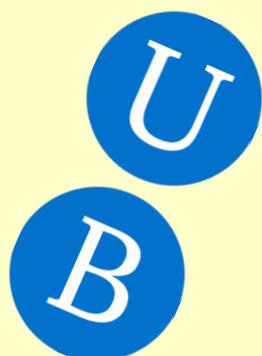
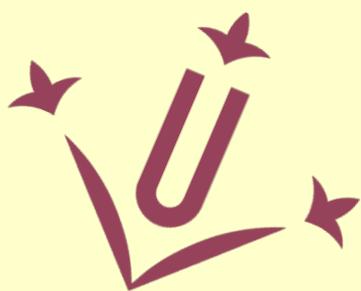


# GUÍA DE CONTENIDO DIGITAL ACCESIBLE: VÍDEO

Introducción a la  
conversión de audio a texto



**Editores**  
Mireia Ribera  
y Toni Granollers

**Autores**  
Bruno Splendiani y  
Mireia Ribera.

# Introducción a la conversión de audio a texto

## Autores

Bruno Splendiani y Mireia Ribera

en

## Guía de contenido digital accesible: vídeo

### Editores

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

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

### Autores

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

### Diseño de la portada

Cat&Cas. Los miembros del proyecto a partir de la idea de la “Guía de contingut digital accessible”

**Primera edición:** marzo 2013

**Depósito legal:** L-102-2013

**ISBN:** 978-84-695-6918-4



Este documento tiene una licencia **Reconocimiento-No Comercial-Compartir-Igual 3.0 de Creative Commons**. La licencia completa se puede consultar en: <http://creativecommons.org/licenses/by-nc-sa/3.0/es/legalcode.es>

Esta guía ha sido desarrollada gracias a los siguientes recursos:

- proyecto “**El vídeo sense barreres com a suport per la docència en l'educació superior**” correspondiente a la convocatoria MQD2010 de la *Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR)* con número de expediente: **2010MQD00008**
- **convenio de colaboración** entre **ACTIVA MÚTUA 2008** y la **Universitat de Lleida** que mediante el grup de recerca GRIHO da soporte al desarrollo de tecnologías y proyectos que posibiliten la adaptación y la reinserción laboral de personas con discapacidad en el ámbito de las TIC.

# Guía de contenido digital accesible: vídeo

Introducción a la conversión de audio a texto (Dragon Naturally Speaking y Audacity)

---

## PRESENTACIÓN

*Este documento es una introducción a las herramientas Dragon Naturally Speaking y Audacity, especializadas en optimizar la transcripción de archivos sonoros.*



# Índice

1. Introducción .....	163
1.1 Motivación .....	163
1.2 Herramientas utilizadas .....	164
1.3 Recomendaciones generales de las herramientas .....	164
2. Planificación .....	165
2.1 Grabación con cámara de vídeo .....	165
3. Grabación y limpieza con Audacity .....	166
3.1 Calidad de sonido con Audacity .....	166
3.2 Limpieza de sonido con audacity .....	168
4. Entrenamiento de Dragon y transcripción .....	174
4.1 Entrenamiento de voz .....	174
4.2 Optimización del diccionario .....	177
4.3 Transcripción con Dragon .....	181
5. Referencias .....	186



## 1. INTRODUCCIÓN

### 1.1 MOTIVACIÓN

Este documento es **una guía de recomendaciones para transcribir de voz a texto.**

La **transcripción** de audio a texto es fundamental para las personas con sordera, ya que permite acceder al contenido hablado de vídeos (como paso previo para crear subtítulos de vídeo) y a grabaciones de voz.

La transcripción de voz también puede resultar beneficiosa para otros colectivos, como por ejemplo los disléxicos, puesto que el acceso al contenido de manera multisensorial permite mejorar el aprendizaje y la memorización.

En general, la transcripción ayuda en situaciones de acceso a archivos sonoros en ambientes con silencio forzado o ambientes ruidosos.

Existen diversos ejemplos de uso de transcripciones. Por ejemplo, se puede dar el caso de que se quiera extraer el audio de un vídeo de una entrevista y transcribirlo para adaptarlo a un grupo de estudiantes con sordera. O puede ser que un profesor tenga la grabación de una clase (o de una conferencia o de cualquier otro contenido sonoro) y necesite transcribirla para publicarla en texto.

La tarea de transcribir la voz puede resultar a veces bastante laboriosa pero los resultados mejoran considerablemente si la voz de la grabación que se quiere transcribir está limpia. Para conseguirlo se han de tener en cuenta una serie de buenas prácticas durante el proceso de grabación, acompañadas de la combinación de dos herramientas que ayudan a automatizar el proceso de transcripción. Con Audacity se puede optimizar el archivo de voz y con Audacity Dragon Naturally Speaking se puede llevar a cabo su transcripción automática.

## 1.2 HERRAMIENTAS UTILIZADAS

Para describir los pasos de esta guía, se han utilizado la versión 10 de Dragon Naturally Speaking, de Nuance, y la versión 2.0.2 de Audacity.

### **Herramienta y versión: Dragon Naturally Speaking versión 10**

- Licencia de pago [1]
- Plataformas web compatibles: Windows 2000, XP, Vista, 7
- Esta versión no admite el reconocimiento vocal del catalán

### **Herramienta y versión: Audacity versión 2.0.1**

- Licencia gratuita [2]
- Plataformas web compatibles: Windows 2000, XP, Vista, 7, Mac OS X 10.4 o superior, i Linux.

## 1.3 RECOMENDACIONES GENERALES DE LAS HERRAMIENTAS

Dragon Naturally Speaking es actualmente el programa de reconocimiento vocal más popular del mercado.

Uno de los aspectos más destacables de esta herramienta es la posibilidad de entrenar el programa para reconocer la voz y la pronunciación de una persona determinada. Este entrenamiento permite una mejora considerable del reconocimiento de palabras. Dragon también permite deletrear palabras que no se encuentran en su diccionario y añadirlas para que las reconozca en posteriores ocasiones.

Hay que tener en cuenta que Dragon no llega a reconocer nunca el 100 % de las palabras dictadas y es imprescindible revisar siempre manualmente el resultado de la transcripción.

Por otro lado, Audacity es un programa de grabación y edición de sonido fácil de usar, compatible con diferentes plataformas, de uso libre y de código abierto.

Permite:

- Grabar sonido.
- Importar y exportar diferentes tipos de archivos sonoros, como por ejemplo WAV, FLACC, AU y Ogg Vorbis, e importar MP3, pero no trabaja con formatos propietarios como WMA o AAC.
- Eliminar ruido estático u otros ruidos de fondo constantes.

## 2. PLANIFICACIÓN

El primer paso a tener en cuenta en la grabación de sonido, ya sea de forma aislada o como parte de un vídeo (por ejemplo, una clase magistral), es planificar de manera cuidadosa esta operación, intentando evitar que el ruido de fondo pueda amortiguar la voz.

### 2.1 GRABACIÓN CON CÁMARA DE VÍDEO

Como regla general y siempre que sea posible, es necesario realizar la grabación en un lugar con eco reducido y aislado de ruido exterior.

Para optimizar la grabación de sonido en el caso de grabación con una cámara de vídeo profesional, se aconseja grabar la voz del hablante con un micrófono de corbata y el sonido de ambiente con el micrófono de la cámara. De esta manera se garantiza que el audio de la voz sea más limpio y de mejor calidad.

Si no se dispone de cámara profesional con canal independiente para la entrada de audio, se puede grabar la voz a parte con una grabadora con micrófono y a posteriori sincronizar la voz con el vídeo, o directamente grabar la voz con un programa en el ordenador. En este último caso, se puede usar Audacity, que se describe en la siguiente sección.

Después de grabar en vídeo la clase, se ha de separar el sonido de la grabación de vídeo. Para hacerlo, existen en el mercado muchas herramientas que extraen archivos sonoros desde archivos de vídeo. Un ejemplo es la aplicación gratuita

Pazera Free Audio Extractor [3], que permite extraer el archivo sonoro sin pérdida de calidad y convertirlo a diferentes formatos (entre otros, WAV).

Una vez que se tiene el archivo sonoro, se puede pasar al proceso de limpieza de ruido con la herramienta Audacity, que en esta guía se explica en el punto 3.2, “Limpieza de sonido con Audacity”.

### 3. GRABACIÓN Y LIMPIEZA CON AUDACITY

Esta tarea solo es necesaria si se quiere grabar directamente con Audacity o si se quiere mejorar la calidad del sonido grabado.

#### 3.1 CALIDAD DE SONIDO CON AUDACITY

Si se graba directamente con el programa Audacity, se debe determinar la calidad de sonido que se quiere obtener antes de empezar la grabación.

#### Pasos a seguir

##### **Paso 1** Abertura de las preferencias del programa Audacity

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

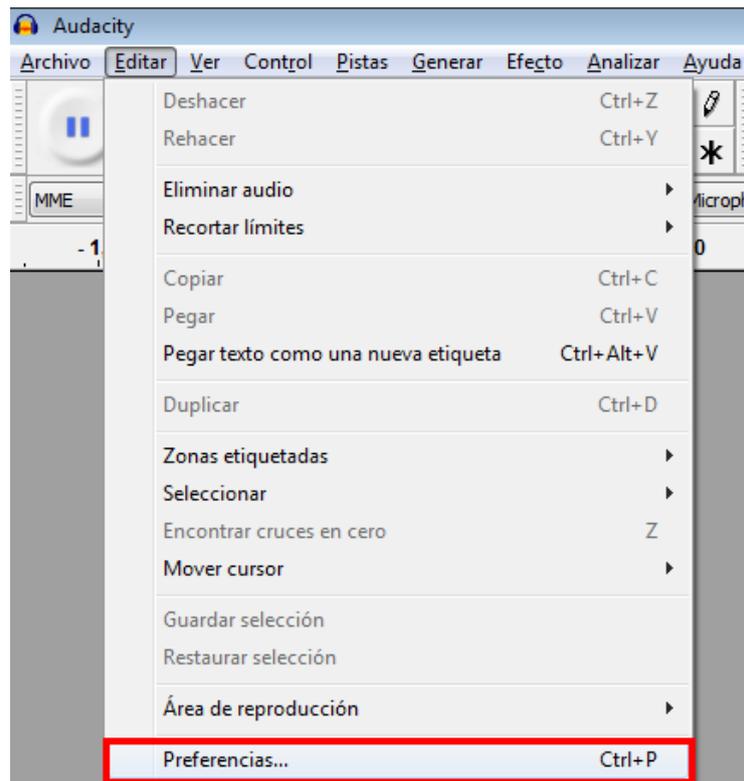


Ilustración 99. Menú para abrir la opción *Preferencias* dentro de Audacity

## **Paso 2** Selección de las preferencias de grabación

Dentro de la pestaña *Calidad*, en la opción *Frecuencia de muestreo predefinida*, escogemos la opción más alta (44100 Mhz) y en la opción *Formato de muestra predeterminada* se recomienda poner *32-bit float*. Estos valores permiten grabar con la mejor calidad de sonido posible con Audacity. Para más detalles sobre estos parámetros, se recomienda consultar la guía de Audacity [4].

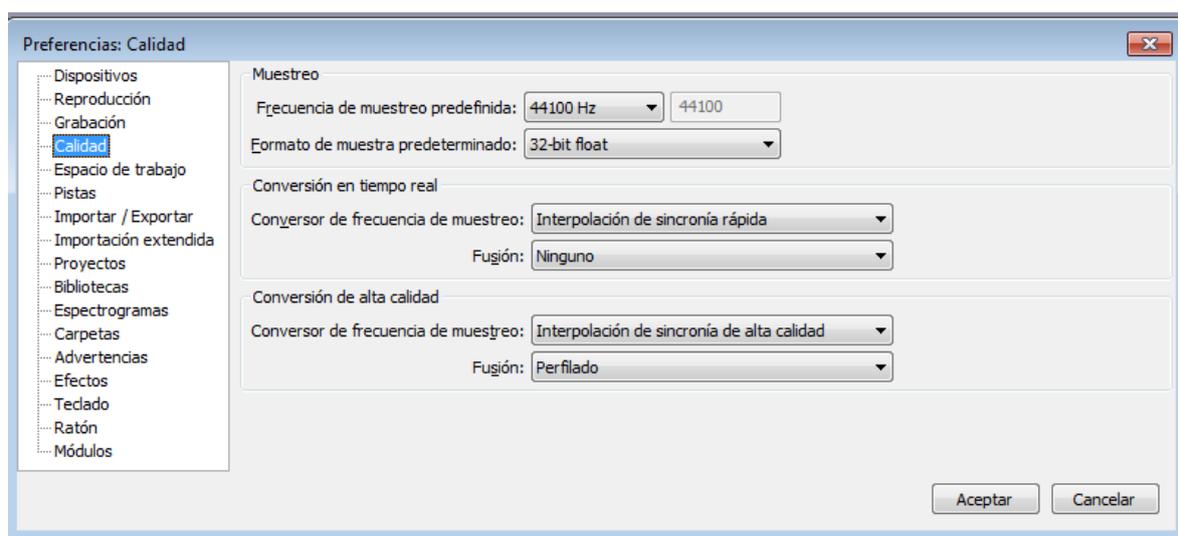


Ilustración 100. Opciones de calidad de sonido dentro de Audacity

### Paso 3 Selección del dispositivo de entrada

En la pestaña *Dispositivos*, escogemos el dispositivo de entrada, que puede ser un micrófono integrado con el ordenador o, idealmente, un micrófono profesional conectado a la entrada para micrófono del ordenador.

### Paso 4 Inicio de la grabación

Ahora se puede empezar la grabación con el comando *Control-Grabar* (tecla R) o con el botón *Grabar*.

## 3.2 LIMPIEZA DE SONIDO CON AUDACITY

Audacity ofrece la posibilidad de aplicar filtros en la fase de postgrabación para aislar la voz del ruido de fondo. Lo que hace en realidad esta opción es eliminar las frecuencias en las que se localiza el ruido. Se podría, pues, dar el caso de que, aplicando este filtro, se elimine alguna parte de la grabación que comparta estas frecuencias. Por este motivo se recomienda utilizar el efecto con precaución y comprobar si el resultado es un sonido limpio o se nota algún efecto metálico en la voz.

## Pasos a seguir

### Paso 1 Abertura del archivo de sonido con Audacity

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

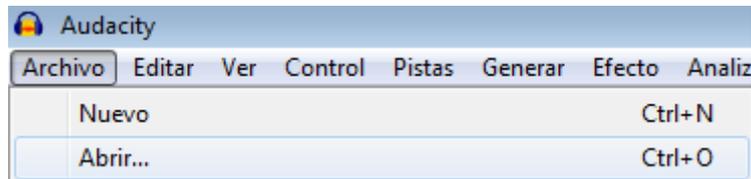


Ilustración 101. Menú para abrir un archivo de sonido con Audacity

### Paso 2 Selección del fragmento con ruido

Para seleccionar un fragmento donde se detecta que hay ruido de fondo, se utiliza la herramienta de selección en el menú de Audacity y se escoge la zona con ruido que se quiere filtrar. La zona seleccionada aparece destacada con un color más oscuro.

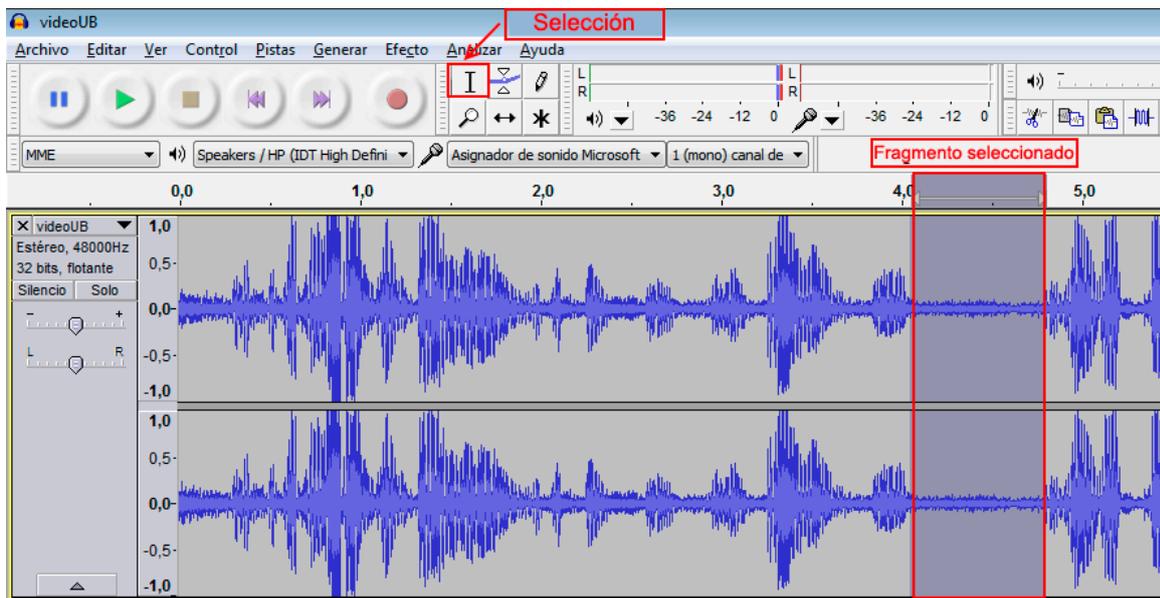


Ilustración 102. Selección de una zona con Audacity

Para poder seleccionar de manera más sencilla la zona con ruido, se usa la herramienta *Zoom*, en la opción del menú *Ver-Ampliar* (o Ctrl + 1).

### Paso 3 Selección de la opción de reducción de ruido

Vamos a *Efecto > Reducción de ruido*.

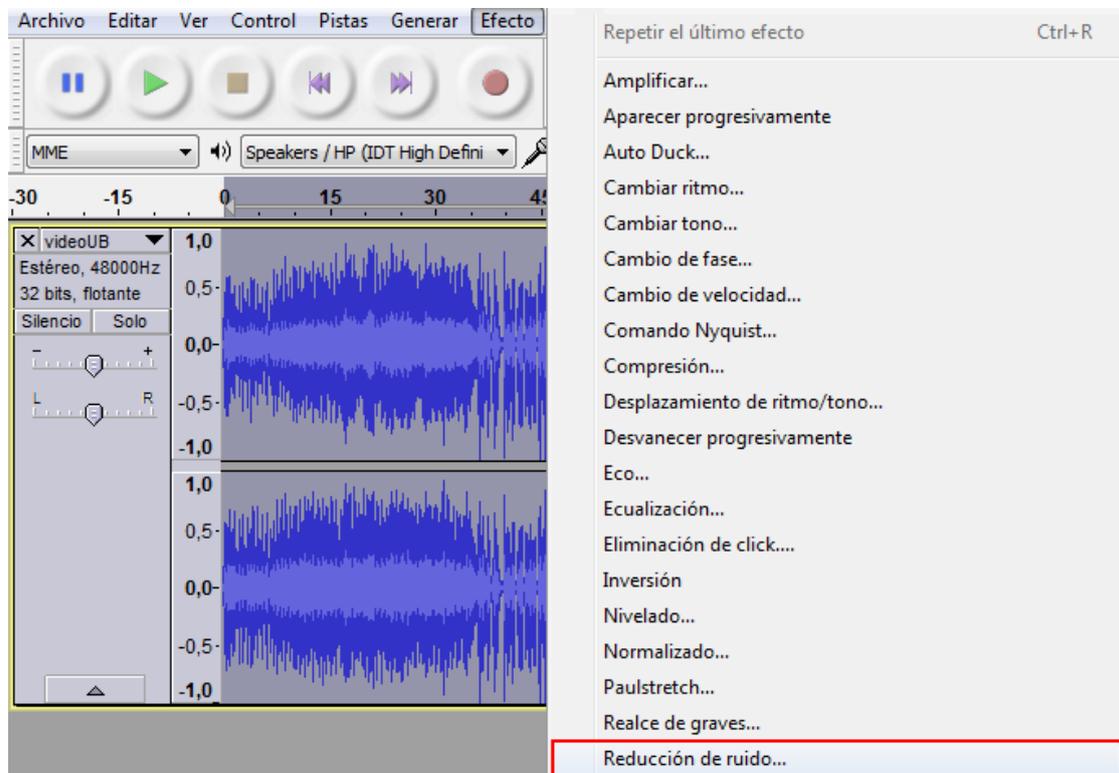


Ilustración 103. Menú para abrir la opción *Reducción de ruido* en Audacity

### Paso 4 Obtención del perfil de ruido

En el paso 1 del cuadro que se abre, si se ha seleccionado previamente la porción del archivo con ruido, hay que apretar el botón *Obtener perfil de ruido*. De esta manera el programa reconoce el perfil de ruido que se quiere filtrar.

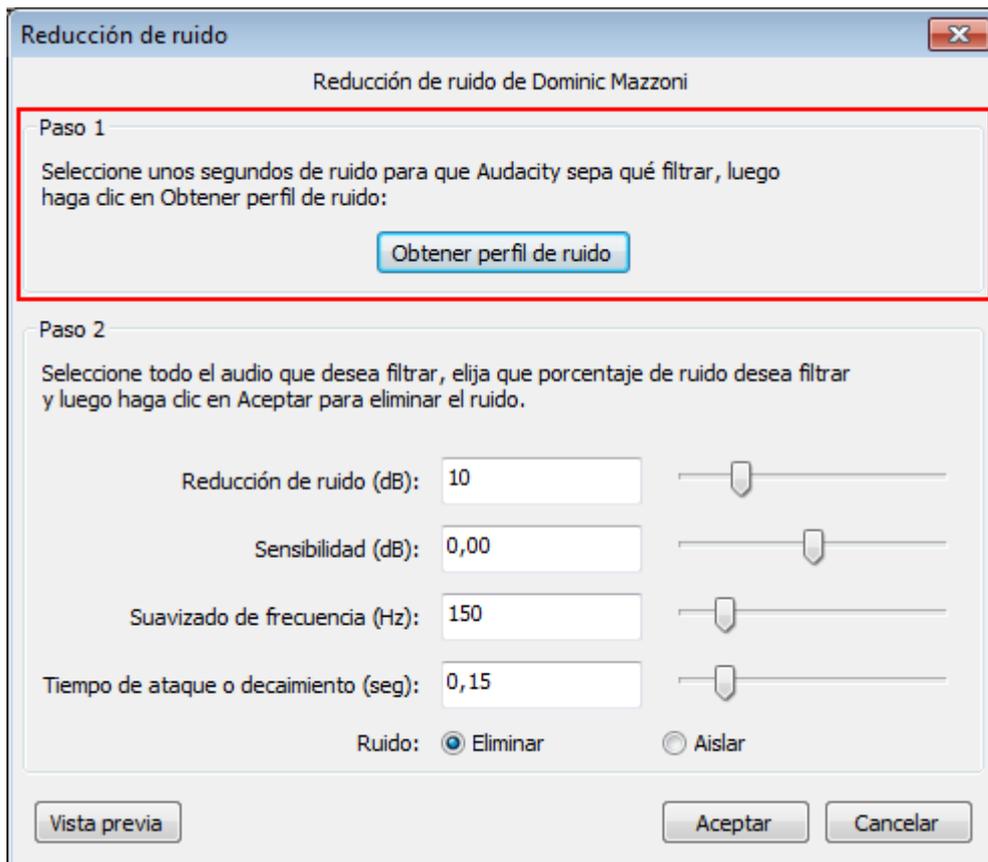


Ilustración 104. Opción de obtención de perfil de ruido con Audacity

A continuación, se retorna a la pantalla con las ondas de sonido y se selecciona todo el archivo, para poder indicar al programa que la limpieza del perfil de ruido que se ha escogido se ha de aplicar a todo el archivo.

Después hay que ir a la opción *Efecto > Reducción de ruido* y continuar con el siguiente paso.

## **Paso 5** Eliminación del ruido

En el paso 2, se puede modificar una serie de parámetros específicos para reducir el ruido.

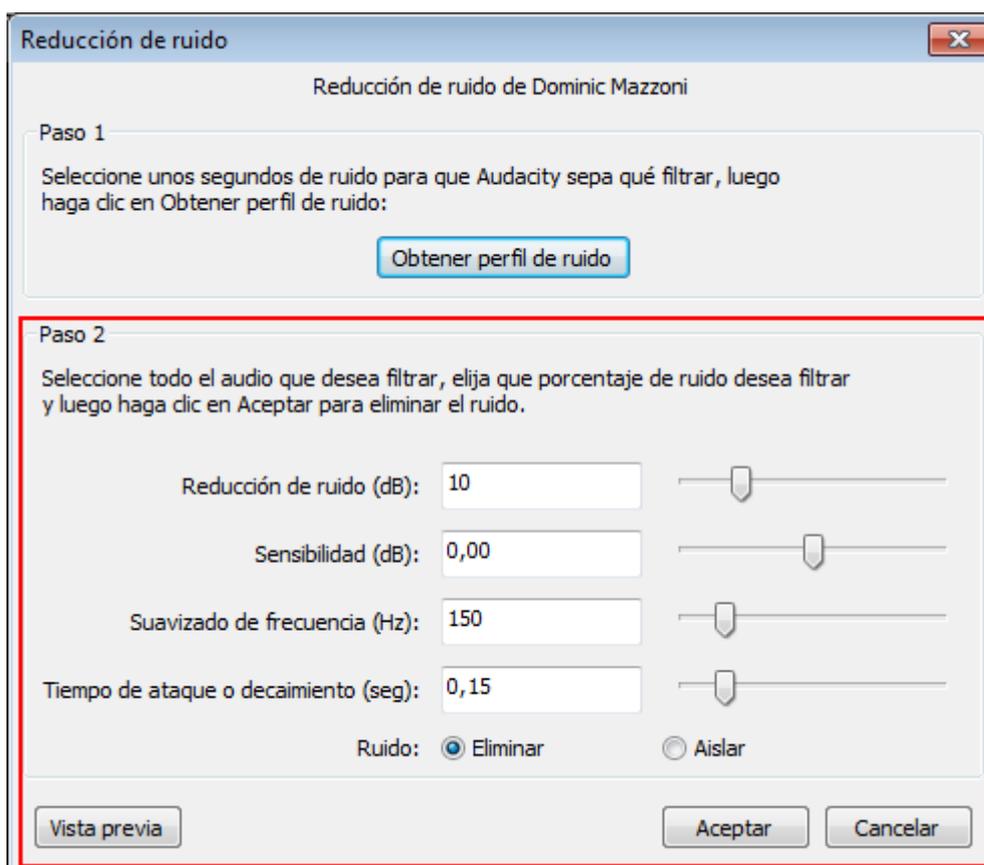


Ilustración 105. Parámetros de filtro de ruido con Audacity

El primer parámetro permite ajustar la intensidad a la que se quiere reducir el ruido (en decibelios). Con un archivo solo de voz, como el caso que se explica en esta guía, el resto de parámetros no resultan particularmente determinantes para mejorar la calidad de la voz. Por tanto, se pueden dejar en los valores predeterminados. (Para más detalles sobre estos parámetros se recomienda consultar la guía de Audacity [5].)

En el paso 2, lo ideal es ir probando las diferentes reducciones de ruido y comprobar los resultados apretando el botón *Vista previa*.

### Atención

El filtro de ruido se ha de aplicar con extrema precaución, ya que podría generar una reducción del sonido de la voz o un efecto metálico, y dificultaría el reconocimiento vocal.

Cuando se esté seguro del resultado, hacemos clic en *Aceptar*. El programa empieza el proceso de limpieza.

## **Paso 6** Guardado y exportación del archivo de sonido

Una vez acabado el proceso de limpieza, guardamos el archivo con la opción *Archivo > Exportar*, seleccionamos el formato de exportación y lo guardamos.

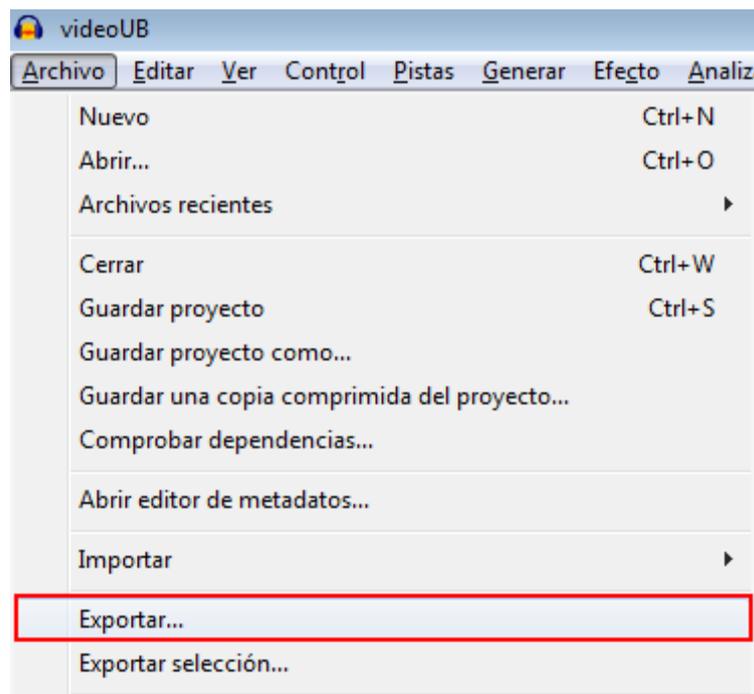


Ilustración 106. Opción para exportar archivos con Audacity

### **Atención**

Se aconseja exportar el archivo a un formato que mantenga inalterada la calidad del sonido, como por ejemplo el formato WAV.

Ahora el archivo de sonido está preparado para comenzar el proceso de transcripción.

## 4. ENTRENAMIENTO DE DRAGON Y TRANSCRIPCIÓN

Dragon Naturally Speaking es un programa de reconocimiento vocal que permite automatizar el proceso de transcripción. La herramienta está diseñada para transcribir la voz de un hablante individual. Para ello se aconseja seguir un proceso previo de entrenamiento (véase la sección 4.1) y de optimización de su diccionario (véase la sección 4.2). Así se consigue adaptar el reconocimiento a una voz específica y optimizar las prestaciones.

Una vez optimizada la herramienta, se puede empezar el proceso de reconocimiento con Dragon (véase la sección 4.3), que da como resultado final una transcripción.

### 4.1 ENTRENAMIENTO DE VOZ

#### Pasos a seguir

#### Paso 1 Abertura de un perfil de usuario

Vamos a *NaturallySpeaking* > *Abrir usuario*.

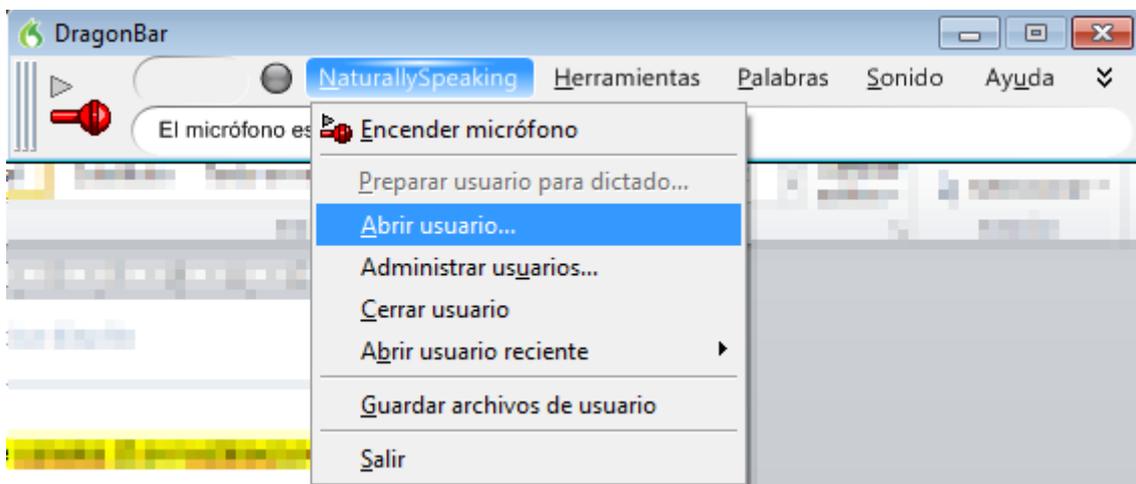


Ilustración 107. Opción para gestionar perfiles de usuarios en Dragon

## Paso 2 Creación de un perfil de usuario

Escogemos la opción *Nuevo* (si es la primera vez que se abre el programa, por defecto pide la creación de un perfil de usuario nuevo).

A continuación se abre un asistente para crear un usuario nuevo. Después de escoger el nombre, el idioma y la fuente de dictado (normalmente la mejor fuente de dictado sería un micrófono con auriculares conectado a la entrada de micrófono del PC), hay que indicar qué tipo de entrenamiento inicial se quiere [6].

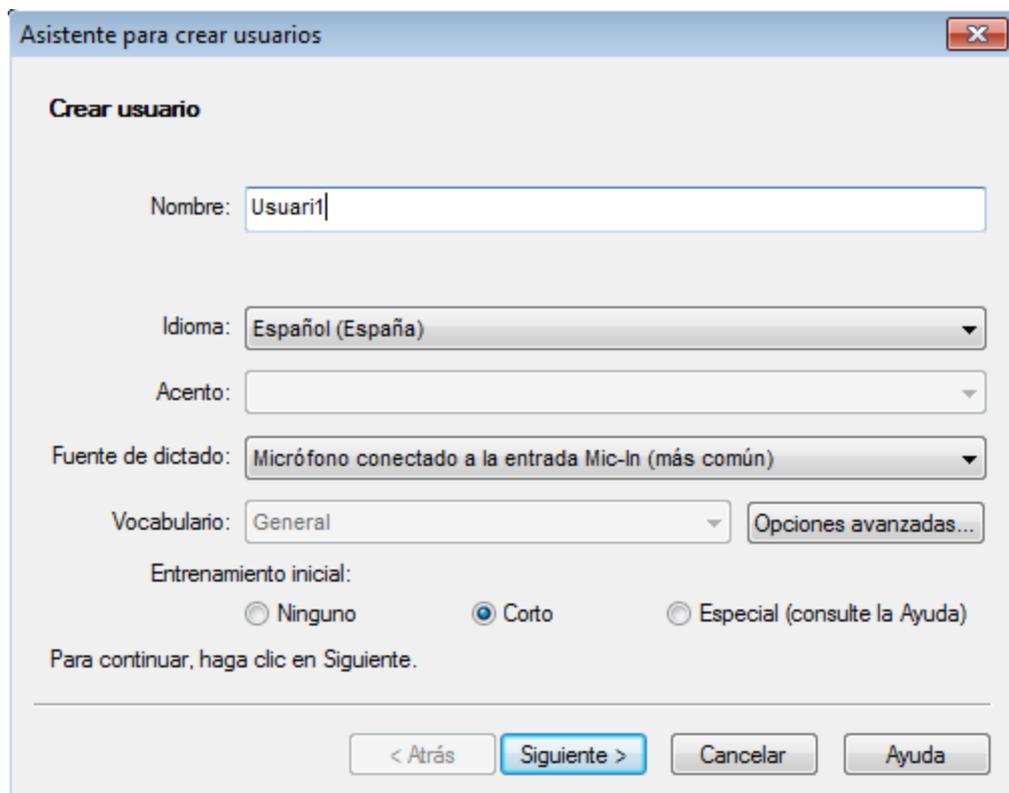


Ilustración 108. Opciones para crear un nuevo usuario en Dragon

Hay tres opciones disponibles:

- *Ninguno*: se omite por completo el entrenamiento y solo se define el volumen del micrófono y la calidad del sistema de sonido del usuario. Opción desaconsejada.
- *Corto* (opción por defecto): se guía al usuario por el programa en la lectura de diferentes textos. Este entrenamiento permite una mayor precisión en el

reconocimiento de palabras respecto a la opción *Ninguno*. Esta es la opción escogida en este manual.

- *Especial*: el usuario lee diferentes textos sin ninguna interacción o indicación de Dragon. Este tipo de entrenamiento está pensado para usuarios con impedimentos en el habla u otras dificultades en la articulación de palabras. Ofrece la posibilidad de imprimir los textos que se han de leer y deja al usuario todo el tiempo que necesite para completar el entrenamiento, incluyendo largas pausas.

### **Paso 3** Optimización del sonido del sistema

Una vez escogida la opción de entrenamiento más adecuada, el asistente guía al usuario en diferentes pasos para elegir el sistema de sonido, el calibrado del volumen y la comprobación de la calidad del sonido del sistema y del micrófono. Después de que el programa haya confirmado que la prueba de sonido ha acabado exitosamente empieza el entrenamiento.

### **Paso 4** Entrenamiento

El entrenamiento consiste en la lectura en voz alta de unas frases y un texto que el programa permite escoger entre diferentes estilos.

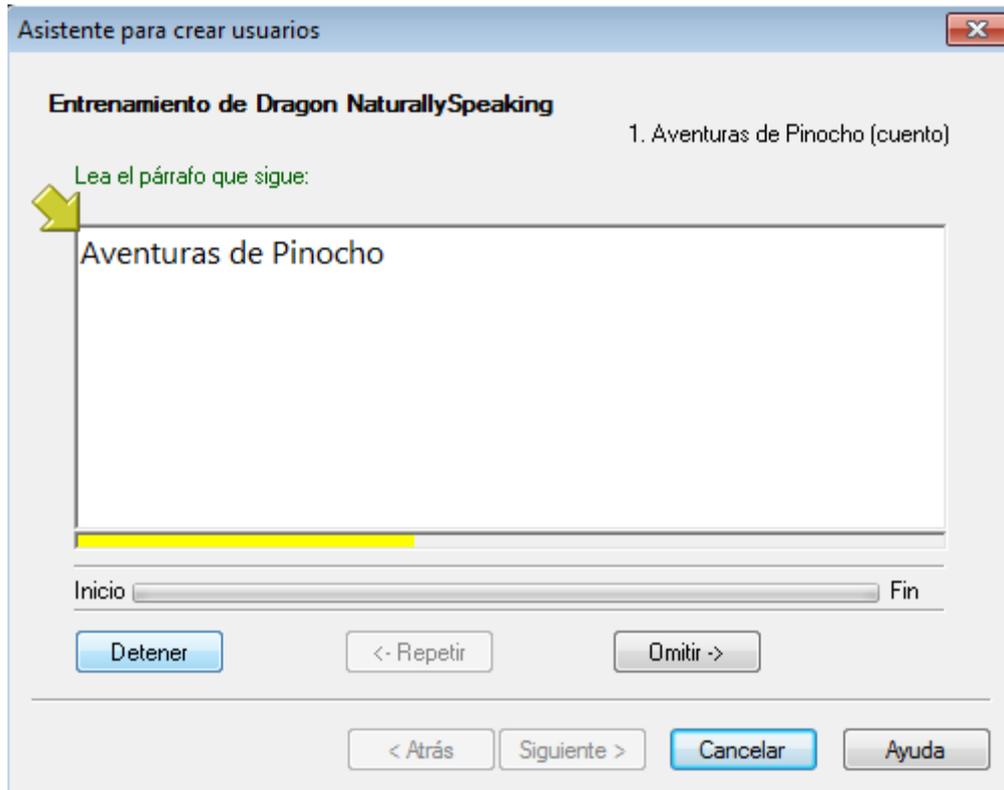


Ilustración 109. Ventana del asistente de entrenamiento de voz de Dragon

Cuando se ha acabado el entrenamiento, el programa tarda unos minutos en adaptar y guardar los archivos del nuevo perfil de usuario. Una vez finalizado este proceso, el programa está optimizado para reconocer el archivo de voz.

## 4.2 OPTIMITZACIÓN DEL DICIONARIO

### Pasos a seguir

#### **Paso 1** Elección de la opción de edición del diccionario

Las opciones relacionadas con el diccionario se encuentran en la pestaña *Palabras* del menú principal de Dragon.

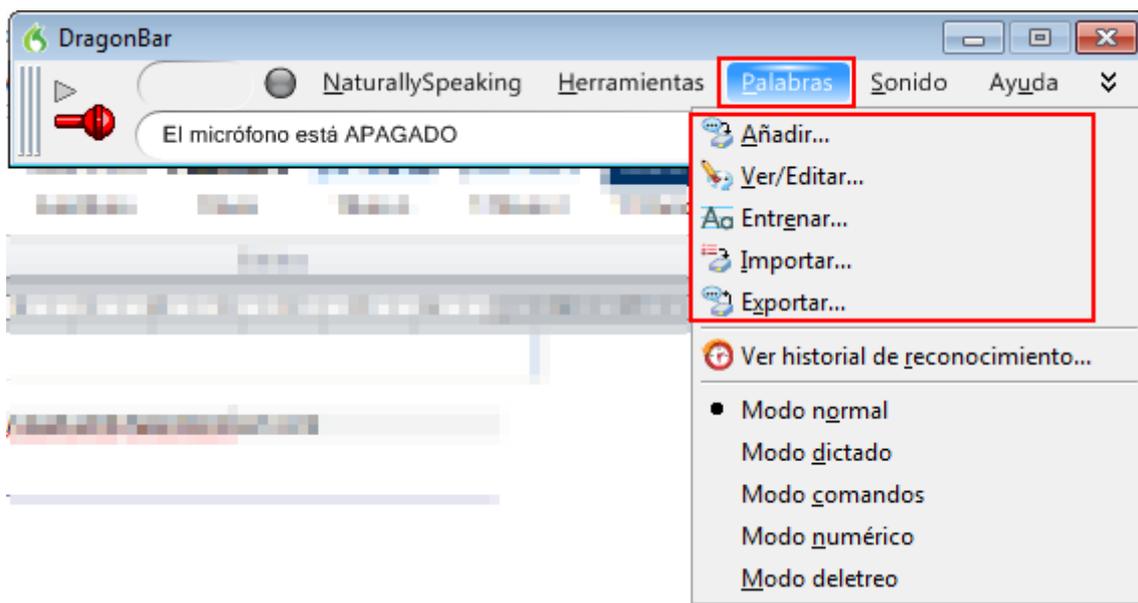


Ilustración 110. Opciones de edición del diccionario de Dragon

Entre las diferentes opciones, se pueden añadir palabras nuevas, importarlas desde archivos de texto o editar las que ya se encuentran en el diccionario.

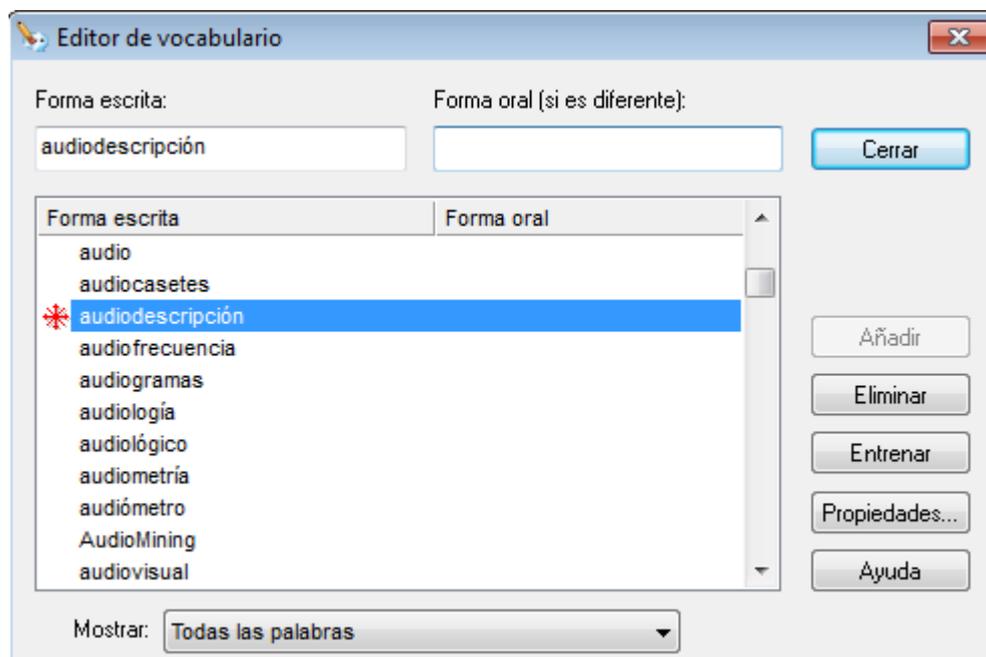


Ilustración 111. Cuadro de diálogo del editor del diccionario de Dragon

## Paso 2 Optimizador automático del diccionario

Otra manera de enriquecer el diccionario es permitir a Dragon analizar nuestros documentos del ordenador para identificar las palabras más usadas. Esta opción en entornos bilingües como el nuestro no es demasiado recomendable.

Ir a *Herramienta > Centro de precisión*.

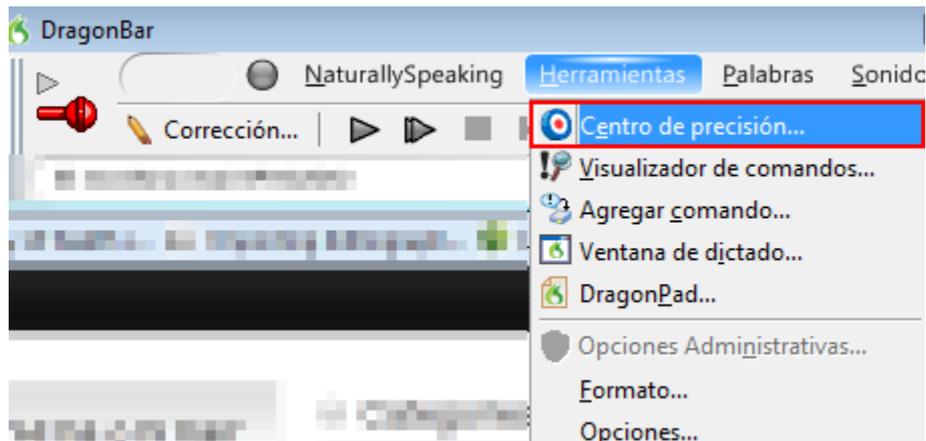


Ilustración 112. Pestaña del menú de Dragon para acceder a la opción *Centro de precisión*

En el cuadro de diálogo *Centro de precisión*, se selecciona la opción *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](#)

Ilustración 113. Opciones del *Centro de precisión* de Dragon

Esta función analiza los correos electrónicos y los documentos presentes en el PC para adaptar el diccionario de Dragon al estilo de escritura del usuario que utiliza el PC y mejorar la precisión del reconocimiento.

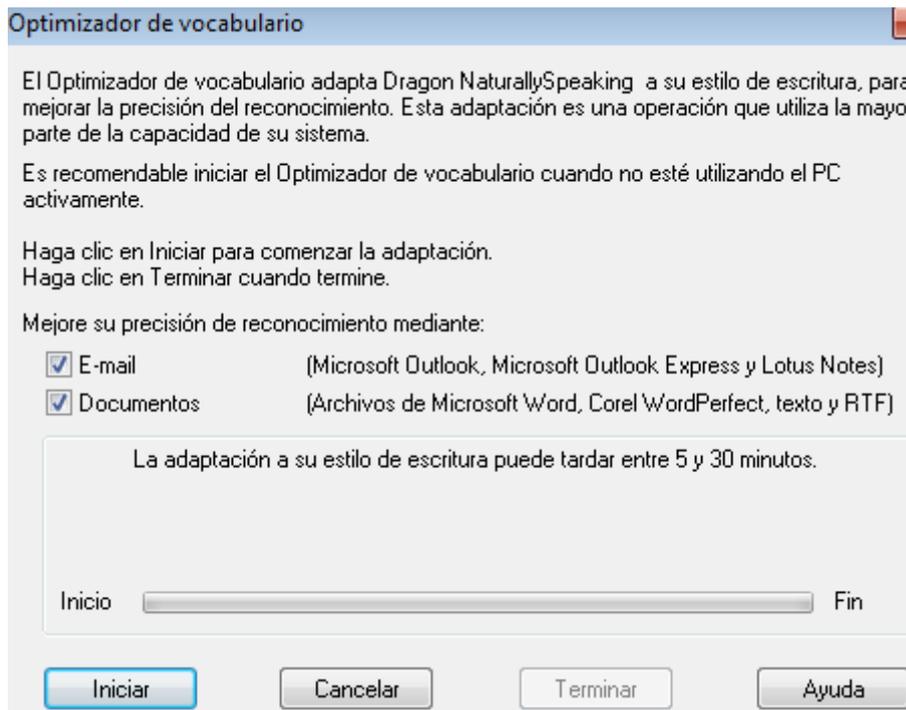


Ilustración 114. Cuadro de diálogo del optimizador de diccionario de Dragon

### 4.3 TRANSCRIPCIÓN CON DRAGON

#### Pasos a seguir

#### Paso 1 Abertura del archivo de sonido con Dragon

Ir a *Sonido > Transcribir grabación*.

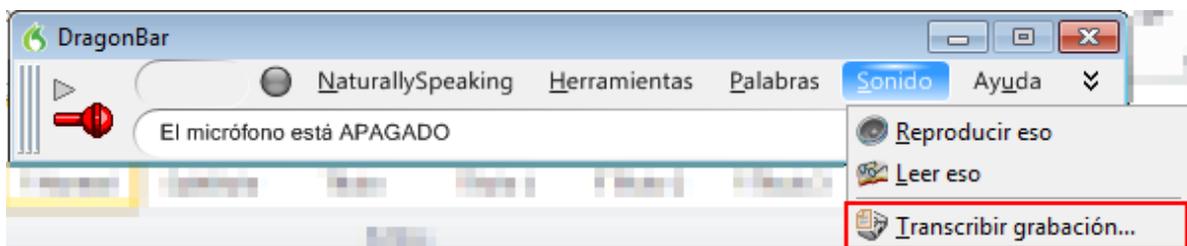


Ilustración 115. Opción de transcripción de grabación de Dragon

Se abre un cuadro de diálogo en el que se han de definir las opciones de transcripción.

## Paso 2 Elección de las opciones de transcripción

Para escoger el archivo de voz que se desea transcribir, se hace clic en el botón *Examinar* en el cuadro de diálogo.



Ilustración 116. Cuadro de diálogo para escoger el archivo a transcribir con Dragon

El botón *Avanzadas* abre un cuadro de diálogo que permite escoger:

- Dónde se transcribe el texto, si en el DragonPad (el programa de edición de texto por defecto de Dragon) o en la ventana de otro editor de texto seleccionado por el usuario.
- Si se desea transcribir solo el texto del dictado o también comandos de Dragon. Si se activa la opción de reconocimiento de comandos, Dragon podría reconocer algunas palabras como comandos de dictado (por ejemplo, *nueva línea*, que activaría el comando *return*). Se aconseja evitar esta opción.

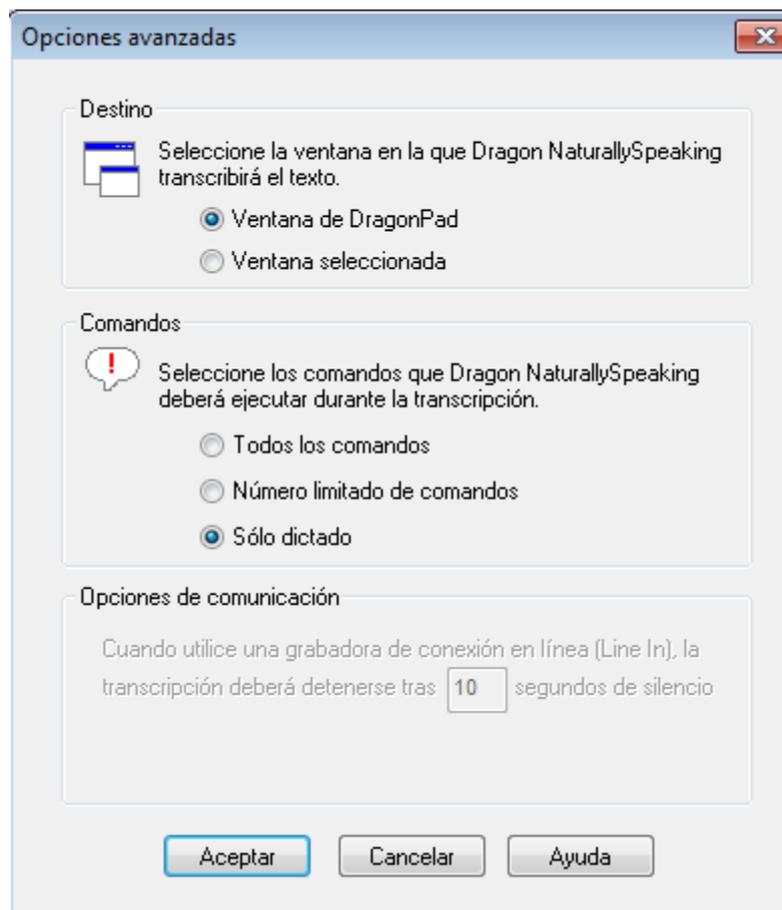


Ilustración 117. Cuadro de diálogo de opciones avanzadas de transcripción de Dragon

### Paso 3 Inicio del proceso de transcripción

Una vez definidas las opciones de transcripción, se puede iniciar la transcripción apretando el botón *Transcribir*. El programa empieza el proceso de transcripción, que se puede controlar directamente en la ventana del editor de texto escogido.

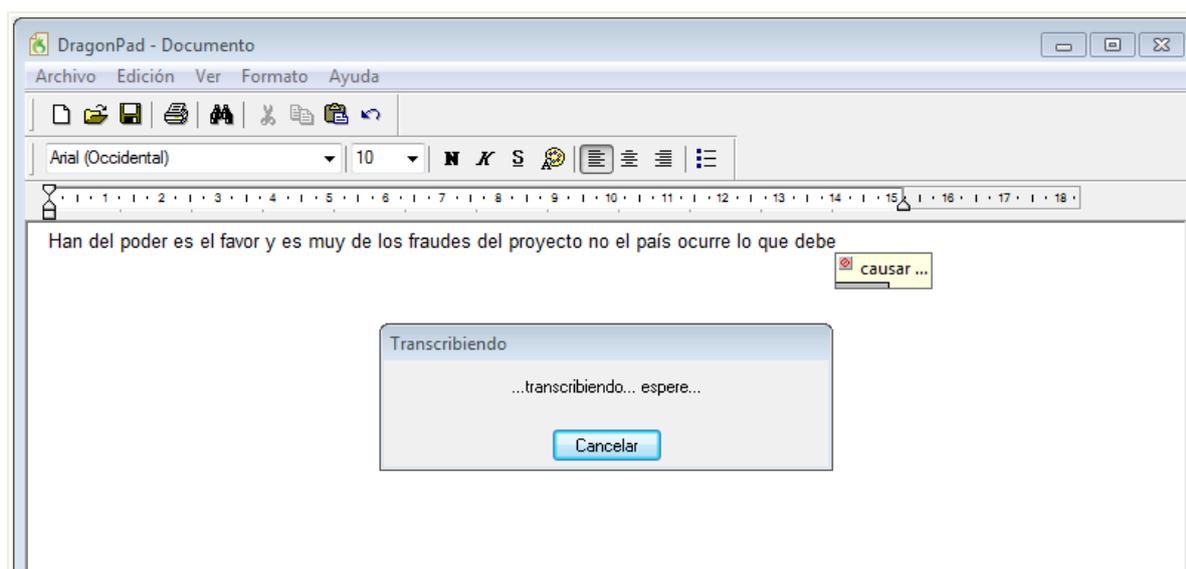


Ilustración 118. Ejemplo de ventana de DragonPad durante el proceso de transcripción

#### Paso 4 Corrección de palabras

Acabada la transcripción, se puede revisar si el texto es correcto y corregirlo dentro del editor de texto. El programa proporciona una serie de opciones de corrección que permiten, entre otras operaciones, especificar a Dragon la palabra adecuada para substituir un error de transcripción, mejorando la precisión del reconocimiento en transcripciones posteriores. Una manera de corregir palabras es situarnos al inicio de la palabra a corregir y después activar el botón *Corrección* en el menú de navegación de Dragon.

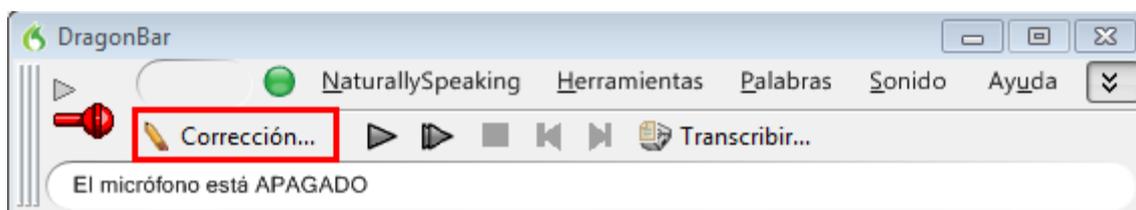


Ilustración 119. Pestaña *Corrección* en el menú Dragon

En el editor de texto se abre un desplegable con diferentes opciones de corrección.

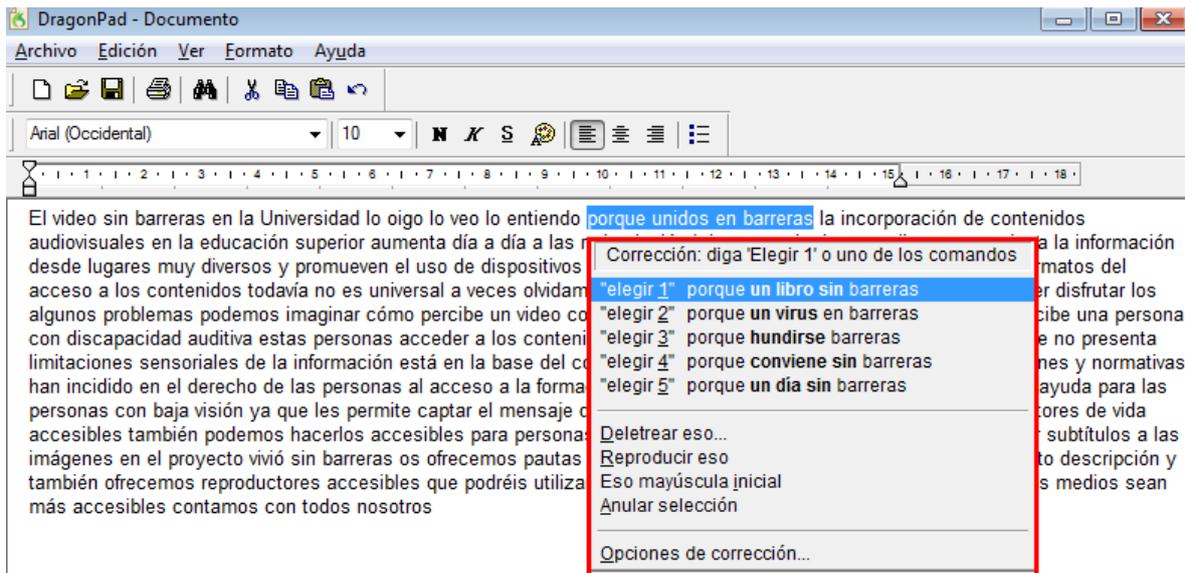


Ilustración 120. Menú de corrección de errores de transcripción de Dragon

Si alguna de las alternativas es correcta, se selecciona para aceptarla. Si no, se puede escribir o dictar el texto correcto.

**Consejo:**

Se recomienda revisar siempre el texto de la transcripción, ya que la capacidad de reconocimiento de Dragon puede llegar como máximo al 98-99 %.

Una vez finalizadas las correcciones en la transcripción, se puede finalmente guardar el archivo de texto, concluyendo así el proceso de transcripción.

## 5. REFERENCIAS

- [1] NUANCE COMMUNICATIONS (2012). *Dragon – Dragon Naturally Speaking – Nuance* [en línea]. Madrid. <<http://www.nuance.es/dragon/>>. [Consulta: 17/10/2012].
- [2] AUDACITY TEAM (2012). *Audacity* [en línea]. <<http://audacity.sourceforge.net/download/windows>>. [Consulta: 17/10/2012].
- [3] *Pazera Free Audio Extractor* (versión 1.4) [en línea]. <<http://www.pazera-software.com/products/audio-extractor/>>. [Consulta: 17/10/2012].
- [4] AUDACITY TEAM (2012). *Audacity 2.0.2 Manual* (en inglés) [en línea]. <<http://manual.audacityteam.org/o/>>. [Consulta: 13/02/2012].
- [5] 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] Las opciones *Acento* y *Vocabulario* no están disponibles en castellano.



Esta obra forma parte de la colección "Guies de Contingut Digital Accessible: vídeo".

Esta guía ha estado creada en el marco del proyecto "El vídeo sense barreres com a suport per la docència en l'educació superior" correspondiente a la convocatòria MQD2010 de la Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) y con el número de expediente: 2010MQD00008, y también gracias al convenio que Agrupació Mútua 2008 ha establecido con el grupo de investigación GRIHO (UdL) para desarrollar tecnologías y proyectos que posibiliten la adaptación y la reinserción laboral a personas con discapacidad en el ámbito de las TIC.



**Universitat de Lleida**



**B** Universitat de Barcelona

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