

# Eines computacionals per a la creació i explotació de corpus orals en català

Juan María Garrido Almiñana (Madrid)

**Summary:** The goal of this article is to review the inventory of computational tools currently available for the development of oral corpora in Catalan, from their design to their exploitation once they have been developed. Several tools for the design, transcription, annotation and exploitation are described and analysed, and an example of oral corpus in Catalan developed and exploited using this kind of tools is also presented. The final conclusion is that the existing tools may help to make easier the creation and exploitation tasks of oral corpora in Catalan, although more efforts are still necessary to improve the existing tools and to create new ones.

**Keywords:** corpora, speech, orality, computational tools, Catalan ■

Received: 24-01-2020 · Accepted: 05-08-2020

## ■ 1 Introducció

El desenvolupament de corpus orals, especialment quan es vol recollir una gran quantitat de dades, és una tasca que requereix una gran inversió d'esforç i temps. El procés de construcció d'un corpus implica habitualment un esforç important que requereix tot un equip de persones i un treball de coordinació i formació importants si es fa de forma manual. Actualment existeixen eines computacionals, ja sigui de pagament o de lliure distribució, que permeten l'automatització total o parcial de determinades tasques en les diferents fases del procés, reduint la feina i els temps de desenvolupament. Moltes d'aquestes eines han estat desenvolupades per a una llengua concreta, o bé requereixen adaptacions per poder ser utilitzades en el processament d'una nova llengua. Això fa que el catàleg d'eines disponibles sigui diferent d'una a l'altra, amb una oferta en general més gran per a les llengües majoritàries.

D'altra banda, l'explotació de corpus orals és una tasca que avui en dia requereix cada cop més d'eines automàtiques, ja sigui per a la cerca de material i la seva recuperació, o per al seu processament amb finalitats de recerca. Els corpus orals de mida gran són cada cop més un instrument



habitual per a la recerca en moltes disciplines, com per exemple la Fonètica, ja que ofereixen la possibilitat d'analitzar gran quantitat de dades per a l'estudi d'un fenomen determinat. Requereixen, però, dels instruments computacionals adequats per recuperar del corpus les dades rellevants per a l'estudi i per processar-los. És possible trobar hores d'ara eines per portar a terme aquestes tasques, però com en el cas del desenvolupament dels corpus, l'oferta pot variar bastant d'una llengua a d'altra.

L'objectiu d'aquest treball és fer un repàs a l'inventari d'eines automàtiques disponibles actualment que poden fer-se servir per a les diferents fases de desenvolupament i explotació de corpus orals en català, amb especial atenció a aquelles que poden ser més útils en l'àmbit de la Fonètica, i presentar un exemple concret d'ús d'aquestes eines per al desenvolupament d'un corpus oral en català.

## ■ 2 Eines per al desenvolupament de corpus

El procés de construcció d'un corpus oral inclou habitualment les mateixes fases: disseny del corpus, recollida del material oral, transcripció ortogràfica i fonètica i anotació. En el cas de corpus especialment orientats a la recerca en Fonètica o a aplicacions tecnològiques, a més, es fa necessari alinear l'ona sonora amb la transcripció i anotació. A dia d'avui tenim al nostre abast una sèrie d'eines automàtiques o semiautomàtiques que ens poden ajudar en cadascuna d'aquestes tasques, també si estem desenvolupant un corpus en català. En les seccions que venen a continuació repassem algunes d'elles, amb una atenció especial a aquelles en què hem col·laborat d'alguna forma en el seu desenvolupament.

### ■ 2.1 Recollida de corpus

Si es tracta d'un corpus de frases, el disseny d'un corpus oral pot implicar crear o escollir frases o textos que siguin representatius d'una sèrie de fenòmens. Així, per exemple, pot ser interessant que les frases continguin exemples de tots els fonemes i al·lòfons d'una llengua, o que siguin representatives de diferents tipus d'estructures entonatives. Si es tracta d'un corpus d'una certa mida, aquesta tasca pot ser complexa de fer manualment.

Els anomenats algorismes voraços (*greedy*) són eines que permeten seleccionar els textos més representatius del conjunt de paràmetres que volem tenir en compte en el nostre corpus d'entre un conjunt de textos d'entrada. Necessiten com entrada un corpus amb els textos candidats

(quants més n’hi hagi, millor serà la selecció), el llistat de criteris que s’han de tenir en compte i el nombre de frases o textos que s’han de seleccionar. L’algorisme proporciona com a sortida un llistat dels textos triats, i la cobertura dels paràmetres que ofereixen.

A Escudero et al. (2010) es descriu un exemple d’aplicació d’aquest tipus d’algorismes per a la selecció de textos paral·lels per a un corpus castellà-català, que tinguessin les mateixes característiques en termes de cobertura fonètica i prosòdica en ambdues llengües, per la qual cosa els textos havien de ser escollits en parells. Els textos originals, en castellà, van ser traduïts al català i després transcrits fonèticament abans de fer servir l’algorisme *greedy* per a seleccionar aquells parells que oferien la millor cobertura. El procés de selecció va requerir diverses revisions manuals per acabar d’optimitzar els textos i diverses iteracions de l’algorisme fins arribar a obtenir una cobertura fonètica i prosòdica adequades.

Aquests algorismes, però, no són aplicacions que estiguin disponibles per a ús general, sinó que requereixen d’adaptació específica per part d’un programador per obtenir els resultats desitjats.

## ■ 2.2 Transcripció ortogràfica

Una altra de les tasques habituals en la recollida d’un corpus oral és la transcripció ortogràfica del material oral recollit. Aquesta tasca sol ser un dels colls d’ampolla de qualsevol projecte de desenvolupament de corpus, ja que implica molt de temps si es realitza de forma completament manual. Aquesta tasca pot alleugerir-se amb l’ús de determinades eines que fan més fàcil la transcripció d’un fitxer d’àudio. Hi ha una bona oferta d’aquest tipus d’eines, algunes de lliure distribució (com ara Transcriber<sup>1</sup> o la seva evolució TranscriberAG<sup>2</sup>) i d’altres de pagament (Transcribe,<sup>3</sup> per exemple). Es tracta d’eines que permeten carregar un fitxer d’àudio i anar desplaçant-se per ell per transcriure o anotar els seu contingut de forma fàcil. Amb aquestes eines, la transcripció l’ha de fer igualment una persona, però de forma més ràpida que escoltant i copiant en un editor de textos. La figura 1 mostra un exemple d’una d’aquestes aplicacions, TranscriberAG.

---

1 Cf. <<http://trans.sourceforge.net/en/presentation.php>>.

2 Cf. <<http://transag.sourceforge.net/>>.

3 Cf. <<https://transcribe.wreally.com/>>.

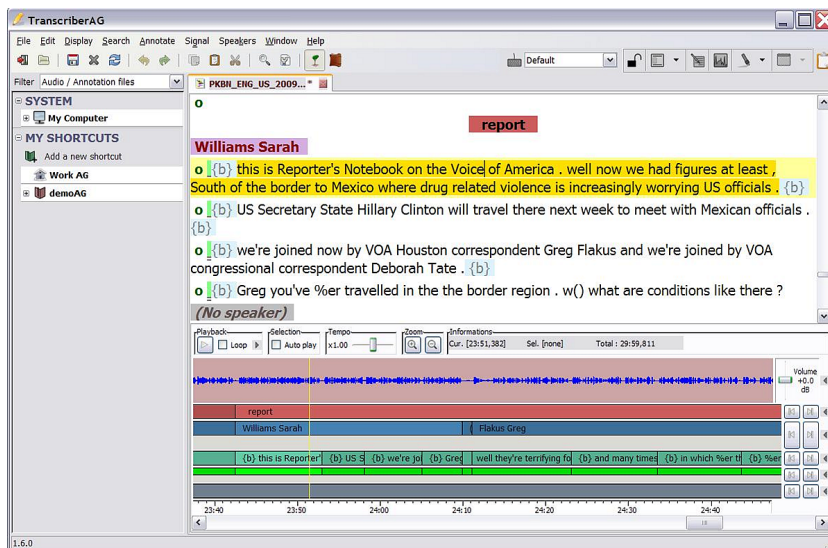


Figura 1. Exemple de pantalla del programa TranscriberAG<sup>4</sup>

La majoria d'aquestes eines són independents de la llengua, per la qual cosa es poden fer servir per a la transcripció de material en català. Només les eines que inclouen la tecnologia de transcripció automàtica no admeten la possibilitat de transcriure en qualsevol llengua, i el català no sol estar entre les llengües admeses.

Les eines que inclouen la transcripció automàtica permeten obtenir la transcripció ortogràfica d'un fitxer d'àudio sense la intervenció d'una persona. Fan servir tecnologia de reconeixement automàtic de parla, que pot convertir en text un àudio sense conèixer-ne el contingut. Com ja hem esmentat, algunes eines de transcripció incorporen aquesta tecnologia (Transcribe n'és un exemple), però les que la inclouen no tenen el català entre els idiomes que es poden processar. De fet, només unes poques empreses disposen de tecnologia de reconeixement de parla en català (les grans companyies, com ara Nuance,<sup>5</sup> Apple,<sup>6</sup> Microsoft<sup>7</sup> o Google,<sup>8</sup> i una empresa catala-

4 Cf. <[http://transag.sourceforge.net/images/screenshot\\_1.jpg](http://transag.sourceforge.net/images/screenshot_1.jpg)>.

5 Cf. <<https://www.nuance.com/es-es/omni-channel-customer-engagement/voice-and-ivr/automatic-speech-recognition.html>>.

6 Cf. <<https://developer.apple.com/documentation/speech>>.

7 Cf. <<https://azure.microsoft.com/es-es/services/cognitive-services/speech-to-text/>>.

8 Cf. <<https://cloud.google.com/speech/?hl=es>>.



na, Verbio). Algunes d'elles ofereixen a través de la web la possibilitat de poder transcriure àudios en diferents idiomes, inclòs el català (Google, Microsoft) però són eines de pagament, algunes pensades específicament per a desenvolupadors i no per al gran públic. A més, tot i que són eines completament automàtiques, la qual cosa no fa necessària la intervenció d'un transcriptor humà, la sortida d'aquests sistemes sol contenir errors, especialment en el cas de parla espontània, amb soroll o disfluències, i per això la transcripció obtinguda ha de ser normalment revisada per una persona.

### ■ 2.3 Transcripció fonètica

En funció de l'ús que es vulgui donar a un corpus, el seu desenvolupament pot implicar també la transcripció fonètica del seu contingut, tasca que, com en el cas de la transcripció ortogràfica, pot ser llarga i pesada si es realitza manualment, especialment en el cas de corpus d'una mida gran. Una alternativa a la transcripció manual pot ser l'ús d'eines de transcripció fonètica automàtica de textos, que permeten obtenir la transcripció fonètica d'un text o d'una llista de paraules sense la intervenció d'una persona.

L'ús d'aquestes eines, però, presenta una sèrie d'inconvenients que és important tenir en compte i que farà que amb molta probabilitat la transcripció hagi de ser revisada manualment si volem tenir una transcripció acurada del que es diu al corpus. El principal és, potser, que la transcripció que s'obté és la 'teòrica', ja que es fa a partir de la transcripció ortogràfica del text, per la qual cosa no podem obtenir la pronunciació real de la persona que va emetre l'enunciat si per algun motiu (s'ha equivocat, per exemple) no ha realitzat la pronunciació esperada pel transcriptor automàtic, que normalment sol ser l'estàndard de la llengua en qüestió, tot i que també existeixen sistemes que permeten transcriure en determinades varietats dialectals.

Existeixen diferents eines de transcripció fonètica a partir de text de lliure distribució que poden fer-se servir per a aquesta tasca, com ara Segre (Pachès et al., 2000), TransText (Garrido, Codina i Fodge, 2019) o TransDic (Garrido, Codina i Fodge, 2018).

TransText és una eina de lliure distribució<sup>9</sup> per a la transcripció de textos en diferents dialectes del castellà i català. El nucli de regles lingüístiques va ser desenvolupat per investigadors del GLiCom (Grup de Lingüística Computacional) de la Universitat Pompeu Fabra, però l'aplicació es va desenvolupar al Laboratori de Fonètica de la Universitat de Barcelona.

---

9 Es pot descarregar des de <<https://sites.google.com/site/juanmariagarrido/research/resources/tools/trans-text-1>>.

TransText funciona per línia de comandes i s'ofereix en versions per a Linux i MacOS. Es pot fer servir amb textos del castellà i del català, i a més permet obtenir la transcripció tant en estàndard com en diferents dialectes, que en el cas del català són de l'àrea de Ribagorça (ca\_ri), del Pallars (ca\_pa), de Tortosa (ca\_to), l'Occidental central (ca\_ac), el de València septentrional (ca\_vs), el de València central (ca\_vc), el de València meridional (ca\_vm) i el d'Alacant (ca\_al). A més, es pot triar l'alfabet per a la transcripció (IPA o SAMPA). La taula 1 presenta un exemple de transcripció obtinguda amb aquesta eina en aquests dos alfabetes.

Text d'entrada	Per raons tècniques, aquest programa no es va emetre ahir, com s'havia anunciat. En el seu lloc es va tornar a emetre el programa anterior. Els demanem disculpes.
Transcripció IPA	pəʀ#rə.ˈoŋs#t'ɛk.ni.kəs ə.k'ɛt#pʀu.ɣr'a.mə#n'o#əs #p'a#ə.m'ɛ.trə#ə'i k'om#sə.β'i.ə#ə.nun.si.'at  ən #ə1#s'ew#l'ok#əs#p'a#tur.n'a#ə#ə.m'ɛ.trə#ə1#pʀ u.ɣr'a.mə#ən.tə.fi.'or  ə1s#tə.mə.n'ɛm#dis.k'ul.pə s
Transcripció SAMPA	p@r#r@."oŋs#t"Ek.ni.k@s...@.k"Et#pʀu.Gr"a.m@# n"o#@#s#p"a#@.m"E.tr@#@i...k"Om#s@.B"i.@#@. nun.si."at...@n#@1#s"ew#L"Ok#@#p"a#tur.n"a#@ #@.m"E.tr@#@1#pʀu.Gr"a.m@#@n.t@.ri."or...@1s# t@.m@.n"Em#dis.k"ul.p@s...

Taula 1. Exemple de transcripció obtinguda amb TransText per a un text de mostra en català estàndard

La figura 2 (pàgina següent) mostra l'estructura interna de TransText, amb els diferents nivells o capes que incorpora: el nucli de processament i els mòduls específics per a cada idioma o dialecte, més interns, i una capa superior d'aplicació, que inclou la interfície d'usuari.

Per la seva banda, TransDic és també una eina de transcripció de lliure distribució<sup>10</sup> que permet la transcripció en castellà i català, tant estàndard com dialectal, però presenta dues diferències fonamentals amb TransText: la primera, que està especialment pensada per al desenvolupament de diccionaris fonetitzats, per la qual cosa admet com a entrada llistes de paraules

10 Es pot descarregar des de <<https://sites.google.com/site/juanmariagarrido/research/resources/tools/transdic>>.



Figura 2. Estructura de TransText

i no text; i la segona, que permet obtenir diferents variants de transcripció per a una mateixa paraula, funcionalitat que la fa especialment útil per al desenvolupament de diccionaris per al reconeixement de parla. Com TransText, és una eina multiplataforma (Linux, Mac OS, Windows), funciona per línia de comandes i permet obtenir la transcripció tant en alfabet IPA com en SAMPA. La taula 2 presenta un exemple del format de sortida de TransDic.

boxejar	bok.sed.3'a	bok.sej.3'a	bok.sed.3'ar	bok.sej.3'ar
boxers	bok.s'es	bok.s'ers		
branca	br'an.ka			
braçalet	bra.sa.l'et			
briòix	bri.'ojs			
brisa	br'i.za			

Taula 2. Mostra de diccionari generat amb TransDic (format per defecte)

## ■ 2.4 Anotació

Els corpus orals poden ser anotats amb informació molt variada, depenent dels usos que se'n vulguin fer. En el cas d'un corpus orientat al treball en Fonètica acústica, aquesta informació pot ser:

- Informació prosòdica:
  - Unitats prosòdiques (síl·labes, grups entonatius majors i menors, paraules fonològiques...)
  - Fenòmens prosòdics (accent, tons, entonació)
- Informació lingüística
  - Informació morfològica (categories lèxiques)
  - Informació sintàctica (estructura de constituents)
  - Informació pragmàtica (modalitat oracional, actes de parla, informació nova/coneguda, focus)
- Informació paralingüística:
  - Emocions

Aquestes tasques es poden automatitzar totalment o parcialment amb l'ajut de diverses eines.

### ■ 2.4.1 Anotació prosòdica

Existeixen, en primer lloc, eines per a la identificació i anotació dels límits de les unitats prosòdiques (síl·labes, grups entonatius...). Aquestes eines permeten anotar les unitats prosòdiques d'un enunciat a partir de la transcripció fonètica de l'enunciat o bé de l'anàlisi acústica del senyal de parla. Un exemple d'aquest tipus d'eines és SegProso (Garrido, 2013), una eina de lliure distribució<sup>11</sup> per a l'anotació automàtica de les unitats prosòdiques de corpus de parla desenvolupada amb Praat (Boersma, 2001). SegProso identifica quatre tipus d'unitats: síl·labes, grups accentuals, grups entonatius majors i grups entonatius menors. L'eina dona com a sortida un TextGrid de Praat amb l'anotació alineada amb el senyal de parla. Desenvolupada inicialment per al castellà i el català, es va adaptar posteriorment al portuguès de Brasil (Silva i Garrido, 2016), el mandarí i el francès. La figura 3 mostra un exemple de TextGrid de sortida obtingut amb aquesta eina.

---

11 Està disponible per a descàrrega des de <<https://sites.google.com/site/juanmaria-garrido/research/resources/tools/segproso>>.

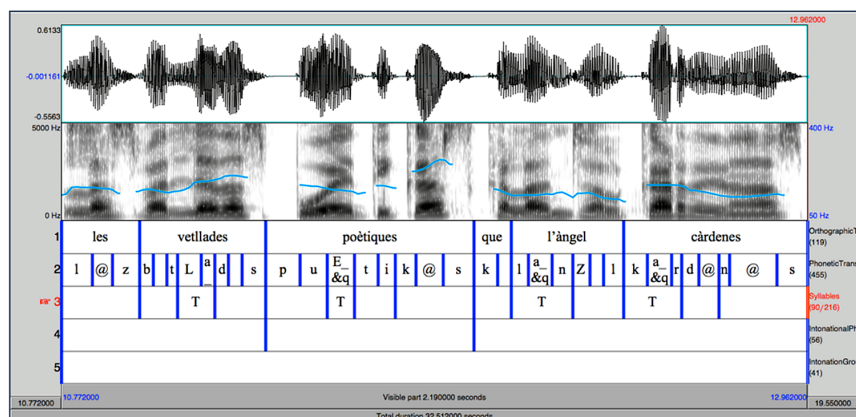


Figura 3. TextGrid i forma d’ona corresponent a l’enunciat “*les vetllades poètiques que l’Àngel Càrdenas*”, pronunciat per una locutora catalana professional (corpus Glissando)

Encara dins del nivell prosòdic, també hi ha eines per a l’anotació automàtica de fenòmens prosòdics com l’entonació. En aquest àmbit és possible trobar eines per a l’anotació de material en català, com *Anàlisi Melòdica del Hable* (Mateo, 2010), MelAn (Garrido, 2010) o FonetToBI (Elvira i Garrido, 2019).

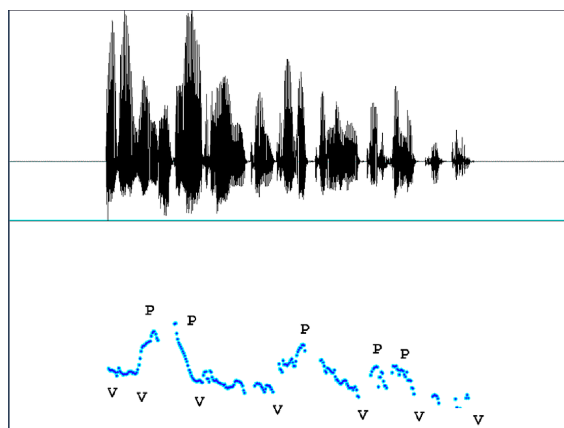


Figura 4. Ona sonora i corba d’F0 anotada de l’enunciat “*Aragón se ha reencontrado como motor del equipo*”, pronunciat per una parlant de castellà

MelAn és una eina, també de lliure distribució,<sup>12</sup> que està formada per un conjunt de scripts de Praat per a la anotació simbòlica de contorns d’F0 basada en el model de descripció prosòdica proposat a Garrido (1996, 2001). La figura 4 (pàgina anterior) mostra un exemple d’anotació d’una corba d’F0 amb el sistema emprat en aquest model, al qual els punts rellevants de la corba són etiquetats amb *P* o *V* segons siguin ‘pics’ o ‘valls’ a la corba.

El procés d’anotació es realitza a MelAn en dues fases: una primera d’estilització de la corba d’F0, que redueix la corba d’F0 original a un conjunt de punts d’inflexió rellevants, i que està basada en la funcionalitat d’estilització de Praat; i una segona d’anotació, durant la qual es realitza l’etiquetatge dels punts d’inflexió obtinguts a la primera fase amb etiquetes que representen la seva alçada tonal dins del grup entonatiu. El resultat es desa en un fitxer TextGrid de Praat com el que es presenta a la figura 5.

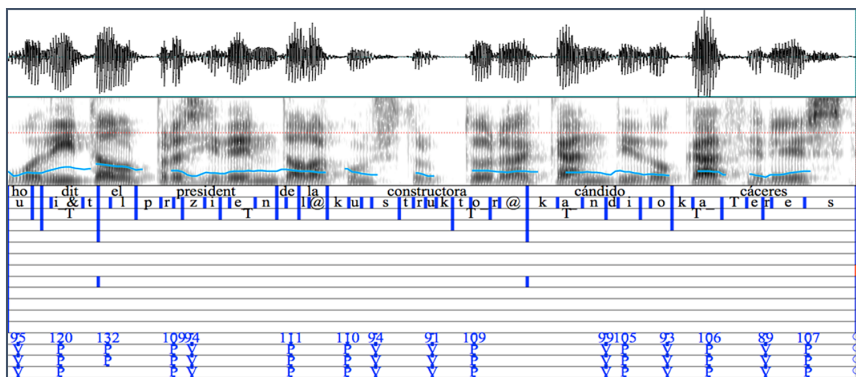


Figura 5. Ona sonora, contorn d’F0, espectrograma i anotació de l’enunciat “*ho ha dit el president de la constructora, Cándido Cáceres*”, pronunciat per una parlant de català central

Una de les característiques d’aquest sistema d’anotació és que pretén capturar els moviments perceptivament rellevants de la corba d’F0, de forma que la corba estilitzada que se’n obté sigui el més similar possible des del punt de vista perceptiu a la corba original. La figura 6 exemplifica aquesta similitud: en ella es presenten dues versions de la corba d’F0 del mateix

<sup>12</sup> Està disponible per a descàrrega des de <[https://github.com/juanmariagarrido/MelAn/tree/Version\\_1\\_1](https://github.com/juanmariagarrido/MelAn/tree/Version_1_1)>.

enunciat, la original (a la esquerra) i l'estilitzada (a la dreta). La similitud perceptiva entre les dues versions va ser verificada experimentalment mitjançant una prova de percepció descrita a Garrido (2010).

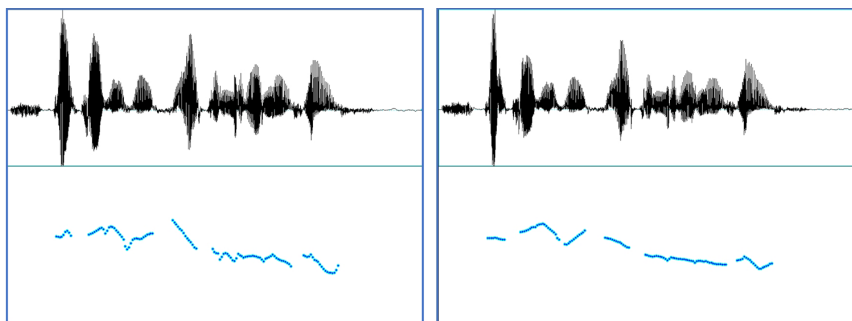


Figura 6. Ones sonores i contorns d’F0 de dues versions sintètiques de l’enunciat “Y cada vez la tendremos más”: amb el contorn d’F0 original (esquerra) i estilitzada (dreta)

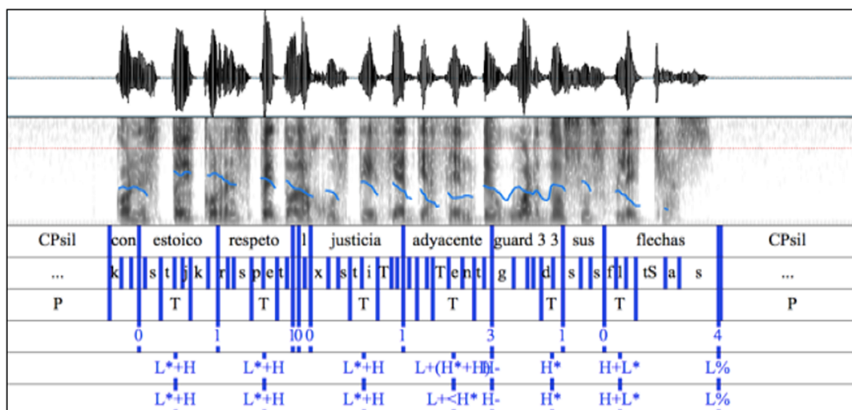


Figura 7. Ona sonora, espectrograma, corba d’fo i anotació generada per FonetíToBI corresponent a l’enunciat del castellà “fue inyectado en el abdomen y en una pierna”, pronunciat per una parlant

FonetíToBI és una eina de lliure distribució<sup>13</sup> per a la anotació prosòdica automàtica de parla en el marc del Model Mètric Autosegmental (AM) (Pierrehumbert, 1980). Es tracta de la primera eina que permet l'anotació automàtica completa dels enunciats en el marc de les convencions ToBI (*Tones and Break Indices*). Ha estat dissenyada per al seu ús tant amb enunciats en castellà com en català, i és el resultat de la integració de dues eines existents prèviament: SegProso i EtíToBI (Elvira et al., 2016), eina per a la anotació de tons. És també una eina de lliure distribució. La figura 7 (pàgina anterior) mostra un exemple de fitxer TextGrid de sortida.

## ■ 2.4.2 Anotació lingüística i paralingüística

L'anotació d'un corpus de parla amb informació lingüística és necessària si es vol explotar el corpus per fer recerca sobre la relació entre el component fònic de les llengües i els diferents nivells gramaticals o pragmàtics. Aquesta anotació pot ser llavors molt variada i molt dependent de l'ús final que es vulgui donar al corpus, i pot incloure anotacions sobre les categories lèxiques de les paraules, l'estructura de constituents o els actes de parla. També pot ser interessant anotar elements no estrictament lingüístics, com les emocions, actituds...

Per automatitzar els processos d'anotació lingüística es pot recórrer a eines de processament de text, que extreguin aquesta informació, en la mesura del possible, de la transcripció ortogràfica del contingut del corpus. Existeixen moltes eines d'aquest tipus per a moltes llengües; no tantes, en canvi, pel que fa al català. Dos exemples d'eines de processament de text que permeten processar text en català són FreeLing (Carreras et al., 2004), desenvolupat a la Universitat Politècnica de Catalunya, i TextAFon (Garrido et al., 2012 i 2014). Ens detindrem en la descripció d'aquesta segona.

TextAFon és un conjunt d'eines de processament automàtic de text per al català i el castellà desenvolupat al GLiCom de la Universitat Pompeu Fabra. Inicialment concebuda com a eina de processament de text per a tasques de síntesi de veu, permet realitzar tasques com ara la normalització de textos, la transcripció fonètica, la sil·labificació, la segmentació d'un text en grups entonatius o la predicció d'accent lèxic. La figura 8 presenta un exemple de la sortida que ofereix aquesta eina en el seu funcionament bàsic.

---

13 Està disponible per a la seva descàrrega des de <<https://github.com/wendylviragarcia/FonetíToBI>>.



(1) Input	Esta es una demostración de TexAFon para IberSpeech 2012. TexAfon ha sido desarrollado conjuntamente por el Grupo de Lingüística Computacional (GLiCom) de la UPF y el Grupo de Voz y Lenguaje de Barcelona Media. Hoy es 11/06/2012, y son las 10:15.
(2) Segmentation into sentences	Esta es una demostración de TexAFon para IberSpeech 2012. TexAfon ha sido desarrollado conjuntamente por el Grupo de Lingüística Computacional (GLiCom) de la UPF y el Grupo de Voz y Lenguaje de Barcelona Media. Hoy es 11/06/2012, y son las 10:15.
(3) Segmentation into IG	Esta es una demostración de TexAFon para IberSpeech 2012 [signo_punto] TexAfon ha sido desarrollado conjuntamente por el Grupo de Lingüística Computacional [signo_abre_parentesis] GLiCom [signo_cierra_parentesis] de la UPF y el Grupo de Voz y Lenguaje de Barcelona Media [signo_punto] Hoy es 11/06/2012 [signo_coma] y son las 10:15 [signo_punto]
(4) Normalisation	esta es una demostración de texafon para iberspeech dos mil doce [signo_punto] texafon ha sido desarrollado conjuntamente por el grupo de lingüística computacional [signo_abre_parentesis] glicom [signo_cierra_parentesis] de la u pe efe y el grupo de voz y lenguaje de barcelona media [signo_punto] hoy es once de junio de dos mil doce [signo_coma] y son las diez quince [signo_punto]
(5) Phonetic transcription	e_&quot; s . t a # e_&quot; s # u_&quot; . n a # d e . m o s . t r a . T j o_&quot; N # d e # t e k . s a_&quot; . f o N # p a . r a # i . b e r s . p e . e_&quot; t S # d o s # m i_&quot; l # d o_&quot; . T e [signo_punto] t e k . s a_&quot; . f o N # a_&quot; # s i_&quot; . d o # d e . s a . r r o . L a_&quot; . d o # k o n . x u_&quot; n . t a . m e_&quot; n . t e # p o r # e l # g r u_&quot; . p o # d e # l i n . g w i_&quot; s . t i . k a # k o m . p u . t a . T j o . n a_&quot; l [signo_abre_parentesis] g l i . k o_&quot; m [signo_cierra_parentesis] d e # l a # u # p e_&quot; # e_&quot; . f e # i # e l # g r u_&quot; . p o # d e # b o_&quot; T # i # l e n . g w a_&quot; . x e # d e # b a r . T e . l o_&quot; . n a # m e_&quot; . d j a [signo_punto] o_&quot; j # e_&quot; s # o_&quot; n . T e # d e # x u_&quot; . n j o # d e # d o s # m i_&quot; l # d o_&quot; . T e [signo_coma] i # s o_&quot; N # l a s # d j e_&quot; T # k i_&quot; n . T e [signo_punto]

Figura 8. Sortida de TexAFon

TexAFon va ser desenvolupat completament en Python per lingüistes computacionals, i s'hi han anat afegint mòduls per a la detecció automàtica en text d'elements com ara la modalitat de les oracions d'un text, els actes de parla o les emocions. Els mòduls encarregats de la predicció dels actes de parla i les emocions, descrits a Garrido et al. (2014) i Kolz, Garrido i Laplaza (2014), són eines experimentals, que encara ofereixen resultats millorables amb un conjunt limitat d'etiquetes d'actes de parla i d'emocions.

### ■ 2.4.3 Alineació de l'anotació amb el senyal de parla

Una tasca habitual en el desenvolupament de corpus de parla, especialment d'aquells especialment orientats als estudis fonètics, és l'alineació de la transcripció i anotació del corpus amb el temps, per poder presentar-lo de forma sincronitzada amb l'ona sonora de l'enregistrament. Aquesta tasca es pot fer manualment amb l'ajut d'eines com Praat, però novament es presenta com una tasca inabordable per aquesta via si es vol fer amb corpus de mides molt grans. Per alleugerir aquest procés existeixen eines per a la alineació automàtica del senyal d'àudio, que permeten alinear la transcripció ortogràfica i fonètica d'un enunciat, i altres tipus d'anotacions, amb el fitxer d'àudio corresponent. Requereixen com entrada un fitxer d'àudio i la seva corresponent transcripció ortogràfica i ofereixen com a sortida un fitxer amb la transcripció fonètica alineada amb el senyal (en format Textgrid de Praat habitualment). Algunes d'aquestes eines permeten també el processament de material en català, com ara MAUS (Schiel, 1999),<sup>14</sup> WebMAUS,<sup>15</sup> EasyAlign (Goldman, 2011)<sup>16</sup> o SPPAS (Bigi, 2012).<sup>17</sup>

SPPAS és una eina desenvolupada al *Laboratoire Parole et Langage* d'Aix-en-Provence (França) per Brigitte Bigi. Permet obtenir diferents nivells d'anotació fonètica, alineats amb el senyal: transcripció fonètica, segmentació sil·làbica, segmentació en grups entonatius i anotació prosòdica (Mo-Mel). La versió en català va ser desenvolupada en col·laboració amb el GLiCom de la Universitat Pompeu Fabra, fent servir la versió en català del corpus Glissando (Garrido et al., 2013). La figura 9 presenta un exemple de fitxer de sortida generat amb aquesta eina.

14 Cf. <<http://www.bas.uni-muenchen.de/Bas/BasMAUS.html>>.

15 Cf. <<https://clarin.phonetik.uni-muenchen.de/BASWebServices/index.html>>.

16 Cf. <<http://latcui.unige.ch/phonetique/easyalign.php>>.

17 Cf. <[http://sldr.org/SLDR\\_data/Disk0/preview/000800/?lang=en](http://sldr.org/SLDR_data/Disk0/preview/000800/?lang=en)>.

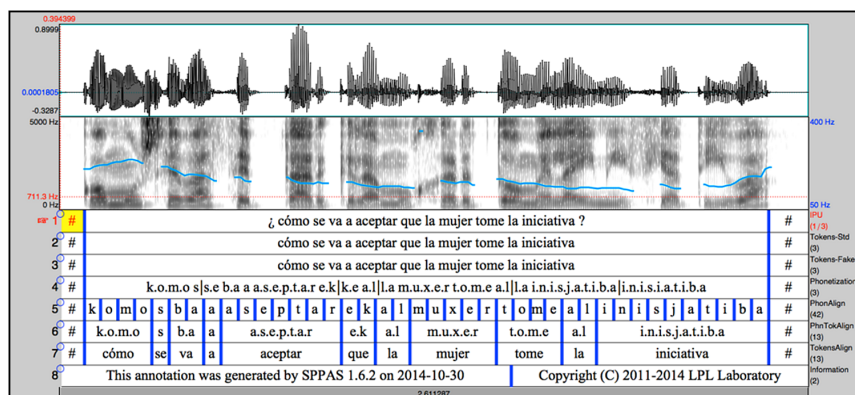


Figura 9. Fitxer TextGrid amb la transcripció fonètica i l’ anotació prosòdica obtinguda amb SPPAS per a l’ enunciat “¿Cómo se va a aceptar que la mujer tome la iniciativa?” pronunciada per una parlant de castellà

Com és habitual amb aquest tipus d’ eines, la seva precisió no és completa, però ofereixen una sortida prou acceptable com per obtenir una primera alineació que després sigui revisada per mitjans manuals.

### ■ 3 Anàlisi de corpus

L’ anàlisi de dades d’ un corpus de forma automatitzada és una línia de recerca emergent en els últims anys en Fonètica, en la línia d’ altres disciplines científiques, que cada cop fan servir més eines computacionals, basades en tecnologies d’ aprenentatge automàtic i intel·ligència artificial, per obtenir resultats a partir del processat de grans quantitats de dades. En l’ àmbit concret de l’ anàlisi de l’ entonació, existeixen ja algunes eines que permeten realitzar estudis a partir del processament de corpus prèviament anotats. Algunes estan preparades per manegar dades del català, com per exemple MelAn (Garrido, 2010) o ProDis (Elvira et al., 2018).

MelAn, a més de les funcionalitats d’ anotació ja descrites en l’ apartat anterior, pot ser utilitzada també per a l’ anàlisi i modelització de contorns entonatius, concretament per a l’ extracció de l’ inventari complet de patrons entonatius del corpus d’ anàlisi. L’ inventari que ofereix es basa en el model descriptiu de l’ entonació proposat a Garrido (1996, 2001), que diferencia entre dos tipus de patrons: globals, que modelen el rang i l’ alçada tonals i es defineixen en l’ àmbit dels grups entonatius; i locals, que mode-

len el patrons tonals lligats a l'accent i l'entonació i es defineixen en l'àmbit del grup accentual. Les figures 10 i 11 presenten exemples d'aquest dos tipus de patrons.

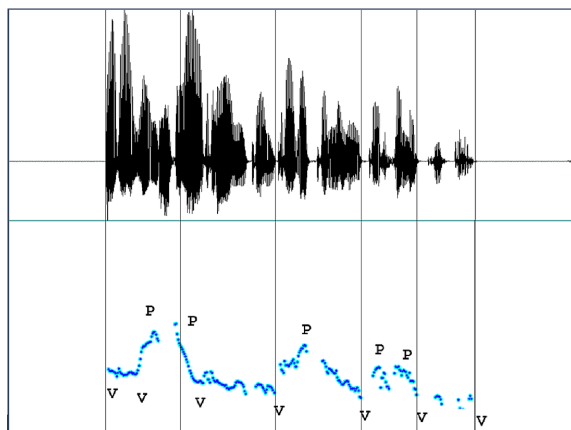


Figura 10. Ona sonora i contorn d'F0 anotat corresponent a l'enunciat “*Aragón se ha reencontrado como motor del equipo*”, pronunciat per una parlant de castellà. Les línies inclinades representen els patrons globals de l'enunciat

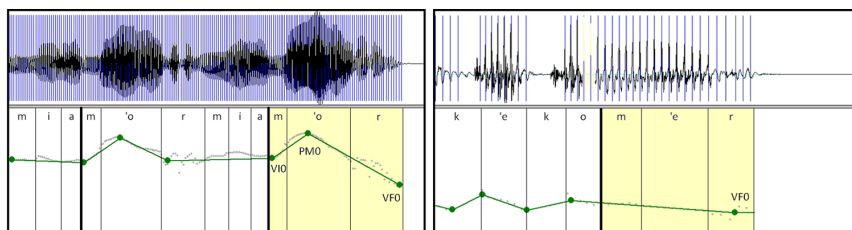


Figura 11. Exemple de dos patrons entonatius finals etiquetats amb MelAn (Garrido et. al., 2012)

La sortida d'aquesta anàlisi es, d'una banda, un inventari dels patrons locals trobats al corpus, amb indicació de la seva freqüència (taula 3) i dels patrons globals de referència, calculats per a diferents condicions, com ara la posició dins de la oració, o la modalitat oracional (taula 4).

Patró	Num. síl·labes	Posició síl·laba accentuada	Modalitat oracional	Emoció	Acte parla	Posició GA	Num. aparicions patró
0	1	1	INDEF	INDEF	INDEF	INTERIOR	382
VI0_PM0	1	1	INDEF	INDEF	INDEF	INTERIOR	172
PM0	1	1	INDEF	INDEF	INDEF	INTERIOR	140
PI0	1	1	INDEF	INDEF	INDEF	INTERIOR	134
VF0	1	1	INDEF	INDEF	INDEF	INTERIOR	123
0	2	1	INDEF	INDEF	INDEF	INTERIOR	205
PI0	2	1	INDEF	INDEF	INDEF	INTERIOR	100
PI1	2	1	INDEF	INDEF	INDEF	INTERIOR	90
VI0_PM0	2	1	INDEF	INDEF	INDEF	INTERIOR	82
PM0	2	1	INDEF	INDEF	INDEF	INTERIOR	69

Taula 3. Fragment de fitxer de sortida amb la llista dels patrons obtinguts de l'anàlisi d'una part del subcorpus de notícies del corpus Glissando

Sentence Mood	IG Position	Number of Cases	Mean Initial Value P	Mean Final Value P	Mean Slope P	Mean Initial Value V	Mean Final Value V	Mean Slope V
ENUNCIATIVA	FINAL	404	141.8	91.1	-37.5	109	69.2	-29.3
ENUNCIATIVA	INICIAL	1946	149.5	109.8	-22.9	112.9	85	-12.0
ENUNCIATIVA	INICIAL_FINAL	363	158.9	90.5	-45.3	120.2	69.1	-33.7
ENUNCIATIVA	INTERIOR	1823	137.5	106.1	-18.1	108.1	80.7	-14.7
EXCLAMATIVA	FINAL	2	238.5	62	-157	129.8	84.6	-44.3
EXCLAMATIVA	INICIAL	84	181.1	124	-49.2	134.5	86	-43.5
EXCLAMATIVA	INICIAL_FINAL	52	204.9	118.9	-93.9	141.1	86.4	-49.3
EXCLAMATIVA	INTERIOR	5	186.8	99.3	-81.3	139.5	80.2	-55s

Taula 4. Mostra d'un fitxer de sortida de MelAn que conté la llista de patrons globals obtinguda a partir de l'anàlisi d'una part del subcorpus de notícies del corpus Glissando

#### ■ 4 Exemple d'aplicació: Glissando

Un exemple d'aplicació d'aquestes eines al desenvolupament d'un corpus de parla és Glissando, un corpus desenvolupat especialment per als estudis prosòdics, del qual es van desenvolupar dues versions paral·leles, una en castellà i una altra en català, per la qual cosa es pot fer servir també per a

estudis comparatius. És un corpus de lliure distribució, disponible a través d'ELRA (Glissando-sp<sup>18</sup> i Glissando-ca<sup>19</sup>).

El corpus Glissando inclou tres subcorpus, corresponents a tres estils de parla diferents: notícies, diàlegs orientats a objectiu i diàlegs informals. El total de material enregistrat per a cada llengua supera les 12 hores. A més dels enregistraments, el corpus inclou també diferents nivells d'anotació, alineada en el temps amb l'ona sonora, en fitxers TextGrid de Praat:

- Transcripció ortogràfica (paraules)
- Transcripció fonètica (SAMPA)
- Segmentació en unitats prosòdiques:
  - Síl·labes
  - Grups accentuals
  - Grups entonatus majors
  - Grups entonatus menors
- Anotació de l'entonació
- Anotació parcial de la informació lingüística

El desenvolupament de tot aquest material no hauria estat possible sense l'ús d'eines automàtiques en diferents fases, com ara la selecció dels textos de notícies o l'anotació. En el cas concret de la transcripció i anotació, el procés va implicar dues fases: una automàtica, durant la qual es van fer servir diferents eines de processament per obtenir una primera versió de la transcripció i anotació; i una altra de manual, en la que un equip de revisors entrenats especialment per a aquesta tasca van revisar i corregir la sortida obtinguda de forma automàtica.

Les eines que es van fer servir per als diferents tipus d'anotació van ser:

- Una eina per a la transcripció, segmentació i alineament proporcionada per Barcelona Media, en el cas de la transcripció fonètica.
- SegProso, per a la segmentació en unitats prosòdiques.

Les figures 12 i 13 mostren exemples de fragments del corpus de notícies i de diàlegs orientats a objectiu, respectivament, amb l'anotació obtinguda amb aquest procediment.

---

18 Cf. <<http://catalog.elra.info/en-us/repository/browse/ELRA-S0406/>>.

19 Cf. <<http://catalog.elra.info/en-us/repository/browse/ELRA-S0407/>>.

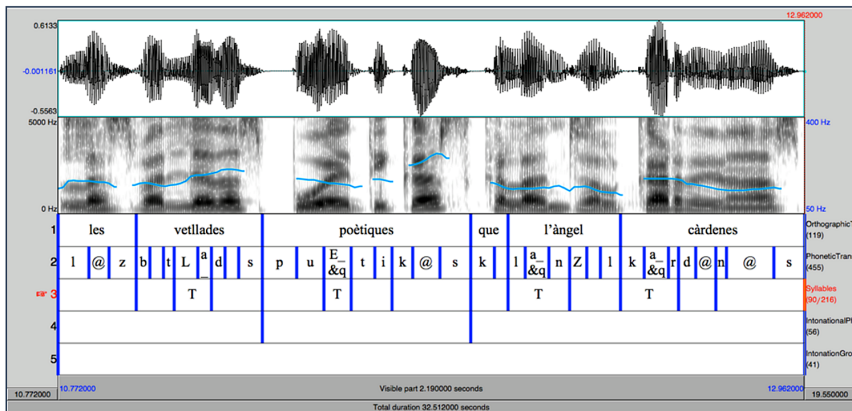


Figura 12. TextGrid i ona sonora corresponent a l'enunciat “*les vetllades poètiques que l'Àngel Càrdenes*”, pronunciat per una parlant de català

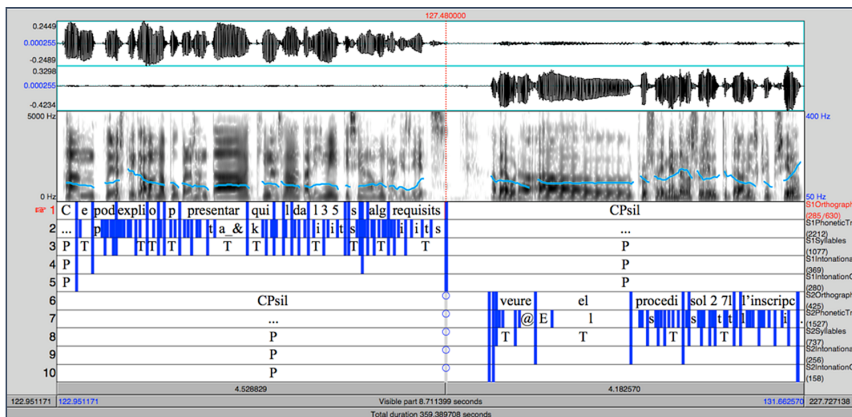


Figura 13. TextGrid i ona sonora corresponent a un fragment de diàleg orientat a objectiu entre dos parlants de català

Per a l'explotació del corpus amb fins de recerca també s'han fet servir eines automàtiques, com ara MelAn, que han permès realitzar anàlisis sobre la entonació del castellà (Garrido i Rustullet, 2011; Garrido, Laplaza i García, 2017) i del català (Garrido, 2012).

## ■ 5 Conclusions

La revisió feta en aquest treball permet concloure que, tot i que poques, hi ha eines automàtiques per a determinades fases del desenvolupament i explotació de corpus orals en català, com ara la transcripció fonètica, l'annotació prosòdica, l'annotació lingüística o l'alineació amb el senyal acústic. Hi ha d'altres àmbits, com la transcripció automàtica, on les eines no estan disponibles per al gran públic. També hi ha eines per a l'anàlisi de dades de corpus orals que permeten la realització d'estudis fonètics basats en gran quantitat de dades, almenys en àmbits com l'anàlisi prosòdica. Aquestes eines permeten alleugerir moltes fases del procés, però en el seu estat de desenvolupament encara requereixen una fase posterior de revisió manual per obtenir un resultat de qualitat. I hi ha àmbits, com l'annotació lingüística, on encara s'ha de recórrer a l'annotació manual donada la manca d'eines mínimament fiables. Cal encara, doncs, molts esforços de recerca per millorar el rendiment de les eines existents si es vol arribar a eliminar les tasques de revisió manual, i també per desenvolupar noves eines que permetin realitzar tasques encara no cobertes, com en el cas de l'annotació lingüística. I especialment pel que fa al català, si es compara el nombre d'eines existents amb el d'altres llengües amb més pes econòmic. ■

## ■ Referències bibliogràfiques

- Apple: “Speech”, <<https://developer.apple.com/documentation/speech>> [08.03.2020].
- Bigi, Brigitte (2012): «SPPAS: a tool for the phonetic segmentations of Speech», *The Eight International Conference on Language Resources and Evaluation, Istanbul (Turkey)*, Istanbul: ELRA, 1748–1755.
- Boersma, Paul (2001): «Praat, a system for doing phonetics by computer», *Glott International* 5:9/10, 341–345.
- Carreras, Xavier / Chao, Isaac / Padró, Lluís / Padró, Muntsa (2004): «FreeLing: an open-source suite of language analyzers», *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*, Lisboa: ELRA, 239–242.
- EasyAlign: “Phonetic alignment with Praat”, <<http://latlucui.unige.ch/phonetique/easyalign.php>> [08.03.2020].



- Elvira-García, Wendy / Roseano, Paolo / Fernández Planas, Ana M. / Martínez Celdrán Eugenio (2016): «A tool for automatic transcription of intonation: Eti-ToBI a ToBI transcriber for Spanish and Catalan», *Language Resources and Evaluation* 50:4, 767–792.
- / Balocco, Simone / Roseano, Paolo / Fernández Planas, Ana M. (2018): «ProDis: A dialectometric tool for acoustic prosodic data», *Speech Communication* 97, 9–18.
- / Garrido, Juan M. (2019): «FonetiToBI, una herramienta para la anotación prosódica automática de corpus», in: Lahoz, José M. / Pérez, Rubén (eds.): *Subsidia: Tools and Resources for Speech Sciences*, Málaga: Universidad de Málaga, 133–142.
- Escudero, David / González, César / Garrido, Juan M. / Rodero, Emma / Aguilar, Lourdes / Bonafonte, Antonio (2010): «Combining greedy algorithms with expert guided manipulation for the definition of a balanced prosodic Spanish-Catalan radio news corpus», *Speech Prosody 2010, Chicago, May 2010*, Chicago: Speech Prosody Special Interest Group, 100061, 1–4.
- FonetiToBi: “Praat-based tool for prosodic annotation in ToBI format”, <<https://github.com/wendyelviragarcia/FonetiToBI>> [08.03.2020].
- Garrido, Juan M. (1996): *Modelling Spanish Intonation for Text-to-Speech Applications*, Barcelona: Universitat Autònoma de Barcelona (tesi doctoral).
- (2001): «La estructura de las curvas melódicas del español: propuesta de modelización», *Lingüística Española Actual* XXIII:2, 173–209.
- (2010): «A Tool for Automatic F0 Stylistation, Annotation and Modelling of Large Corpora», *Speech Prosody 2010, Chicago, May 2010*, Chicago: Speech Prosody Special Interest Group, 100041, 1–4.
- (2012): «Glissando, un corpus anotat per a l’anàlisi de la prosòdia del català i del castellà. Descripció i primers resultats d’exploració», *Phonica* 8, 3–16.
- (2013): «SegProso: A Praat-Based tool for the Automatic Detection and Annotation of Prosodic Boundaries», *Proceedings of TRASP 2013*, 74–77.
- / Rustullet, Sílvia (2011): «Patrones melódicos en el habla de diálogo en español: un primer análisis del corpus Glissando», *Oralia: Análisis del discurso oral* 14, 129–160.

- / Laplaza, Yésika / Marquina, Montserrat (2012): «On the use of melodic patterns as prosodic correlates of emotion in Spanish», *Speech Prosody 2012, Shanghai*, Xangai: Tongji University Press, 645–648.
- / — / — / Schoenfelder, Claudia / Rustullet, Sílvia (2012): «TexAFon: a multilingual text processing tool for text-to-speech applications», *Proceedings of IberSpeech 2012, Madrid, Spain, November 21–23, 2012*, Madrid: Universidad Autónoma de Madrid, 281–289.
- / Escudero, David / Aguilar, Lourdes / Cardenoso, Valentín / Roderó, Emma / de-la-Mota, Carme / González, César / Rustullet, Sílvia / Larrera, Olatz / Laplaza, Yésika / Vizcaíno, Francisco / Cabrera, Mercedes / Bonafonte, Antonio (2013): «Glissando: a corpus for multidisciplinary prosodic studies in Spanish and Catalan», *Language Resources and Evaluation* 47:4, 945–971.
- / Laplaza, Yésika / Kolz, Benjamin / Cornudella, Miquel (2014): «TexAFon 2.0: A text processing tool for the generation of expressive speech in TTS applications», *Proceedings of LREC 2014, Ninth International Conference on Language Resources and Evaluation, Reykjavik (Iceland), 26–31 May 2014*, París: ELRA, 3494–3500.
- / — / García, Clara Luna (2017): «Caracterización pragmática y prosódica de la vocalización “mmm” en español», in: Marrero, Victoria / Estebas, Eva (coords.): *Tendencias actuales en fonética experimental. Cruce de disciplinas en el centenario del Manual de Pronunciación Española (Tomás Navarro Tomás)*, Madrid: UNED, 125–129.
- / Codina, Marta / Fodge, Kimber (2018): «TransDic, a public domain tool for the generation of phonetic dictionaries in standard and dialectal Spanish and Catalan», *IberSPEECH 2018 Proceedings*, Barcelona: Telefónica / Universitat Politècnica de Catalunya / La Salle, 291–295.
- / — / — (2019): «TransText, un transcriptor fonético automático de libre distribución para español y catalán», in: Lahoz, José M. / Pérez, Rubén (eds.): *Subsidia: Tools and Resources for Speech Sciences*, Málaga: Universidad de Málaga, 27–31.
- Glissando-ca: <<http://catalog.elra.info/en-us/repository/browse/ELRA-S0407/>> [08.03.2020]
- Glissando-sp: <<http://catalog.elra.info/en-us/repository/browse/ELRA-S0406/>> [08.03.2020].

- Goldman, Philippe (2011): «EasyAlign: an automatic phonetic alignment tool under Praat», *Proceedings of InterSpeech, September 2011*, Firenze: International Speech Communication Association, 3233–3236.
- Google: “Reconocimiento de voz avanzado”, <<https://cloud.google.com/speech/?hl=es>> [08.03.2020].
- Kolz, Benjamin / Garrido, Juan M. / Laplaza, Yésika (2014): «Automatic prediction of emotions from text in Spanish for expressive speech synthesis in the chat domain», *Procesamiento del Lenguaje Natural* 52, 61–68.
- Mateo, Miguel (2010): *Análisis melódico del habla: protocolo para la automatización de la obtención de los datos de la curva estándar*, Barcelona: Universitat de Barcelona (treball de fi de Màster).
- MAUS: “The Munich Automatic Segmentation System”, <<http://www.bas.uni-muenchen.de/Bas/BasMAUS.html>> [08.03.2020].
- MelAn: “Script to label the F0 contour of a wav file with tonal events labels and to extract its melodic patterns”, <[https://github.com/juanmariagarrido/MelAn/tree/Version\\_1\\_1](https://github.com/juanmariagarrido/MelAn/tree/Version_1_1)> [08.03.2020].
- Microsoft: “Reconocimiento de voz avanzado”, <<https://azure.microsoft.com/es-es/services/cognitive-services/speech-to-text/>> [08.03.2020].
- Nuance: “Reconocimiento de voz automático: soluciones de reconocimiento de voz perfeccionadas por IA”, <<https://www.nuance.com/es-es/omni-channel-customer-engagement/voice-and-ivr/automatic-speech-recognition.html>> [08.03.2020].
- Pachès, Pau et al. (2000): «Segre: An automatic tool for grapheme-to-allophone transcription in Catalan», *Proceedings of the Workshop on Developing Language Resources for Minority Languages: Reusability and Strategic Priorities (LREC-2000)*, Atenes: Institute for Language and Speech Processing (ILSP) / National Technical University of Athens, 52–61.
- Pierrehumbert, Janet (1980): *The phonology and phonetics of English intonation*. Cambridge, MA: Massachusetts Institute of Technology (tesi doctoral).
- Schiel, Florian (1999): «Automatic phonetic transcription of non-prompted speech», *Proceedings of the ICPhS 1999. San Francisco, August 1999*, Berkeley: University of California, 607–610.
- SegProso: “Praat-based tool for the automatic segmentation of speech corpora into prosodic units”, <<https://sites.google.com/site/juanmariagarrido/research/resources/tools/segproso>> [08.03.2020].

Silva, Cristiane / Garrido, Juan M. (2016): «Validación perceptiva de dos procedimientos de representación de la melodía aplicados al español, al portugués brasileño y al español como lengua extranjera», *Onomázein* 34:2, 242–260.

SPPAS: “The automatic annotation and analysis speech”, <<http://www.sppas.org/>> [08.03.2020].

Transcriber: “Transcriber. A tool a tool for segmenting, labeling and transcribing speech”, <<http://trans.sourceforge.net/en/presentation.php>> [08.03.2020].

TranscriberAG: “TranscriberAG. A tool a tool for segmenting, labeling and transcribing speech”, <<http://transag.sourceforge.net/>> [08.03.2020].

Transcribe: “Transcribe”, <<https://transcribe.wreally.com/>> [08.03.2020].

TransDic: “GNU tool to create phonetic dictionaries from word lists in Spanish or Catalan”, <<https://sites.google.com/site/juanmariagarrido/research/resources/tools/transdic>> [08.03.2020].

TransText: “GNU tool to generate phonetic transcriptions of texts in Spanish or Catalan”, <<https://sites.google.com/site/juanmariagarrido/research/resources/tools/transtext-1>> [08.03.2020].

WebMAUS: “Bavarian Archive for Speech Signals”, <<https://clarin.phonetik.uni-muenchen.de/BASWebServices/index.html>> [08.03.2020].

■ Juan María Garrido Almiñana, UNED, Laboratorio de Fonética Antonio Quilis, Facultad de Filología, Paseo Senda del Rey, 7, E-28040 Madrid, <[jmgarrido@flog.uned.es](mailto:jmgarrido@flog.uned.es)>.