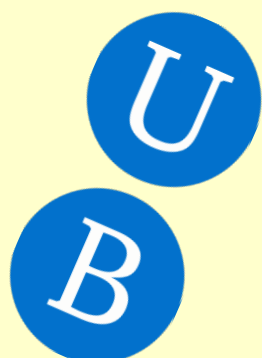


GUIA DE CONTINGUT DIGITAL ACCESSIBLE: VÍDEO

*Introducció a la
conversió d'àudio a text*



Editors

Mireia Ribera
i Toni Granollers

Autors

Bruno Splendiani i
Mireia Ribera.

Introducció a la conversió d'àudio a text

Autors

Bruno Splendiani i Mireia Ribera

dins

Guia de contingut digital accessible: vídeo

Editors

Mireia Ribera Turró, Universitat de Barcelona. Membre del grup de treball dedicat a l'accessibilitat digital en docència, recerca i innovació docent (Adaptabit).

Toni Granollers i Saltiveri, Universitat de Lleida. Membre del grup de recerca en interacció persona-ordinador i integració de dades (GRIHO).

Autors

Xavier Carrera Farran; Miquel Centelles Velilla; Jordi Coiduras Rodríguez; Roberto García González; Rosa M. Gil Iranzo; Víctor Gil Lloret; Juan Manuel Gimeno Illa; Toni Granollers i Saltiveri; Marta Oliva Solé; Afra Pascual Almenara; Mireia Ribera Turró; Josep M. Ribó Balust; Marina Salse Rovira; Montserrat Sendín Veloso; Bruno Splendiani

Disseny de la portada

Cat&Cas. Els membres del projecte a partir de la idea de la "Guia de contingut digital accessible"

Primera edició: març 2013

Dipòsit legal: L-102-2013

ISBN: 978-84-695-6918-4



Aquest document té una llicència **Reconeixement-No Comercial-Compartir-Igual 3.0 de Creative Commons**. La llicència completa es pot consultar a: <http://creativecommons.org/licenses/by-nc-sa/3.0/es/legalcode.ca>

Aquesta guia ha estat desenvolupada gràcies als següents recursos:

- projecte "**El vídeo sense barreres com a suport per la docència en l'educació superior**" corresponent a la convocatòria MQD2010 de l'Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) i amb el número d'expedient: **2010MQD00008**
- **conveni de col·laboració** entre **ACTIVA MÚTUA 2008** i la **Universitat de Lleida** que mitjançant el grup de recerca GRIHO dóna suport al desenvolupament de projectes i de tecnologies que possibilitin l'adaptació i reinserció laboral de persones amb discapacitat, en l'àmbit de les Tecnologies de la Informació i Sistemes Informàtics.

Guia de contingut digital accessible: vídeo

Introducció a la conversió d'àudio a text (Dragon Naturally Speaking i Audacity)

PRESENTACIÓ

Aquest document és una introducció a les eines Dragon Naturally Speaking i Audacity, especialitzades a optimitzar la transcripció d'arxius sonors.

Sumari

1	Introducció.....	157
1.1	Motivació	157
1.2	Eines utilitzades.....	157
1.3	Recomanacions generals de les eines	158
2	Planificació.....	159
2.1	Gravació amb càmera de vídeo	159
3	Enregistrament i neteja amb l'audacity	160
3.1	Qualitat amb audacity.....	160
3.2	Neteja de so amb l'audacity	162
4	Entrenament del dragon i transcripció	168
4.1	Entrenament de veu	168
4.2	Optimització del diccionari.....	171
4.3	Transcripció amb el dragon	175
5	Referències.....	179

1. INTRODUCCIÓ

1.1 MOTIVACIÓ

Aquest document és **una guia de recomanacions per transcriure de veu a text**.

La **transcripció** d'àudio a text és fonamental per a les persones amb sordesa, ja que permet accedir al contingut parlat de vídeos (com a pas previ per crear subtítols de vídeo) i a enregistraments de veu.

La transcripció de veu també pot resultar beneficiosa per a altres col·lectius, per exemple els dislèctics, ja que l'accés al contingut de manera multisensorial permet millorar l'aprenentatge i la memorització.

En general, la transcripció ajuda en situacions d'accés a arxius sonors en ambients amb silenci forçat o ambients sorollosos.

Hi ha diversos mostres d'ús de transcripcions. Per exemple, es podria donar el cas que es vulgui extreure l'àudio d'un vídeo d'una entrevista i fer-ne una transcripció per adaptar-lo a un públic d'estudiants amb sordesa. O pot ser que un professor tingui la gravació d'una classe (o d'una conferència o de qualsevol contingut sonor) i necessiti fer-ne la transcripció per publicar-la en text.

La tasca de transcriure la veu pot resultar a vegades bastant laboriosa però els resultats milloren considerablement si la veu de la gravació que es vol transcriure està neta. Per aconseguir-ho s'ha de partir d'un seguit de bones pràctiques en el procés de gravació acompanyades d'una combinació de dues eines que ajuden a automatitzar aquest procés de transcripció. Amb l'Audacity es pot optimitzar l'arxiu de veu i amb l'Audacity Dragon Naturally Speaking es pot fer la transcripció automàtica.

1.2 EINES UTILITZADES

Per descriure els passos d'aquesta guia, s'han utilitzat la versió 10 del Dragon Naturally Speaking, de Nuance, i la versió 2.0.2 de l'Audacity.

Eina i versió: Dragon Naturally Speaking versió 10

- Llicència de pagament [1]
- Plataformes web compatibles: Windows 2000, XP, Vista, 7
- Aquesta versió no suporta el reconeixement vocal del català

Eina i versió: Audacity versió 2.0.1

- Llicència gratuïta [2]
- Plataformes web compatibles: Windows 2000, XP, Vista, 7, Mac OS X 10.4 o superior, i Linux.

1.3 RECOMANACIONS GENERALS DE LES EINES

El Dragon Naturally Speaking és actualment el programa de reconeixement vocal més popular al mercat.

Un dels aspectes més destacables d'aquesta eina és la possibilitat d'entrenar el programa per reconèixer la veu i la pronúncia d'una persona determinada. Aquest entrenament permet una millora considerable del reconeixement de paraules. El Dragon també permet lletrejar paraules que no es troben al seu diccionari i afegir-les-hi perquè les reconegui en ocasions següents.

S'ha de tenir en compte que el Dragon no arriba mai a reconèixer el 100 % de les paraules dictades i és imprescindible revisar sempre manualment el resultat de la transcripció.

D'altra banda, l'Audacity és un programa d'enregistrament i edició de so senzill d'utilitzar, que suporten diferent plataformes, d'ús lliure i de codi obert. Permet:

- Enregistrar so.
- Importar i exportar diferents tipus d'arxius sonors, com ara WAV, FLAC, AU i Ogg Vorbis, i importar MP3, però no treballa amb formats propietaris com ara WMA o AAC.
- Eliminar soroll estàtic o altres sorolls de fons constants.

2. PLANIFICACIÓ

El primer pas a tenir en compte en l'enregistrament de so, ja sigui de forma aïllada o com a part d'un vídeo (per exemple, una classe magistral), és planificar de manera acurada aquesta operació, tot intentant evitar que el soroll de fons pugui esmorteir la veu.

2.1 GRAVACIÓ AMB CÀMERA DE VÍDEO

Com a regla general i sempre que sigui possible, és necessari efectuar l'enregistrament en un lloc amb eco reduït i aïllat de soroll exterior.

Per optimitzar l'enregistrament de so en el cas de gravació amb una càmera de vídeo professional, s'aconsella gravar la veu del parlant amb un micròfon de tipus "corbata" i el so de l'ambient amb el micròfon de la càmera. D'aquesta manera es garanteix que l'àudio de la veu sigui més net i de millor qualitat.

Si no es disposa de càmera professional amb canal independent per a l'entrada d'àudio, es pot enregistrar la veu a banda amb una gravadora amb micròfon i a posteriori sincronitzar la veu amb el vídeo, o directament enregistrar la veu amb un programa a l'ordinador. En aquest últim cas, es pot usar l'Audacity, que es descriu a la secció següent.

Després de gravar en vídeo la classe, s'ha de separar el so de l'enregistrament vídeo. Per fer-ho, hi ha al mercat moltes eines que extreuen arxius sonors des

d'arxius de vídeo. Un exemple és l'aplicació gratuïta Pazera Free Audio Extractor [3], que permet extreure l'arxiu sonor sense pèrdua de qualitat i convertir-lo a diferents formats (entre d'altres, WAV).

Un cop es té l'arxiu sonor, es pot passar al procés de neteja de soroll amb l'eina Audacity, que en aquesta guia s'explica en el punt 3.2, "Neteja de so amb l'Audacity".

3. ENREGISTRAMENT I NETEJA AMB L'AUDACITY

Aquesta tasca només és necessària si es vol enregistrar directament amb l'Audacity o si es vol millorar la qualitat del so enregistrat.

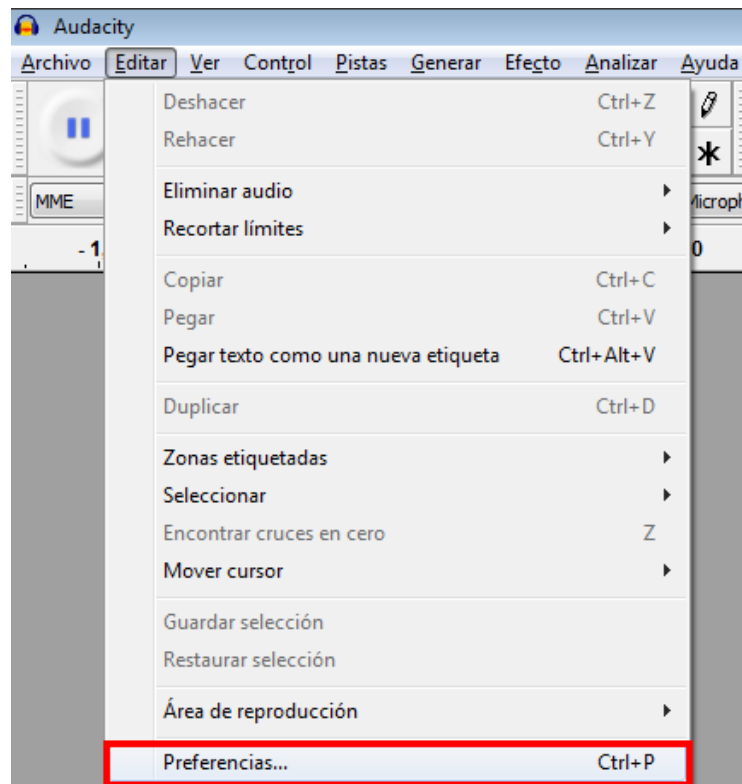
3.1 QUALITAT DE SO AMB AUDACITY

Si s'enregistra directament amb el programa Audacity, s'ha de determinar la qualitat de so que es vol obtenir abans de començar la gravació.

Passos que s'han de seguir

Pas 1 Obertura de les preferències del programa Audacity

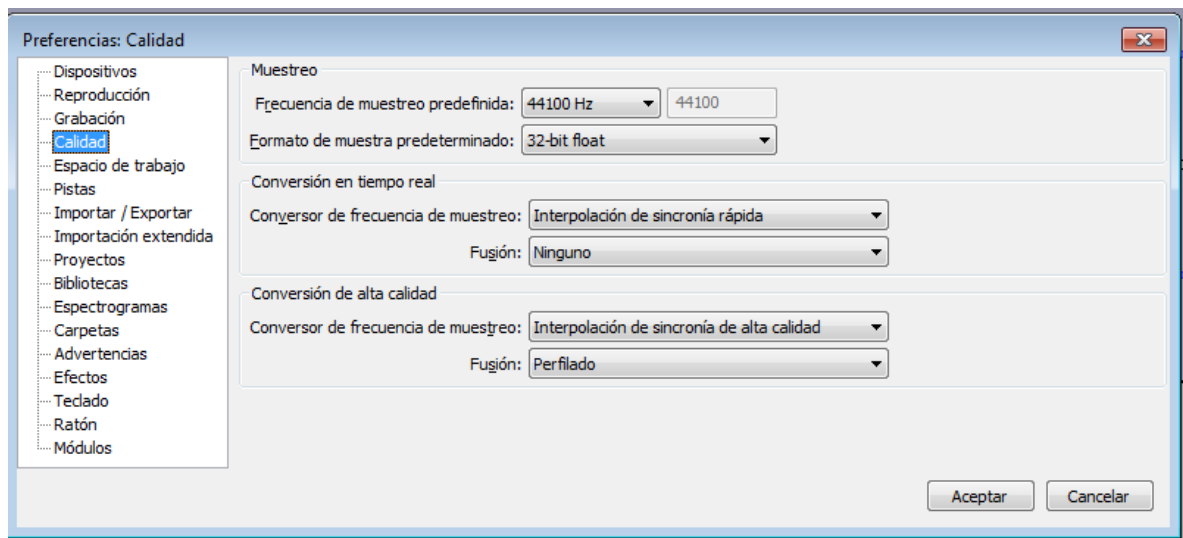
Aneu a *Editar > Preferencias* (o Ctrl + P).



Imatge 99. Menú per obrir l'opció *Preferencias* dins de l'Audacity

Pas 2 Selecció de les preferències d'enregistrament

Dins de la pestanya *Calidad*, a l'opció *Frecuencia de muestreo predefinida*, hauríeu d'escollir l'opció més alta (44100 Mhz) i a l'opció *Formato de muestra predeterminada* es recomana posar-hi *32-bit float*. Aquests valors permeten gravar amb la millor qualitat de so possible amb l'Audacity. Per a més detalls sobre aquests paràmetres, es recomana consultar la guia de l'Audacity [4].



Imatge 100. Opcions de qualitat de so dins de l'Audacity

Pas 3 Selecció del dispositiu d'entrada

A la pestanya *Dispositivos*, escolliu el dispositiu d'entrada, que pot ser un micròfon integrat amb l'ordinador o, idealment, un micròfon professional endollat a l'entrada per a micròfon de l'ordinador.

Pas 4 Inici de l'enregistrament

Ara es pot començar la gravació amb el comandament *Control-Grabar* (teclat R) o amb el botó *Grabar*.

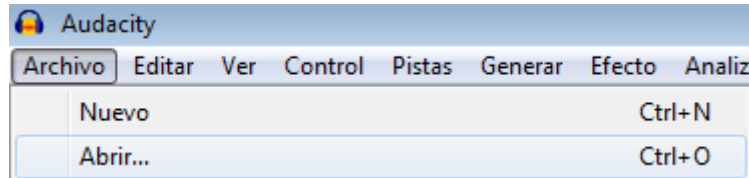
3.2 NETEJA DE SO AMB L'AUDACITY

L'Audacity ofereix la possibilitat d'aplicar filtres en fase de postenregistrament per aïllar la veu del soroll de fons. Aquesta opció en realitat el que fa és eliminar les freqüències en què es localitza el soroll. Es podria, doncs, donar el cas que, aplicant aquest filtre, s'elimini alguna part de la gravació que comparteixi aquestes freqüències. Per aquest motiu es recomana d'utilitzar l'efecte amb precaució i comprovar si el resultat és un so net o es nota algun efecte metàl·lic en la veu.

Passos a seguir

Pas 1 Obertura de l'arxiu de so amb l'Audacity

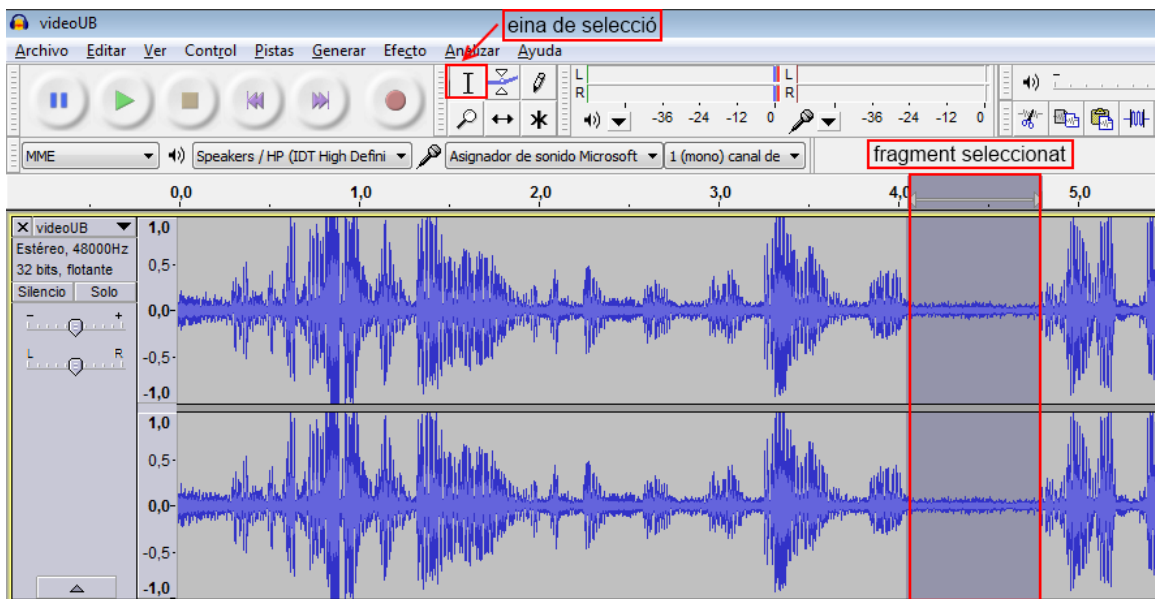
Aneu a *Archivo* > *Abrir* (o Ctrl + O).



Imatge 101. Menú per obrir un arxiu de so amb l'Audacity

Pas 2 Selecció del fragment amb soroll

Per seleccionar un fragment on es detecta que hi ha soroll de fons, s'utilitza l'eina de selecció en el menú de l'Audacity i, es tria la zona amb soroll que es vol filtrar. La zona seleccionada apareix destacada amb color més fosc.

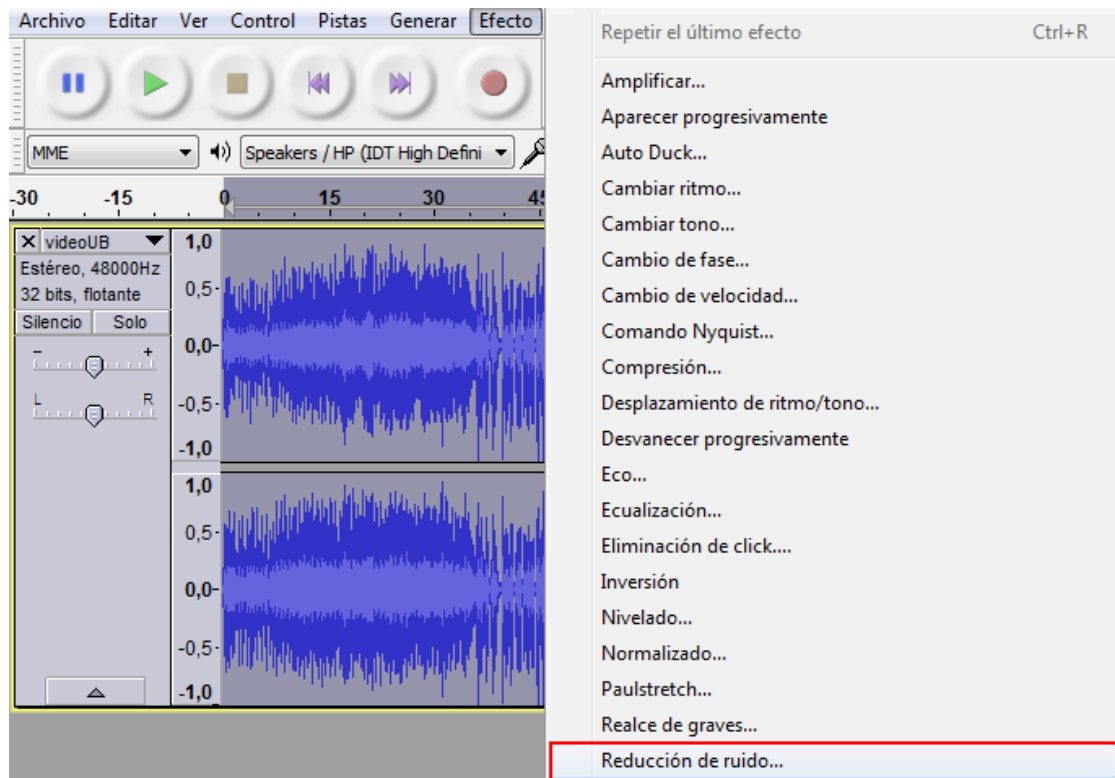


Imatge 102. Selecció d'una zona amb l'Audacity. Text de la imatge: eina de selecció – fragment seleccionat

Per poder seleccionar de manera més senzilla la zona amb soroll, utilitzeu l'eina *Zoom*, a l'opció del menú *Ver-Ampliar* (o Ctrl + 1).

Pas 3 Selecció de l'opció de reducció de soroll

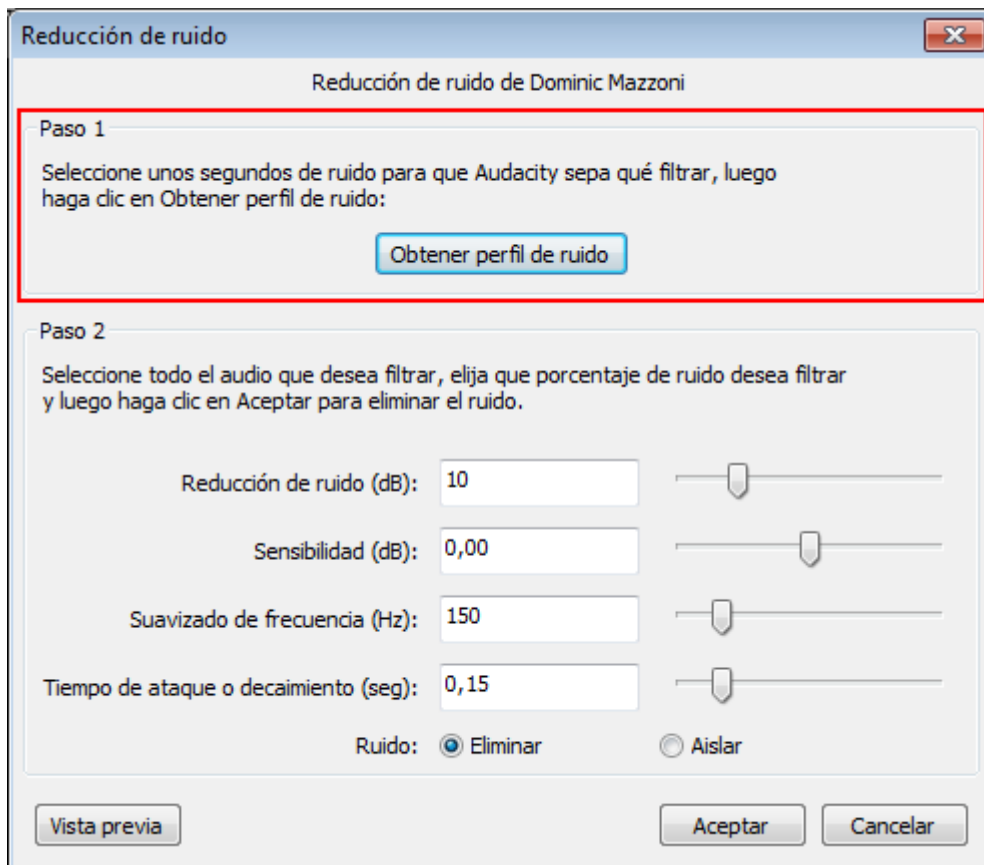
Aneu a *Efecto* > *Reducción de ruido*.



Imatge 103. Menú per obrir l'opció *Reducción de ruido* a l'Audacity

Pas 4 Obtenció del perfil de soroll

En el pas 1 del quadre que s'obre, si s'ha seleccionat prèviament la porció de l'arxiu amb soroll, s'ha de prémer el botó *Obtener perfil de ruido*. D'aquesta manera el programa reconeix el perfil de soroll que es vol filtrar.



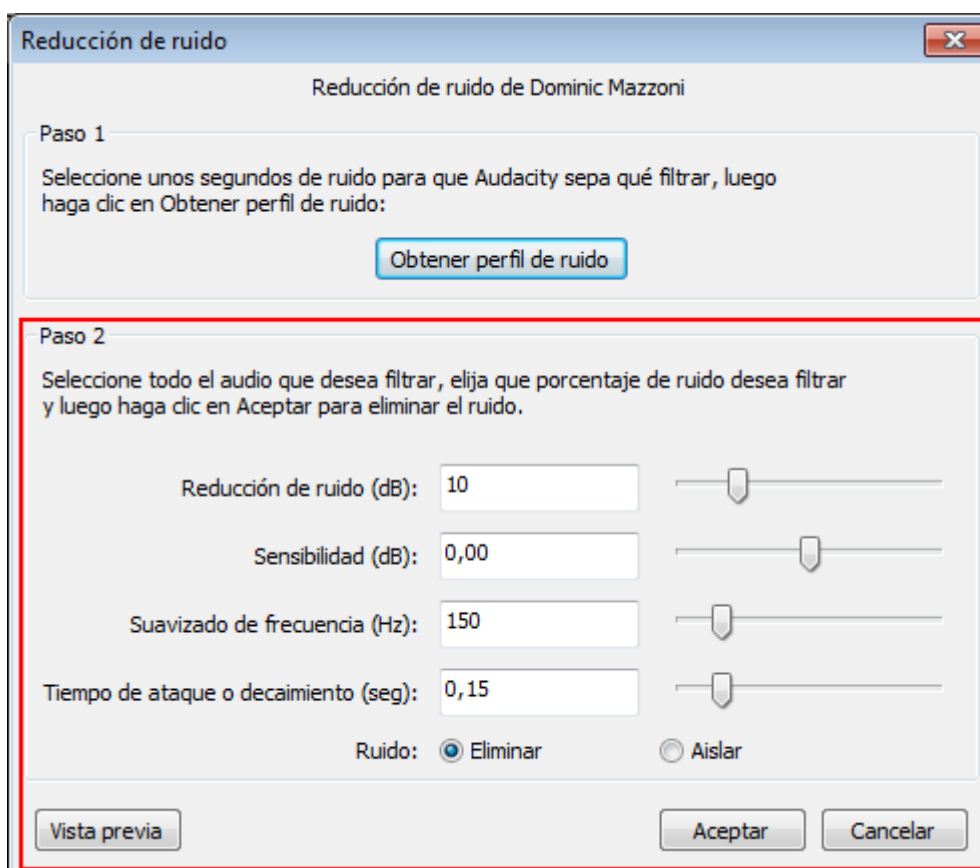
Imatge 104. Opció d'obtenció de perfil amb soroll amb l'Audacity

Ara s'ha de tornar a la pantalla amb les ones de so i seleccionar tot l'arxiu, per poder indicar al programa que la neteja del perfil de soroll que s'ha escollit s'ha d'aplicar a tot l'arxiu.

Després s'ha d'anar a l'opció *Efecto > Reducción de ruido* i continuar amb el pas següent.

Pas 5 Eliminació del soroll

En el pas 2, es pot modificar una sèrie de paràmetres específics per reduir el soroll.



Imatge 105. Paràmetres de filtre de soroll amb l'Audacity

El primer paràmetre permet ajustar la intensitat a què es vol reduir el soroll (en decibels). Amb un arxiu només de veu, com el cas que s'explica en aquesta guia, la resta de paràmetres no resulten particularment determinants per millorar la qualitat de la veu. Per tant, es poden deixar en els valors predeterminats. (Per a més detalls sobre aquests paràmetres es recomana consultar la guia de l'Audacity [5].)

En el pas 2, l'ideal és anar provant les diferents reduccions de soroll i comprovar els resultats prement el botó *Vista previa*.

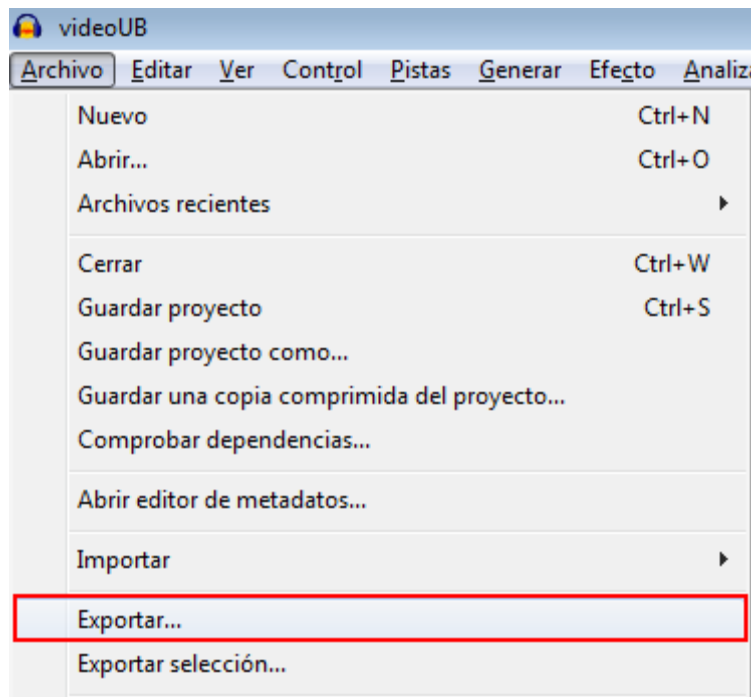
Atenció

El filtre de soroll s'ha d'aplicar amb precaució extrema, ja que podria generar una reducció del so de la veu o un efecte metàl·lic, i dificultaria així el reconeixement vocal.

Quan s'estigui segur del resultat, feu *Aceptar*. El programa comença el procés de neteja.

Pas 6 Desament i exportació de l'arxiu de so

Un cop acabat el procés de neteja, deseu l'arxiu amb l'opció *Archivo > Exportar*, seleccioneu el format d'exportació i deseu-lo.



Imatge 106. Opció per exportar arxius amb l'Audacity

Atenció

S'aconsella exportar l'arxiu a un format que mantingui inalterada la qualitat del so, com ara el format WAV.

Ara l'arxiu de so està preparat per començar el procés de transcripció.

4. ENTRENAMENT DEL DRAGON I TRANSCRIPCIÓ

El Dragon Naturally Speaking és un programa de reconeixement vocal que permet automatitzar el procés de transcripció. L'eina està dissenyada per transcriure la veu d'un parlant individual. Per això s'aconsella seguir un procés previ d'entrenament (vegeu la secció 4.1) i d'optimització del seu diccionari (vegeu la secció 4.2). Així s'aconsegueix adaptar el reconeixement a una veu específica i optimitzar les prestacions.

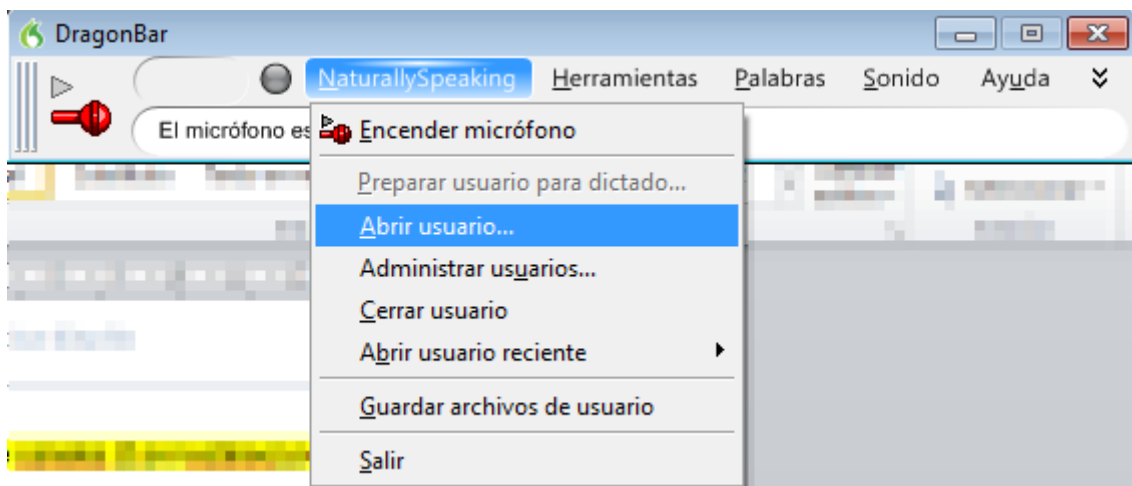
Un cop optimitzada l'eina, es pot començar el procés de reconeixement amb Dragon (vegeu la secció 4.3), que dóna com a resultat final una transcripció.

4.1 ENTRENAMENT DE VEU

Passos a seguir

Pas 1 Obertura d'un perfil d'usuari

Aneu a *NaturallySpeaking* > *Abrir usuario*.



Imatge 107. Opció per gestionar perfils d'usuaris al Dragon

Pas 2 Creació d'un perfil d'usuari

Escolliu l'opció *Nuevo* (si és el primer cop que s'obre el programa, per defecte demana la creació d'un perfil d'usuari nou).

Tot seguit s'obre un assistent per crear un usuari nou. Després d'escollir el nom, l'idioma i la font de dictat (normalment la millor font de dictat seria un dispositiu com ara un micròfon amb auriculars connectat a l'entrada de micròfon del PC), heu d'indicar quin tipus d'entrenament inicial voleu fer [6].

Imatge 108. Opcions per crear un nou usuari al Dragon

Hi ha tres opcions disponibles:

- *Ninguno*: s'omet per complet l'entrenament i només es defineix el volum del micròfon i la qualitat del sistema de so de l'usuari. Opció desaconsellada.

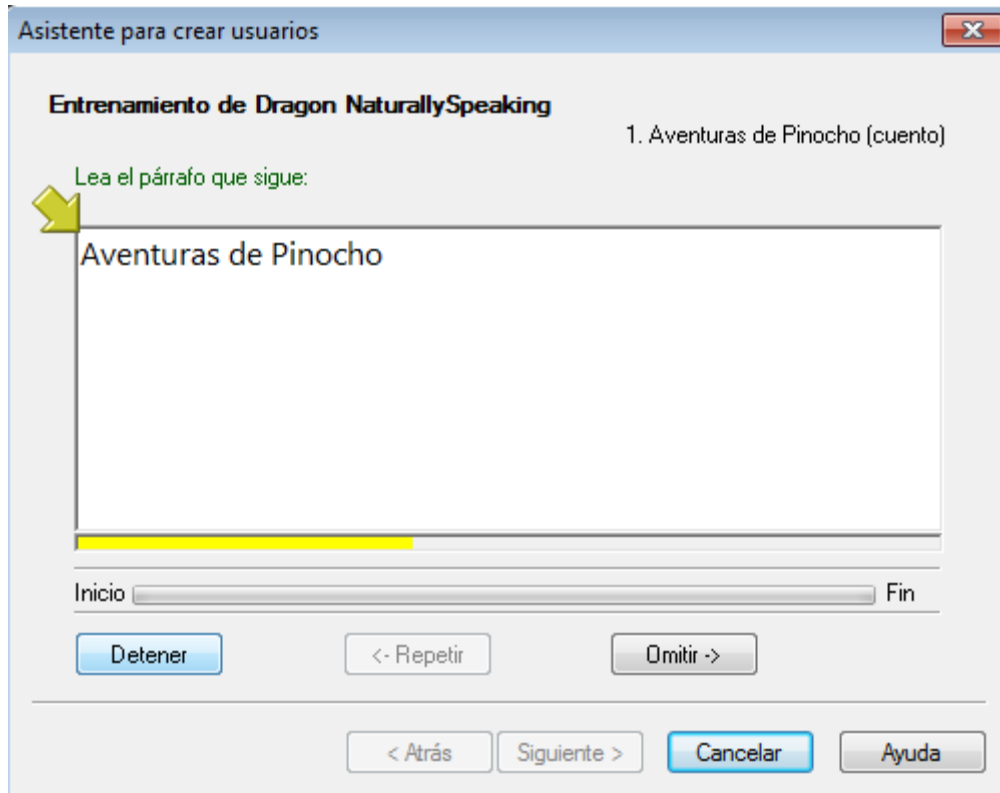
- *Corto* (opció per defecte): es guia l'usuari pel programa en la lectura de diferents textos. Aquest entrenament permet una millor precisió en el reconeixement de paraules respecte a l'opció *Ninguno*. Aquesta és l'opció escollida en aquest manual.
- *Especial*: l'usuari llegeix diferents textos sense cap interacció o indicació del Dragon. Aquest tipus d'entrenament està pensat per a usuaris amb impediment en la parla o altres dificultats en l'articulació de paraules. Ofereix la possibilitat d'imprimir els textos que s'han de llegir i deixa a l'usuari tot el temps que necessiti per completar l'entrenament, incloent-hi llargues pauses.

Pas 3 Optimització del so del sistema

Un cop escollida l'opció d'entrenament més adequada, l'assistent guia l'usuari en diferents passos per elegir el sistema de so, el calibratge del volum i la comprovació de la qualitat del so del sistema i del micròfon. Després que el programa hagi confirmat que la prova de so ha acabat exitosament comença l'entrenament.

Pas 4 Entrenament

L'entrenament consisteix en la lectura en veu alta d'unes frases i un text que el programa permet escollir entre diferents tipus d'estils.



Imatge 109. Finestra de l'assistent d'entrenament de veu del Dragon

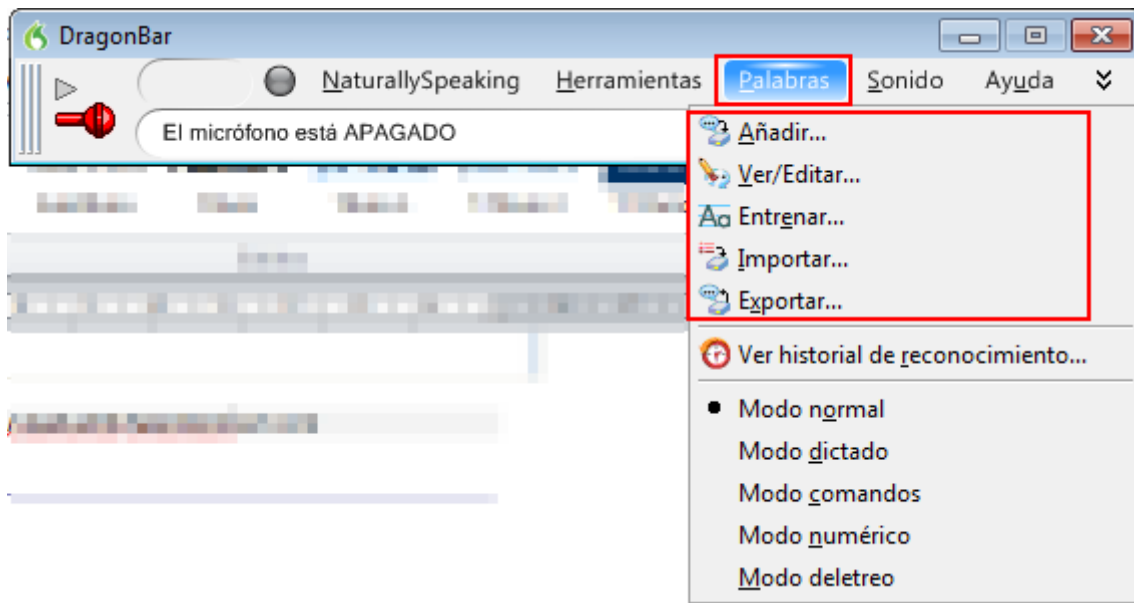
Quan s'ha acabat l'entrenament, el programa triga uns minuts a adaptar i desar els arxius del nou perfil d'usuari. Finalitzat aquest procés, el programa està optimitzat per reconèixer l'arxiu de veu.

4.2 OPTIMITZACIÓ DEL DICCIONARI

Passos a seguir

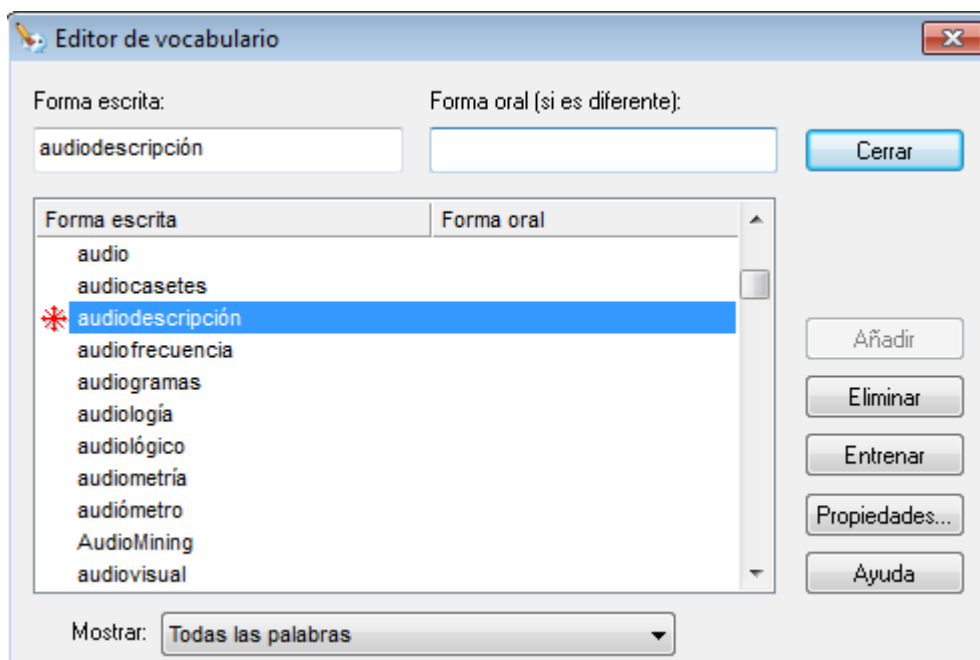
Pas 1 Tria de l'opció d'edició del diccionari

Les opcions relacionades amb el diccionari es troben a la pestanya *Palabras* del menú principal del Dragon.



Imatge 110. Opcions d'edició del diccionari del Dragon

Entre les diferents opcions, es poden afegir noves paraules, importar-les des d'arxius de text o editar les mateixes paraules que ja es troben al diccionari.

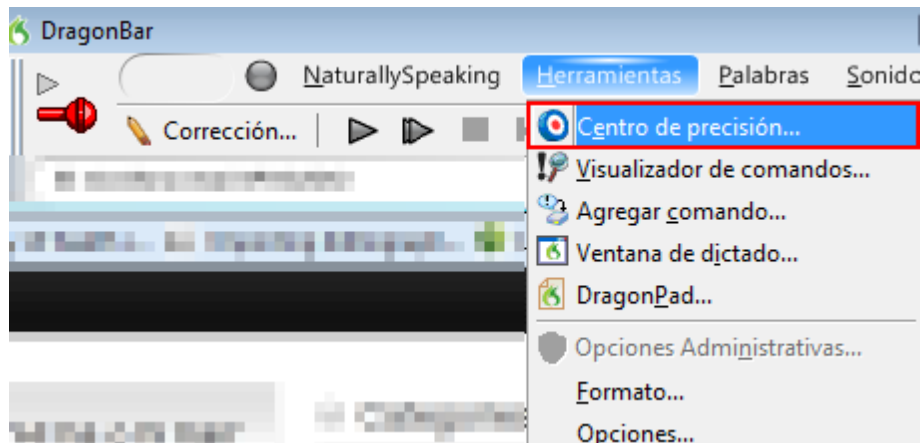


Imatge 111. Quadre de diàleg de l'editor del diccionari del Dragon

Pas 2 Optimitzador automàtic del diccionari

Una altra manera d'enriquir el diccionari és permetre al Dragon analitzar els nostres documents de l'ordinador per identificar les paraules més usades. Aquesta opció en entorns bilingües com el nostre no és gaire recomanable.

Aneu a *Herramienta* > *Centro de precisión*.




Imatge 112. Pestanya del menú Dragon per accedir a l'opció *Centro de precisión*

En el quadre de diàleg *Centro de precisión*, seleccioneu l'opció *Ejecutar el optimizador de vocabulario*.





¡Bienvenido al Centro de precisión!

El Centro de precisión le proporciona acceso a las herramientas que le ayudarán a mejorar la forma en que Dragon NaturallySpeaking oye y entiende lo que le dice.



Haga clic o lea el vínculo que le interese.

 [¿Qué herramienta debo usar?](#)







Mejorar la acústica

-  [Comprobar la configuración del sonido](#)
-  [Ejecutar el Optimizador acústico y de lenguaje](#)
-  [Programar el Optimizador acústico y de lenguaje a ejecutar después](#)
-  [Realizar entrenamiento adicional](#)


Administrar los comandos

-  [Iniciar el Visualizador de comandos](#)
-  [Salir](#)

Mejorar el vocabulario

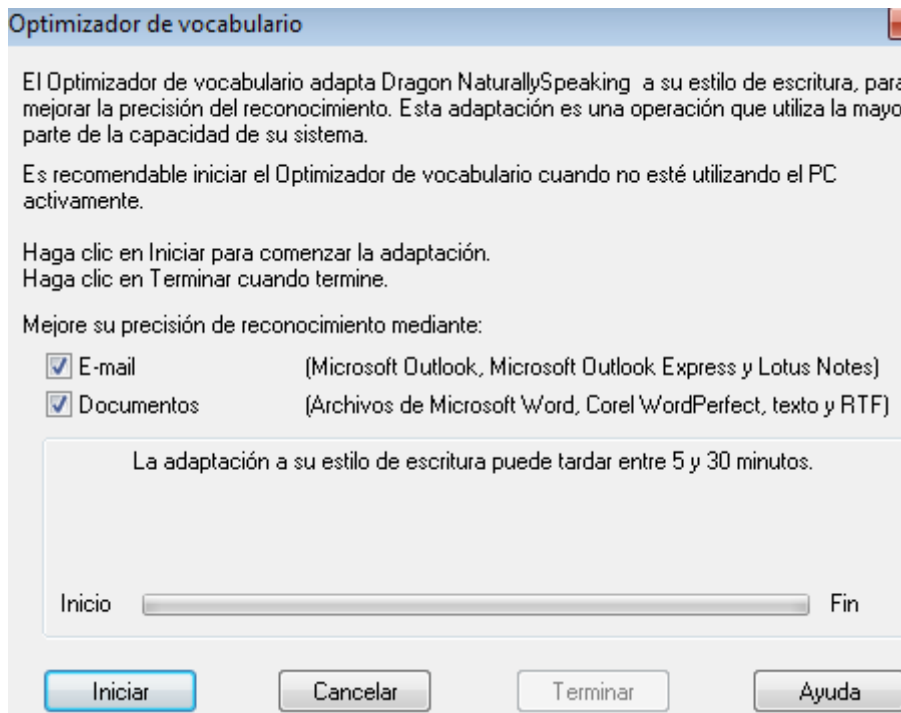
-  [Añadir una palabra al vocabulario](#)
-  [Añadir palabras de documentos al vocabulario](#)
-  [Aumentar la precisión usando e-mail](#)
-  [Añadir una lista de palabras al vocabulario](#)
-  [Ver o modificar el vocabulario](#)
-  [Ejecutar el Optimizador de vocabulario](#)

Obtener más información

-  [Cómo administrar y personalizar vocabularios](#)

Imatge 113. Opcions del *Centro de precisión* del Dragon

Aquesta funció analitza els correus electrònics i els documents presents al PC per adaptar el diccionari del Dragon a l'estil d'escriptura de l'usuari que utilitza el PC i millorar la precisió del reconeixement.



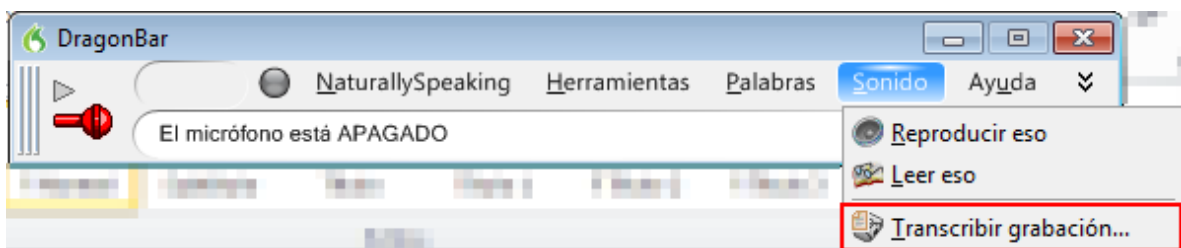
Imatge 114. Quadre de diàleg de l'optimitzador de diccionari del Dragon

4.3 TRANSCRIPCIÓ AMB EL DRAGON

Passos a seguir

Pas 1 Obertura de l'arxiu de so amb el Dragon

Aneu a *Sonido* > *Transcribir grabación*.

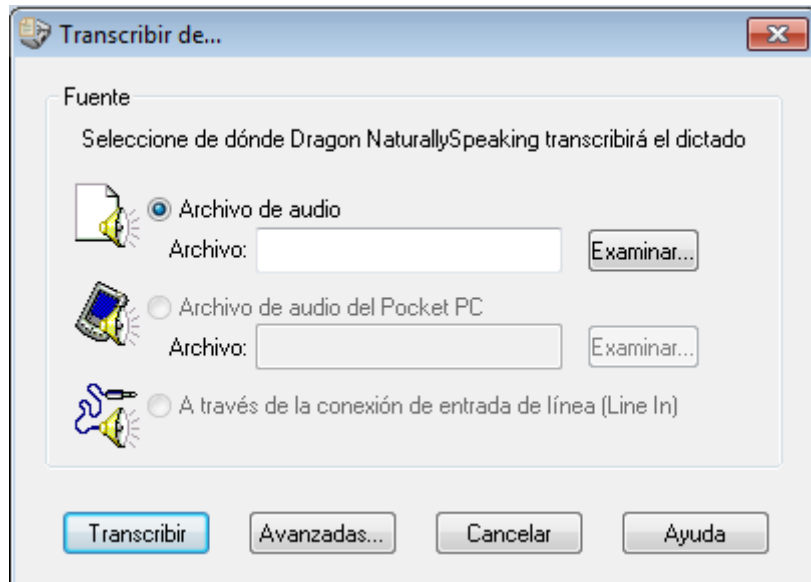


Imatge 115. Opció de transcripció de gravació del Dragon

S'obre un quadre de diàleg on s'han de definir les opcions de transcripció.

Pas 2 Tria de les opcions de transcripció

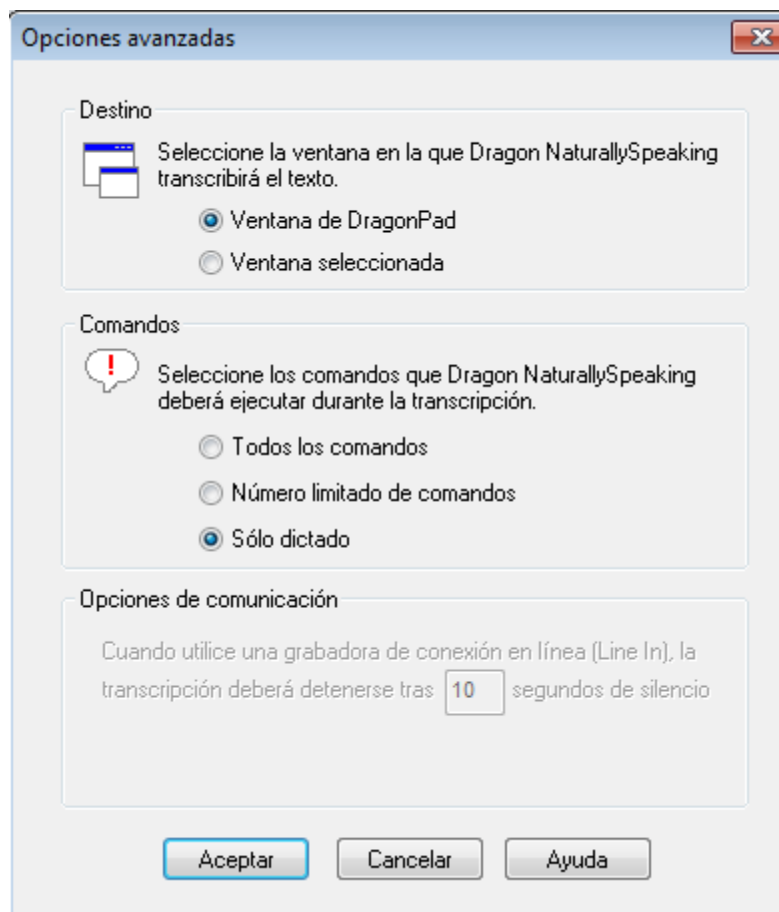
Per escollir l'arxiu de veu que es vol transcriure, premeu el botó *Examinar* en el quadre de diàleg.



Imatge 116. Quadre de diàleg per escollir l'arxiu a transcriure amb el Dragon

El botó *Avanzadas* obre un quadre de diàleg que permet escollir:

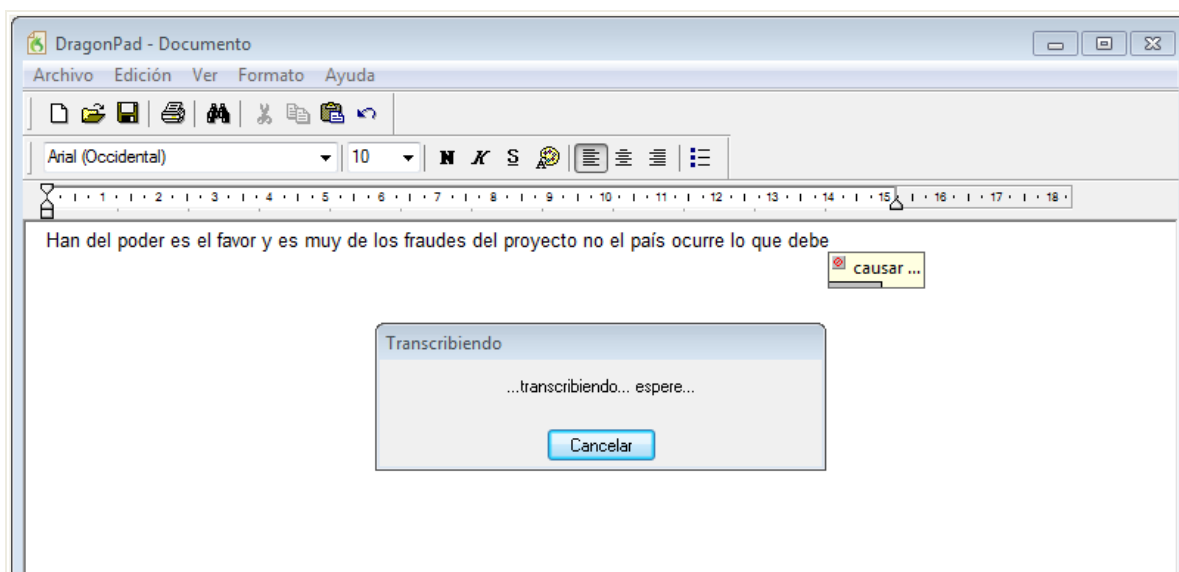
- On es transcriu el text, si al DragonPad (el programa d'edició de text per defecte del Dragon) o a la finestra d'un altre editor de text seleccionat per l'usuari.
- Si es vol transcriure només el text del dictat o també comandaments del Dragon. Si activeu l'opció de reconeixement de comandaments, el Dragon podria reconèixer algunes paraules com a comandaments de dictat (per exemple, *Nueva línea*, que activaria el comandament retorn). S'aconsella evitar aquesta opció.



Imatge 117. Quadre de diàleg d'opcions avançades de transcripció del Dragon

Pas 3 Inici del procés de transcripció

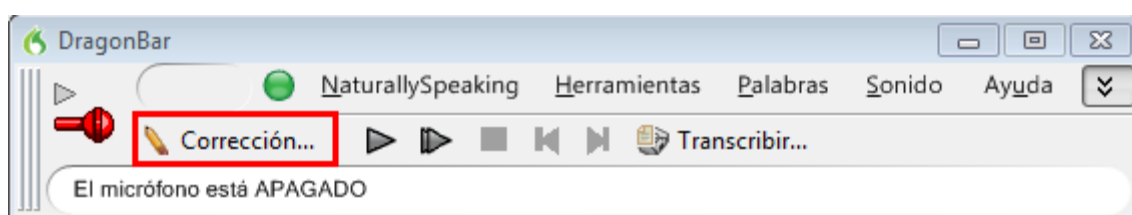
Un cop definides les opcions de transcripció, es pot començar la transcripció prement el botó *Transcribir*. El programa comença el procés de transcripció, que es pot controlar directament a la finestra de l'editor de text escollit.



Imatge 118. Exemple de finestra del DragonPad durant el procés de transcripció

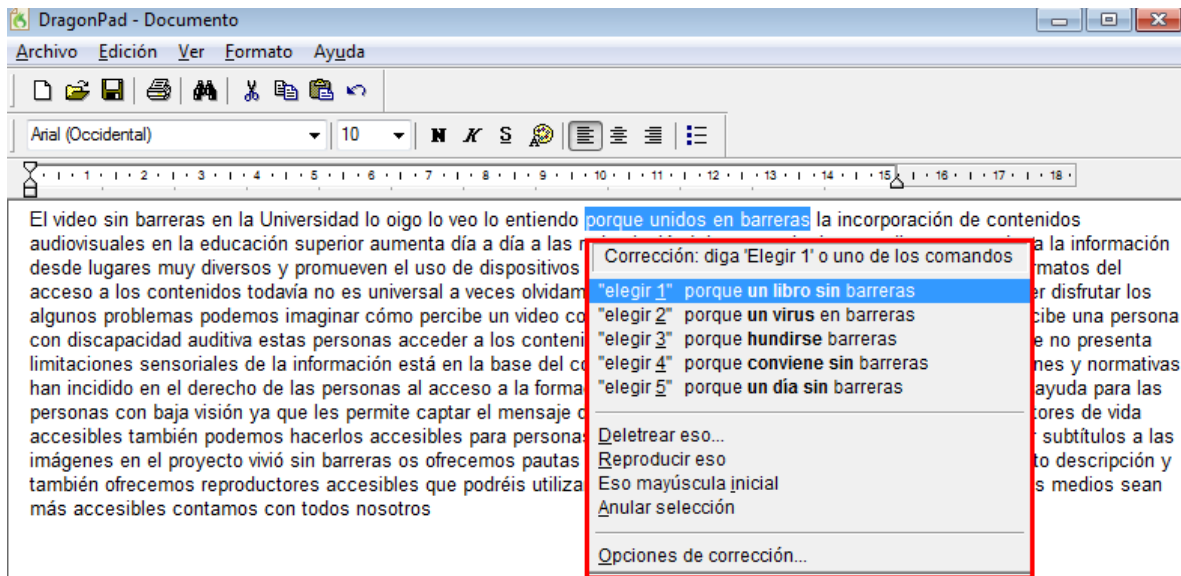
Pas 4 Correcció de paraules

Acabada la transcripció, es pot revisar si el text és correcte i corregir-lo dins l'editor de text. El programa proporciona un seguit d'opcions de correcció que permeten, entre d'altres, especificar al Dragon la paraula adequada per substituir un error de transcripció, millorant la precisió del reconeixement en transcripcions següents. Una manera de corregir paraules és situar-nos a l'inici de la paraula a corregir i després activar el botó *Correcció* en el menú de navegació del Dragon.



Imatge 119. Pestanya *Correcció* en el menú Dragon

En l'editor de text s'obre un desplegable amb diferents opcions de correcció.



Imatge 120. Menú de correcció d'errors de transcripció del Dragon

Si alguna de les alternatives és correcta, seleccioneu-la per acceptar. Si no, escriviu o dicteu el text correcte.

Consell:

Es recomana sempre revisar el text de la transcripció, ja que la capacitat de reconeixement del Dragon pot arribar com a màxim al 98-99 %.

Acabades les correccions en la transcripció, es pot finalment guardar l'arxiu de text, concloent així el procés de transcripció.

5. REFERÈNCIES

[1] NUANCE COMMUNICATIONS (2012). *Dragon – Dragon Naturally Speaking – Nuance* [en línia]. Madrid. <<http://www.nuance.es/dragon/>>. [Consulta: 17/10/2012].

[2] AUDACITY TEAM (2012). *Audacity* [en línia]. <<http://audacity.sourceforge.net/download/windows>>. [Consulta: 17/10/2012].

- [4] *Pazera Free Audio Extractor* (versió 1.4) [en línia].
<<http://www.pazera-software.com/products/audio-extractor/>>.
[Consulta: 17/10/2012].
- [5] AUDACITY TEAM (2012). *Audacity 2.0.2 Manual* (en anglès) [en línia].
<<http://manual.audacityteam.org/o/>>. [Consulta: 13/02/2012].
- [6] SPEECH EMPOWERED COMPUTING ([s. d]). "Tips for achieving high recognition accuracy". *Speech Empowered Computing Blog*.
<<http://speechempoweredcomputing.co.uk/Newsletter/>>. [Consulta: 17/10/2012].
- [6] Les opcions *Acento* i *Vocabulario* no estan disponibles per al castellà.



Aquesta obra forma part de la col·lecció "Guies de Contingut Digital Accessible: vídeo".

Aquesta guia ha estat creada en el marc del projecte "El vídeo sense barreres com a suport per la docència en l'educació superior" corresponent a la convocatòria MQD2010 de l'Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) i amb el número d'expedient: 2010MQD00008, i també gràcies al conveni que Agrupació Mútua 2008 ha establert amb el grup de recerca GRIHO (UdL) per desenvolupar tecnologies i projectes que possibilitin l'adaptació i reinserció laboral de persones amb discapacitat en l'àmbit de les TIC.



Universitat de Lleida



B Universitat de Barcelona

GRIHO
Campus de Cappont
c. Jaume II, 69
25001 Lleida
griho.udl.cat
info@griho.net