquest cas jeràrquic, que utilitza la duració de la síl·laba com un paràmetre intermedi: s'estima primer mitjançant una xarxa neuronal la duració de la síl·laba en funció d'alguns factors (tonicitat, número de fonemes de la síl·laba...) i llavors es divideix aquesta duració entre els diferents fonemes de la síl·laba, atenent a la identitat de cada al·lòfon.



### 3.3.3 Intensitat

La intensitat dels sons varia, igual com hem vist que passa amb la duració, en funció de diferents factors:

→ El tipus de segment. Els sons de la parla es caracteritzen per tenir una intensitat intrínseca que depèn de l'energia del impuls generador del so.

→ La tonicitat de la síl·laba. La intensitat de les vocals és major en síl·labes tòniques que no pas en síl·labes àtones.

→ La posició dins la síl·laba. En les seqüències que formen part de la mateixa síl·laba, les vocals que estan en posició de nucli són més intenses que les que estan fora d'aquesta posició. En el cas de les consonants, la intensitat és major en posicions d'atac.

→ La posició del segment dins del grup fònic. Els sons acostumen a tenir una menor intensitat en posicions que es troben abans d'una pausa que en posicions de grup fònic.

Tanmateix, en la conversió de text a parla, la intensitat no acostuma a modelar-se, ja que es considera que els canvis de duració i de freqüència fonamental poden ser suficients per caracteritzar els fenòmens prosòdics que solen tenir-se en compte: l'accent, l'entonació o el ritme. Així doncs, la majoria dels sistemes de síntesi, únicament normalitzen la intensitat de les unitats concatenades per evitar discontinuïtats en el senyal.

No obstant això, durant els últims anys, alguns sistemes de síntesi han incorporat un mòdul d'assignació de la intensitat amb la finalitat de realçar informació, ja que aquest paràmetre es considera de gran importància en la focalització d'un element de l'enunciat, per exemple si és vol remarcar a on s'està realitzant una determinada acció per evitar confusions.

Un altre camp en el que també s'ha tingut en compte la intensitat és en la síntesi d'emocions, a on un canvi a la corba de la intensitat pot servir per manifestar com es troba el locutor: sorprès, trist, alegre...

### 3.3.4 Pausas

Quan s'està llegint un text el més normal és fer pauses, ja siguin de tipus fisiològiques o bé lingüístiques. Les primeres són les que es fan per necessitat, les que es produeixen quan el locutor para un moment per agafar aire i així poder respirar. Les segones serveixen per delimitar i estructurar els enunciats. Un bon lector intenta fer coincidir els dos tipus de pauses.

Els signes de puntuació són un bon indicador per realitzar l'assignació de les pauses de tipus lingüístic en la conversió de text a veu:

*El primer equip del Barça, que jugava a casa contra el Madrid, va guanyar per un clar i contundent 5-0.*

Cal tenir en compte també que en determinades ocasions no es realitzen pauses davant d'alguns signes, especialment quan l'última pausa es troba en la paraula immediatament anterior, com passa freqüentment en les enumeracions:

*Per obtenir resultats a l'hort cal seguir els següents passos: llaurar, sembrar i recollir.*

En aquests casos la funció *demarcativa* s'acostuma a realitzar utilitzant altres correlats prosòdics com ara la duració o la freqüència fonamental, per citar-ne dos exemples.

No obstant això, les pauses marcades ortogràficament no resulten suficients per emular la parla humana i, encara que els convertidors no necessitin realitzar pauses per respirar, el seu encaix en llocs estratègics contribueix a una major naturalitat de la parla sintetitzada. Per exemple la oració

*El primer equip del Barça va jugar a casa contra el Madrid.*

es pot entonar sense cap pausa, ara bé, resulta més natural i més fàcil d'entendre si realitzem alguna de les pauses marcades a continuació tot usant el distintiu *<break>*:

*El primer equip del Barça <break> va jugar a casa <break> contra el Madrid.*

Com es pot observar, existeix més d'una possibilitat pel que fa al número i posició de les pauses, ja que hi ha un gran nombre de combinacions possibles a l'hora de crear grups de síl·labes, segons si es volen enunciats més curts o més llargs.

Tot i la gran facilitat de combinacions a l'hora de col·locar els espais, hi ha algunes posicions en les quals afegir-hi un espai és del tot antinatural. Per exemple, no seria admissible la inserció d'una pausa després d'un article o d'una preposició; ni tampoc entre elements que mantenen un alt grau de cohesió sintàctic i semàntic: entre un nom i un adjectiu, entre un adverbi i un verb... Així doncs, en l'exemple anterior qualsevol de les pauses aquí ficades seria incorrecte:

*El <break> primer equip del Barça va jugar a casa contra <break> el Madrid.*

Per tant, encara que el fet d'afegir pauses cada determinat període de temps sigui bo i doti d'una major naturalitat la síntesi de veu, ha d'existir una certa correlació entre l'estructura sintàctica i la posició de les pauses. També cal determinar la duració d'aquestes, temps inversament proporcional a la velocitat d'elocució, al número de pauses i al grau de cohesió existent entre els constituents sintàctics que separen.

En la conversió de text a veu, el tractament de les pauses se sol realitzar en dues etapes: durant la primera es determina la posició d'aquestes per així a la segona assignar-les-hi una duració. La informació que se sol considerar és la següent:

→ Les etiquetes morfosintàctiques assignades a cada una de les paraules, inclosos signes de puntuació que se solen considerar com una paraula més, portant associada la seva corresponent etiqueta morfosintàctica.

→ Un anàlisi sintàctic superficial, en el qual se sol determinar l'agrupació de paraules en sintagmes.

→ La longitud, en número de paraules o de síl·labes, entre pausa i pausa.

Tenint en compte aquestes consideracions, es poden dissenyar regles o utilitzar algorismes de classificació que aprenguin les regles d'assignació partint d'un corpus en el qual s'incloguin les característiques mencionades i les posicions correctes de les pauses. Així mateix, mitjançant algorismes de regressió aplicats a un corpus oral es pot determinar la duració de cada tipus de pausa en la parla natural.

### 3.4 Síntesi de veu

Entrem a parlar de l'última etapa que porten a terme els sistemes de conversió de text a veu. Una vegada determinats quins segments s'han de produir i amb quines característiques, cal generar o sintetitzar -terme que a vegades també s'utilitza per referir-se al sistema complet- la veu. Per crear la veu se solen utilitzar els següents mètodes: síntesi basada en regles o també coneguda com a *síntesi per formants*, síntesi articulatòria i síntesi per concatenació. De tots aquests mètodes, el més utilitzat en els sistemes d'avui en dia és la síntesi per concatenació, ja que és el que permet obtenir una veu més natural.

### 3.4.1 Síntesi per formants

Aquest mètode es basa en un model de producció de font i filtre. El senyal *font* o *excitació* modela fonamentalment l'efecte de sortida de l'aire dels pulmons i dels moviments de les cordes vocals. Per modelar les excitacions sonores s'utilitza un tren d'impulsos, separats segons la freqüència fonamental (indicada per la corba melòdica), que passa per uns filtres que conformen el pols glotal. La font per sons sords es genera mitjançant un soroll blanc (totes les freqüències tenen el mateix contingut) organitzat mitjançant un filtre passa baixos. Pels sons africats i, també amb menor mesura, per tots els sons sonors, es barregen els dos tipus d'excitació. Referent al filtre, que representa el tracte vocal i nasal, es modela mitjançant un banc de filtres, en cascada o en paral·lel, cada un d'ells modelant un dels formants del tracte vocal. Els filtres van modulant el contingut freqüencial de l'excitació mitjançant aquestes ressonàncies, de la mateixa manera que el tracte vocal modula l'ona que li arriba provinent de la laringe.

Els diferents paràmetres que controlen el sistema, concretament la freqüència i l'ample de banda dels formants, varien lentament seguint unes regles que indiquen les trajectòries. La idea és que cada segment (o cada al·lòfon en un determinat context fonètic), per cada locutor, té associat una realització *objectiu* i, per tant, uns paràmetres *objectiu* del model: freqüències, amplituds i amplituds de banda. Els paràmetres segueixen una trajectòria suau entre aquests valors, però amb un límit en la velocitat de canvi provocada per la inèrcia dels articuladors. Per això, si la duració d'un so és curta, no s'arriba al valor objectiu, sinó que abans s'inicia la transició cap a un altre so.

Aquest mètode és molt flexible, ja que amb pocs paràmetres es pot controlar la generació de la veu. Una vegada dissenyades les regles per una veu, és fàcil generar veus i efectes diferents. La veu que es genera és intel·ligible, ara bé, no és molt natural ni es reconeix la identitat de la veu utilitzada per derivar els paràmetres. La limitació fonamental no sembla estar en el model, sinó més aviat en derivar els paràmetres que s'apliquen a aquest model.

Aquest mètode, tot i no ser dels més utilitzats, encara no ha desaparegut, és més, actualment encara s'estan investigant maneres per obtenir els paràmetres automàticament a partir d'una veu gravada. Un dels grans avantatges que presenta la síntesi per formants és la poca memòria que requereix, un fet molt estimat si es vol fer funcionar un sintetitzador d'aquests en un dispositiu amb grans restriccions computacionals, com per exemple un mòbil.

### 3.4.2 Síntesi articulatòria

En aquest mètode, la veu s'obté mitjançant un modelat explícit dels articuladors: forma i posició de la llengua, mandíbula, obertura del conducte nasal, volum d'aire que arriba dels pulmons... Aquest model presenta una gran complexitat, tant en el seu plantejament com en el cost computacional, a més de que resulta difícil aproximar els paràmetres que regeixen els moviments dels articuladors. La qualitat de la veu obtinguda no arriba al nivell de la resultant de la síntesi per formants, encara que avui en dia encara hi ha una línia d'investigació activa de la qual alguns n'esperen obtenir més coneixement, més flexibilitat i una major qualitat que la presentada per els altres tipus de síntesi.

### 3.4.3 Síntesi per concatenació

La manera que té la síntesi per concatenació de crear veu és agafar segments de veu natural i, tal i com el seu nom indica, concatenar-los de manera que se suavitzin les discontinuïtats. Aquests segments, en principi, haurien de tenir les característiques prosòdiques modificades per tal d'atendre els requisits del mòdul prosòdic. Aquest fet causa certa degradació a la qualitat de la veu. Encara que en un principi es pot pensar que els al·lòfons constitueixen la unitat lingüística més adequada per aquest tipus de sistemes, la veritat és que la concatenació d'al·lòfons presenta molts problemes, ja que les seves fronteres són precisament els punts de major variabilitat i, en conseqüència, els llocs en els quals la discontinuïtat serà més gran. Les unitats més utilitzades són el que es coneix com a *difonemes*, unitats de la mateixa dimensió que un al·lòfon però que contenen la transició a la part central del segment, és a dir, que emmagatzemen almenys la part final de cada al·lòfon juntament amb el principi de qualsevol altre al·lòfon que pugui seguir.

Els sistemes de referència actuals utilitzen grans corpus de veu (de desenes d'hores) i seleccionen les unitats afavorint aquells segments de major longitud, reduint així el nombre de concatenacions i prioritzant les unitats amb unes característiques més acords al que exigeix el modelat prosòdic; d'aquesta manera s'evita o es redueix la modificació del senyal.

L'encarregat de seleccionar aquestes unitats és un algorisme que busca aquelles que fan que la diferència entre les característiques prosòdiques de les unitats respecte l'objectiu fixat pels mòduls anteriors (apartats 3.2 i 3.3) sigui mínima. Aquesta diferència es quantifica mitjançant una sèrie de factors basats en l'adequació de les unitats escollides a la corba melòdica, a la duració, a la intensitat i també, en ocasions, a un determinat tipus d'informació fonològica com pot ser el context fonètic o el caràcter accentuat

de la síl·laba. A més a més, la mateixa cerca també inclou un criteri de selecció que afavoreix aquelles seqüències d'unitats que es van pronunciar juntes al gravar el corpus i que presenten una menor discontinuïtat en les unions. Quan es dissenya el corpus, es fa maximitzant la variabilitat fonètica i prosòdica de les unitats, de manera que l'algoritme de síntesi hagi de realitzar poques concatenacions i poques modificacions prosòdiques.

Una vegada seleccionades les unitats de veu que es necessiten per formar la frase és necessari, primerament, modificar els paràmetres acústics relacionats amb l'entonació per satisfer les especificacions dels mòduls anteriors. Un enfocament molt conegut es basa en el model *Linear Prediction Coding* (LPC), utilitzat en molts codificadors per comprimir la veu. En aquest model se separa, igual que en el sintetitzador de Klatt, la font del filtre, separació basada en propietats del senyal més que no pas en consideracions relacionades amb la producció de la veu. El què aquí es denomina filtre engloba, no només el tracte vocal i nasal, sinó també l'efecte de radiació i la forma del pols glotal. L'excitació conserva informació sobre la periodicitat dels senyals sonors (freqüència fonamental). Per dur a terme aquesta descomposició excitació-filtre existeixen algorismes simples, però eficients, que actuen de forma local cada 10-20 mili-segons. El filtre pot representar-se, a part de pels coeficients, per altres paràmetres -com els *Line Spectrum Frequencies*- que ofereixen molt bones propietats d'interpolació. És a dir, si al unir dues unitats de síntesi s'utilitza un filtre que és una mitja dels filtres que representen els extrems de les unitats que s'han de concatenar, el senyal mostra, des del punt de vista espectral, una transició suau en el punt d'unió. Les manipulacions prosòdiques han de realitzar-se sobre l'excitació, sent aquest l'aspecte que més atenció ha centrat tant en codificació com en síntesi de veu. El model més senzill, conegut com *vocoder LPC* substitueix l'excitació d'aquells sons sonors per uns polsos separats en funció de la freqüència fonamental requerida pel perfil de la corba melòdica. Pels sons sords, l'excitació és soroll blanc (amb una intensitat igual en totes les freqüències). Per modificar la intensitat s'ha d'ajustar el guany del filtre. Finalment, per variar la duració d'una unitat, s'ha de generar el senyal d'excitació per aquesta duració i separar o apropar, segons sigui, els punts en els quals es canvia el filtre. És a dir, si tenim un senyal de 100 ms de duració representat per 10 filtres espaiats equitativament cada 10 ms, per tenir un senyal de duració doble l'excitació durarà el doble i s'aplicaran els 10 filtres, però canviant-los cada 20 ms. La limitació del vocoder LPC resideix en la baixa qualitat que ofereix el propi codificador, encara que s'utilitzi només per comprimir la veu, sense modificar ni la freqüència fonamental, ni la intensitat ni la duració. Temps enrere es van proposar millores en l'excitació consistents en separar el pols glotal del tracte vocal. Simultàniament, en codificació, es van anar estudiant

millors representacions de l'excitació, com el multi-pols, polsos regulars o *Code Excited Linear Prediction* (CELP). Aquests codificadors, en els quals es basen la majoria d'estàndards de telefonia mòbil s'han aplicat amb èxit relatiu a la síntesi de veu.

Cap a finals dels anys 90 es va introduir l'algoritme *Pitch Synchronous Overlap and Add* (PSOLA), que permet canviar la duració i la freqüència fonamental d'un segment de veu sense necessitat d'un model de producció. L'ús d'aquest algoritme es va estendre ràpidament gràcies a la seva simplicitat i als seus bons resultats a l'hora de realitzar canvis relativament moderats de duració i freqüència. La base del seu funcionament és la següent: se suposa que el senyal és periòdic, la qual cosa és una aproximació raonable per la part sonora i es divideix en trams presos regularment segons el període del senyal ( $1/\text{Freqüència fonamental}$ ); per evitar discontinuïtats, les finestres que s'utilitzen per seleccionar els trams tenen una duració més gran que el període del senyal, normalment el doble, encara que les mostres s'atenuen segons s'allunyen del centre. Per reconstruir el senyal (sense cap modificació) s'han de sumar aquests trams de senyal sobreposada que s'han extret en sincronia amb les marques de la freqüència. Si es desitja un senyal de major duració cal repetir els períodes; i al contrari, si el què es vol és un senyal d'una duració menor cal eliminar-ne. Per modificar la freqüència fonamental s'han d'apropar o allunyar els trams de senyal abans de sumar-se. L'algoritme PSOLA pot realitzar canvis significatius en aquesta freqüència i en la duració, de manera que es manté una excel·lent qualitat en la veu.

No obstant, aquest algoritme presenta alguns problemes en la concatenació de segments. En general, el problema es deu a que mentre que l'algoritme assumeix que el senyal és perfectament periòdic i funciona bé per a senyals gairebé periòdics, aquesta hipòtesi pot ser que no es compleixi al agafar períodes de dos segments diferents, inclús en el cas de que ambdós representin el mateix soroll. En primer lloc, pot haver-hi un canvi de fase entre els dos segments, és a dir, els períodes de senyal que se sumen poden no estar perfectament sincronitzats, cosa que farà que al sumar-los, el senyal quedi degradat. També pot haver-hi diferències en la freqüència fonamental d'ambdós segments, fet pel qual les finestres no duren el mateix i la suma tampoc és coherent. Finalment, pot haver-hi diferències espectrals entre els segments (desplaçaments dels formants) i aquest algoritme no té cap mecanisme per suavitzar les transicions. S'han proposat variants que milloren el comportament del PSOLA, tot i que n'augmenten la complexitat. Una primera alternativa amb la qual s'han obtingut bons resultats consisteix en aplicar aquest algoritme al residu resultant d'un anàlisi LPC. L'avantatge d'aquest algoritme és que permet suavitzar l'espectre mitjançant interpolació del filtre de síntesi. Una altra alternativa és l'algoritme MBROLA, desen-

volupat per Dutoit i Leich, que realitza una normalització de les unitats de veu perquè, abans d'aplicar l'algoritme PSOLA, aquestes tinguin la mateixa longitud de període i estiguin en fase. A més a més, en aquestes condicions l'ajustament temporal equival a la interpolació espectral, fet que fa reduir els tres tipus de discontinuïtats: de fase, de freqüència fonamental i espectral.

També s'han proposat algoritmes que es basen en una representació sinusoidal o harmònica del senyal (Stylianou). En cada tram de senyal, la part periòdica d'aquest es representa mitjançant els components freqüencials que modelen els harmònics. Al generar el senyal, les amplituds, les fases i les freqüències d'aquests components evolucionen per acostar-se a les del següent tram, amb lo qual es produeixen transicions suaus. Aquest procediment per obtenir transicions suaves és vàlid tant per trams de veu que originalment eren consecutius com pels que no ho eren, tal com passa en la concatenació de les unitats de síntesi. Aquests algoritmes poden oferir millors prestacions, encara que són molt complexes tant des del punt de vista del desenvolupament com del cost computacional.

Encara que la utilització d'algoritmes per modificar la prosòdia és un aspecte fonamental en síntesi per concatenació, cal indicar que, si es seleccionen les unitats d'un corpus gran, aquest aspecte és menys important. La raó és que les unitats seleccionades ja solen tenir uns valors prosòdics propers als indicats pel model prosòdic, fet pel qual les modificacions són menors i, per tant, menys problemàtiques. De fet, molts dels sistemes actuals utilitzen aquests algoritmes de manera molt limitada; en ocasions s'utilitzen únicament per reduir les discontinuïtats de la concatenació, sense modificar la prosòdia. No obstant, els algoritmes de modificació prosòdica segueixen sent objecte d'estudi. En primer lloc perquè si es millora la qualitat de la veu modificada no seran necessaris corpus tan grans, reduint així la memòria utilitzada, el temps de cerca i el cost de producció d'aquests. A més a més, a mesura que es continuï ampliant el coneixement sobre l'entonació, es vagin tocant estils de parla més variats o es consideri la síntesi de les emocions, augmentarà l'exigència del control prosòdic sobre el senyal generat.

### **3.5 La conversió de text a veu dins els sistemes de diàleg**

El fet d'utilitzar la conversió de text a veu dins d'un sistema de diàleg té associat una sèrie de particularitats que seran comentades en aquest apartat. La principal diferència que hi ha respecte al seu ús convencional rau en que, tot i que un sistema de diàleg pot recórrer a un convertidor



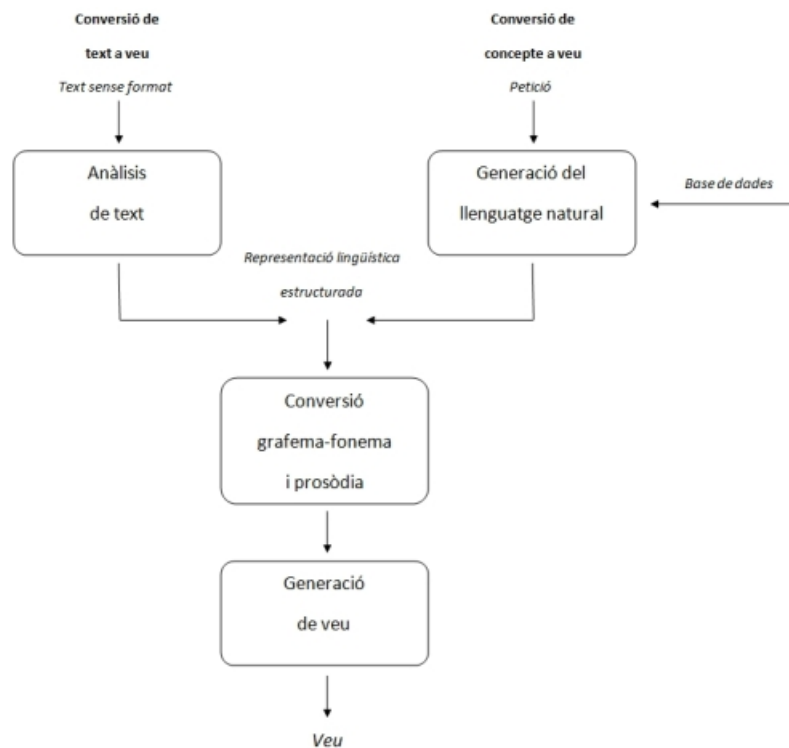


Figura 3.3: Conversió de text a veu i de concepte a veu

d'aquests per llegir textos generals (correus electrònics, pàgines d'Internet, notícies...), en la majoria d'ocasions els missatges estan elaborats per un mòdul de generació de llenguatge -l'ordinador crea el text que serà llegit pel mateix ordinador, cosa que facilita l'introducció de millores i precisió-, cas en què la informació relativa al missatge no es limita al text que genera aquest mòdul, sinó que pot incloure també informació semàntica, sintàctica i inclús pragmàtica, que es pot utilitzar per facilitar les tasques pròpies dels diferents mòduls del mateix convertidor i per generar la veu amb una major expressivitat. Quan es presenta el cas que l'entrada no és un text lliure, sinó que és un missatge generat per una màquina, s'utilitza el terme *conversió de concepte en veu*. A continuació veurem algunes de les característiques d'aquest tipus de conversió, entre les quals destaquen l'increment de l'eficàcia en els processos lingüístics, la millora en la generació de veu al tractar-se de dominis restringits i un major control de la prosòdia. La figura 3.3 mostra, en línies generals, la diferència entre els sistemes de conversió de text a veu i els de conversió de concepte a veu. En el primer sistema cal analitzar el text, com ja s'ha explicat, mentre que en el segon el text és generat a partir d'una determinada petició, accedint a la base de dades.

Analitzar el text és una tasca complexa, així que si hi ha un mètode que produeix text que inclou marques estàndards o bé específiques per tal indicar l'estructura -paràgrafs, llistes, frases...- i millorar així la feina de convertir cal aprofitar-ho, ja que s'evita haver d'utilitzar heurística amb totes les errades que això pot comportar. Un exemple de les marques de format que poden ser afegides pels sistemes de generació de llenguatge natural per tal d'ajudar als mòduls de conversió de grafia a text, prosòdia i generació de veu el podem veure a continuació:

*El sopar anual que s'havia de celebrar el 29 de juny a casa dels senyors Despuig ha quedat posposat fins a nova convocatòria.*

*<sentence> El sopar anual que s'havia de celebrar el <say-as interpret-as="date" format="dm">29/6</say-as> a casa dels srs. Despuig ha quedat posposat fins a nova convocatòria. </sentence>*

El primer exemple és la frase tal i com s'escriu normalment en la nostra llengua i el segon és la sentència creada per un sistema de generació de llenguatge. En aquest cas, les marques *<sentence>* i *</sentence>* serveixen per indicar l'inici i la fi de l'oració i les marques *<say-as interpret-as="date" format="dm">* i *</say-as>* s'utilitzen per assenyalar que 29/6 és una data expressada en el format dia/mes, evitant així confusions i mals pronunciats, com ara vint-i-nou barra sis. També ajuden a resoldre l'ambigüitat del punt darrera de l'abreviatura Srs. És cert que la data es pot crear escrita, tal i com està a la primera frase, però representada així pot aportar avantatges, ja que d'aquesta manera els mòduls comentats anteriorment disposen de més informació sense haver d'analitzar el text, apart de que el convertidor ja està preparat per afrontar les dates en aquest format degut a que en els sistemes informàtics es representen així. Per exemple, els components prosòdics (determinació de grups fònics, entonació...) poden estar preparats específicament per a la lectura d'entitats del tipus *data*.

Una altra avantatge que tenen els sistemes de conversió de concepte a veu sobre els sistemes de conversió de text a veu es basa en que, moltes vegades, en els primers el lèxic (o una gran part d'ell) és conegut. Això pot utilitzar-se per adaptar els diccionaris del sistema, tant per a la interpretació de sigles, abreviatures, símbols i expressions numèriques com per a assegurar una correcta pronunciació de totes les paraules.

En la conversió de text a parla és fonamental la determinació dels grups fònics (fragments compresos entre dues pauses), especialment quan no apareixen signes de puntuació. Mitjançant les marques SSML es pot deixar constància en el text de quan s'han de realitzar les pauses i quina ha de ser

la seva duració i, inclús, poden associar-s'hi corbes melòdiques.

La possibilitat d'incloure marques amb informació prosòdica rellevant ha inspirat l'aparició del sistema de marcat Stem-ML (Kochanski, Shih 2003), que permet controlar els fenòmens prosòdics amb el conseqüent efecte en la corba d'entonació. Amb aquest tipus de sistemes es pretén millorar la naturalitat controlant, amb precisió, la prosòdia.

En els sistemes generals de conversió de text a veu és molt difícil explotar totes les capacitats potencials de la prosòdia (distinció d'enunciats, èmfasis, expressió d'actituds...). Això és degut, fonamentalment, a la dificultat que suposa obtenir de forma automàtica, a partir del text d'entrada, tota la informació necessària per generar els missatges amb una prosòdia adequada. Quan l'entrada del convertidor és un text sense etiquetes, el sistema ha d'interpretar el significat d'aquest per a poder inserir correctament les pauses o realçar una part de l'enunciat, per exemple. En el nou escenari que suposa la conversió de concepte a veu s'incrementa el control i el nombre de funcions prosòdiques que poden tenir-se en compte, fet que millora la naturalitat dels sistemes.

Un dels aspectes funcionals de la prosòdia és el fet que permet eliminar l'ambigüitat d'enunciats com el següent:

*En cas d'enfonsament, les dones i els nens que no sàpiguen nedar han de ser els primers en abandonar el vaixell.*

L'usuari que hagi d'interpretar la frase ho pot fer de moltes maneres, segons quina prosòdia porti associada aquesta. Si es vol afavorir una lectura en la qual "que no sàpiguen nedar" només faci referència a "els nens", podem afegir una pausa o una inflexió de la freqüència fonamental abans de la conjunció i":

*En cas d'enfonsament, les dones <pausa> i els nens que no sàpiguen nedar <pausa> han de ser els primers en abandonar el vaixell.*

En canvi, si el què es vol és que el fet "que no sàpiguen nedar" faci referència tant a les dones com a "els nens", cal afegir la pausa just abans del complement:

*En cas d'enfonsament, les dones i els nens <pausa> que no sàpiguen nedar han de ser els primers en abandonar el vaixell.*

També hi ha l'opció de no afegir-hi cap pausa, tal i com està escrit al principi, i deixar que l'usuari ho interpreti al seu gust.

Un altre aspecte de la prosòdia, encara que aquest no s'hagi empleat tant sovint en la conversió de text a veu, és la funció emfàtica o centrar el focus d'atenció sobre una determinada part del text. Les estratègies utilitzades per dur a terme aquesta pràctica són: inserció de pauses abans i/o després de la informació emfatitzada, disminució de la velocitat d'elocució durant la realització del element assenyalat, inhibició del desplaçament del pic de la freqüència fonamental corresponent a la síl·laba accentuada amb un valor més alt del normal seguit d'un to greu sostingut després del focus o bé un augment significatiu de la intensitat sobre l'element.

En els sistemes de diàleg és freqüent que l'usuari focalitzi un element de l'enunciat amb la finalitat de corregir els errors que pot presentar el reconeixedor del sistema:

*USUARI: Trucava per fer una comanda de nous.*

*SISTEMA: Quantes dotzenes d'ous desitja?*

*USUARI: NOUS, vull NOUS no ous.*

Un possible ús que el sistema podria fer del realçament prosòdic seria el de contrastar informació quan el gestor de diàleg decideix confirmar una dada, bé perquè el propi sistema de reconeixement detecta que una paraula no és del tot fiable o bé perquè aquest gestor constata una inconsistència en els previs. L'exemple anterior vist en aquesta situació quedaria de la següent manera:

*USUARI: Trucava per fer una comanda de nous.*

*SISTEMA: Perdoni, ha dit OUS o NOUS?*

*USUARI: Nous.*

Apart de la funció emfàtica, la prosòdia pot ser útil en altres situacions del diàleg, com per exemple a l'hora de regular els torns de paraula entre el sistema i l'usuari. En ocasions, l'usuari s'avança a l'hora de contestar i interromp el torn del sistema. És un error molt comú que es podria arreglar amb el correcte control de la prosòdia. Observem el següent exemple:

*El preu de l'entrada és de 7 euros. Ara bé, si disposa del carnet jove té un 5% de descompte.*

Si no s'introdueixen marques prosòdiques especials, el més habitual és que la generació de l'enunciat es realitzi mitjançant una corba melòdica amb un final descendent per a la primera part de la frase. Degut a aquest fet l'usuari pot pensar que el missatge ja s'ha acabat i intenti contestar alguna cosa, quan

en realitat el sistema encara continuarà parlant. Aquest fet es pot evitar escurçant la pausa entre les dues frases o modificant el perfil d'entonació del final de la primera, donant a entendre així que el missatge encara no està complet.

Finalment, una altra aplicació de la prosòdia, pràcticament desatesa fins ara, té a veure amb el correcte control d'aquesta per ajudar a que l'actitud de l'usuari davant del sistema sigui més positiva. Els models d'entonació que s'usen en els sistemes de conversió de text a veu estan dissenyats perquè es puguin utilitzar en un gran nombre d'aplicacions variades; per aquest motiu les intervencions del sistema solen ser neutres i, molt sovint, excessivament monòtones. No obstant, un diàleg requereix adaptar-se a les situacions que es vagin produint, evitant així que l'usuari tingui una sensació d'excessiva neutralitat i es cansi, es desmotivi o es mostri incrèdul. Aquest problema té solució en els convertidors de concepte a veu, en els quals les intervencions del sistema estan més controlades. Així, l'usuari sent que es troba davant d'una veu amable i té una bona acceptació del sistema.

Pel que fa al mòdul de generació de veu, s'ha de tenir en compte que l'àmbit d'aplicació dels diàlegs sol restringir el número d'enunciats que llegeix el sistema. Quan el domini és restringit, la qualitat de la parla pot millorar sensiblement. Una opció que s'utilitza freqüentment és generar la contesta mitjançant frases pre-gravades. Si a priori es coneixen totes les possibles respostes, una solució consisteix en gravar-les, tot utilitzant un locutor professional, i prescindir així de la síntesi. Quan el cos dels missatges sol ser el mateix i només varien certes parts d'aquest, es recorre al ús de plantilles, és a dir, frases portadores en les quals existeixen elements fixes i elements variables, com es pot observar en el següent cas:

*Pròxima circulació per via \$NUMVIA tren amb destinació \$DESTI.*

En aquestes plantilles s'ha de gravar, per una part la frase portadora, amb uns valors arbitraris en les variables i per l'altra tots els possibles valors de les variables, les paraules en majúscula i que comencen pel signe \$. Aquesta manera de crear els missatges de contestació fa que tinguin una major qualitat de so, ara bé, els que surten dels convertidors de text a veu poden adaptar-se, d'una forma flexible, a diversos escenaris.

El compromís entre qualitat i flexibilitat fa que els desenvolupadors recorrin a tècniques que permeten millorar la resposta amb poc esforç. Així, per exemple, en la tècnica d'omplir plantilles, si alguna de les variables apareix en varies plantilles, els valors d'aquestes solen haver-se de repetir varies vegades per obtenir la pronunciació més adequada en cada context prosòdic. Perquè la concatenació entre els segments no presenti discontinuïtat

espectral, és habitual incloure silencis abans i després dels segments associats a les variables. Encara que aquest mètode és molt utilitzat i proporciona una veu d'una gran qualitat, presenta l'inconvenient de que és molt rígid, ja que una vegada gravades les plantilles i els valors, si es vol canviar o afegir una nova frase o variable s'ha de realitzar una nova gravació, amb totes les despeses que això comporta. Un altre problema d'aquest procediment és que, a vegades, encara que s'hagin gravat variables per a diferents contexts, poden presentar-se discontinuïtats en la corba melòdica.

Una solució de compromís consisteix en afegir frases portadores, amb variables incloses, dissenyades especialment per a l'aplicació al corpus general que es grava per extreure'n les unitats de síntesi. Així, l'algoritme del sistema de síntesi necessita realitzar menys concatenacions degut a que dins del corpus ja s'hi troben paraules o segments llargs que coincideixen amb les frases que genera el sistema de diàleg. No obstant, pot passar que el mòdul prosòdic assigni a alguna frase uns valors allunyats als de la frase portadora i que una part del segment d'aquesta frase sigui rebutjat per poder escollir altres segments amb més similitud prosòdica. Aquest comportament és inadequat, ja que si la frase es va gravar correctament, no hauria de refusar-se per les característiques prosòdiques que presenta. Per evitar-ho es pot realitzar una cerca en dos passos: primer es busca el grup fònic i llavors, només en el cas de que no s'hagi trobat, es busquen unitats sub-lèxiques. Si els segments que s'uneixen ho fan sense una pausa molt llarga, s'ha de suavitzar la corba de la freqüència fonamental en els extrems per evitar discontinuïtats. Aquest mètode permet que la qualitat s'apropi a la obtinguda amb les frases portadores gravades i que no es perdi la flexibilitat pròpia d'un convertidor de text a veu, ja que es poden modificar lleugerament els missatges o afegir-n'hi de nous, sense necessitat d'haver de gravar noves frases. Veiem doncs que amb un petit corpus d'aplicació (10-20 minuts) la qualitat pot augmentar significativament respecte al ús d'un corpus genèric.

### **3.6 Eines corresponents a la part de TTS que seran utilitzades en aquest projecte**

Per muntar la part corresponent a la síntesi de veu dins de la nostra aplicació s'utilitzarà Festival, una eina distribuïda com a programari lliure amb llicència de tipus X11 que ofereix un marc general per a la creació, sense cap tipus de restricció comercial, de sistemes de síntesi de veu, incloent-hi exemples, i que va ser desenvolupada originàriament pel Centre d'Investigació de Tecnologies del Llenguatge de la Universitat d'Edimburg.

Està programada en C++ i implementada com un intèrpret de comandes, fet que li permet connectar-se amb diversos mòduls i aplicacions d'una manera fàcil i eficaç. Es tracta d'una eina multilingüe, arribant a suportar una gran varietat d'idiomes com poden ser, entre d'altres, l'anglès (el més avançat), el castellà i fins i tot el català, gràcies a grups com el Centre de Tecnologies i Aplicacions del Llenguatge i la Parla (TALP), un centre específic de recerca interdepartamental de la Universitat Politècnica de Catalunya (UPC) que treballa en la recerca i desenvolupament de sistemes de síntesi de veu en la nostra llengua.

Un dels projectes que ha dut a terme aquest centre i que precisament s'utilitzarà en aquesta aplicació és el *FestCat* (*Festival parla català*). Aquest paquet consisteix en unes llibreries per tal d'estendre Festival i fer-lo parlar en català. Està format per dos components principals:

1. Dades lingüístiques i codi per estendre Festival al català: diccionaris, transcripció fonètica, etiquetador morfo-sintàctic...

Hi trobem dues carpetes:

**dicts/upc:** bàsicament diccionaris.

**upc\_catalan:** bàsicament codi.

2. Veus: dades dependents del locutor.

Hi ha una carpeta per cada veu:

**voices/catalan/upc\_ca\_'nom-locutor'**

# Capítol 4

## Estudi de viabilitat

### 4.1 Introducció

A continuació es presenta l'estudi de viabilitat del projecte *Tecnologies de la parla: Creació d'un punt d'informació*. En ell s'hi podran trobar els diferents punts bàsics que permetran concloure si el projecte és viable o no i per tant si cal o no començar-lo.

#### 4.1.1 Tipologia i paraules clau

**Tipologia:** Desenvolupament

**Paraules clau:** punt d'informació, ordinador, veu, turisme, ATK, HTK, Festival, ASR, TTS, síntesi, reconeixedor.

#### 4.1.2 Descripció

Són molts els municipis del país (sobretot municipis petits que no tenen els diners suficients com per permetre-s'ho) que no disposen de cap punt d'informació amb una persona treballant-hi per donar servei i contestar les preguntes i les curiositats que puguin tenir els visitants del poble respecte les festes, llocs per visitar, on menjar... Molts ajuntaments el que fan és respondre les qüestions a l'ajuntament mateix, però això suposa una limitació, ja que moltes cases consistorials a les tardes no obren i els caps de setmana, que és quan se solen produir les visites juntament amb dies festius, tampoc. És a dir, o tenen sort i els turistes troben l'ajuntament obert o es queden sense informació que els podria ser de gran interès a l'hora de dur a terme la visita turística i de conèixer en profunditat el municipi i els seus costums. Si que és veritat que hi ha panells d'informació repartits pel poble amb els llocs de més interès, una breu explicació de les festes, fotografies...



però potser no amb el detall suficient com per aconseguir al visitant o com per resoldre-l'hi totes les preguntes que es pugui fer. Tanmateix hi ha la possibilitat que el visitant només vulgui saber una determinada dada i que s'hagi de llegir tot el panell per trobar el que busca. En aquest cas hi trobarà informació extra que es podria estalviar de llegir.

Avui en dia en que les noves tecnologies es troben per tot arreu gràcies al fet de que no tenen un cost gaire elevat d'implementació i que hi ha un gran nombre de turistes que viatgen pel país disposats a descobrir nous indrets, el projecte Aplicació com a punt d'informació utilitzant la comunicació oral el que vol és satisfer les demandes d'aquests turistes, que ells preguntin el que vulguin saber sobre el municipi i que la màquina els contesti a través de la comunicació oral. Més que una necessitat (no és una màquina que faci arribar més visitants al poble ni és imprescindible, de fet no n'hi sol haver a cap poble) vindria a ser una oportunitat de crear un punt d'informació que faciliti als turistes la visita al municipi, que els respongui els dubtes/qüestions més en profunditat i sempre els contesti el que ells demanin, sense informació extra que no vulguin saber (això permetria solucionar el fet que l'usuari s'hagi de llegir, o en aquest cas escoltar, coses que no li interessin).

#### 4.1.3 Objectius del projecte

1. Crear una aplicació capaç d'establir comunicació oral amb els usuaris que la utilitzin. ***Crític***
2. Desenvolupar un software que gestioni el intercanvi d'informació (mitjançant la parla) entre els usuaris i la base de dades. ***Crític***
3. Muntar un punt d'informació disponible les 24 hores del dia els 365 dies de l'any. ***Crític***
4. Aconseguir que el punt d'informació pugui satisfer totes les demandes dels usuaris. ***Crític***
5. Estalviar a l'ajuntament la col·locació de panells informatius i la contractació d'una treballadora que estigui a un punt d'informació. ***Prioritari***
6. Millorar la forma en què l'usuari rep la informació. ***Prioritari***
7. Gestionar la introducció de noves dades que s'han preguntat i no s'han pogut respondre perquè la cerca a la base de dades ha sigut nul·la. ***Secundari***

#### 4.1.4 Definicions, acrònims i abreviacions

1. **HTK**: Hidden Markov Model Toolkit. Són un conjunt d'eines portables que permeten construir i manipular models de Markov ocults. HTK s'utilitza principalment per al reconeixement de veu.
2. **API**: Application Programming Interface (Interfície de Programació d'Aplicacions). És el conjunt de funcions i procediments (mètodes en la programació orientada a objectes) que ofereixen certes biblioteques i que poden ser utilitzats per un altre software com una capa d'abstracció, evitant feina al programador.
3. **ATK**: API for HTK. És una API dissenyada per facilitar la creació d'aplicacions experimentals de HTK.
4. **Client**: Ajuntament que compra l'aplicació informàtica i la instal·la al seu municipi.
5. **Interactuar**: Realitzar una pregunta al sistema informàtic i que aquest et respongui.
6. **Usuaris**: persones físiques (turistes o fins i tot gent del poble) que interaccionaran amb l'aplicació informàtica.
7. **PMI PMBOK**: Project Managment Body Of Knowledge. És una metodologia en l'administració de projectes.
8. **Microsoft Project**: És un software d'administració de projectes que ajuda en el desenvolupament de plans, en l'assignació de recursos a les tasques, en donar seguiment al progrés, en administrar el pressupost i en analitzar les càrregues de treball.
9. **WBS**: Work Breakdown Structure. És una estructura exhaustiva, jeràrquica i descendent (normalment en forma d'arbre) que descriu les fases, activitats i tasques del projecte. El seu propòsit és documentar l'abast del projecte i serveix com la base per a la planificació del projecte. Tot treball a realitzar en el projecte ha de provenir d'una o més entrades de la WBS.
10. **Milestone**: Són els punts de control del projecte.
11. **Diagrama de Gantt**: És un mètode gràfic de planificació i control en el que un projecte es divideix en diferents activitats i es realitzen estimacions sobre quant temps requereix cadascuna d'elles, així com el total de temps necessari per acabar la totalitat del projecte.

### 4.1.5 Parts interessades

#### Stakeholders:

La taula 4.1 mostra quines són les parts interessades en què aquest projecte es dugui a terme.

| Nom                | Descripció                | Responsabilitat                                                                                                                              |
|--------------------|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| A                  | Ajuntament                | Definició de requisits i funcionalitats, realització del seguiment del projecte i aprovació si és correcte i també aportació d'informació    |
| Xavier Closa Serra | Administrador del sistema | Participar en la definició de requisits, aportar coneixements i informació d'interès, aportar el punt de vista de l'usuari final del sistema |
| Montse Meneses     | Directora del projecte    | Supervisar la feina de l'alumne, avaluar el projecte, afegir-hi noves funcionalitats o limitar-les                                           |

Taula 4.1: Stakeholders

#### Perfils d'usuari del sistema:

Si observem la taula 4.2 podrem veure quins perfils d'usuari tindrà definits l'aplicació.

| Nom  | Perfil                    | Responsabilitat                                                |
|------|---------------------------|----------------------------------------------------------------|
| UA   | Administrador del sistema | Gestió i control del sistema i manteniment de la base de dades |
| User | Usuari                    | Consulta d'informació                                          |

Taula 4.2: Perfils d'usuari del sistema

**Equip de desenvolupament:**

La taula 4.3 defineix els rols de cada persona i explica la responsabilitat que tindran dins d'aquest projecte.

| Nom                | Descripció             | Responsabilitat                                                                                                                                                                  |
|--------------------|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Xavier Closa Serra | Cap del projecte       | Definir, gestionar, planificar i controlar el projecte, realitzar l'estudi de viabilitat i analitzar-ne les conclusions                                                          |
| Xavier Closa Serra | Analista               | Col·laboració en la realització de l'estudi de viabilitat i en la planificació del projecte, analitzar l'aplicació (metodologia, arquitectura...) i participar en el seu disseny |
| Xavier Closa Serra | Programador            | Participar també en el disseny de l'aplicació i programar-la seguint els criteris establerts. També realitzarà els documents necessaris (memòria...)                             |
| Xavier Closa Serra | Tècnic de proves       | Realitzar tota mena de proves per comprovar el correcte funcionament de l'aplicació. Control de qualitat del software creat                                                      |
| Montse Meneses     | Directora del projecte | Supervisar feina de l'alumne, que en aquest cas farà de cap de projecte, analista, programador i tècnic de proves                                                                |

Taula 4.3: Equip de desenvolupament

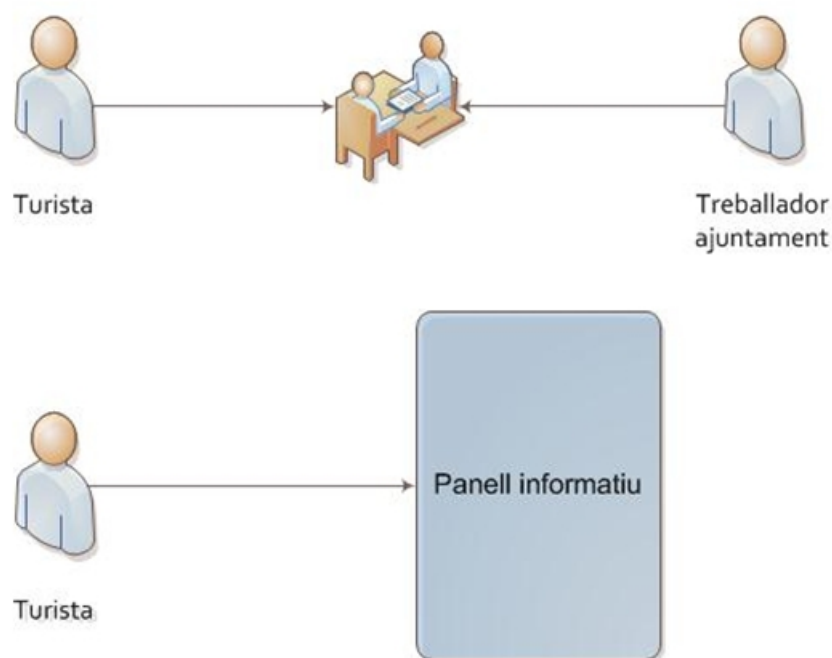


Figura 4.1: Situació actual

## 4.2 Estudi de la situació actual

### 4.2.1 Context

Actualment el municipi només disposa d'un punt on anar a demanar informació: les oficines de l'ajuntament. Aquestes només estan obertes en horari d'oficina, que és força reduït (tots els matins de la setmana i una tarda, amb els caps de setmana i festius tancat). Les persones que treballen en aquestes oficines compaginen la feina que els comporta gestionar l'ajuntament, realitzar els tràmits... amb la tasca de proporcionar informació a tot aquell que en demani. També hi ha repartits per diferents llocs de la població alguns panells amb informació bàsica sobre el municipi, acompanyada d'alguna fotografia.

### 4.2.2 Lògica del sistema

La situació actual del municipi queda descrita a la imatge 4.1

### 4.2.3 Descripció física

Actualment a les oficines de l'ajuntament hi ha, disponible per a tot el públic, tríptics amb informació detallada que és el que fan servir els/les treballadors/es per proporcionar la informació als turistes, juntament amb els coneixements del municipi dels quals disposa cadascú. No hi ha cap sistema físic creat expressament per donar servei d'informació a turistes.

### 4.2.4 Usuaris i/o personal del sistema

Bàsicament el personal i els usuaris que intervenen en el sistema actual es poden dividir en dos grups:

1. **Turistes:** persones que arriben al municipi amb la intenció de visitar-lo i van en busca d'informació que els faciliti aquesta visita.
2. **Oficinistes:** treballen a les oficines de l'ajuntament realitzant els tràmits, gestions... per contribuir al correcte funcionament d'aquest. A més, també proporcionen informació a tota aquella gent que s'acosti a la casa consistorial a demanar-ne.

### 4.2.5 Diagnòstic del sistema

#### Deficiències

1. El fet que els treballadors hagin de compaginar les tasques de gestionar l'ajuntament amb el proporcionar informació als turistes pot implicar dues coses: la primera és que si hi ha molta gent esperant per fer algun tràmit a l'ajuntament, el turista que arribi en busca d'informació s'haurà d'esperar estona abans no el puguin atendre. L'altre situació és que els/les treballadors/es hagin de parar de realitzar la feina que els correspon a l'ajuntament per aportar a un nouvingut tota la informació que demani. És la seva feina però aquest temps en què atenen dubtes sobre el municipi per part de turistes, és temps perdut pel que fa a la gestió del poble.
2. El sistema actual depèn molt de la feina que hi hagi a les oficines de l'ajuntament.
3. Com que el personal que treballa a les oficines de l'ajuntament (per molt que sigui nascut al poble o hi visqui des de fa molts anys) no està llicenciat en turisme, segurament no sigui el més adequat per aportar la informació necessària que cal aportar en un punt d'informació.

4. Hi ha la possibilitat que els treballadors no dominin el idioma del turista.
5. Els cartells que hi ha pel poble poden ser molt simples i no aportar tota la informació que el turista està interessat a saber sobre el municipi.
6. Horari d'atenció al públic limitat.

### Millores

1. Separar la feina de l'ajuntament del fet de proporcionar informació al turista.
2. Evitar que el turista s'hagi d'esperar per rebre informació o evitar que el fet de demanar informació faci perdre temps al personal de l'ajuntament.
3. Proporcionar la informació desitjada d'una forma completa sense dependre de cap horari: sigui l'hora que sigui de qualsevol dia hi ha d'haver el punt d'informació disposat a satisfer les necessitats dels usuaris.
4. Aportar més informació que la que hi ha impresa als cartells (en una base de dades hi cap molta més informació que no pas en un cartell, on l'espai per imprimir-hi text és reduït).

## 4.2.6 Normatives i legislació

1. Normativa de projectes Enginyeries Tècniques en Informàtica.

## 4.3 Requisits del sistema

### 4.3.1 Requisits funcionals

1. L'aplicació ha de proporcionar correctament tota la informació que els usuaris li demanin. ***Essencial***
2. Gestió i manteniment (introducció, modificació i eliminació) del contingut dels fitxers. ***Essencial***
3. Si no es troba la resposta d'una pregunta als fitxers, s'ha d'enviar un missatge al administrador del sistema perquè tan aviat com sigui possible la hi introdueixi. ***Opcional***

4. Si l'aplicació deixa de funcionar, s'ha d'enviar un missatge al administrador del sistema perquè tan aviat com sigui possible l'arregli.  
**Essencial**
5. Desar les preguntes i fer un estudi de quines són les més freqüents.  
**Opcional**
6. Poder observar l'històric de preguntes. **Opcional**
7. Sistema d'ajuda. **Opcional**
8. Controlar el tipus de preguntes fetes. **Opcional**

#### 4.3.2 Requisits no funcionals

1. Tolerància a preguntes mal fetes. **Essencial**
2. El temps de resposta ha de ser ràpid. **Essencial**
3. Ha de presentar una bona seguretat per evitar que persones alienes al administrador del sistema puguin modificar els fitxers de configuració, resposta... **Essencial**
4. Facilitat d'ús per a tots els públics, independentment de l'edat i sexe de la persona i del grau de coneixements que tinguin en informàtica.  
**Essencial**
5. Sistema fiable que tingui una bona recuperació davant d'errors.  
**Essencial**
6. Que en un futur pugui ser ampliable amb altres característiques.  
**Condiciona**

#### 4.3.3 Restriccions del sistema

1. L'aplicació s'ha d'implementar en un ordinador que tingui instal·lat un sistema operatiu Linux.
2. L'aplicació ha de ser creada tenint en compte el hardware sobre el qual serà implementada, aconseguint així un bon rendiment durant l'execució.
3. L'idioma utilitzat per l'aplicació ha de ser el català.
4. El projecte ha d'estar finalitzat abans del 19 de setembre del 2012.



### 4.3.4 Catalogació i priorització dels requisits

Dins la taula 4.4 es pot veure com es cataloguen i prioritzen els requisits funcionals. La 4.5 mostra la catalogació i priorització dels requisits no funcionals.

|    | RF1 | RF2 | RF3 | RF4 | RF5 | RF6 | RF7 | RF8 |
|----|-----|-----|-----|-----|-----|-----|-----|-----|
| O1 | X   |     |     |     |     |     |     |     |
| O2 | X   |     |     |     |     |     |     |     |
| O3 |     |     |     | X   |     |     |     |     |
| O4 |     |     | X   |     | X   | X   | X   |     |
| O5 | X   |     |     |     |     |     |     |     |
| O6 |     |     |     |     |     |     |     |     |
| O7 |     | X   | X   |     |     |     |     | X   |

Taula 4.4: Priorització de requisits funcionals

|    | RNF1 | RNF2 | RNF3 | RNF4 | RNF5 | RNF6 |
|----|------|------|------|------|------|------|
| O1 | X    |      |      | X    |      |      |
| O2 |      | X    |      | X    |      |      |
| O3 |      |      | X    |      | X    |      |
| O4 |      |      |      |      |      |      |
| O5 |      |      |      |      |      |      |
| O6 |      | X    |      | X    | X    |      |
| O7 |      |      |      |      |      |      |

Taula 4.5: Priorització de requisits no funcionals

## 4.4 Alternatives i selecció de la solució

### 4.4.1 Alternativa 1

#### Contractació d'una o més persones que atenguin al punt d'informació

Aquesta alternativa suposa tenir treballant a una persona 8 hores cada dia (dies laborables) més a dues altres que treballin els caps de setmana i festius (en torns diferents) per aconseguir que el punt d'informació romangui obert gairebé tot el temps.

*Costos d'aquesta alternativa:*

Hi ha 3 persones contractades, 1 que treballa entre setmana 8 hores i les altres dues que treballen els caps de setmana i festius 7 hores cada una en dos torns diferents. Això suposa un total de 3 sous a pagar, un de dia laborable i dos més de festius. Suposant que les hores de treball dels dies laborables es paguen a 6€ i les dels festius o caps de setmana a 8,5€, el cost derivat dels sous (suposant un mes normal sense gaires festius per exemple Juliol) es pot trobar a la taula 4.6.

| Preu/hora | Hores laborables per dia | Cost per dia            | Cost per mes                                                                    | Cost total                                    |
|-----------|--------------------------|-------------------------|---------------------------------------------------------------------------------|-----------------------------------------------|
| 6€/hora   | 8h                       | $6€ \cdot 8h = 48€$     | $48€ \text{ per dia} \cdot 22 \text{ dies laborables} = 1056€$                  | $1 \text{ treballador} \cdot 1056€ = 1056€$   |
| 8,5€/hora | 7h                       | $8,5€ \cdot 7h = 59,5€$ | $59,5€ \text{ per dia} \cdot 9 \text{ dies festius o caps de setmana} = 535,5€$ | $2 \text{ treballadors} \cdot 535,5€ = 1071€$ |
|           |                          |                         | Cost derivat dels sous                                                          | <b>2127€</b>                                  |

Taula 4.6: Cost mensual total de l'alternativa 1

**Avantatges:** El turista disposarà d'una bona i personalitzada atenció proporcionada per un ésser humà llicenciat en turisme i expert en la zona, que sempre resulta més atractiu que no pas tractar amb un ordinador. També permet l'opció de comunicar-se amb els treballadors en diferents idiomes, ja que se suposa que aquests necessiten saber idiomes per accedir a la feina i així poder respondre a turistes de qualsevol nacionalitat.

**Inconvenients:** Hi ha la limitació de què el punt d'informació romandrà tancat algunes hores durant la nit: els dies laborables ho estarà de 7 de la tarda a 11 del matí i els dies festius i caps de setmana ho estarà de 10 de la nit a 8 del matí. Mai, durant els 365 dies de l'any, hi haurà el punt d'informació obert les 24 hores del dia. Els costos mensuals són molt elevats (veure taula 4.6).

### 4.4.2 Alternativa 2

Desenvolupar una aplicació que permeti interaccionar amb un ordinador fent servir la veu

Aquesta alternativa consisteix en crear una aplicació que sigui capaç d'interaccionar amb un usuari fent servir la parla, és a dir, que l'usuari formuli la pregunta i l'aplicació l'assimili, en busqui la resposta i li contesti.

Es busca poder crear un punt d'informació a través de la parla que no requereixi tenir a cap persona física allà aportant la informació.

*Costos d'aquesta alternativa:*

El cost d'aquesta alternativa ve especificat en la planificació.

**Avantatges:** Al fer-se l'aplicació des de 0 es pot enfocar segons les necessitats i desitjos del client (ajuntament en aquest cas). Només suposa invertir diners al principi, a l'hora de crear l'aplicació, llavors un cop muntat el punt d'informació no hi ha cap més cost apart de la corrent on s'endollarà l'ordinador. Estarà disponible les 24 hores del dia els 365 dies de l'any. Inclou suport i contracte de manteniment.

**Inconvenients:** Parlar amb una màquina no és el mateix que fer-ho amb una persona humana. El tracte és diferent, ara bé, la informació que pot proporcionar la màquina és idèntica a la que pot facilitar-te la persona. Aquesta versió del projecte només funciona amb català, un inconvenient a l'hora de que la puguin utilitzar turistes estrangers.

### 4.4.3 Solució proposada

Com es pot observar s'han proposat dues alternatives que permeten millorar la forma actual de resoldre el problema descrit. Totes dues tenen els seus avantatges i els seus inconvenients i si s'analitzen en profunditat pensant en quina és la que millor s'ajusta a les necessitats de l'ajuntament a l'hora de crear un punt d'informació es pot anunciar que: l'alternativa 2 serà l'escollida.

Justificació de l'elecció: s'ha escollit aquesta alternativa ja que és la que té un cost més baix (a la llarga) i a més s'ajusta al que el client vol, que és tenir un punt d'informació disponible les 24 hores del dia els 365 dies de l'any. El fet de que es creï des de 0 basant-se en els requisits del client també ha ajudat en la elecció.

## 4.5 Work Breakdown Structure (WBS)

### 4.5.1 Fases i activitats del projecte

La taula 4.7 mostra totes les fases per les quals passarà el projecte al llarg del desenvolupament juntament amb una breu descripció de cada una.

| Fases                  | Descripció                                                                                                                                                                                                               |
|------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Iniciació              | En aquesta fase cal consultar els projectes oferts pels professors, escollir-ne un, ficar-se en contacte amb el que serà el director del projecte per tal de realitzar l'assignació i finalment formalitzar la matrícula |
| Planificació           | En aquesta fase és on es realitzen l'Estudi de Viabilitat i el Pla del Projecte                                                                                                                                          |
| Anàlisi                | Aquí és on s'analitzen els requisits funcionals i no funcionals i també l'arquitectura del sistema                                                                                                                       |
| Disseny                | En aquesta fase és on es dissenya tot el software. Això implica establir com serà la base de dades, com es durà a terme el control i manteniment d'aquesta, com serà la interfície...                                    |
| Desenvolupament        | Fase de desenvolupament de l'aplicació                                                                                                                                                                                   |
| Test i proves          | Cal provar el sistema. Es realitzen tests per comprovar el correcte funcionament de l'aplicació abans d'implementar-la definitivament                                                                                    |
| Implantació            | L'aplicació s'instal·la al seu entorn real. Inclou la formació de l'usuari administrador                                                                                                                                 |
| Generació de documents | Fase de documentació. Aquí és on es realitza la memòria del projecte                                                                                                                                                     |
| Tancament del projecte | El director del projecte signa l'acceptació i tancament del projecte                                                                                                                                                     |
| Defensa del projecte   | Defensa del projecte davant la comissió avaluadora                                                                                                                                                                       |

Taula 4.7: Diferents fases del projecte

### 4.5.2 Diagrama WBS

La imatge 4.2 mostra el diagrama *Work Breakdown Structure* d'aquest projecte.

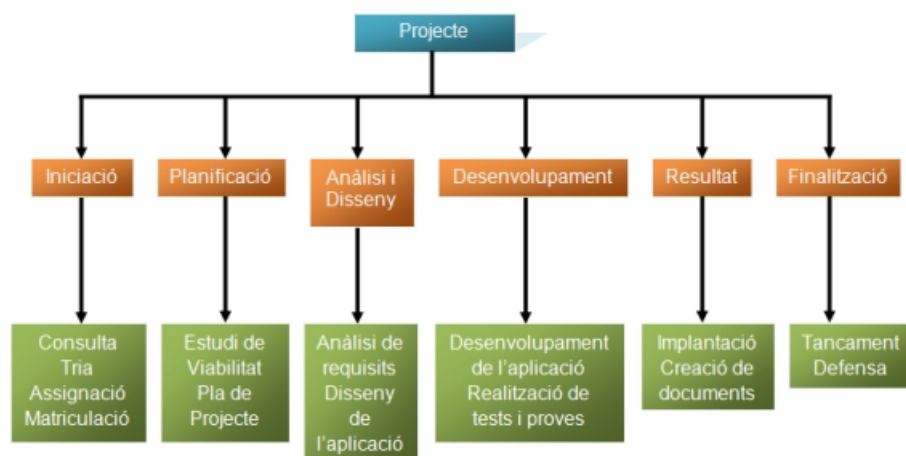


Figura 4.2: Diagrama WBS

### 4.5.3 Milestones

A la taula 4.8 s'hi poden trobar les fites juntament amb la seva descripció i la *due date* corresponent.

| Nom                    | Descripció                             | Data       |
|------------------------|----------------------------------------|------------|
| Iniciació_1            | Tria                                   | 29/09/2011 |
| Iniciació_2            | Assignació                             | 13/10/2011 |
| Iniciació_3            | Matriculació                           | 28/10/2011 |
| Estudi de viabilitat   | Aprovació i lliurament de documentació | 21/11/2011 |
| Pla del projecte       | Aprovació i lliurament de documentació | 13/12/2011 |
| Anàlisi                | Aprovació                              | 05/01/2012 |
| Disseny                | Aprovació                              | 31/01/2012 |
| Desenvolupament        | Assoliment d'un objectiu               | 14/05/2012 |
| Test i proves          | Assoliment d'un objectiu               | 31/05/2012 |
| Implantació            | Assoliment d'un objectiu               | 04/06/2012 |
| Generació de documents | Assoliment d'un objectiu               | 23/06/2012 |
| Tancament              | Acceptació                             | 19/09/2012 |
| Defensa                | Avaluació                              | 27/09/2012 |

Taula 4.8: Milestones

## 4.6 Recursos del projecte

### 4.6.1 Recursos del projecte

Els recursos que es destinaran a elaborar aquest projecte poden ser dividits en dos grups: els humans i els materials (equipament). A la taula 4.9 s'hi poden veure definits els primers i a la 4.10 els segons.

| Recursos humans      | Valoració   |
|----------------------|-------------|
| Cap de projecte      | 50 € / hora |
| Analista             | 40 € / hora |
| Programador          | 35 € / hora |
| Tècnic de proves     | 15 € / hora |
| Director de projecte | 75 € / hora |

Taula 4.9: Recursos humans

| Recursos materials i equipament                                                                                                                                                                                                                                                                                                                                                                                             | Valoració |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| El projecte es durà a terme entre l'escola i la casa particular del cap de projecte, analista, programador i tècnic de proves, per tant no hi ha cap cost d'equipament. S'utilitzaran els recursos disponibles que hi ha en aquests dos equipaments, cosa que evitarà una despesa de material. El software utilitzat serà lliure i l'aplicació s'implantarà sobre un ordinador que l'ajuntament ja té reservat expressament | 0 €       |

Taula 4.10: Recursos materials i equipament

### 4.6.2 Calendari dels recursos

Els recursos humans s'utilitzaran al llarg de tot el projecte. Cadascun intervindrà durant el temps en que s'estigui duent a terme la fase que li pertoca realitzar.

A continuació es mostra la relació entre recurs humà i fase del projecte en la qual intervé:

1. **Cap de projecte:** Intervé en la iniciació, la planificació, ajuda també en l'anàlisi, genera els documents i finalment és l'encarregat de defensar el projecte. També coordina i planifica els milestones.

2. **Analista:** Treballa en la planificació, l'anàlisi i participa en el disseny.
3. **Programador:** A part de participar en el disseny, treballa en el desenvolupament i la implantació.
4. **Tècnic de proves:** Intervé en la fase de test i proves, en una fase del disseny i en la implantació.
5. **Director de projecte:** Majoritàriament pren part en la iniciació i en el tancament del projecte, podent intervenir també en qualsevol fase per aportar alguna modificació o supervisar la feina feta.

L'equipament estarà present des del inici fins a la finalització del projecte, ja que és el lloc on es durà a terme aquest.

Els recursos materials començaran a intervenir a partir de la planificació i llavors, en les fases de desenvolupament, test i proves, implantació i generació de documents tindran un paper important per a poder dur a terme el projecte correctament.

## 4.7 Calendari del projecte

El projecte es realitzarà entre els mesos de setembre de 2011 i setembre de 2012, tenint la defensa d'aquest a finals del mes de setembre de 2012.

Es treballarà 2 hores diàries en el projecte, de dilluns a dissabte, el que suposarà un total de 12 hores de dedicació setmanal.

**Data començament:** 29 de setembre de 2011.

**Data finalització:** 27 de setembre de 2012.

**Eines de planificació i control:** Microsoft Project.

### 4.7.1 Dependències

Totes les fases es desenvolupen utilitzant un model lineal, cosa que vol dir que cap fase es començarà sense que s'hagi completat l'anterior, tal i com es pot observar en la taula 4.11.

| Fase                   | Dependència            |
|------------------------|------------------------|
| Iniciació              | -                      |
| Planificació           | Iniciació              |
| Anàlisi                | Planificació           |
| Disseny                | Anàlisi                |
| Desenvolupament        | Disseny                |
| Test i proves          | Desenvolupament        |
| Implantació            | Test i proves          |
| Generació de documents | Implantació            |
| Tancament del projecte | Generació de documents |
| Defensa del projecte   | Tancament del projecte |

Taula 4.11: Dependències entre les diferents fases del projecte

## 4.7.2 Quadre de tasques del projecte

La imatge 4.3 mostra amb detall totes les tasques del projecte, especificant el nom de cadascuna, la duració total (amb data d'inici i fi), les dependències, quin cost tenen i els recursos destinats per dur-les a terme.

| Nombre de tarea                                                    | Duración  | Comienzo     | Fin          | Predecesoras | Costo       | Nombres de los recursos                            |
|--------------------------------------------------------------------|-----------|--------------|--------------|--------------|-------------|----------------------------------------------------|
| 1 - <b>Tecnologies de la parla: Creació d'un punt d'informació</b> | 212 dies? | jue 29/09/11 | jue 27/09/12 |              | 13.813,00 € |                                                    |
| 2 - <b>Iniciació</b>                                               | 23 dies?  | jue 29/09/11 | vie 28/10/11 |              | 326,00 €    |                                                    |
| 3 Tria                                                             | 2 horas?  | jue 29/09/11 | jue 29/09/11 |              | 100,00 €    | Cap de projecte                                    |
| 4 Assignació                                                       | 1 dia?    | jue 13/10/11 | jue 13/10/11 | 3            | 125,00 €    | Cap de projecte[50%]Director de projecte[50%]      |
| 5 Matriculació                                                     | 1 dia?    | vie 28/10/11 | vie 28/10/11 | 4            | 100,00 €    | Cap de projecte                                    |
| 6 - <b>Planificació</b>                                            | 33 dies?  | mié 02/11/11 | mar 13/12/11 | 2            | 3.063,00 €  |                                                    |
| 7 Estudi de Viabilitat                                             | 16 dias   | mié 02/11/11 | sáb 18/11/11 | 5            | 1.488,00 €  | Cap de projecte[85%]Analista[35%]                  |
| 8 Aprobació Estudi de Viabilitat (Checkpoint)                      | 1 dia?    | lun 21/11/11 | lun 21/11/11 | 7            | 100,00 €    | Cap de projecte                                    |
| 9 Pla de Projecte                                                  | 15 dias   | mar 22/11/11 | lun 12/12/11 | 8            | 1.395,00 €  | Cap de projecte[85%]Analista[35%]                  |
| 10 Aprobació Pla de Projecte (Checkpoint)                          | 1 dia?    | mar 13/12/11 | mar 13/12/11 | 9            | 100,00 €    | Cap de projecte                                    |
| 11 - <b>Anàlisi</b>                                                | 17 dies?  | mié 14/12/11 | jue 06/01/12 | 6            | 1.428,00 €  |                                                    |
| 12 Anàlisi de requisits (casos d'ús)                               | 5 dias?   | mié 14/12/11 | lun 19/12/11 | 10           | 420,00 €    | Cap de projecte[20%]Analista[80%]                  |
| 13 Anàlisi de dades                                                | 5 dias?   | mar 20/12/11 | mar 27/12/11 | 12           | 420,00 €    | Cap de projecte[20%]Analista[80%]                  |
| 14 Anàlisi de la seguretat i de la legalitat                       | 2 dias?   | mié 28/12/11 | jue 29/12/11 | 13           | 168,00 €    | Cap de projecte[20%]Analista[80%]                  |
| 15 Documentació de l'anàlisi                                       | 4 dias?   | vie 30/12/11 | mié 04/01/12 | 14           | 320,00 €    | Analista                                           |
| 16 Aprobació de l'anàlisi (Checkpoint)                             | 1 dia     | jue 05/01/12 | jue 05/01/12 | 15           | 100,00 €    | Cap de projecte                                    |
| 17 - <b>Disseny</b>                                                | 20 dies?  | jue 09/01/12 | mar 21/01/12 | 11           | 1.483,00 €  |                                                    |
| 18 Disseny de les dades                                            | 5 dias?   | lun 09/01/12 | vie 13/01/12 | 16           | 375,00 €    | Analista[50%]Programador[50%]                      |
| 19 Disseny modular de l'aplicació                                  | 3 dias?   | sáb 14/01/12 | mar 17/01/12 | 18           | 225,00 €    | Analista[50%]Programador[50%]                      |
| 20 Disseny de la interfície                                        | 4 dias?   | mié 18/01/12 | sáb 21/01/12 | 19           | 300,00 €    | Analista[50%]Programador[50%]                      |
| 21 Disseny de les proves                                           | 2 dias?   | lun 23/01/12 | mar 24/01/12 | 20           | 118,00 €    | Analista[50%]Tècnic de proves[40%]Programador[10%] |
| 22 Documentació del disseny                                        | 5 dias?   | mié 25/01/12 | lun 30/01/12 | 21           | 375,00 €    | Analista[50%]Programador[50%]                      |
| 23 Aprobació del disseny (Checkpoint)                              | 1 dia?    | mar 31/01/12 | mar 31/01/12 | 22           | 100,00 €    | Cap de projecte                                    |
| 24 - <b>Desenvolupament</b>                                        | 75 dies?  | mié 01/02/12 | lun 14/06/12 | 17           | 5.250,00 €  |                                                    |
| 25 Preparació entorn de desenvolupament                            | 2 dias?   | mié 01/02/12 | jue 02/02/12 | 23           | 140,00 €    | Programador                                        |
| 26 Configuració de les dades                                       | 5 dias?   | vie 03/02/12 | mié 15/02/12 | 25           | 350,00 €    | Programador                                        |
| 27 Desenvolupament de l'aplicació                                  | 68 dias?  | jue 16/02/12 | lun 14/05/12 | 26           | 4.760,00 €  | Programador                                        |
| 28 - <b>Test i proves</b>                                          | 15 dies?  | mar 15/05/12 | jue 31/05/12 | 24           | 520,00 €    |                                                    |
| 29 Proves unitàries                                                | 4 dias?   | mar 15/05/12 | vie 18/05/12 | 27           | 120,00 €    | Tècnic de proves                                   |
| 30 Proves d'integració                                             | 4 dias?   | sáb 19/05/12 | mié 23/05/12 | 29           | 120,00 €    | Tècnic de proves                                   |
| 31 Proves d'estrènia (incidències i racos)                         | 2 dias?   | jue 24/05/12 | vie 25/05/12 | 30           | 60,00 €     | Tècnic de proves                                   |
| 32 Documentació de desenvolupament i test                          | 4 dias?   | sáb 26/05/12 | mié 30/05/12 | 31           | 120,00 €    | Tècnic de proves                                   |
| 33 Aprobació del desenvolupament i proves (Checkpoint)             | 1 dia?    | jue 31/05/12 | jue 31/05/12 | 32           | 100,00 €    | Cap de projecte                                    |
| 34 - <b>Implantació</b>                                            | 3 dias?   | vie 01/06/12 | lun 04/06/12 | 28           | 174,00 €    |                                                    |
| 35 Instal·lació i proves reals                                     | 3 dias?   | vie 01/06/12 | lun 04/06/12 | 33           | 174,00 €    | Programador[70%]Tècnic de proves[30%]              |
| 36 - <b>Generació de documents</b>                                 | 13 dias?  | sáb 09/06/12 | sáb 23/06/12 | 34           | 1.300,00 €  |                                                    |
| 37 Creació de la memòria                                           | 13 dias?  | sáb 09/06/12 | sáb 23/06/12 | 35           | 1.300,00 €  | Cap de projecte                                    |
| 38 Tancament del projecte                                          | 1 dia?    | mié 19/09/12 | mié 19/09/12 | 36           | 140,00 €    | Cap de projecte[20%]Director de projecte[80%]      |
| 39 Defensa del projecte                                            | 1 dia?    | jue 27/09/12 | jue 27/09/12 | 38           | 100,00 €    | Cap de projecte                                    |

Figura 4.3: Quadre de tasques del projecte

El total de dies que durarà el desenvolupament del PFC és de 212. Per



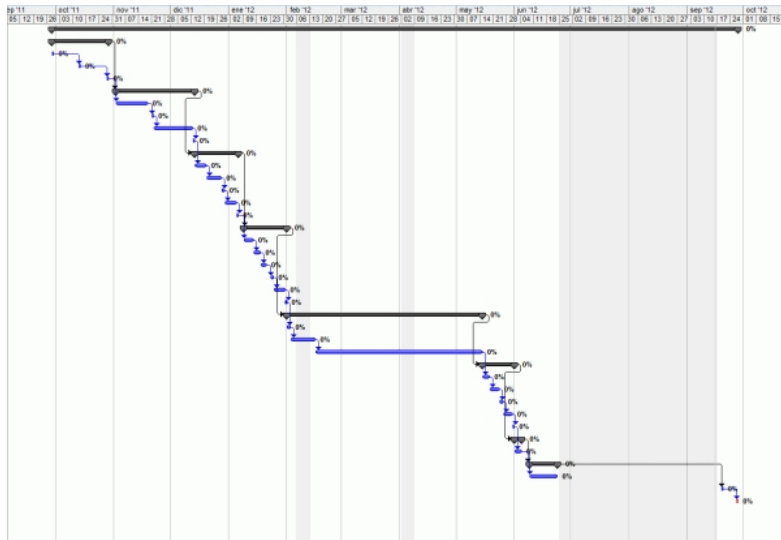


Figura 4.4: Calendari temporal del projecte

tant, si s'hi treballa durant 2 hores cada dia obtenim que el total d'hores dedicades al projecte és de **424**.

### 4.7.3 Calendari temporal

A la figura 4.4 es pot apreciar el calendari d'execució del projecte. En ell s'hi pot veure quan comença i quan dura una tasca juntament amb les dependències que la limiten (fins que no s'acaba la seva predecessora aquesta no es pot començar). Això permet veure amb claredat que si alguna tasca es retarda, el calendari del projecte es pot veure afectat negativament.

## 4.8 Avaluació de riscos

A continuació es procedeix a avaluar els riscos que hi ha o que poden sorgir durant el desenvolupament del projecte. El que es pretén és identificar-los i proposar-hi una solució abans de que surtin a la llum, o si això no és possible, tenir preparada una solució per aplicar el més aviat possible, ja que l'únic que fan és retardar el desenvolupament.

### 4.8.1 Llista de riscos

La taula 4.12 defineix tots els riscos amb els quals ens podem trobar durant l'execució d'aquest projecte, n'estableix l'origen i explica, a més, els efectes

| Risc | Descripció                                         | Origen                                         | Efectes que genera                                                                                                                                                                                                   |
|------|----------------------------------------------------|------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| r1   | Planificació temporal optimista                    | Pla de projecte                                | Alguna tasca programada o fins i tot el projecte no s'acaba en la data prevista                                                                                                                                      |
| r2   | Manca alguna tasca necessària                      | Pla de projecte                                | No es compleixen els objectius marcats del projecte                                                                                                                                                                  |
| r3   | Pressupost poc ajustat                             | Pla de projecte                                | O s'inverteixen més diners per aconseguir el que es tenia previst fer (per tant hi ha un augment de cost) o s'obté un producte de menys qualitat del que s'esperava                                                  |
| r4   | Canvi de requisits                                 | Estudi de viabilitat, anàlisi                  | Reestructurar els requisits, comportant això un endarreriment en els terminis establerts                                                                                                                             |
| r5   | Equip del projecte massa reduït                    | Pla de projecte                                | Pot produir-se un endarreriment en la finalització del projecte, una sobreassignació de recursos per acabar el projecte dins del límit establert o que no es compleixin els objectius proposats                      |
| r6   | Eines de desenvolupament inadequades               | Desenvolupament                                | Poca qualitat del software creat, endarreriment en la finalització del projecte                                                                                                                                      |
| r7   | Dificultat per accedir als stakeholders            | Estudi de viabilitat, anàlisi, proves          | No hi ha prou informació sobre els requisits que es tenen, cosa que pot provocar que el client quedi insatisfet amb el producte final                                                                                |
| r8   | No es realitza correctament la fase de test        | Test proves, implantació                       | Manca de qualitat, errors inesperats, haver de tornar a programar l'aplicació (el que suposa una pèrdua de diners i de temps), que els usuaris no estiguin satisfets amb l'aplicació                                 |
| r9   | Manca d'adopció de mesures de seguretat            | Estudi de viabilitat, anàlisi, desenvolupament | Pot suposar pèrdues, tan d'informació com econòmiques                                                                                                                                                                |
| r10  | Abandonament del projecte abans de la finalització | Pot passar en qualsevol fase                   | Pèrdua de tota la feina feta fins al moment, el que també suposa una pèrdua de temps i diners i, a més, pot fer disminuir la confiança que els clients tenen cap a l'empresa encarregada de desenvolupar el projecte |

Taula 4.12: Possibles riscos

que hi genera.

## 4.8.2 Catalogació de riscos

La taula 4.13 cataloga els riscos, és a dir, per a cada possible risc determina quina probabilitat hi ha de que succeeixi i el grau d'impacte que té sobre el projecte si arriba a passar.

## 4.8.3 Pla de contingència

La taula 4.14 busca possibles solucions per evitar que aquests riscos arribin a afectar el projecte.

| Risc | Probabilitat | Impacte     |
|------|--------------|-------------|
| r1   | Alta         | Crític      |
| r2   | Alta         | Crític      |
| r3   | Mitjana      | Crític      |
| r4   | Alta         | Marginal    |
| r5   | Mitjana      | Crític      |
| r6   | Baixa        | Crític      |
| r7   | Mitjana      | Crític      |
| r8   | Mitjana      | Crític      |
| r9   | Mitjana      | Crític      |
| r10  | Baixa        | Catastròfic |

Taula 4.13: Afectació dels riscos al projecte

## 4.9 Pressupost

### 4.9.1 Estimació cost de personal

La taula 4.15 calcula els costos del personal que participarà en el desenvolupament del projecte.

### 4.9.2 Estimació cost de recursos

Si observem la taula 4.16 veurem que el total del cost corresponent a recursos és relativament baix. Això és degut a que només es conta el cost d'amortització, ja que pel desenvolupament es fan servir recursos que ja hi ha disponibles i no s'han de comprar expressament pel projecte. Del MSProject disposem d'una llicència gratuïta d'un any pel sol fet de ser estudiants del EUIS.

### 4.9.3 Estimació costos indirectes

En aquest pressupost no es té en compte el cost de l'energia que pot gastar l'ordinador mentre es realitza el projecte, com tampoc es té en compte el cost dels viatges des de casa fins a la universitat ni el cost que pot suposar el imprimir les memòries (fulls, tinta...).

### 4.9.4 Resum i anàlisi cost-benefici

Els costos totals del desenvolupament d'aquest projecte pugen a un total de:

| Risc | Solució                                                                                                                                                                 |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| r1   | - Augmentar els recursos que s'assignen a la tasca perquè es pugui acabar dins del termini previst<br>- Ajornar la feina per més endavant                               |
| r2   | - Reescriure el pla de projecte i repassar-lo bé per evitar que en un futur es trobi a faltar una altra tasca                                                           |
| r3   | - Renegociar el pressupost<br>- Afrontar possibles pèrdues<br>- Crear una assegurança                                                                                   |
| r4   | - Demanar un ajornament<br>- Comentar amb el client si és molt important introduir/eliminar el requisit o si de veritat cal modificar-lo (veure si aporta molts canvis) |
| r5   | - Contractar més personal                                                                                                                                               |
| r6   | - Comprar o crear eines que s'ajustin més al desenvolupament que s'ha de fer                                                                                            |
| r7   | - Fixar un calendari de reunions<br>- Si no és possible veure's en persona, fixar un mitjà de comunicació clar, útil i fàcil de fer servir per establir contacte        |
| r8   | - Dissenyar bé els tests<br>- Realitzar tests automàtics<br>- Donar garanties                                                                                           |
| r9   | - Crear assegurança<br>- Aplicar polítiques de seguretat actives                                                                                                        |
| r10  | - No té solució                                                                                                                                                         |

Taula 4.14: Pla de contingència

|                          |                    |
|--------------------------|--------------------|
| Cost de personal .....   | 13.813 €           |
| Cost dels recursos ..... | 181,57 €           |
| <b>TOTAL .....</b>       | <b>13.994,57 €</b> |

És un cost força elevat, el que suposa una inversió inicial alta, però amb el temps s'acaba amortitzant i surt rentable.

Tots els diners que s'inverteixen en crear aquesta aplicació és temps que es guanya a les oficines de l'ajuntament evitant haver de donar informació als turistes, que pot ser aprofitat per gestionar el municipi, que és un guany per a l'ajuntament i el poble.

| Recurs humà           | Hores treballades | Cost            |
|-----------------------|-------------------|-----------------|
| Cap de projecte       | 88,5 hores        | 4.425 €         |
| Analista              | 67,9 hores        | 2.716 €         |
| Programador           | 171,6 hores       | 6.006 €         |
| Tècnic de proves      | 31,4 hores        | 471 €           |
| Director del projecte | 2,6 hores         | 195 €           |
| <b>Total</b>          |                   | <b>13.813 €</b> |

Taula 4.15: Cost total del personal

| Recurs                      | Cost amortització | Cost unitari | Període amortització | Període utilització |
|-----------------------------|-------------------|--------------|----------------------|---------------------|
| Amortització PC programador | 158,4 €           | 950 €        | 60 mesos (5 anys)    | 10 mesos            |
| Amortització MSOffice       | 23,17 €           | 139 €        | 60 mesos (5 anys)    | 10 mesos            |
| Amortització MSPProject     | 0 €               | 0 €          | 12 mesos             | 10 mesos            |
| <b>Total</b>                | <b>181,57 €</b>   |              |                      |                     |

Taula 4.16: Cost total dels recursos

## 4.10 Conclusions finals del estudi de viabilitat

### Beneficis

1. Aconseguir que el municipi disposi d'un punt d'informació disponible les 24 hores del dia els 365 dies de l'any
2. Ajudar als turistes a que la seva visita a la població sigui més entretinguda, completa i divertida.
3. Reducció del temps perdut pels treballadors de l'ajuntament en donar informació als nouvinguts.
4. Bona comunicació entre usuari i màquina.

Inconvenients:

1. Inversió al comprar l'aplicació, però són diners que amb el temps s'acaben amortitzant.
2. Algun turista pot no veure bé el fet de demanar informació a una màquina.
3. Necessitat de protegir la màquina de vàndals que la puguin destrossar.

Ara que s'ha arribat al final de l'estudi de viabilitat, podem assegurar, tot mirant el document i les diferents parts de les quals consta, com per exemple els beneficis que presenta l'alternativa escollida, els objectius que es poden assolir, les millores que suposarà respecte la situació actual..., que el projecte **és viable, és rendible** i que per tant ja podem **iniciar-lo**.

**Estudi de Viabilitat aprovat en data 13 de desembre de 2011**

# Capítol 5

## Disseny de l'aplicació

En aquest capítol, tal i com deixa entreveure el títol, s'explicarà com s'ha creat, des de zero, l'aplicació que dóna lloc a tota aquesta memòria. Pas a pas es podrà observar com tot el què s'ha explicat en les seccions anteriors és ficat dins del nostre sistema de diàleg persona-màquina, començant amb el disseny del diàleg, continuant amb la part del reconeixement de veu i acabant finalment amb la part de la síntesi. De cada un d'aquests dos mètodes es veurà quines eines s'han utilitzat, quina és la seva funcionalitat dins del programa, com s'ha pensat i construït, quines conseqüències comporta...

### 5.1 Diàleg

El punt d'informació a crear creixerà sobre la base del diàleg que el turista mantindrà amb la màquina per demanar i rebre totes les dades que ell desitgi del municipi que visita. Així doncs, és una part molt important d'aquest projecte i, per tant, s'ha de dissenyar amb molta cura, pensant bé quines dades podrà oferir la màquina, com es poden agrupar per simplificar les coses a l'usuari final, quines frases podrà utilitzar el visitant (les que no estiguin dins del diàleg l'aplicació no les reconeixerà), qui ha de portar la iniciativa de la conversa: la màquina o l'usuari, s'ha de definir si es pot interrompre a la màquina mentre xarra o bé fer que aquesta no "sentí" res fins que hagi acabat la seva explicació i un llarg etcètera de successos que convé tenir ben estudiats i lligats.

L'aplicació es dissenya pensant que estarà operativa les vint-i-quatre hores del dia els 365 dies de l'any, per tant sempre haurà d'estar engegada, però això no impedeix que la màquina, quan no tingui activitat, quedi en un estat d'espera per evitar que qualsevol soroll sigui interpretat com a paraula i comenci a funcionar automàticament sense tenir a ningú a davant. Així

doncs, quan arribi un turista caldrà que, mitjançant certes paraules, l'activi. Aquestes només poden ser:

1. *Hola*
2. *Bon dia*
3. *Bona tarda*
4. *Bona nit*

Només es permeten quatre opcions per mirar de limitar una mica el diàleg que l'usuari pot tenir amb la màquina, ja que si no es limita aquest camp podria arribar a ser molt extens, fet que complica la gestió de l'aplicació. Aquestes quatre possibilitats s'han escollit d'entre una llarga llista d'opcions després d'observar i demanar quines són les paraules o frases més comunes que un turista pronuncia quan arriba a un punt d'informació turística i vol reclamar l'atenció de la persona que hi ha treballant per començar a demanar-li la informació.

Un cop l'usuari hagi fet el primer pas, la màquina, amablement, li tornarà la salutació i li donarà la benvinguda al poble. Seguidament li oferirà totes les categories d'informació disponibles perquè el turista pugui escollir la que més s'escaigui al què vulgui saber. Aquestes categories inclouen:

1. *Informació general sobre el poble*
2. *Informació sobre la cultura i les festes del municipi*
3. *Conèixer els llocs d'interès que hi ha a la vila*
4. *Llocs on menjar*
5. *Llocs on dormir*
6. *Informació sobre serveis i equipaments del municipi*
7. *Informació esportiva*

Aquestes agrupacions són a un nivell molt alt, però ja aporten prou informació per permetre a l'usuari anar descartant opcions i començar-se a encaminar cap a una direcció o una altra. A dins de cadascuna d'elles hi ha varies subcategories, excepte en els llocs on menjar o dormir, que si l'usuari ho escull, directament rebrà la informació.

Tal i com s'ha comentat, hi ha una sèrie de frases úniques que són les que l'usuari haurà d'utilitzar en aquesta conversa amb la màquina per demanar



l'informació. Com que el turista no coneix aquesta condició de que només pot pronunciar unes determinades frases perquè la màquina l'entengui, s'han pensat les més comunes que se solen utilitzar per demanar informació a qualsevol lloc i se'n han escrit forces perquè l'usuari no trobi problemes parlant normal, tal i com ho faria davant d'una persona. Entre aquestes frases hi trobem:

1. *(Escullo / Trio) [la opció] [l'entitat esportiva de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
2. *[(Escullo / Trio / M'interessa)] la categoria ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
3. *Explica'm ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
4. *M'agradaria saber coses (de / del / sobre) [l'equip de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
5. *(M'agradaria / Vull) (obtenir / rebre) ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
6. *Informació [general] sobre [la] [(els / les)] [l'entitat esportiva de] [l'equip de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
7. *Quines i quan són ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
8. *[(M'agradaria / Vull)] saber [quins] [quines] [de] [quina] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*

9. [(M'agradaria / Vull / Escullo / Trio)] conèixer [quines són i quan se celebren] [(els / les)] [l'entitat esportiva de] [l'equip de] [equips esportius de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]
10. (M'agradaria / Vull) que m'informis (de / del / sobre) (els / les) [(de / del / sobre)] [de la categoria] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]
11. [(M'agradaria / Vull) visitar [(els / les)] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]
12. M'agraden (els / les) ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]
13. M'agrada l'època medieval informa'm sobre ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]
14. M'interessen més (els / les) ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]
15. [(M'agrada el tèxtil / M'agrada la mineria)] quines ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]
16. On [em] puc [anar a] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]
17. M'interessen [el] [els] [de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]

18. *Anomena'm els [de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
19. *[M'agradaria saber] [de] quins [serveis] [i] [equipaments] [de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
20. *Equipaments i serveis ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
21. *Serveis i equipaments ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
22. *[(M'agradaria / Vull)] (informar-me / saber coses / rebre informació / que m'informés) (de / del / sobre) [(els / les)] [l'entitat esportiva de] [l'equip de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
23. *[(M'agradaria / Vull)] [(obtenir / rebre)] informació (de / del / sobre) [(els / les)] [l'entitat esportiva de] [l'equip de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
24. *Dóna'm informació (de / del / sobre) [(els / les)] [l'entitat esportiva de] [l'equip de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
25. *(Parla'm / Informa'm) (de / del / sobre) [(els / les)] [l'entitat esportiva de] [l'equip de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*
26. *[De] quins [equips esportius de] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*

27. *M'agrada el ...*

28. *De l'equip de ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*

29. *Del de ...*

Cada un d'aquests vint-i-nou punts són frases que poden ser formades mitjançant la combinació de moltes parts, és a dir, no només hi ha vint-i-nou frases possibles a pronunciar, sinó que gràcies a totes les possibilitats, l'usuari no ha de trobar cap problema a l'hora de fer-se entendre. Anem a veure què significa cada cosa: les paraules o sentències que es troben dins de "[]" són paraules/frases opcionals, és a dir, que l'usuari tant les pot dir com no (la màquina si són dites les processa i si no continua endavant sense cap mena de problema). El que hi ha dins dels parèntesi són paraules a escollir (separades per /), és a dir, que l'usuari té la possibilitat d'utilitzar qualsevol de les que hi ha dins i la màquina l'entendrà. Si a fora del parèntesi no hi ha cap [] (que ja hem vist què significa) l'usuari haurà de dir per força alguna d'aquestes opcions perquè la màquina l'entengui. Les frases que hi ha escrites sense cap signe són d'obligada pronunciació. Ja per acabar on hi ha els punts suspensius és a on l'usuari haurà de pronunciar o bé una de les categories descrites anteriorment o bé alguna paraula que hi tingui a veure. Observem que la frase número 8:

*[(M'agradaria / Vull)] saber [quins] [quines] [de] [quina] ... [més importants] [que hi ha] [se celebren] [(té / disposa)] [hi ha per] [que] [puc] [visitar] [teniu aquí] [hi ha aquí] [a prop] [local] [(de / del / a / al / el / en aquest) (Balsareny / municipi)]*

pot crear, entre moltes d'altres, les següents frases:

1. *M'agradaria saber de la història del municipi*
2. *M'agradaria saber quina cultura té el municipi*
3. *M'agradaria saber quines festes se celebren en aquest municipi*
4. *M'agradaria saber quins llocs d'interès hi ha per Balsareny*
5. *M'agradaria saber quins llocs d'interès puc visitar al municipi*
6. *M'agradaria saber quins llocs d'interès té el municipi*
7. *M'agradaria saber quins llocs d'interès teniu aquí al municipi*

8. *M'agradaria saber quins castells teniu aquí a Balsareny*
9. *M'agradaria saber quines esglésies puc visitar en aquest municipi*
10. *M'agradaria saber quines colònies hi ha al municipi*
11. *M'agradaria saber quins llocs per menjar hi ha aquí a prop de Balsareny*
12. *M'agradaria saber quins llocs per dormir té el municipi*
13. *M'agradaria saber quins serveis i equipaments teniu aquí en aquest municipi*
14. *M'agradaria saber quines entitats esportives hi ha per Balsareny*
15. *Vull saber quina cultura local té Balsareny*
16. *Vull saber quines festes se celebren a Balsareny*
17. *Vull saber quins llocs d'interès puc visitar de Balsareny*
18. *Vull saber quins llocs per menjar té Balsareny*
19. *Vull saber quins serveis i equipaments hi ha aquí al municipi*
20. *Vull saber quines entitats esportives hi ha aquí al municipi*
21. ...

Això, encara que només és un petit recull de la gran quantitat de locucions que es poden formar amb un dels punts de la llista, ja mostra la gran varietat de frases disponibles que té el turista per fer funcionar aquesta aplicació.

Ara que hem vist les categories globals que hi ha i com l'usuari pot endinsar-se en elles, anem a veure amb detall les subcategories de cada una. Si el desig del turista és conèixer amb tota mena de detall el municipi i, per tant, escull la primera opció, es trobarà que la màquina li oferirà les següents subcategories per tal de guiar-lo cap a un tipus d'informació o cap a un altre:

1. *Informació sobre l'ajuntament*
2. *Informació sobre el municipi*
3. *Conèixer la història del municipi*

La informació sobre l'ajuntament fa referència a la casa consistorial, com molt bé indica el nom (s'ha intentat escollir noms o frases que siguin entenedores, és a dir, que l'usuari només sentir-les ja sàpiga quin tipus d'informació podrà trobar-hi dins), i explicarà al turista qui és l'alcalde, a quin partit polític pertany, número de regidors que va treure a les últimes eleccions, informarà sobre la resta de partits polítics (tant si són els que governen com els que estan a l'oposició), mostrarà l'organigrama de l'ajuntament, on es troba localitzat, el telèfon...

Escollir informació sobre el municipi permetrà al turista rebre explicacions de a on es troba geogràficament situat aquest, quanta superfície ocupa, a quina altitud està, en què es basa la seva economia, quin tipus de vegetació hi ha, quants habitants hi ha actualment censats al poble, quin és el seu gentilici... L'últim apartat aportarà al usuari coneixements sobre la història del poble, des dels seus inicis (primeres restes trobades que fan referència al municipi) fins a la actualitat, passant per com ha anat evolucionant la població, l'economia, quines guerres hi han fet estralls, quines construccions s'han fet i quin era el seu propòsit, s'anomenen algunes efemèrides...

Si el turista el què vol és rebre informació sobre la cultura i les festes del municipi escollirà la segona opció i, seguidament, l'aplicació li deixarà concretar un pèl més:

1. *Conèixer quines són i quan se celebren les festes més importants del poble*
2. *Conèixer la cultura del municipi*

Com és lògic, si escull festes la màquina li enumerarà cada una de les diades festives que tenen lloc al municipi, posant un especial èmfasi en la data de celebració, en honor a què o a qui se celebra, un breu repàs a la seva història i evolució al llarg dels anys i com no, una bona explicació dels actes característics que tenen lloc durant la festa i de les activitats que es duen a terme.

Si l'opció triada no és festes vol dir que l'usuari el què vol és conèixer la cultura del municipi, així que la màquina li proporcionarà una explicació del ric patrimoni cultural que té el poble. El turista podrà descobrir de primera mà que aquí hi ha dos balls típics, el de bastons i el de la faixa. De cadascun d'ells n'aprendrà quan es representa, quin és el seu origen, quina és la vestimenta típica que porten els dansaires i una mica d'història. El visitant descobrirà també que al poble s'hi representen els Pastorets des d'un munt de temps enrere (lloc de representació, dates...). Ja per acabar s'informarà amb tota mena de detalls (com pot ser la data de construcció, qui els va construir, què representen...) de les figures de les quals disposen diverses entitats del

poble, com ara el gegant de la colla de diables o els quatre gegants de la colla gegantera.

En el cas de que el turista vulgui visitar el municipi i esculli rebre informació sobre els llocs d'interès que hi ha, rebrà per part de la màquina aquestes tres opcions per escollir concretament què és el què realment vol anar a conèixer:

1. *Castells*
2. *Esglésies*
3. *Colònies tèxtils / mineres*

La primera categoria inclou totes els forteses que hi ha al poble. D'aquestes s'indicarà el lloc, l'any i l'estil de construcció, quin és l'estat actual de conservació i una mica d'informació addicional com ara la forma que té, les habitacions, quants merlets té, el motiu pel qual va ser construït, els horaris de visita...

El segon apartat farà que l'aplicació parli de totes les esglésies que hi ha construïdes al terme municipal. D'elles se'n donarà a conèixer el nom, la seva situació geogràfica, a quin any van ser construïdes, quin estil van seguir per fer-les, l'estat de conservació, s'explicaran les reformes que s'hi han dut a terme i quan s'han fet, s'informarà de si disposa d'alguna obra d'art, a quins sants està dedicada i es donarà certa informació complementària que pugui ser rellevant.

Si el què vol el turista és visitar les colònies aixecades als voltants del poble escollirà la tercera via. D'aquesta manera la màquina li oferirà informació sobre les dues colònies tèxtils i la minera que hi ha construïdes dins del terme municipal. De cada una en pronunciarà el nom (per si l'usuari la vol buscar al navegador), la seva situació geogràfica (per situar-la sobre el mapa si el turista no disposa de GPS o vol saber, sense haver de fer cap consulta al navegador, si és molt lluny), a quin any i per qui va ser fundada i un resum de la seva història, des de l'activitat que tenia en els seus inicis fins a com es troba en l'actualitat.

Una informació que pot desitjar conèixer el turista que visita el poble és de quins serveis i equipaments disposen els habitants que viuen aquí. No és una informació que convidi a l'usuari a visitar alguna part del poble com passava amb les anteriors opcions, sinó que es tracta d'aportar coneixement a qui esculli aquesta part. La màquina està programada per, un cop l'usuari ja hagi escollit, preguntar per quina de les categories en què estan dividits els serveis i equipaments està més interessat el turista. Les possibilitats són les següents:

1. *Culturals i educatius*
2. *Esportius*
3. *Medi Ambient*
4. *Neteja*
5. *Sanitat*
6. *D'interès ciutadà*
7. *Diversos*

Dins la primera categoria s'hi troben, com molt bé indica el nom, els serveis i instal·lacions destinats a oferir cultura i educació als vilatans. Aquí el turista coneixerà on es troba situada la biblioteca municipal i quines activitats s'hi duen a terme: explicació de contes pels més menuts, tallers, cursos, xarrades, exposicions... També rebrà informació sobre la localització de la llar d'infants i l'escola pública, juntament amb informació sobre quins estudis s'hi imparteixen. L'aplicació també explicarà que hi ha una escola municipal de música que ofereix una àmplia gamma de cursos per aprendre solfeig i tocar una gran varietat d'instruments i informarà que existeix una escola de teatre per qui vulgui començar els passos per esdevenir actor/actriu. La segona categoria té a veure amb les instal·lacions esportives de les quals disposa el municipi. L'aplicació donarà a conèixer la localització del camp de futbol, del camp de futbol 7, del pavelló d'esports municipal, de la pista poliesportiva, de la pista de skate, de les pistes de tennis, de la pista de voleibol i de la piscina municipal. A més a més de la situació geogràfica, també s'informarà al turista de coses com ara la superfície del terreny de joc, quins esports s'hi practiquen (lògicament a les pistes de tennis, per exemple, si practicarà únicament tennis, però al pavelló pot ser que s'hi practiquin diferents modalitats esportives), si les instal·lacions són de lliure accés o el tenen restringit, horaris d'obertura...

El tercer apartat aporta al turista la informació necessària que ha de conèixer sobre els serveis i els equipaments mediambientals del poble. Es parla de la deixalleria (localització, informació d'interès i productes que s'hi recullen), de la depuradora d'aigua potable (localització i funcionalitat) i de l'estació depuradora d'aigües residuals (localització, funcions i informació interessant). Si l'usuari vol estar informat de qui neteja el poble escollirà la quarta via, fet que li permetrà rebre dades sobre el servei de neteja viària, el servei de recollida de mobles vells, el servei de recollida d'escombraries, el servei de recollida selectiva de matèria orgànica i el servei de recollida selectiva de



paper, cartró, envasos i vidre.

Tal i com deixa entreveure el seu nom, el cinquè apartat tracta sobre els equipaments sanitaris dels quals disposa el poble. En aquest cas només se'n veurà un, l'únic que hi ha: el Centre d'atenció primària, del qual se'n oferirà la localització, per qui és gestionat i quins horaris té.

El penúltim apartat mostra aquells serveis i equipaments destinats a fer més fàcil la vida de la gent del poble. Aquí s'hi inclouen instal·lacions com el casal de la gent gran, el casal Verge de Montserrat, l'oficina de correus, el tanatori i serveis com el de taxi. De tots ells s'explicarà qui se n'encarrega de la gestió, quan estan oberts, on estan localitzats i una mica d'informació d'interès si s'escau.

L'última agrupació de totes inclou tots aquells serveis i equipaments del municipi que no estan inclosos en cap subgrup en concret, com ara el cementiri, el jutjat de Pau, el mercat municipal de venda no sedentària, l'oficina mòbil d'atenció al consumidor i la Sala Sindicat (sala polivalent on s'hi realitzen diferents activitats).

Ja per acabar, l'última informació que pot oferir l'aplicació al turista tracta sobre els esports que es practiquen al municipi. Si l'usuari vol saber quines entitats esportives hi ha al poble haurà d'escollir aquesta opció i la màquina li deixarà escollir les següents opcions per acabar de concretar més quin esport desitja conèixer:

1. *Bàsquet*
2. *Futbol*
3. *Futbol sala*
4. *Handbol*

Al municipi només hi ha quatre entitats esportives (Club Bàsquet Balsareny, Unió Esportiva Balsareny, Futbol Sala Balsareny i Club Handbol Balsareny) que agrupen diferents equips, tal i com es veurà quan es triï un esport o un altre.

D'ells se'n donarà a conèixer l'any de formació, per qui va ser fundat, quina vestimenta utilitzen en els seus partits, a quin camp i en quin horari juguen quan ho fan com a locals, un breu repàs a la seva trajectòria esportiva i si disposen d'equips base.

## 5.2 Reconeixement automàtic de la parla

L'eina que utilitzarem en aquest projecte per tal de reconèixer les paraules que diu el turista és l'ATK.

Primer de tot cal descarregar el paquet. Per fer-ho, és necessari anar a la web de HTK i buscar-lo dins de l'apartat corresponent. Quan ja s'hagi baixat només cal desar-lo a la ruta desitjada, segurament dins una carpeta que contindrà tota la aplicació, sense haver d'instal·lar res. Dins la carpeta principal hi ha moltes subcarpetes, la majoria de les quals contenen arxius de configuració o bé arxius del mateix sistema necessaris pel correcte funcionament del paquet que no s'han de tocar.

D'entre les moltes carpetes que hi ha dins, n'hi ha una que té per nom *ATKApps* que conté alguns exemples d'aplicacions de reconeixement de veu que poden servir de base a l'hora de crear el sistema. Aquí dins és on es treballarà i es desaran els fitxers a mesura que l'aplicació vagi agafant forma. Per tal de fer-ho ordenat, s'ha creat una carpeta específica pel projecte. Cal tenir-ho ben estructurat per simplificar i facilitar el treball, ja que si l'aplicació va creixent en mida pot arribar un moment en el qual no se sàpiga on és o bé per què serveix cada fitxer.

Els fitxers necessaris per fer funcionar l'aplicació els trobarem dins una subcarpeta anomenada *Test*. A fora només s'hi deixa el codi programat.

Tal i com s'ha comentat en el tema 2, en un sistema de diàleg hi ha, com a mínim, un fitxer de gramàtica que conté les paraules estructurades de la mateixa manera en que l'usuari les pot pronunciar en qualsevol moment mentre duri la conversa amb la màquina. Normalment n'hi sol haver més d'un, ja que un diàleg, sobretot si és llarg, és millor estructurar-lo en varies gramàtiques, més endavant s'explicaran les raons del perquè. Concretament en aquesta aplicació n'hi ha fins a 10 de diferents, una per cada *node* del diàleg. Com a node entenem les diferents *fases* per les quals passa el diàleg. La primera, aquí, és la salutació. Un cop passada arriba la part en què l'usuari pot escollir entre les diferents categories d'informació que la màquina li ofereix saber (informació general, esports, llocs per menjar o dormir...), a continuació, segons el què l'usuari esculli, s'activa la gramàtica específica d'aquella opció, ja que s'ha creat un fitxer d'aquest tipus per a cada una d'aquestes categories. A dins contenen subcategories d'informació que l'usuari podrà escollir per acabar de concretar més a fons el què vol conèixer (si tria llocs per visitar s'activa el fitxer *llocs\_interes.gram* que disposa de les subcategories: castells, esglésies o colònies tèxtils/mineres). També hi ha una gramàtica exclusiva perquè quan es demani a l'usuari si vol rebre més informació aquest pugui respondre sí o no.

A la imatge 5.1 es pot veure com és un fitxer de gramàtica. S'ha escollit el fitxer *salutacio.gram* perquè és senzill i permet entendre correctament l'estructura d'aquest. La resta de fitxers són iguals, ja que l'estructura d'aquests sempre segueix la mateixa forma. Primerament hi ha l'opció de definir una o més variables. En aquest cas n'hi ha una, a la qual se li assignen

```
$salutacio = (hola | bon dia | bona tarda | bona nit);

(
$salutacio
)
```

Figura 5.1: Contingut d'un fitxer gramàtica senzill

```
$art = (els | les);
$prep = (de | del | sobre);
$verb1 = (obtenir | rebre);
$verb2 = (parla'm | informa'm);
$verb3 = (m'agradaria | vull);
$verb4 = (escullo | trio);
$opcio = (m'agrada el tèxtil | m'agrada la mineria);

(
informació [general] sobre [la] [$art] [l'entitat esportiva de] [l'equip de] | $verb2 $prep [$art] [l'equip de] [l'entitat esportiva de] | $verb3 saber
coses $prep [$art] [l'equip de] [l'entitat esportiva de] | dona'm informació $prep [$art] [l'equip de] [l'entitat esportiva de] | $verb4 [la opció]
[l'entitat esportiva de] | $verb3 rebre informació $prep [$art] [l'equip de] [l'entitat esportiva de] | $verb3 que m'informés $prep [$art] [l'equip de]
[l'entitat esportiva de] | $verb3 informar-me $prep [$art] [l'equip de] [l'entitat esportiva de] | $verb3 $verb1 informació $prep [$art] [l'equip de]
[l'entitat esportiva de] | $verb3 saber coses $prep [$art] [l'equip de] [l'entitat esportiva de] | $verb3 conèixer [$art] [l'equip de] [l'entitat esportiva
de] | explica'm | $verb4 conèixer [quines són i quan se celebren] [$art] [l'entitat esportiva de] | $verb3 conèixer [quines són i quan se celebren] [$art]
[l'entitat esportiva de] | $verb3 saber [quins] [quines] [de] [quina] | quines i quan són | [$verb3] conèixer els [equips esportius de] | [de] quins [equips
esportius de] | $verb3 visitar [$art] | m'agraden [$art] | m'agrada l'època medieval informa'm sobre | m'interessen més [$art] | [$opcio] quines | on [em] puc
[anar a] | [$verb4] la categoria | m'interessa la categoria | m'interessen els [de] | [$verb3] que m'informés $prep [$art] $prep [de la categoria] |
anomena'm els [de] | m'agradaria saber de quins serveis i equipaments [de] | [de] quins [serveis] [i] [equipaments] | equipaments i serveis | serveis i
equipaments | m'interessa el | m'agrada el | de l'equip de | del de | m'agradaria saber coses $prep [l'equip de] | $verb3 $verb1
)
```

Figura 5.2: Contingut d'un fitxer gramàtica més complicat

tots els possibles valors que pot prendre, és a dir, totes les possibles paraules o frases que l'usuari pot pronunciar per referir-se a aquella variable en concret. En el cas de la imatge 5.1, per saludar a la màquina l'usuari pot fer-ho, tal i com ja s'ha vist, dient les paraules *Hola*, *Bon dia*, *Bona tarda* o *Bona nit*.

El que hi ha escrit entremig del parèntesi és el què realment es tindrà en compte d'aquesta gramàtica. En aquest cas per exemple, només valdran les 4 expressions definides a *\$salutacio*, ja que només hi ha escrita aquesta variable dins el parèntesi, però hi ha casos en què aquí s'hi escriuen frases, amb variables al mig, que no canviaran mai de valor, exceptuant les variables, que lògicament aquestes sí que ho faran. Per veure més clara aquesta última explicació observar la figura 5.2. En ella s'hi pot veure el fitxer de gramàtica *preàmbul.gram*, és a dir, la gramàtica on hi ha la primera part de les frases que l'usuari pot pronunciar per tal d'arribar a demanar la informació que ell desitja i que es pot considerar brossa, ja que no té cap mena de valor. El què realment importa i que servirà per escollir quin camí seguir està ficat en altres fitxers de gramàtiques.

Tot i que ja s'ha vist abans, tornem-ho a recordar ara per ajudar a entendre el contingut d'aquests fitxers. Tot el què s'escrigui dins del símbol *//* és opcional, l'usuari no està obligat a dir-ho. Si ho diu es reconeix i si no la frase continua igual. Per exemple:

*\$verb3 conèixer [quines són i quan se celebren] [\$art]*

permet que l'usuari comenci dient *m'agrada* o *vull*, que és el contingut que pot tenir la variable *\$verb3*, continuï obligat amb la paraula *conèixer* i llavors pot escollir dir *quines són i quan se celebren els/les*, dir només *quines són i quan se celebren*, passar directament de *conèixer* a *els/les* o bé quedar-se amb l'obligatori i no dir res més, excepte quina és la informació que vol, però això ja està dins d'una altra gramàtica i no influeix en aquesta.

Destacar que si la persona no diu cap d'aquestes paraules escrites dins del fitxer gramàtica que hi hagi activat en aquell moment, el reconeixedor no serà capaç d'entendre què vol dir. Aquí es pot veure la importància de fer un bon disseny del diàleg i de les gramàtiques, ja que el més mínim descuit pot portar com a conseqüència que l'aplicació no funcioni correctament perquè no és capaç de reconèixer les paraules que li fa arribar l'usuari.

Una vegada s'ha utilitzat la gramàtica es pot desactivar i aparcar-la, ja que no conté cap informació útil que es pugui fer servir en el que resta de conversa, ja que l'aplicació està dissenyada de manera lineal i no permet tornar enrere si no és que s'arriba al final. D'aquesta manera s'allibera memòria i es fan els fitxers més simples i fàcils de comprendre (no hi ha tantes frases, paraules...). Aquesta és una de les raons per les quals els diàlegs es divideixen en diverses gramàtiques.

El mòdul ATK no utilitza els arxius *.gram*. Aquests que hem vist i explicat fins ara són només una representació d'alt nivell de la gramàtica de l'aplicació que es vol crear. Cal passar-los a fitxers *.slf* que són els que realment ATK farà servir a l'hora de reconèixer el text que li xarri l'usuari. Per fer-ho, HTK disposa de la comanda *HParse*:

*HParse salutacio.gram salutacio.slf*

Els fitxers resultats contindran una sèrie de nodes numerats (cada paraula d'aquella gramàtica) amb totes les possibles transicions entre ells. Un exemple de fitxer d'aquest tipus el podem veure a la imatge 5.3.

Els camps *N* i *L* indiquen, respectivament, el nombre total de nodes i arcs (transicions). Cada node està numerat, començant des de 0, i identificat amb la lletra *I*. Els arcs també ho estan, però en aquest cas amb la lletra *J*. La *W* té assignades les paraules. Si s'observen amb atenció les paraules de la imatge 5.3, es podrà veure que n'hi ha una de nova que no apareix definida al fitxer de gramàtica creat anteriorment. Aquesta és la paraula *!NULL*, que es fa servir per indicar el principi i el final de les transicions (mai es passa directament de l'inici a una paraula de la gramàtica ni tampoc d'una paraula al node fi, observar figura 5.4).

```

SUBLAT=salutacio
N=10  L=12
I=0    W=NULL
I=1    W=NULL
I=2    W=hola
I=3    W=NULL
I=4    W=bon
I=5    W=dia
I=6    W=bona
I=7    W=tarda
I=8    W=bona
I=9    W=nit
J=0     S=3   E=1
J=1     S=0   E=2
J=2     S=2   E=3
J=3     S=5   E=3
J=4     S=7   E=3
J=5     S=9   E=3
J=6     S=0   E=4
J=7     S=4   E=5
J=8     S=0   E=6
J=9     S=6   E=7
J=10    S=0   E=8
J=11    S=8   E=9
.
SUBLAT=confirm
N=3  L=2
I=0    W=NULL
I=1    W=NULL
I=2    W=si
J=0    S=0   E=2
J=1    S=2   E=1
.
# Main toplevel network
N=3  L=2
I=0    W=NULL
I=1    L=salutacio s=salutacio
I=2    W=NULL
J=0    S=0   E=1
J=1    S=1   E=2

```

Figura 5.3: Contingut d'un fitxer *.slf*

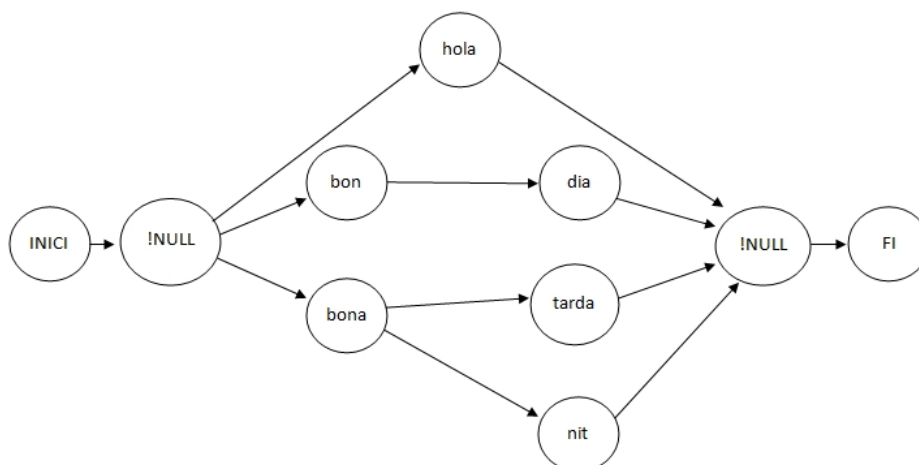


Figura 5.4: Relacions/combinacions entre paraules per construir frases

El què hi ha assignat a la lletra *S* (*start-node*) és el número de node des d'on surt l'arc i el què hi ha a la lletra *E* (*end-node*) és a on arriba aquest. Ficant com a exemple el fitxer de la imatge 5.3, l'arc 6 surt del node 0 (paraula *!NULL*) i va a parar al node 4 (paraula *bon*). L'arc 7 surt precisament d'aquest node 4 i va a parar al 5 (paraula *dia*) i l'arc 3 (no necessàriament han d'estar per ordre, bàsicament perquè seria impossible) surt del node 5 i va a parar al 3 (paraules *dia* i *!NULL*). D'aquesta manera es crea la frase *Bon dia*.

El fitxer està dividit en tres parts: la primera i la que ha estat explicada ara es podrà trobar a tots els fitxers .slf, lògicament amb diferents mides (tant de nodes com d'arcs) i transicions. La segona sempre serà la mateixa sigui quin sigui el fitxer, ja que és la confirmació. La tercera i última és el *menú* que identifica aquell fitxer. Aquí és on es pot definir si abans i/o després d'aquesta gramàtica que representa el fitxer .slf n'hi ha d'haver alguna o no (en aquest cas no, així que *!NULL*). En alguns casos, com per exemple en el fitxer *serveis\_equipaments.slf*, veurem que abans de la gramàtica *serveis\_equipaments.gram* hi ha d'anar la gramàtica *preamble.gram* i a darrera la *postamble.gram*.

Un reconeixedor de la parla també necessita disposar d'un diccionari, un fitxer on hi ha una llista d'absolutament totes les paraules que en algun moment del diàleg poden sortir pronunciades, ordenades alfabèticament i amb la transcripció fonètica al costat.

Com que l'aplicació està desenvolupada per enginyers tècnics informàtics, realitzar la transcripció fonètica de les paraules a mà, com és lògic, és un problema (sortiran errors, es perdrà temps...), així que s'utilitzarà una eina creada específicament per a realitzar transcripcions fonètiques per a totes les variants de la llengua catalana. Aquesta aplicació s'anomena *Segre* i va ser desenvolupada al Centre TALP de la UPC amb la col·laboració del Laboratori de Fonètica del Departament de Filologia Espanyola de la UAB i també del Grup de Recerca en Variació Dialectal del Departament de Filologia Catalana de la UB. El seu funcionament és senzill, només se li ha de passar un text ortogràfic i ell retorna una cadena d'al·lòfons amb sil·labització i accentuació. Està basat en regles externes que se subministren al programa en forma de fitxers.

Veiem pas a pas la creació d'aquest fitxer diccionari. Primerament és necessari agafar gramàtica per gramàtica i apuntar en una llista totes les paraules que hi surten. Aquesta llista cal ordenar-la alfabèticament i un cop llest, eliminar les paraules repetides. És important fer-ho en aquest ordre ja que així es veu d'una manera clara totes les paraules que s'han d'esborrar i és difícil deixar-se'n alguna.

Si l'arxiu de text que conté la llista està en format UTF cal passar-lo a ISO,

|              |                     |
|--------------|---------------------|
| m'agrada     | m @ G r a D @       |
| m'agradaria  | m @ G r @ D e r i @ |
| m'agraden    | m @ G r a D @ n     |
| medi         | m E D i             |
| medieval     | m @ D i @ B a l     |
| menjar       | m @ n Z a           |
| més          | m e s               |
| mineres      | m i n e r @ s       |
| mineria      | m i n @ r i @       |
| m'informés   | m i m f u r r m e s |
| m'informis   | m i m f o r r m i s |
| m'interessa  | m i n t @ r E s @   |
| m'interessen | m i n t @ r E s @ n |
| municipi     | m u n i s i p i     |
| neteja       | n @ t E Z @         |
| nit          | n i t               |
| no           | n o                 |
| obtenir      | u p t @ n i         |
| on           | o n                 |
| opció        | u p s i o           |
| parla'm      | p a r r l @ m       |
| passar       | p @ s a             |
| per          | p @ r r             |
| practiquen   | p r @ k t i k @ n   |
| prendre      | p E n d r @         |
| prop         | p r O p             |
| puc          | p u k               |
| quan         | k w a n             |
| que          | k @                 |
| quina        | k i n @             |
| quines       | k i n @ s           |
| quins        | k i n s             |
| rebre        | r r E B r @         |
| repetir      | r r @ p @ t i       |
| saber        | s @ B E             |
| sala         | s a l @             |
| sanitat      | s @ n i t a t       |
| se           | s @                 |
| serveis      | s @ r B E j s       |
| si           | s i                 |
| SIL          | s i l               |
| sobre        | s o B r @           |
| són          | s o n               |
| sopar        | s u p a             |
| tarda        | t a r D @           |

Figura 5.5: Contingut del fitxer diccionari

ja que Segre treballa amb aquest format. La mateixa aplicació ja disposa d'un script que fa el canvi automàticament.

Per fer arribar els fitxers que contenen les regles a l'aplicació existeix un arxiu on hi ha definides les diferents rutes i noms d'aquests fitxers i també alguns paràmetres que s'han d'ajustar, com ara el nom del fitxer on hi ha la llista ISO de paraules i el nom que ha de tenir el fitxer de sortida.

Una vegada executat el programa, cal ajuntar el fitxer que ha creat, on només hi ha transcripcions fonètiques (això si, ordenades), amb el fitxer on hi ha la llista de paraules ordenada i sense repeticions de manera que quedi un fitxer resultant (diccionari) amb les paraules escrites normals i al costat la seva transcripció.

En aquest arxiu diccionari cal afegir-hi manualment, al lloc que li pertoqui, la paraula SIL (amb majúscula perquè ve establert així per ATK) que és la representació del silenci.

Ja per acabar amb l'explicació de tot el necessari per a la construcció d'un reconeixedor de la parla englobat dins d'un sistema de diàleg, parlar dels models acústics. Aquests, al ser l'aplicació en català, també han de ser específics d'aquesta llengua. La UAB disposa d'aquests models per a les seves investigacions i els seus projectes, així que s'han utilitzat els seus.

| E                                   | Paquet                     | Versió instal·lada | Darrera versió      | Descripció                                                    |
|-------------------------------------|----------------------------|--------------------|---------------------|---------------------------------------------------------------|
| <input checked="" type="checkbox"/> | festival                   | 1.97~svn20100328   | 1.97~svn20100328    | General multi-lingual speech synthesis system                 |
| <input type="checkbox"/>            | festival-doc               |                    | 1.4.2-8             | Documentation for Festival                                    |
| <input type="checkbox"/>            | festival-czech             |                    | 0.3-1               | Czech support for Festival speech synthesis system            |
| <input type="checkbox"/>            | festival-freebsoft-utils   |                    | 0.10-2              | Festival extensions and utilities                             |
| <input checked="" type="checkbox"/> | festival-dev               | 1.97~svn20100328   | 1.97~svn20100328    | Development kit for the Festival speech synthesis system      |
| <input type="checkbox"/>            | festival-te                |                    | 0.3.3-1             | festival text to speech synthesizer for Telugu (te) language  |
| <input type="checkbox"/>            | festival-mr                |                    | 0.1-5               | festival text to speech synthesizer for Marathi language      |
| <input type="checkbox"/>            | festival-hi                |                    | 0.1-5               | festival text to speech synthesizer for Hindi language        |
| <input type="checkbox"/>            | festlex-iffd               |                    | 2.0+debian0-3       | Italian support for Festival                                  |
| <input type="checkbox"/>            | speech-dispatcher-festival |                    | 0.6.8~unofficial~rc | Festival support for Speech Dispatcher                        |
| <input type="checkbox"/>            | festlex-oald               |                    | 1.4.0-3.1           | Festival lexicon from Oxford Advanced Learners' Dictionary    |
| <input type="checkbox"/>            | pidgin-festival            |                    | 2.4-1               | pidgin plugin to hear incoming messages using voice synthesis |
| <input type="checkbox"/>            | eflite                     |                    | 0.4.1-3             | Festival-Lite based emacspeak speech server                   |
| <input type="checkbox"/>            | stardict-plugin-festival   |                    | 3.0.1-6             | International dictionary - Festival TTS plugin                |
| <input type="checkbox"/>            | festvox-rablpc8k           |                    | 1.4.0-2             | British English male speaker for festival, 8khz sample rate   |
| <input type="checkbox"/>            | festvox-rablpc16k          |                    | 1.4.0-2             | British English male speaker for festival, 16khz sample rate  |
| <input type="checkbox"/>            | festvox-ca-upc-hts-teo     |                    | 0.90                | Catalan male speaker for festival, 16kHz hts                  |
| <input type="checkbox"/>            | festvox-ca-upc-hts-bet     |                    | 0.91                | Catalan female speaker for festival, 16kHz hts                |
| <input type="checkbox"/>            | festvox-ca-upc-hts-pol     |                    | 0.90                | Catalan male speaker for festival, 16kHz hts                  |
| <input type="checkbox"/>            | festvox-ca-upc-hts-jan     |                    | 0.90                | Catalan male speaker for festival, 16kHz hts                  |
| <input type="checkbox"/>            | festvox-ca-upc-clunits-bet |                    | 0.90                | Catalan female speaker for festival, 16kHz clunits            |
| <input type="checkbox"/>            | festvox-ca-upc-clunits-pep |                    | 0.90                | Catalan male speaker for festival, 16kHz clunits              |
| <input type="checkbox"/>            | festvox-ca-upc-hts-eva     |                    | 0.90                | Catalan female speaker for festival, 16kHz hts                |
| <input type="checkbox"/>            | festvox-ca-upc-hts-pau     |                    | 0.90                | Catalan male speaker for festival, 16kHz hts                  |
| <input type="checkbox"/>            | festvox-ca-upc-hts-uri     |                    | 0.90                | Catalan kid speaker for festival, 16kHz hts                   |
| <input type="checkbox"/>            | festvox-ca-upc-hts-mar     |                    | 0.90                | Catalan female speaker for festival, 16kHz hts                |
| <input type="checkbox"/>            | festvox-ellpc11k           |                    | 1.4.0-3             | Castilian Spanish male speaker for Festival                   |
| <input checked="" type="checkbox"/> | festlex-cmu                | 1.4.0-6            | 1.4.0-6             | CMU dictionary for Festival                                   |
| <input type="checkbox"/>            | festvox-suopuhe-common     |                    | 1.0g-20051204-3     | Common files for Festival Finnish speakers                    |
| <input type="checkbox"/>            | festvox-czech-ph           |                    | 0.1-2               | Czech male speaker for Festival                               |

Figura 5.6: Gestor de paquets Synaptic amb el *festival* marcat

## 5.3 Síntesi de veu

L'eina que utilitzarem en aquest projecte per tal de transmetre la informació al turista és el Festival.

A diferència dels mòduls d'ATK i HTK, aquest mòdul sí que s'ha d'instal·lar a l'ordinador. Per fer-ho, però, només cal seguir unes senzilles i intuïtives instruccions. Primerament cal instal·lar el Festival: això es pot fer des del gestor de paquets Synaptic (veure imatge 5.6) o des de la consola mateix. La primera opció és més fàcil i clara, ja que un cop a dins d'aquest gestor només cal seleccionar el paquet que es vol instal·lar (en aquest cas *festival*) i automàticament es marquen la resta de paquets dels quals depèn, havent d'aplicar només els canvis i fent així que comenci la instal·lació. La segona és més ràpida però s'ha d'estar un pèl informat ja que cal escriure a la línia de comandes cada paquet que es vol instal·lar, procurant no deixar-se'n cap. Quedaria, en aquest cas, una sentència com la següent:

```
$ sudo apt-get install festival festlex-cmu festlex-poslex festvox-kallpc16k
libestools1.2 festival-dev
```

S'instal·la a la següent ruta: */usr/share/festival*.



El pròxim pas consisteix en comprovar que funcioni. Si al escriure, a la mateixa línia de comandes, la següent sentència:

```
$ echo 'hello world!' — festival -tts
```

se sent bé, és que la instal·lació s'ha realitzat correctament (és necessari tenir la tarja d'àudio ben configurada).

Efectivament, l'idioma predeterminat és l'anglès, així doncs cal instal·lar el paquet en català:

```
$ sudo apt-get install festival-ca
```

si s'escriu la comanda al terminal o bé s'ha de buscar el paquet *festivalca* al gestor Synaptic. Per defecte, el Festival en el nostre idioma ja porta incorporada una veu: *upc\_ca\_ona.hts*. És la que s'utilitzarà mentre no se'n descarregui una altra i es forci el canvi.

Per comprovar que funciona l'idioma n'hi ha prou amb executar la següent línia:

```
$ echo 'Bon dia, Catalunya' — festival -language catalan -tts
```

Si la resposta és afirmativa es pot donar per acabat el procés d'instal·lació del paquet que permetrà que l'aplicació parli en català.

Com s'ha comentat, hi ha un grapat de veus creades per aquest idioma. Al web del TALP s'hi pot trobar un apartat que permet escoltar totes aquestes veus per així poder escollir la més adequada al sistema que es crearà (home, dona, veu més aguda, més greu...). Un cop triada, només s'ha de baixar el paquet i desempaquetar-lo:

```
$ wget http://www.talp.cat/festcat/download/upc_ca_pau.hts.tgz
```

```
$ sudo tar -C /usr/share/festival -xzf upc_ca_pau.hts.tgz
```

Ara ve el pas més important de tots, cridar el Festival des de l'aplicació, és a dir, fer que el codi C++ que executa el programa i disposa del text l'envii a aquest paquet i així pugui ser sintetitzat. La figura 5.7 mostra un exemple bàsic del que és necessari (biblioteques, paràmetres de configuració, funcions...) perquè funcioni i l'usuari pugui escoltar la veu que surt creada de l'ordinador.

Es pot observar que les dues primeres línies són per incloure biblioteques,

```

#include "/usr/include/festival/festival.h"
#include "/usr/include/speech_tools/EST_String.h"

int main(int argc, char **argv)
{
    //EST_Wave wave;
    const EST_String language = "catalan";
    int heap_size = 210000; // default scheme heap size
    int load_init_files = 1; // we want the festival init files loaded
    festival_initialize(load_init_files, heap_size);
    festival_init_lang(language);

    // Say some text;
    festival_say_text("Hola món!");

    festival_wait_for_spooler();

    return 0;
}

```

Figura 5.7: Com cridar *Festival* des d'un fitxer C++

fitxers necessaris perquè el codi que ve a continuació pugui ser executat sense cap mena de problema. A dins la funció principal (*main*) el primer que es pot trobar és una declaració i inicialització de variables.

Seguidament es criden un parell de funcions: una anomenada *festival\_initialize*, que com molt bé deixa entreveure el seu nom, té la funció d'inicialitzar el sintetitzador de veu amb els valors de les variables que s'han definit i l'altra *festival\_init\_lang* que serveix per definir l'idioma sobre el qual ha de funcionar l'aplicació.

Ara que ja hi ha en marxa el Festival només falta passar-li el text que es desitja convertir a veu. Per fer-ho s'utilitza la funció *festival\_say\_text* que, com a paràmetre, s'hi ha de ficar la frase a sintetitzar. En aquest senzill exemple l'aplicació només dirà *Hola món!*, però en l'aplicació creada, al haver-hi tant text per sintetitzar i ser molt variat (primer s'ha de dir una frase, llavors una altra de diferent...) pot ser que com a paràmetre hi hagi una variable que cada cop que es crida aquesta funció té una cadena de caràcters diferent:

```

festival_say_text(text.c_str());

```

Com es pot comprovar, implementar la síntesi de veu no és molt complicat, ja que només s'ha d'instal·lar un paquet amb l'idioma

corresponent, escollir la veu que més agradi i introduir al codi la crida a aquestes quatre funcions que són la base de que l'ordinador pugui dir, a través de la veu, el text que l'aplicació, i per tant el dissenyador d'aquesta, li hagi volgut fer dir.

# Capítol 6

## Proves

Primer de tot comentar que al tractar-se d'un projecte que utilitza les tecnologies de la parla tant per a l'entrada com per a la sortida de dades, és molt difícil plasmar sobre el paper les proves que s'han dut a terme.

Una vegada dit això, explicar que la finalitat d'aquest apartat és assegurar el correcte funcionament de l'aplicació: que s'activi quan toca, que *escolti* i respongui el què l'usuari ha demanat, que totes les funcionalitats estiguin disponibles i sense errades... En definitiva, que l'aplicació estigui preparada per ser instal·lada al carrer.

Per començar les proves cal engegar el sistema. Aquest és un pas que només es farà una vegada, la primera que s'obri l'aplicació després de ser instal·lada, ja que tal i com es porta comentant, un dels objectius del projecte és que estigui operatiu les 24 hores del dia els 365 dies de l'any, per tant, a no ser que hi hagi factors externs que obliguin a fer-ho, no es parará mai. Quan cridem l'executable apareix el què es mostra a la figura 6.1. Només cal prémer el botó *Start* i l'aplicació començarà a *escoltar*. La veu que contesta

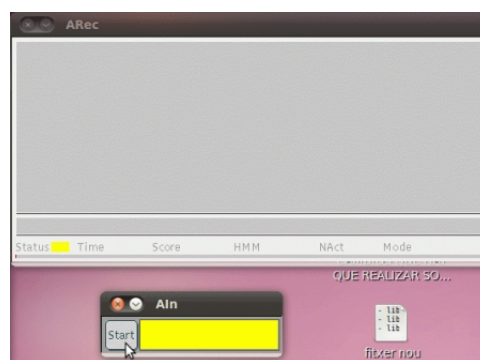


Figura 6.1: Activació del ASR

```

SSDS: ATK Simple Spoken Dialog System
=====

Punt d'informació Balsarenyenc
#####

Esperant activació...

Usegroup: salutacio-ask 0

```

Figura 6.2: Aplicació a l'espera que l'usuari l'activi

```

SSDS: ATK Simple Spoken Dialog System
=====

Punt d'informació Balsarenyenc
#####

Bon dia,
Benvingut a Balsareny!

Quina informació li agradaria rebre? Pot escollir entre:
1. Informació general sobre Balsareny.
2. Informació sobre la cultura i les festes del municipi.
3. Conèixer els llocs d'interès que hi ha a Balsareny.
4. Llocs on menjar.
5. Llocs on dormir.
6. Informació sobre serveis i equipaments del municipi.
7. Informació esportiva.

Usegroup: escollir_usuari-ask 0

```

Figura 6.3: Salutació, benvinguda i menú principal

no cal activar-la, ja que ho està sempre.

Seguidament l'aplicació entrarà en mode d'espera fins que algú l'activi. Quan està en aquest estat l'usuari que la vulgui fer servir veurà una pantalla exactament igual que la que apareix a la figura 6.2.

Per a realitzar la prova la *despertem* utilitzant la sentència *bon dia*. La màquina s'ha d'activar i ha de contestar, tant per veu com per pantalla, amb la mateixa salutació, així que hauria de sortir un Bon dia, la benvinguda a Balsareny i el menú principal.

Si apreciem la figura 6.3 veiem que l'aplicació respon tal i com ha de fer-ho (la veu també se sent). Ara és el torn de provar, al atzar, una opció del menú per veure si permet anar endavant a la recerca d'informació.

S'escull la segona opció, la de cultura i festes. Com veiem a la imatge 6.4

```

SSDS: ATK Simple Spoken Dialog System
=====

    Punt d'informació Balsarenyenc
    #####

    Bon dia,
    Benvingut a Balsareny!

    Quina informació li agradaria rebre? Pot escollir entre:
    1. Informació general sobre Balsareny.
    2. Informació sobre la cultura i les festes del municipi.
    3. Conèixer els llocs d'interès que hi ha a Balsareny.
    4. Llocs on menjar.
    5. Llocs on dormir.
    6. Informació sobre serveis i equipaments del municipi.
    7. Informació esportiva.

Usgroup: escollir_usuari-ask 0

    Ha escollit rebre informació sobre la cultura i les festes del municipi.
    Si us plau, triï ara una de les dues opcions:
    1. Conèixer quines són i quan se celebren les festes més importants de
    Balsareny
    2. Conèixer la cultura del municipi (els balls típics, els Pastorets, els
    gegants...).

Usgroup: cultura_festes-ask 0

```

Figura 6.4: Menú secundari que apareix quan s'ha escollit *cultura i festes*

una vegada el sistema ha rebut i processat el què hem escollit treu un altre menú a la part inferior de la pantalla (a més de la veu que contesta) que permet tornar a fer una segona tria, per encaminar més el que es vol saber.

Per continuar endavant amb la prova s'escull l'opció *cultura*. La idea és que seguidament aparegui per pantalla tota la informació que té el sistema sobre el tema escollit i es comenci a sentir pels altaveus.

Es pot apreciar a la figura 6.5 com es mostra tota la informació que rebrà l'usuari per pantalla, independentment de l'acústic. Quan el sistema ha acabat de transmetre totes les dades, ha de preguntar al turista si desitja fer alguna altra consulta, tal i com s'aprecia a la part inferior de l'anterior imatge (6.5).

Si la resposta és afirmativa, el sistema ha de tornar a mostrar el missatge de benvinguda que conté el menú principal, en canvi si és negativa s'hauria d'acomiarar i tornar-se a ficar en mode d'espera. Per continuar amb més proves escollim que sí, que volem més informació i anem a parar al menú principal. Aquesta vegada, però, escollim l'opció número tres, la dels llocs d'interès. Hauria de succeir exactament igual que abans, que a la part inferior

A continuació rebrà tota la informació relacionada amb la cultura de Balsareny, tal com ha sol·licitat.

Al poble s'hi representen dos balls:

1. Ball de Bastons:
  - Quan es representa? Normalment per la Festa Major, tot i que pot ser representat en altres dates assenyalades si s'escau.
  - Origen: Es remonta a finals del segle dinou, quan un professor de música balsarenenc va recopilar antigues melodies i va crear-ne alguna de nova per aquest ball.
  - Vestimenta: Els balladors vesteixen camisa i pantalons blancs amb una faldilla adornada, duen cascavells, set vetes i un barret penjat del coll, a part dels bastons.
2. Ball de la Faixa:
  - Quan es representa? Normalment per la Festa Major, tot i que pot ser representat en altres dates assenyalades si s'escau.
  - Origen: Dansa masculina pròpia de Balsareny i d'origen desconegut.
  - Història: S'havia ballat fins a principis del segle vint, llavors va desaparèixer. A finals dels anys quaranta, un balsarenenc va anunciar que recordava els compassos i la música d'aquest i, després d'assejar-lo molt, el dia de Sant Marc de 1951 es va tornar a ballar.

També tenim una representació dels Pastorets:

- Origen: Les primeres representacions daten de principis del segle vint, de la mà del grup del Centre Catòlic.
- On es representen? A la Sala Sindicat.
- Dates de representació: Durant les festes nadalenques.

A Balsareny també podem trobar una colla de diables anomenada Peta-Fluix:

- Figures de la colla: El Peta-Fluix, un gegant ràpid i lleuger que representa una bèstia infernal que, amb els seus 280 centímetres i 35 quilos de pes lluita entre el bé i el mal.

Ja per acabar, informar que al poble també hi ha una Colla Gegantera:

- Figures de la colla: Actualment la colla disposa de quatre figures, dos gegants i dos gegantons. Els dos gegants, el Marc i la Maria, van ser construïts l'any 1985 per Manel Casserres i representen a un antic traginer del poble i a una pagesa. Els dos gegantons, el Joan i la Mariona van ser construïts l'any 2011 per Toni Muià. Ell és el rector del poble i ella representa una filadora.

Desitja fer alguna altra consulta?

Usegroup: si\_no-ask 0

☐

Figura 6.5: Informació cultural del municipi

de la pantalla aparegués un nou menú amb més opcions relacionades per escollir.

Ha funcionat correctament, tal i com es pot veure a la figura 6.6. Ara, de les noves opcions, escollim la primera: *castells*. Ara és hora de veure per pantalla i escoltar pels altaveus què ens ha de dir el sistema sobre els castells del municipi. Observem la imatge 6.7 a veure si ho fa.

```

SSDS: ATK Simple Spoken Dialog System
=====

    Punt d'informació Balsarenyenc
#####

Bon dia,
Benvingut a Balsareny!

Quina informació li agradaria rebre? Pot escollir entre:
1. Informació general sobre Balsareny.
2. Informació sobre la cultura i les festes del municipi.
3. Conèixer els llocs d'interès que hi ha a Balsareny.
4. Llocs on menjar.
5. Llocs on dormir.
6. Informació sobre serveis i equipaments del municipi.
7. Informació esportiva.

Usegroup: escollir_usuari-ask 0

Ha escollit rebre informació sobre els llocs d'interès que hi ha a Balsareny.
A continuació pot triar, per categories, què li fa més gràcia visitar:
1. Castells.
2. Esglésies.
3. Colònies tèxtils/mineres.

Usegroup: llocs_interes-ask 0

```

Figura 6.6: Menú secundari que apareix quan s'ha escollit *llocs d'interès*

```

A continuació rebrà la informació sobre castells que ha sol·licitat.
Al municipi hi ha un castell i un fortí carlí:

1. Castell de Balsareny (monument més representatiu i carismàtic del poble):
- Lloc de construcció: Al nord del municipi, sobre d'un turó a 420 metres
d'altitud. És ven visible des de gairebé tota la població.
- Any de construcció: Data del segle catorze (fet amb una sola campanya
constructiva).
- Estil construcció: Gòtic civil català, amb planta pentagonal.
- Conservació: Bon estat de conservació.
- Informació adicional: El castell té dos nivells: l'inferior, que és on hi
ha el pati amb un pou i a més és on hi devia haver els cellers, les quadres
i les cambres dels servidors, i el superior, amb la sala d'armes, la cuina,
el menjador i les habitacions. El castell està coronat pels merlets amb una
espitllera al mig.
- Horaris de visita: Dissabtes, diumenges i festius al matí a les 11, a les
12 o a la 1.

2. Fortí Carlí:
- Lloc de construcció: Al cap damunt del serrat del Maurici, al costat de la
carretera que va cap a Súria i a l'altra banda de l'actual Eix del Llobregat.
- Any de construcció: Segle passat.
- Estil construcció: Arquitectura militar (n'és un dels millors exemples que
hi ha).
- Conservació: Tot i que no es va arribar a construir del tot, encara és
manté ferm.
- Motiu construcció: Va ser edificat per a defensar-se dels carlins.
- Horaris de visita: Es pot visitar lliurement.

Desitja fer alguna altra consulta?

Usegroup: si_no-ask 0

```

Figura 6.7: Informació sobre els castells del poble



```
Moltes gràcies per utilitzar el sistema.  
Que gaudeixi de la visita a Balsareny. Adéu!
```

Figura 6.8: Acomiadament

```
SSDS: ATK Simple Spoken Dialog System  
=====
```

Punt d'informació Balsarenyenc  
#####

Esperant activació...

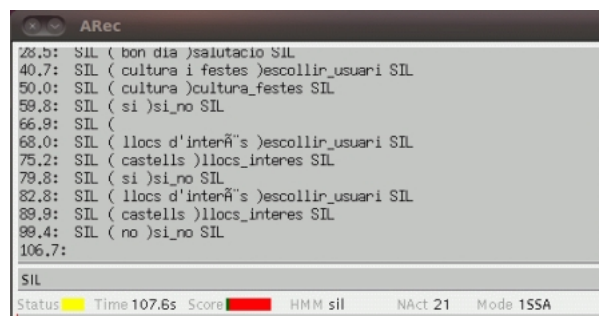
Usegroup: salutacio-ask 0

Figura 6.9: Aplicació en estat d'espera

Aquesta segona prova ha tornat a funcionar. Ara només falta comprovar si quan l'usuari ja no desitja fer cap més consulta el sistema diu adéu i passa a mode d'espera. Així que per dur a terme la prova diem que no a la màquina i observem el resultat a la imatge 6.8.

Hi podem observar el missatge d'adéu que el sistema proporciona al turista i, seguidament, veiem aparèixer per pantalla exactament el què mostra la figura 6.9, que com podem comprovar és el sistema en estat d'espera, tal i com l'ha d'haver trobat el turista quan s'hi ha acostat a demanar.

Ja per acabar, mostrar una última imatge (6.10) que reflexa tot el què l'usuari (nosaltres en aquest cas quan hem fet les proves) ha demanat a la màquina (això el turista no ho veurà) i en quin temps de la conversa ho ha fet.



```
ARec  
28.5: SIL ( bon dia )salutacio SIL  
40.7: SIL ( cultura i festes )escollir_usuari SIL  
50.0: SIL ( cultura )cultura_festes SIL  
59.8: SIL ( si )si_no SIL  
66.9: SIL (  
68.0: SIL ( llocs d'interès )escollir_usuari SIL  
75.2: SIL ( castells )llocs_interes SIL  
79.8: SIL ( si )si_no SIL  
82.8: SIL ( llocs d'interès )escollir_usuari SIL  
89.9: SIL ( castells )llocs_interes SIL  
99.4: SIL ( no )si_no SIL  
106.7:  
SIL  
Status Time 107.6s Score HMM sil NAct 21 Mode 1SSA
```

Figura 6.10: Paraules o frases pronunciades en algun moment de la conversa per l'usuari

# Capítol 7

## Conclusions

Des d'un començament la intenció d'aquest projecte va ser crear una aplicació on un usuari pogués preguntar a un ordinador, fent servir la seva veu, quelcom que li fes gràcia saber i aquest li respongués, també parlant. És a dir, crear el què es coneix com a sistema de diàleg persona-ordinador. Mai es van ficar uns objectius relacionats amb la qualitat d'aquesta aplicació (com podria ser que l'ordinador entengués un 100% de les vegades tot el què li diuen) ja que serien molt difícils d'assolir amb tants pocs recursos i temps. Bàsicament es volia aprofundir una mica en el funcionament d'aquestes tecnologies que eren noves per a mi. Així doncs, deixant clar això des d'un principi, ara que ja s'ha acabat el projecte podem extreure'n les conclusions i les sensacions obtingudes durant la realització d'aquest. Ho veurem en dues parts, separant per una banda les conclusions del projecte i per l'altra les personals.

### 7.1 Conclusions del projecte

Aquesta part de les conclusions fa referència estrictament a l'aplicació. Ara que ja esta finalitzada, si mirem el principal objectiu inicial podem dir que s'ha complert a la perfecció: s'ha creat un sistema de diàleg persona-màquina utilitzant ASR i TTS que té els fonaments sobre la idea de ser un punt d'informació turístic.

Altres objectius planejats eren que funcionés les 24 hores del dia els 365 dies de l'any, cosa que és capaç de fer (encara que només s'hagi provat màxim una hora, s'assegura completament el compliment d'aquest objectiu). També és clar que aquest sistema millora molt la forma en què l'usuari rep la informació (tenint en compte que abans només hi havia panells informatius).

Un objectiu que no es va marcar va ser que té a veure amb la qualitat de l'aplicació (per exemple que el tant per cent de preguntes enteses i ben

contestades fos un 100)), ja que degut a l'escassetat de recursos i a la manca inicial de coneixement es va veure que podria ser que no s'aconseguís.

Al mercat es poden trobar molts i millors sistemes de diàleg entre persones i màquines, la gran majoria fets per empreses que disposen de més recursos que jo, per tant és impossible competir-hi i treure l'aplicació al mercat amb afany de guanyar una gran quantitat de diners. No obstant, aquest projecte es podria implementar perfectament en un municipi petit i no gaire turístic al qual no hi vagin molts visitants estrangers, satisfent gairebé en un 100% les curiositats dels catalans que la utilitzin.

Ara bé, el fet que s'hagin assolit els objectius inicials i s'hagi acabat el projecte, no vol dir que s'hagi d'acabar aquí l'aplicació. Encara queden moltes coses per fer: polir detalls, canviar/millorar certs aspectes, implementar-ne de nous... En el següent subapartat s'explicaran aquests nous possibles objectius.

### 7.1.1 Millores futures

Aquí seran descrites les línies a seguir en futures versions d'aquest projecte i que ara no s'han pogut implementar per falta de temps, diners... Tot el què aquí s'expliqui serà per intentar millorar, cada vegada més, l'aplicació creada i així obtenir més bons resultats.

Per començar, l'aplicació presentada està dissenyada només per fer servir el català com a idioma durant la conversa, suposant això una limitació important ja que sent un projecte pensat per funcionar com a punt d'informació turístic només podria ser utilitzat per gent catalana o bé amb uns coneixements importants d'aquesta llengua. Així doncs, les noves versions d'aquesta aplicació haurien de ser multilingües, incorporant, a part del català, el castellà i l'anglès com a mínim. De cara a un futur més llunyà i segons a on vagi ubicat aquest punt d'informació pensar en incorporar-n'hi més que siguin útils.

Una altra millora consistiria en ajustar les mesures de confiança per tal que la màquina sigui menys sensible i no reconegui qualsevol so de fons com a paraula pronunciada per l'usuari. En aquests moments, per exemple, si l'usuari fa una pausa i per darrera seu passa una moto, o inclús només si fa vent, el micròfon rep un senyal que el sistema interpreta com a paraula i reconeix una frase que no és el què hauria de reconèixer.

Si mentre el sistema està responent una pregunta de l'usuari aquest xarra, encara que no tingui cap intenció de dirigir-se a la màquina, quan el programa acaba de sintetitzar la veu processa aquestes paraules que ha *escoltat* i treu resultats inesperats. Això és degut a que mentre Festival està sintetitzant àudio, la resta de components del reconeixedor, al tractar-se de

*threads* independents, continuen funcionant en un segon pla. Quan s'acaba de pronunciar tot el text, el programa principal es desbloqueja i ATK processa de cop tots els sons que ha rebut i té acumulats, provocant així aquestes situacions. Una solució a això seria intentar que mentre hi ha el Festival activat, el reconeixedor estigui en *stand-by* o bé desactivat.

Ja per acabar amb la proposta de millores, comentar-ne una altra que està una mica relacionada amb l'anterior. Actualment fins que el sistema no ha acabat de *xarrar* l'usuari no pot comunicar-se amb la màquina. Així doncs, seria bo que la persona que faci servir la màquina la pogués interrompre en qualsevol moment, parant aquesta de xarrar (sintetitzar àudio) i posant-se a reconèixer el què l'usuari li vol fer arribar.

## 7.2 Conclusions personals

Comentar, com a conclusió personal, que la sensació que m'ha quedat ha estat bona, una sensació d'haver adquirit nous coneixements i d'haver aconseguit els objectius planejats, ja que començant des de zero ho he après pràcticament tot del reconeixement automàtic de la parla i de la síntesi de veu i, a més, he estat capaç de crear un sistema de diàleg que funciona correctament, a un nivell bàsic això sí. També, des del meu punt de vista, crec que he mostrat el procés seguit d'una manera clara i entenedora perquè qualsevol persona que, com em passava a mi abans de començar, no sàpiga de què va el tema pugui entendre què és i què fa cada part, sent capaç fins i tot de crear ell una aplicació nova seguint aquests passos.

Hi he dedicat hores i esforç però ho recompensa el fet d'haver apres un munt de coses noves que abans de començar ni tant sols sabia que existien. Comentar també que el fet d'estar realitzant aquest projecte em va ajudar a aconseguir una feina remunerada, què tal i com estan les coses en aquest context de crisi que afecta el país és una gran alegria. Durant els mesos que va durar la feina vaig poder adquirir nous coneixements que he intentat reflectir en l'aplicació.

# Bibliografia

- [1] C. LLamas, V. Cardenoso. *Reconocimiento automático del habla. Técnicas y aplicación*. 3a edició. Valladolid: Secretariado de Publicaciones e Intercambio Científico, Universitat de Valladolid, 1997.
- [2] J.N. Holmes, W. Holmes. *Speech Synthesis and Recognition*. 2a edició. London: Taylor & Francis, 2001.
- [3] A. Becerra, M. Gómez, M. F. Ordoñez, B. Macas. *Reconocimiento de voz mediante Modelos Ocultos de Markov*. Agost 2009.
- [4] P. Pachès, C. de la Mota, M. Riera, M. P. Perea, A. Febrer, M. Estruch, J. M. Garrido, M. J. Machuca, A. Ríos, J. Llisterri, I. Esquerra, J. Hernando, J. Padrell i C. Nadeu. *SEGRE: An automatic tool for grapheme-to-allophone transcription in catalan*. Maig 2000.
- [5] S. Furui. *Digital speech processing, synthesis and recognition*. 2a edició. New York: Marcel Dekker, 2001.
- [6] M. R. Schroeder. *Computer speech. Recognition, compression, synthesis. With introductions to hearing and signal analysis and a glossary of speech and computer terms*. 2a edició. Berlin - Heidelberg - New York: Springer, 2004.
- [7] M. A. Anusuya, S. K. Katti. (2009) *Speech recognition by machine: A review*. International Journal of Computer Science and Information Society. Disponible a: <http://arxiv.org/pdf/1001.2267> (Maig-Juny 2012).
- [8] S. Lemmetty. (1999). *Review of speech synthesis technology. Master's Thesis*. Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology. Disponible a: [http://www.acoustics.hut.fi/publications/files/theses/lemmetty\\_mst/index.html](http://www.acoustics.hut.fi/publications/files/theses/lemmetty_mst/index.html) (Maig-Juny 2012).

- [9] J. B. Mariño, C. Nadeu, J. Llisterri. (1987). *Síntesis automática del habla. Inteligencia artificial: Conceptos, técnicas y aplicaciones*. Barcelona: Marcombo. Disponible a: [http://liceu.uab.cat/joaquim/publicacions/Marino\\_Nadeu\\_Llisterri\\_87\\_Sintesis-Automatica\\_Habla.pdf](http://liceu.uab.cat/joaquim/publicacions/Marino_Nadeu_Llisterri_87_Sintesis-Automatica_Habla.pdf) (Maig-Juny 2012).
- [10] Normativa de projectes Enginyeries Tècniques en Informàtica. <[http://www.uab.cat/Document/541/595/Normativa\\_PFCNovembre2010.pdf](http://www.uab.cat/Document/541/595/Normativa_PFCNovembre2010.pdf)>
- [11] HTK. <<http://htk.eng.cam.ac.uk/>>
- [12] Microsoft Project. <<http://www.microsoft.com/project/>>
- [13] PMI PMBOK. <<http://www.pmi.org/>>
- [14] Tecnologies de la parla en català. <<http://www.talp.upc.edu/tecnoparla/>>
- [15] Instal·lació veus catalanes de Festival. <<https://wiki.ubuntu.com/CatalanTeam/Tutorials/S%C3%ADntesiDeVeu>>
- [16] Viquipèdia. <<http://ca.wikipedia.org/wiki/Portada>>

---

Signat: ***Xavier Closa Serra***

Sabadell, ***setembre*** de ***2012***