PROTOCOLO DE TRABAJO EQUIPOS AMERESCO (ELAN)

(Versión actualizada enero 2020)

Es.Vag.Atenuación

La atenuación pragmática en su variación genérica: géneros discursivos escritos y orales en el español de España y América

Proyecto Mineco FFI2016-75249-P

http://esvaratenuacion.es/

http://corpusameresco.com/

Protocolo de trabajo para equipos Ameresco para transcripción en ELAN

Equipo de redacción:

Briz Gómez, Antonio
Albelda Marco, Marta
Estellés Arguedas, Maria
Cabedo Nebot, Adrián
García Ramón, Amparo
Soler Bonafont, Amparo
Uclés Ramada, Gloria
Carcelén Guerrero, Andrea

Protocolo de trabajo equipos AMERESCO by Briz et al. is licensed under a Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional License.
Creado a partir de la obra en http://corpusameresco.com/.

1. EL CORPUS AMERESCO 3
2. CRONOGRAMA 3
3. GRABACIONES 4
3.1. Metodología y cuestiones prácticas sobre las grabaciones 4
3.2. Criterio de selección de hablantes 4
3.3. Características de las grabaciones 5
4. SISTEMA DE TRANSCRIPCIÓN 8
4.1. Características del sistema de transcripción 8
4.2. Transcripción Val.Es.Co. para ELAN 9
4.3. Símbolos y convenciones del sistema Val.Es.Co. que se mantienen 11
4.4. Símbolos y convenciones del sistema Val.Es.Co. no necesarios para ELAN 11
4.5. Preguntas frecuentes 12
5. Anonimización del audio CON AUDACITY 14
5.1. ¿Qué es Audacity? 14
5.2. Anonimización semiautomática con ELAN y Audacity (opción recomendada). 14
5.3. Anonimización manual con Audacity 18
6. FICHA TÉCNICA 22
7. ENTREGA DE ARCHIVOS 24
8. MODO DE CITAR EL CORPUS 24
ANEXOS 25

1. EL CORPUS AMERESCO

El proyecto Ameresco (América y España español coloquial) surge de la mano de Antonio Briz como extensión natural del corpus Val.Es.Co. en 2010, con el fin de profundizar en el estudio de la variedad coloquial del español en geolectos europeos y americanos. El principal resultado de trabajo del proyecto Ameresco es la recopilación del Ameresco (Albelda y Estellés, en línea).

El corpus Ameresco es un conjunto de subcorpus de conversaciones coloquiales obtenidas en las distintas ciudades que se integran en el proyecto. Cada equipo de trabajo será propietario de su corpus, aunque se compromete a cederlo para que sea integrado en una base de datos común.

Esta base de datos es de acceso libre para todos los investigadores que lo deseen. Cada corpus aparecerá firmado por el equipo de trabajo responsables y podrá ser usado también para el análisis de fenómenos no relacionados con el proyecto.

Una de las características de la conversación coloquial es que la alternancia de turnos no está predeterminada; es decir, en ella los participantes toman el turno espontánea y libremente, y no hay ningún hablante que tenga el rol de hacer preguntas ni dirigir la interacción.

Es importante, por tanto, que el corpus sea de conversaciones espontáneas y no de entrevistas, género en el que el grado de planificación es más algo y existe mayor predeterminación en la alternancia de turnos. Así pues, se tratará de conversaciones informales (coloquiales) grabadas de manera secreta en lugares cotidianos para los hablantes, en una situación de familiaridad o amistad.

El corpus Ameresco sigue el mismo protocolo de obtención y recogida de datos que se estableció para el primer corpus del Grupo Val.Es.Co. (Briz et alii 2002), al cual pertenece Ameresco. Una de las principales novedades del corpus Ameresco es la incorporación de un etiquetado pragmático-conversacional y el alineado de la transcripción y el audio.

2. CRONOGRAMA

El objetivo propuesto es que todas las conversaciones estén grabadas y transcritas en la fecha prevista por cada equipo. El cronograma se concretará con cada uno de forma individual según sus posibilidades.

Sugerimos realizar este proceso (de grabación, por un lado y de transcripción, por otro) en dos fases cronológicas[1].

Cada equipo habrá de proponer, por tanto, su propio cronograma de trabajo, adaptado a sus circunstancias y enviarlo a la coordinación del proyecto. Desde el equipo central proponemos dos fechas de entrega divididas de la siguiente manera:

- Primera entrega de archivos: grabación y transcripción de 12 conversaciones.
- Segunda entrega de archivos: grabación y transcripción de las 15 conversaciones restantes.

[1] No obstante, los equipos pueden considerar otra posibilidad, en función de la disponibilidad de sus recursos.

3. GRABACIONES

3.1. Metodología y cuestiones prácticas sobre las grabaciones

La muestra que deberá recogerse será, en total y a ser posible, de 27 conversaciones y seguirá la distribución por rasgos sociolingüísticos (sexo, edad y nivel sociocultural) que se especifica en el *Cuadro 1*.

Se recomienda a los investigadores responsables de cada ciudad que creen un equipo de trabajo para la grabación y transcripción de las conversaciones. La recogida y confección del corpus es paulatina, es decir, comenzaremos con un número pequeño de grabaciones y poco a poco se irán sumando otras hasta obtener el número previsto.

Se recomienda, también, que cada equipo lleve un registro de las muestras extraídas en cada fase de elaboración del corpus.

3.2. Criterio de selección de hablantes

CUADRO 1. RESUMEN DE LA MUESTRA EXTRAÍDA												
EDAD	*NIVEL SOCIOCULTURAL*									*TOTAL*		
	Alto			Medio			Bajo					
		Partici-pantes	Conver-saciones		Partici-pantes	Conver-saciones		Partici-pantes	Conver-saciones		Partici-pantes	Conver-saciones
18-25	V* 4 M** 4	8	3	V 4 M 4	8	3	V 4 M 4	8	3	V 12 M 12	24	9
26-55	V 4 M 4	8	3	V 4 M 4	8	3	V 4 M 4	8	3	V 12 M 12	24	9
≥56	V 4 M 4	8	3	V 4 M 4	8	3	V 4 M 4	8	3	V 12 M 12	24	9
Total	V 12 M 12	24	9	V 12 M 12	24	9	V 12 M 12	24	9	V 36 M 36	**72**	**27**

* Varón
** Mujer

En cuanto al nivel sociocultural, tendremos en cuenta únicamente el nivel de instrucción:

- **Nivel alto**: estudios superiores
- **Nivel medio**: estudios de secundaria y formación profesional
- **Nivel bajo**: estudios primarios o sin estudios

Según los totales del cuadro, tenemos **una tabla de mínimos con 27 conversaciones** con las que, sumando los participantes necesarios para cubrir las diferentes celdas (los estratos sociolingüísticos), obtenemos un total de **72 hablantes**, 8 por cada grupo etario y nivel de estudios (4 mujeres y 4 varones).

Esta cantidad podrá ser mayor ya que el plan de grabaciones no es rígido y no siempre podremos cuadrar a los hablantes que necesitamos para cumplir con cada casilla.

Los hablantes se pueden cruzar, es decir, podemos tener en una misma grabación a hablantes de diferentes grupos etarios, sexo y nivel sociocultural.

Los hablantes que actúen pasivamente en la conversación no serán tenidos en cuenta en el cómputo ya que su participación es mínima y no cuenta a efectos de representatividad.

3.3. Características de las grabaciones

Recomendación previa

Antes de comenzar a grabar masivamente, se recomienda encarecidamente que se graben y transcriban 1 o 2 conversaciones y que se envíen a la coordinación del proyecto para que se revisen, de manera que, si se detectara algún inconveniente sobre la naturaleza de la grabación, se pueda paliar antes de que se grabe todo el corpus.

Carácter secreto de la grabación. Autorizaciones

Las **grabaciones serán secretas**, con petición de autorización[2]. En todos los casos se solicitará autorización a cada uno de los participantes, incluido el investigador o persona que graba[3]. En caso de no obtener dicha autorización, la grabación deberá ser destruida.

Esta autorización **necesita tres firmas por hablante**: una primera firma antes de realizar la grabación para demostrar que fueron avisados con antelación pero que no sabrán en qué momento será grabados; la segunda se recogerá una vez realizada la grabación. Por último, el consentimiento para el tratamiento de datos personales de acuerdo con la Ley Orgánica de Protección de Datos 15/1999, de 13 de diciembre y la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de datos personales y garantía de los derechos digitales.

Soporte digital

Hay que asegurarse de que la grabadora sea de buena calidad. Si es posible, es mejor grabar en **formato wav**[4]. Se pueden utilizar *smartphones* si realizan grabaciones óptimas.

Espacios

La persona que realice la grabación, sea o no el investigador, debe guardar con los participantes una relación vivencial de proximidad y el marco interaccional debe ser familiar, para asegurar un mayor control de los factores contextuales.

Convendrá que las grabaciones se realicen en espacios sin ruido (o con el menor ruido posible), para facilitar la transcripción. Por ello, es recomendable que las grabaciones se hagan en espacios interiores[5].

---

[2] El modelo de autorización se adjunta como anexo al final de este documento.
[3] El investigador encargado de la grabación deberá firmar tantas autorizaciones como grabaciones en las que participe, no es válido que solo haya una autorización del investigador para todas las grabaciones.
[4] El programa ELAN exige este formato. En caso de que la grabación esté en formato mp3 o mp4, deberá convertirse a wav. Existen conversores gratuitos en línea.
[5] Hay que evitar que haya ruido de fondo constante durante toda la grabación ya que impediría un posterior análisis prosódico. Por ejemplo, no debe grabarse en bares o cafeterías concurridas o dentro de un coche con la radio encendida, en casa con la música, la televisión puesta, perros ladrando constantemente u otros casos de este estilo.

Número de interlocutores

Como **mínimo 2 y como máximo 4**. Un número mayor de participantes dificulta la transcripción. Además, este número es suficiente para que puedan producirse conversaciones paralelas.

No serán válidas aquellas conversaciones en las que haya dos hablantes si uno de ellos es el encargado de la grabación puesto que el grado de espontaneidad y el carácter secreto se ve comprometido.

Duración de las grabaciones

Dadas las características de la conversación coloquial espontánea la media de duración de las grabaciones no se puede prever. Deberá tenerse en cuenta que no se deberá forzar la conversación para cumplir con una duración fija ni se deberá cortar una grabación que sea una buena muestra de conversación espontánea por haber superado un límite de tiempo concreto.

Como tiempos únicamente orientativos, la mayor parte de las conversaciones recogidas suele ocupar entre 20 y 60 minutos.

Identificación de las grabaciones (ID)

El equipo central adjudicará una ID a cada archivo según este sistema:

VLC_001_03_18

- Tres o cuatro primeras letras: nombre de la ciudad. Se tomarán las abreviaturas de los códigos oficiales de los aeropuertos[6].

  VLC (Valencia)

- Tres dígitos siguientes: código de grabación interno de cada ciudad, por orden correlativo a medida que se vayan grabando.

  001 indica que es la primera grabación.

- Dos dígitos siguientes: número de participantes en la conversación.

  03 quiere decir que hay 3 hablantes.

- Dos últimos dígitos: año de recogida de la grabación, las dos últimas cifras teniendo en cuenta que se graba después del año 2000.

  18 indica que fue grabado en el 2018.

- Entre cada bloque añadiremos una barra baja _

[6] En el caso de no existir ningún aeropuerto, se toman las tres primeras letras de la ciudad.

Otros ejemplos:

MTY_012_03_16: corresponde a la ciudad de Monterrey, grabación número 12 de acuerdo con el número interno de grabaciones de este equipo, 3 participantes y grabada en 2016.

TUC_001_02_14: corresponde a la ciudad de Tucumán, grabación número 1 de acuerdo con el número interno de grabaciones de este equipo, hay 2 participantes y se grabó en 2014.

4. SISTEMA DE TRANSCRIPCIÓN

El objetivo final de este trabajo es conseguir que todas las grabaciones estén transcritas siguiendo las convenciones del sistema de transcripción utilizado por el grupo de investigación Val.Es.Co.[7] combinado con el sistema de etiquetas que se detalla a continuación. Para ello se utilizará el programa ELAN[8].

ELAN está diseñado para la anotación lingüística y permite alinear la transcripción con el audio. Se distribuye como *software* libre y es de código abierto. Puede descargarse en el siguiente enlace: https://tla.mpi.nl/tools/tla-tools/elan/download/

Recomendación previa

Antes de transcribir todos los archivos se recomienda encarecidamente que se transcriban 1 o 2 conversaciones y se envíen a la coordinación del proyecto para que sean revisadas, de manera que, si se detectara algún inconveniente, se pueda solucionar antes de que se transcriba todo el corpus.

4.1. Características del sistema de transcripción

Para transcribir en ELAN se emplearán símbolos del sistema de transcripción de Val.Es.Co. y etiquetas basadas en el sistema desarrollado por el Proyecto PRESEEA, formato compatible con XML que garantiza la recuperación posterior de la información por medio del motor de búsqueda. Este sistema está concebido como codificación interna.

Diferenciamos dos tipos de etiquetas:

- **etiquetas simples** (compuestas por un solo elemento <ininteligible/>)
- **etiquetas dobles** (compuestas por dos elementos, uno de apertura y otro de cierre como vemos en <cita></cita>).

Además, algunas etiquetas contienen **atributos**, información que debe aparecer dentro de las comillas (es el caso de <énfasis t=“”></énfasis> u <obs t=“”/>).

Las etiquetas simples van entre espacios, a excepción de <alargamiento/>, que se escribe pegada a la palabra a la que afecta.

A: me dijo tu primo que<alargamiento/> bue<alargamiento/>no que no sabía

B: bueno es que él nunca sabe <risas/> mi primo nunca sabe

Las etiquetas de inicio y fin de las etiquetas dobles irán pegadas al fragmento al que acompañan.

A: ayer vi una película que <énfasis t=“silabeo”>madre mía</énfasis> era buenísima

[7] A. Briz y Grupo Val.Es.Co. (2002), *Corpus de conversaciones coloquiales*, Anejo 1 Oralia, Madrid, Arco Libros.
[8] Consúltese el documento de Uclés (2019), *Transcripción de conversaciones con ELAN*.

4.2. Transcripción Val.Es.Co. para ELAN

<alargamiento/>

Marca tanto alargamientos vocálicos como consonánticos.

Ejemplo: camió<alargamiento/>n en<alargamiento/>

<ininteligible/>

Marcar fragmentos ininteligibles.

Ejemplo: parece que <ininteligible/> mañana

<extranjero t=“ ”></extranjero>

Señala el uso de un extranjerismo. Entre las dos etiquetas se introduce la voz extranjera a la que se hace referencia y en el campo atributo la pronunciación de lo que se oye.

Ejemplo: <extranjero t="pílic">peeling</extranjero>

<siglas t=“ ”></siglas>

Marca el uso de siglas. Entre las dos etiquetas se introduce la sigla y en el campo atributo la pronunciación de lo que se oye.

Ejemplo: huyo de los captadores de las <siglas t=“o ene ges”>ONG</siglas>

<sic></sic>

Marca pronunciación errónea que no es un error de transcripción.

Ejemplo: cómete las <sic>almóndigas</sic>

<entre_risas></entre_risas>

Delimita fragmentos entre risas.

Ejemplo: mañana <entre_risas>vamos a la playa</entre_risas>

<risas/>

Señala que alguno de los hablantes ríe.

<tos/>

Señala que alguno de los hablantes tose.

<gritos/>

Señala que alguno de los hablantes chilla.

<énfasis t=“ ”></énfasis>

Expresa pronunciación marcada[9] o silabeo.

Ejemplo: eres un <énfasis t=“silabeo”>pesado</énfasis>

eres un <énfasis t=“pronunciación_marcada”>pesado</énfasis>

<susurro></susurro>

Etiqueta para marcar voz baja o susurros.

Ejemplo: mañana es <susurro>la fiesta sorpresa</susurro>

<anónimo></anónimo>

Marca las anonimizaciones. El protocolo para anonimizar los nombres de personas y de lugares que puedan poner en peligro de confidencialidad de los participantes consiste en cambiar por nombres parecidos e introducir esta etiqueta (ver apartado 4.5.)

Ejemplo sin anonimizar: ayer vi a María y a Pedro

Anonimizado: ayer vi a <anónimo>Marta</anónimo> y a <anónimo>Pablo</anónimo>

<cita></cita>

Indica que un se está reproduciendo el estilo directo.

Ejemplo: y luego me dijo <cita>mañana no iré a trabajar</cita>

<fsr t=“ ”></fsr>

Recoge fenómenos de fonética sintáctica y en general aquellos casos en los que la ortografía y la pronunciación de una palabra no coinciden.

Ejemplo: voy <fsr t=“pa’l”>para el</fsr> supermercado

<fsr t=“tonces”>entonces</fsr> le dijo que no lo hiciera

<fsr t=“toa”>toda</fsr> la noche

<fsr t=”po”>pues</fsr> entonces no sé

<obs t=“ ”></obs>

Recoge cualquier tipo de información relevante para la transcripción que no se recoge en el resto de las etiquetas. Se utiliza con fenómenos que solo afecten a un hablante.

Ejemplo: <obs t=“da palmas”>madre mía</obs>

veníamos de la <obs t=“lugar de trabajo de la hablante B”>cueva</obs>

[9] El atributo para pronunciación marcada lleva una barra baja: “pronunciación_marcada”.

<obs t=“ ”/>

Recoge cualquier tipo de información relevante para la transcripción que no se recoge en el resto de las etiquetas.

Siempre debe aparecer en una línea dedicada exclusivamente a este tipo de observaciones. Se utiliza con fenómenos que afectan más de una intervención.

4.3. Símbolos y convenciones del sistema Val.Es.Co. que se mantienen

[	lugar donde se inicia un solapamiento o superposición.
]	final del habla simultánea.
-	reinicios y autointerrupciones sin pausa.
((siempre))	transcripción dudosa.
¿ ?	interrogaciones.
¿¡ !?	interrogaciones exclamativas.
¡ !	exclamaciones.
?	cuando no se reconozca el interlocutor se creará una línea en ELAN

4.4. Símbolos y convenciones del sistema Val.Es.Co. no necesarios para ELAN

Pausas

Puesto que las transcripciones se van a alinear con el sonido con el programa ELAN, no es necesario marcarlas.

Tonemas

No se tienen que marcar los tonemas, aunque no coincidan con la pronunciación habitual.

Sucesión inmediata entre interlocutores

El sistema de transcripción Val.Es.Co. utiliza el símbolo § para indicar una sucesión inmediata, sin pausas, entre interlocutores. No es necesario marcar este fenómeno, puesto que las transcripciones se van a alinear con el sonido con el programa ELAN.

Mantenimiento del turno de un participante en un solapamiento

Representado por el símbolo =. ELAN permite ver quién mantiene el turno sin necesidad de marcarlo con un símbolo adicional en la transcripción.

4.5. Preguntas frecuentes

Detallamos a continuación aquellas cuestiones relacionadas con las convenciones de transcripción que han generado más dudas hasta ahora.

Identificación de hablantes

Los hablantes deben identificarse con una letra mayúscula (A, B, C, etc.). No puede utilizarse la inicial del nombre real.

En la ficha técnica se identificará a cada hablante con esta misma letra. No puede aparecer el nombre real, este solo aparecerá en las autorizaciones.

Correcto	Incorrecto
A: ¿puedes darme eso? B: claro/ toma C: déjamelo cuando acabes/ por favor	M: ¿puedes darme eso? J: claro/ toma S: déjamelo cuando acabes/ por favor
A: [¿o sea ya juntos?] B: [o sea iban a hacer]	Carmen: [¿o sea ya juntos?] Antonio: [o sea iban a hacer]

Signos de puntuación

No se utilizan signos de puntuación en la transcripción a excepción de signos de interrogación y exclamación.

Correcto	Incorrecto
A: me trajo a casa se despidió y se fue	A: me trajo a casa, se despidió y se fue.

Mayúsculas

No se utilizan las mayúsculas por razones de puntuación. Solamente se utilizan para los nombres propios y siglas.

Correcto	Incorrecto
A: ¿y con quién vive?/ ¿con su tía Ana?	A: ¿Y con quién vive? ¿Con su tía Ana?

Vacilaciones y cambio de plan sintáctico

Se marca mediante guion -, no con raya —. No debe indicarse de manera textual.

Si en una vacilación la misma palabra se repite varias veces, se transcribirá tantas veces como sea pronunciada.

Correcto	Incorrecto
A: tres diosas ¿no? pero os vei- os veis muy poco	A: tres diosas ¿no? pero os vei— os veis muy poco
A: o sea el ancho la- la- la- medida exacta	A: o sea el ancho la (vacilación) medida exacta A: o sea el ancho la (repetición de palabra) medida exacta

Aclaraciones metalingüísticas

No se incluyen dentro de la transcripción. De necesitar añadirlas, se hará utilizando las etiquetas <obs t=“”></obs> u <obs t=“”/> según afecten a uno o a todos los hablantes.

Números y símbolos

Se transcriben con letras.

Correcto	Incorrecto
cien mil dólares	$ 100.000

Anonimización de datos personales

No pueden aparecer datos reales de ninguna persona ni lugares reconocibles que permitan identificar a los hablantes. Habrá que sustituirlo por un nombre ficticio.

Se deberá utilizar, en la medida de lo posible, referentes más o menos equivalentes al que se está sustituyendo.

Sí es necesario anonimizar: nombres de los hablantes[10], amigos, familiares y otras personas que se mencionen, lugares de residencia, de trabajo u otros datos que permitan identificarlos, nombres de mascotas.

No es necesario anonimizar: nombres de personajes públicos (políticos, famosos, etc.), lugares comunes[11], marcas.

[10] El nombre ficticio elegido deberá mantenerse durante toda la transcripción.

[11] Si, por ejemplo, están diciendo que ayer comieron en Burger King, no haría falta anonimizarlo, pero si dijeran *ayer comimos en el Burger King que está debajo de mi casa*, en ese caso sí. Si hablan de una obra de García Márquez no habría que anonimizarlo pero si dice que su hermana trabajó durante 10 años para García Márquez, en ese caso sí sería necesario.

5. ANONIMIZACIÓN DEL AUDIO CON AUDACITY

Existen varias maneras de realizar la anonimización del audio y varios programas informáticos para este fin. Exponemos a continuación dos de las formas posibles, semiautomática y manual, para realizar esta tarea y que no presentan complicaciones.

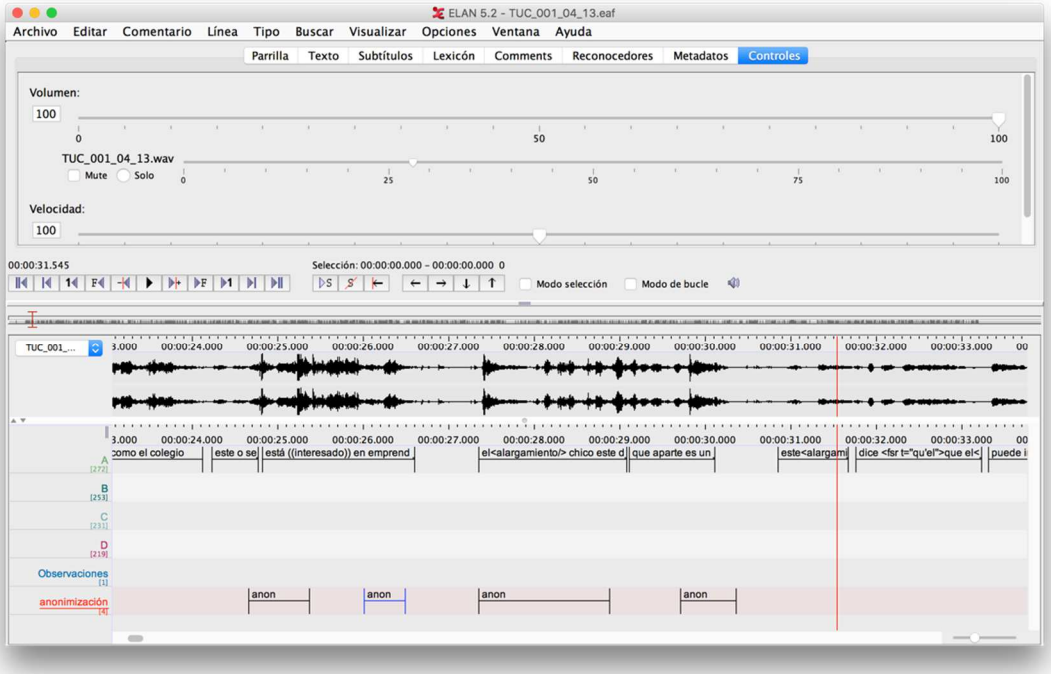
5.1. ¿Qué es Audacity?

Audacity es una aplicación informática libre que se puede usar para la edición de audio. Puede descargarse de manera gratuita en el siguiente enlace: https://www.audacityteam.org/

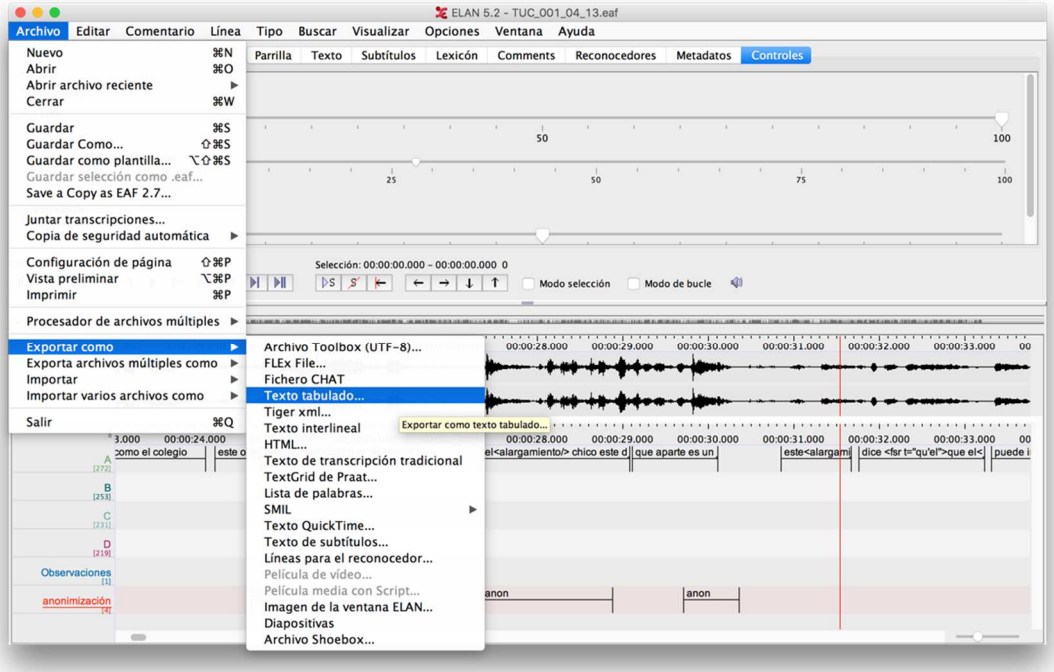
5.2. Anonimización semiautomática con ELAN y Audacity (opción recomendada)

La anonimización de archivos de audio asociados a una transcripción se puede hacer de forma semiautomática con el uso de ELAN y Audacity. En primer lugar, crearemos una nueva línea (véase sección 1.5.1) en el archivo de ELAN en la que se seleccionan todos los fragmentos que requieren anonimización.

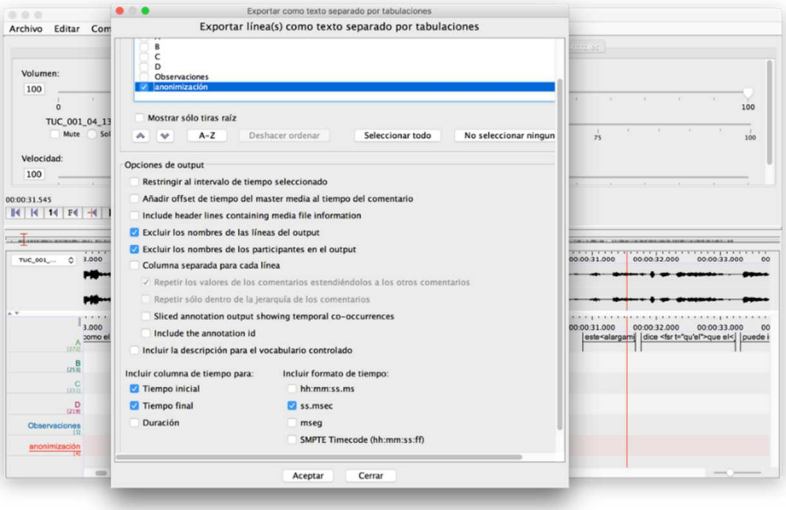
IMPORTANTE: esta línea solo se utilizará para delimitar fragmentos donde aparezca la información que se debe silenciar en el archivo de audio. La anonimización para la transcripción se debe hacer en la línea del hablante correspondiente y supone, pues, una tarea diferente.



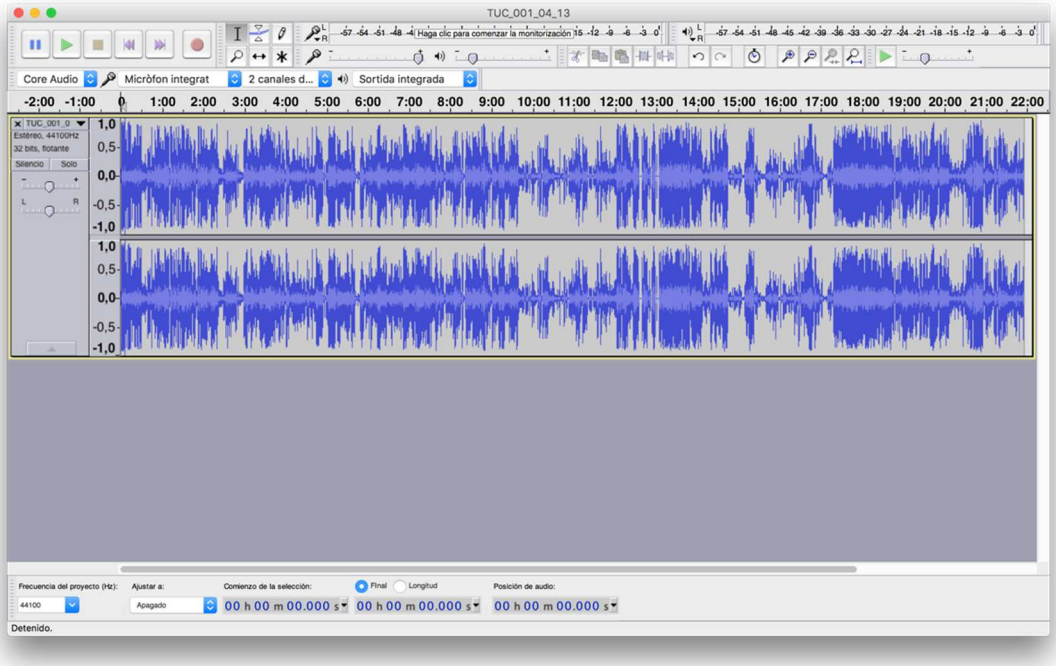
Una vez se hayan seleccionado todos los fragmentos que queremos silenciar en el audio final, se exporta solamente la línea de anonimización con el formato de texto tabulado.



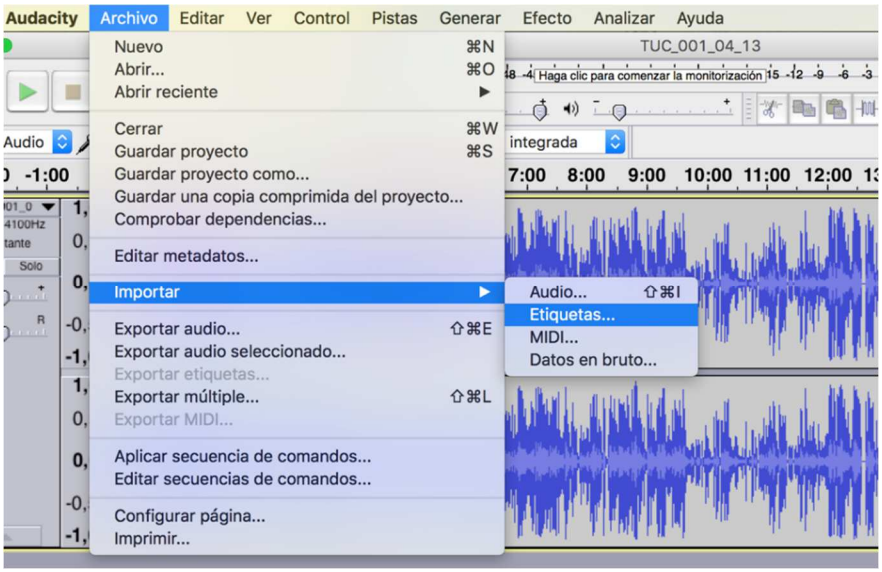
En las opciones de exportación, se debe seleccionar únicamente la línea usada para seleccionar los fragmentos que se deben anonimizar. Debe estar también marcado *Excluir los nombres de las líneas del output* y *Excluir los nombres de los participantes en el output*. El tiempo debe mostrarse en el formato ss.msec y se debe incluir las opciones de columna con el tiempo inicial y con el tiempo final. Es decir, la configuración para exportar la línea debe coincidir la que se muestra en la imagen siguiente.



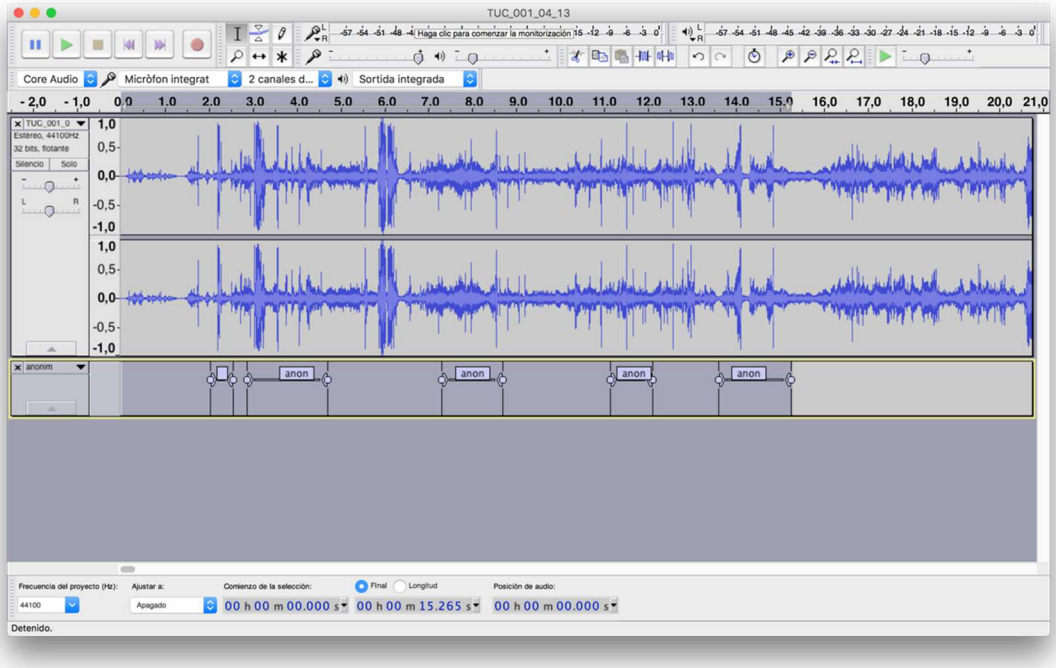
A continuación, se nos mostrará una ventana que nos permitirá nombrar y guardar el archivo en nuestro ordenador. Una vez hayamos concluido este paso, nuestra tarea en ELAN habrá acabado y proseguiremos con Audacity. En primer lugar, abriremos con este programa el archivo de audio que requiere anonimización.



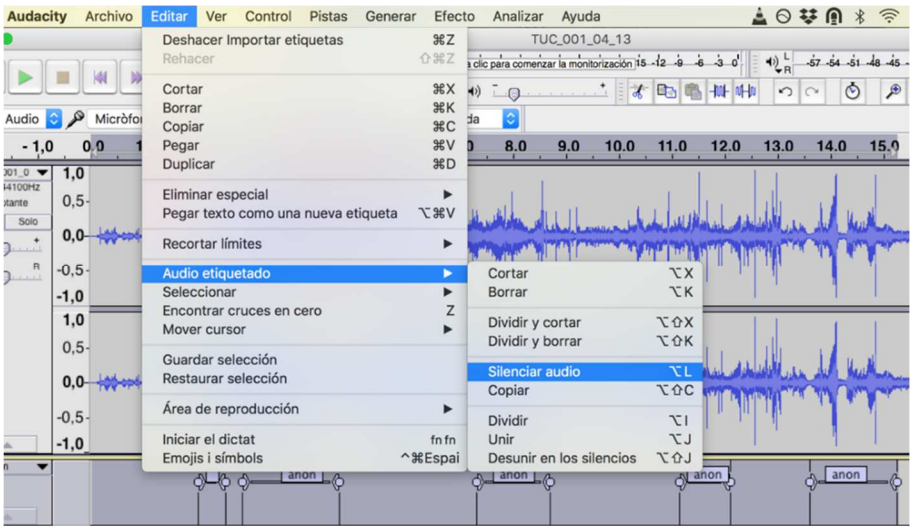
Junto con el archivo de audio, cargaremos la línea de anonimización creada en ELAN que previamente hemos exportado a un documento .txt. Para ello nos dirigimos a la pestaña *Archivo*, buscamos la opción *Importar* y seleccionamos del menú desplegable *Etiquetas…* Se abrirá una ventana con el explorador de archivo y seleccionaremos el archivo que hemos definido con las marcas de tiempo de la anonimización.



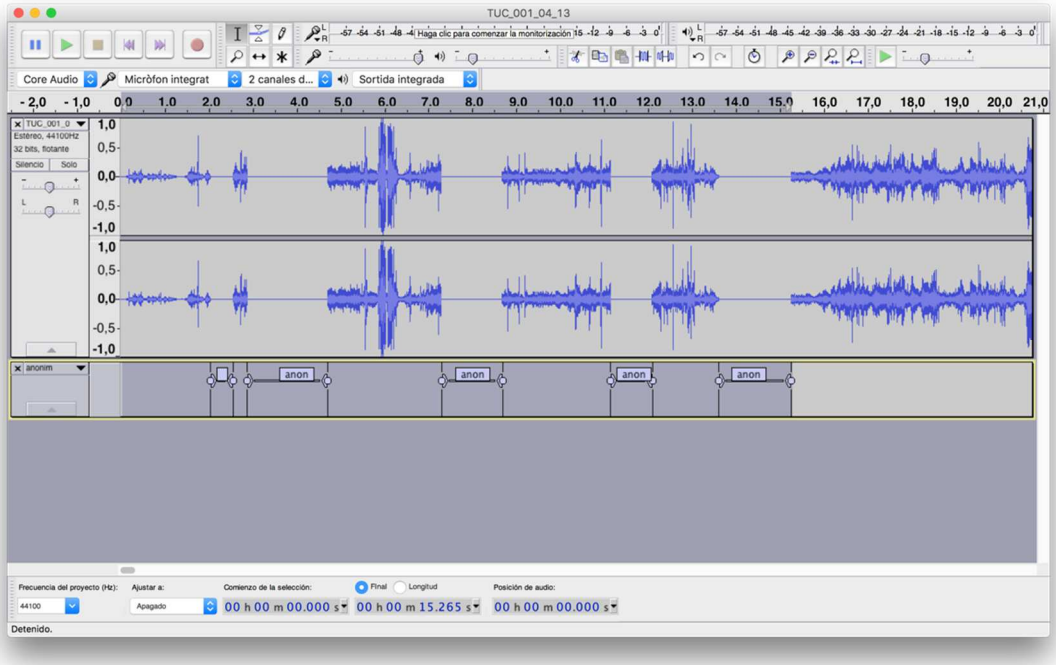
Una vez cargado el documento de texto, obtendremos un resultado similar al de la imagen que se muestra a continuación. Aparecerán, pues, los fragmentos que se han definido en ELAN representados como pequeñas cajas justo debajo del espectrograma del audio, en la franja temporal definida.



Para anonimizar todos estos fragmentos a la vez, en primer lugar, seleccionaremos el cuadro donde aparece la anonimización haciendo clic con el ratón sobre el mismo. Sabremos que está seleccionado porque, como se ve en la imagen siguiente, los bordes de dicho cuadro aparecen sombreados de color amarillo. Una vez hecho esto, nos dirigiremos a la pestaña *Editar* y buscaremos la opción *Audio etiquetado*. Dentro del menú desplegable haremos clic en *Silenciar audio*.



Al hacer clic en esa opción todos los fragmentos previamente delimitados en ELAN quedarán silenciados. De esta manera, pues, se puede anonimizar el audio de toda una conversación en Audacity con un solo clic.

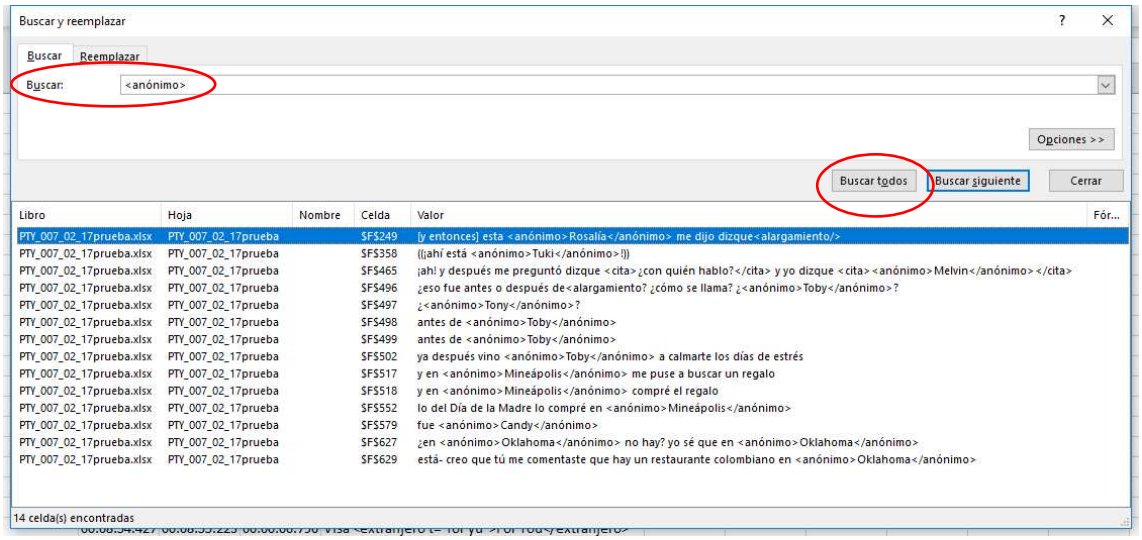


5.3. Anonimización manual con Audacity

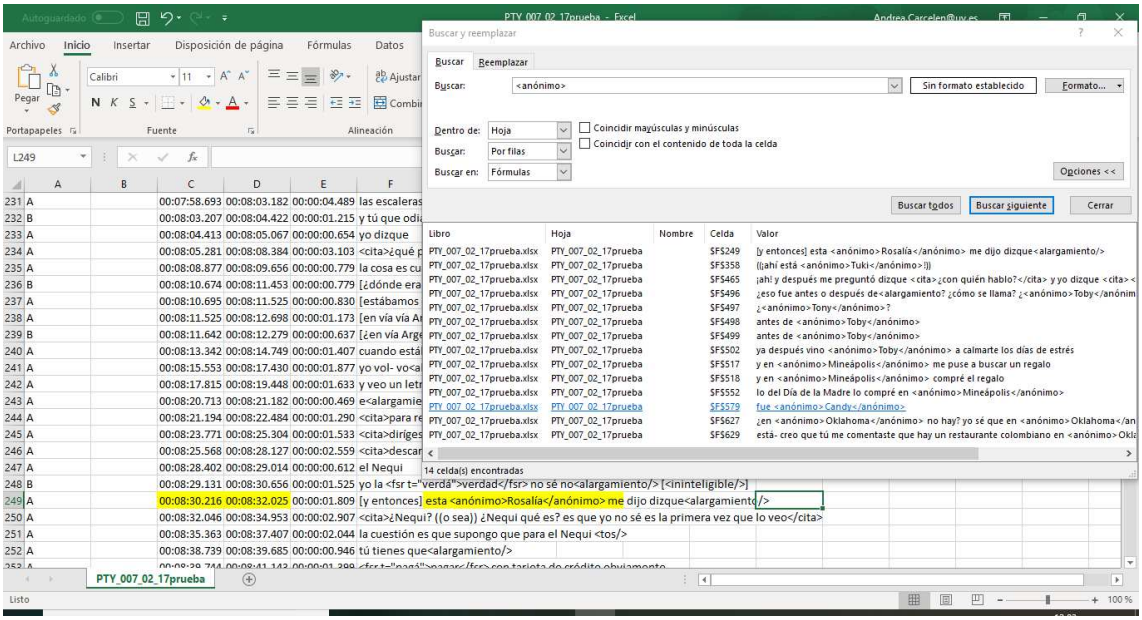
Una vez transcrito el audio en ELAN y exportada la transcripción a un documento de Excel siguiendo los pasos del documento *Transcripción de conversaciones en ELAN*, debemos:

- Abrir el audio en Audacity: Archivo > abrir > seleccionar el archivo deseado
- Utilizar el archivo de Excel para localizar los elementos anonimizados. Para ello, utilizamos la opción Buscar <anónimo> y seleccionamos Buscar todos.

  Ya tenemos en pantalla el listado de todos los elementos con esta etiqueta.

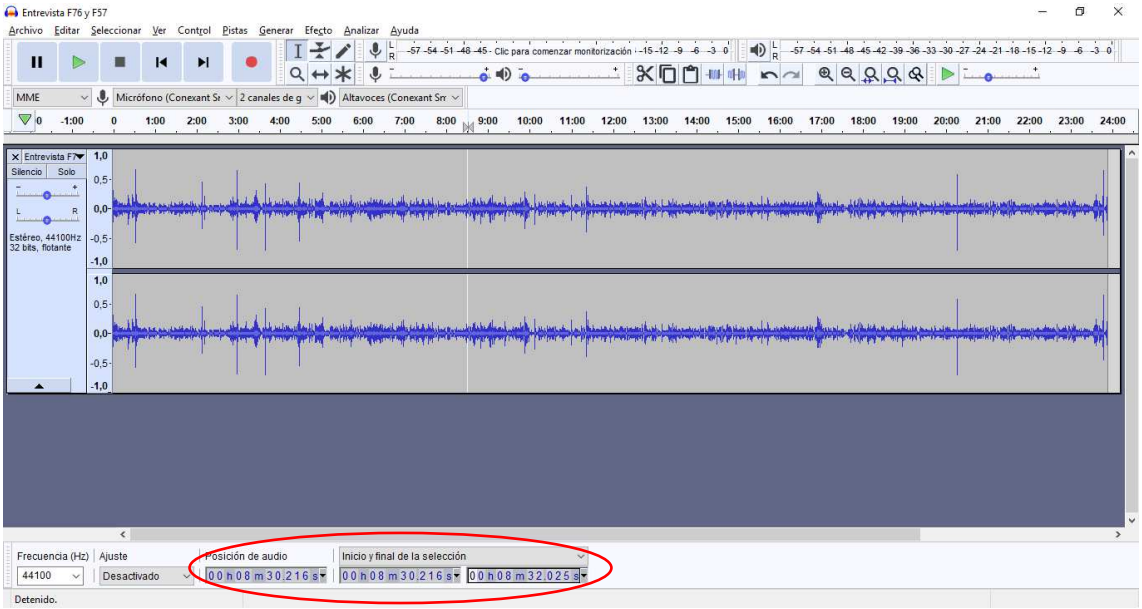


- Ir caso por caso con cada elemento anonimizado y localizar el tiempo de inicio y fin.

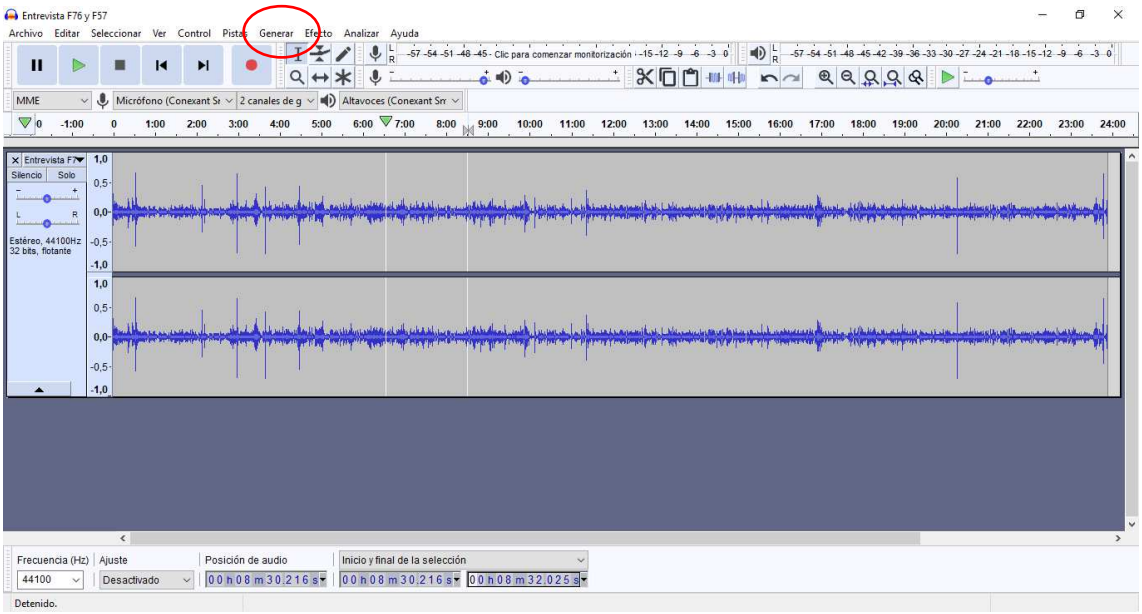


- Localizar el tiempo de inicio y fin en el audio abierto con Audacity y buscar el momento exacto que necesita anonimización. Hay que tener en cuenta que este fragmento se corresponde con una anotación completa, no con la palabra anonimizada en concreto.

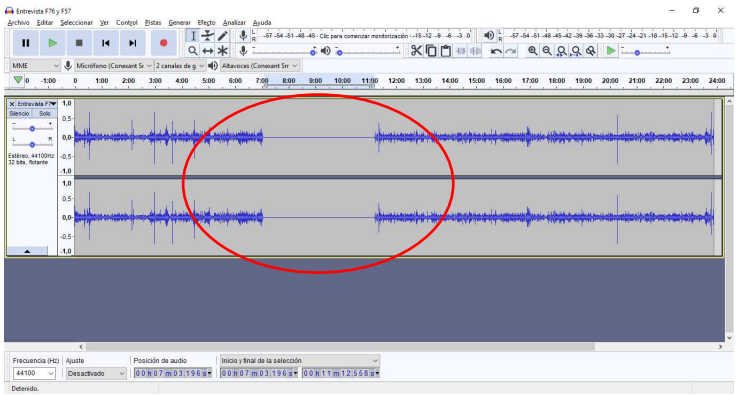
  No podemos silenciar todo el fragmento, deberemos escucharlo hasta dar con el momento exacto en que aparece el elemento que queremos anonimizar.



- Desplegamos la ventana Generar y seleccionamos la opción Silencio. Pulsar Aceptar.



En la siguiente captura se puede observar el fragmento que hemos silenciado (línea recta del espectrograma).



- Una vez hecho esto con todos los elementos encontrados, abrimos la ventana Archivo> Exportar> Exportar como wav y pulsamos Guardar.
- Por último, volvemos a abrir el archivo eaf en ELAN. Abrimos la ventana Editar>Ficheros vinculados y quitamos (desvinculamos) el audio que utilizamos en primer lugar. Vincularemos el nuevo audio anonimizado. Pulsamos Aplicar y Guardar.

IMPORTANTE: Revisar la grabación completa y asegurarse de que todo ha sido anonimizado correctamente, tanto en el audio como en la transcripción.

6. FICHA TÉCNICA

Cada archivo deberá ir acompañado de una ficha técnica[12] que deberá rellenarse siguiendo este modelo de ejemplo:

a) Investigador: *María García García*

b) Datos identificadores de la grabación:

- Fecha de la grabación: *20 de junio de 2015*

- Tiempo de la grabación

Duración del audio	*25:18*
Momento de inicio de la transcripción (minuto:segundo)	*00:05*
Momento de finalización de la transcripción (minuto:segundo)	*25:12*

- Lugar de grabación**: no puede aparecer la dirección exacta**

Municipio	*Valencia capital*
Espacio concreto (casa, bar, aula…)	*Casa particular de hablante B*

c) Situación comunicativa:

- Temas: *exámenes, fiesta fin de curso y vacaciones de verano*.

- Propósito o tenor funcional predominante (marcar con una **X** una de las dos opciones):

Transaccional[13]	
Interpersonal[14]	*X*

- Tono: *informal* **(se refiere a si es formal o informal).**

- Modo o canal: **puesto que trabajamos con conversaciones coloquiales espontáneas debería ser *oral espontáneo.***

d) Tipo de discurso (conversación, debate…): *conversación*, ya que es un corpus de conversaciones coloquiales.

e) Técnica de grabación (seleccionar una de las dos opciones de cada fila):

Conversación libre / semidirigida	*Conversación libre*
Investigador participante / no participante	*Investigador no participante*
Grabación secreta / no secreta	*Grabación cursiva*

[12] Tomamos como modelo la ficha técnica del grupo Val.Es.Co. Los equipos locales pueden proporcionar información adicional a la que se recoge en la ficha si corresponde a sus intereses de estudio.
[13] Su finalidad es la comunicación por la comunicación, hablar por hablar.
[14] Su finalidad es concreta, por ejemplo, cuando hablamos con el dependiente de una tienda o con el médico.

f) Descripción de los participantes:

- Número de participantes: *3*
- Tipo de relación que los une (familia, amigos, hermanos, compañeros de piso…): *compañeros de clase*
- Cuadro de rasgos sociolingüísticos:

Clave hablante	**Sexo (V/M)**	**Edad (especificar edad exacta dentro de una de las tres franjas etarias)**			**Nivel de instrucción**			**Activo/Pasivo**	**Monolingüe cast./Bilingüe**	**Profesión**	**Residencia habitual (municipio)**
		18-25	**26-55**	**≥56**	**Bajo**	**Medio**	**Alto**				
A	*M*	*24*					*X*	*Activo*	*Monolingüe*	*Estudiante último curso de grado*	*Alboraya, Valencia*
B	*M*	*24*					*X*	*Activo*	*Monolingüe*	*Estudiante último curso de grado*	*Valencia capital*
C	*V*	*23*					*X*	*Activo*	*Monolingüe*	*Estudiante último curso de grado*	*Valencia*
D											
E											
F											
G											

g) Grado de prototipicidad coloquial (marcar con una **X** una de las dos opciones)**:**

Conversación coloquial prototípica[15]	*X*
Conversación coloquial periférica[16]	

[15] Cumple con las características siguientes: fin interpersonal, tono informal, ausencia de planificación, relación de igualdad entre los interlocutores, relación vivencial de proximidad, marco de interacción familiar y cotidiano, temática no especializada.
[16] No cumple con alguno de los rasgos enumerados anteriormente.

7. ENTREGA DE ARCHIVOS

Una vez terminado el trabajo, los archivos que deberán enviarse[17] al equipo central son:

1. **Audio anonimizado en formato wav**
2. **Transcripción en ELAN (formato eaf)[18]**
3. **Autorizaciones**
4. **Ficha técnica debidamente cumplimentada**
5. **Transcripción archivo en Excel**

Estos cinco archivos deberán ir en una misma carpeta con el mismo nombre del archivo:

Carpeta audio 1. Contendrá:

Audio 1

Transcripción audio 1

Autorización A audio 1

Autorización B audio 1

Autorización C audio 1 (tantas como participantes)

Ficha Técnica audio 1

Así, continuaríamos la numeración sucesivamente (audio 2, audio 3…). Se recomienda continuar la numeración consecutiva de los archivos, aunque se entreguen en diferentes plazos.

La ficha técnica debe ir en un documento diferente a la transcripción.

8. MODO DE CITAR EL CORPUS

Nombre de la/s persona/s coordinadora/s del equipo (en línea): "Corpus de conversaciones Ameresco-Ciudad", en Albelda y Estellés (coords.): Corpus Ameresco, www.corpusameresco.com, Universitat de València, ISSN: 2659-8337.

Ejemplo:

González Mafud, Ana María (en línea): "Corpus de conversaciones Ameresco-La Habana", en Albelda y Estellés (coords.): Corpus Ameresco, www.corpusameresco.com, Universitat de València, ISSN: 2659-8337.

[17] Si bien los archivos pueden enviarse por correo electrónico como archivos adjuntos, es recomendable que se comparta la carpeta de cada archivo completa a través de alguna plataforma de almacenamiento en la nube (Google Drive, One Drive, Dropbox…). Los audios suelen tener demasiado peso para compartirlos como adjuntos, con la nube conseguimos que cada audio vaya acompañado siempre de su transcripción, ficha técnica y autorización y que no se extravíe ninguno de los archivos.

[18] Junto al archivo eaf, ELAN genera automáticamente un archivo con extensión .rar y otro .pfsx. que no es necesario que se adjunten.

ANEXOS

AUTORIZACIÓN PARA EMPLEAR LA GRABACIÓN Y LA TRANSCRIPCIÓN DEL MATERIAL CON FINES INVESTIGADORES EN LINGÜÍSTICA

(Proyecto MINECO FFI2016-75249P)

A. Autorización previa a la grabación

Dña./D. _____ con documento de identificación o pasaporte número _____

DECLARO

1) que se me ha informado de que voy a ser grabado/a de forma secreta en las próximas semanas;

2) que, posteriormente a la grabación, podré escuchar el contenido de mi grabación;

3) que, en caso de no estar de acuerdo, puedo ejercer mi derecho a retirar la grabación.

A los efectos oportunos, firmo la presente autorización en _____ a _____de 20____.

Fdo. _____

B. Autorización posterior a la grabación

Dña./D. _____ con documento de identificación o pasaporte número _____

DECLARO

1) que se me ha informado de que he sido grabado/a secretamente y he escuchado el contenido de mi grabación;

2) que se me ha informado de que puedo ejercer mi derecho a retirar la grabación.

Y, por tanto, AUTORIZO al uso de la grabación de su contenido, previamente anonimizados texto y audio, para fines estrictamente de investigación.

A los efectos oportunos, firmo la presente autorización, en _____ a _____de 20____.

Fdo._____

1. Datos personales

Los datos personales obtenidos mediante el presente formulario se incorporarán a los sistemas de información de la Universitat de València – Estudi General (links.uv.es/lopd/dpo) en el marco del Proyecto Es.Vag.Atenuación. La atenuación pragmática en su variación genérica: géneros discursivos escritos y orales en el español de América y de España (Proyecto MINECO FFI2016-75249P), dirigido por las doctoras Marta Albelda Marco y Maria Estellés Arguedas.

La información objeto de tratamiento será utilizada para el desarrollo de funciones docentes y académicas propias de la Universitat de València como la investigación, la creación, desarrollo, transmisión y crítica de la ciencia, de la técnica y de la cultura y la difusión, la valorización y la transferencia del conocimiento.

En concreto, estas grabaciones formarán parte del corpus oral del español coloquial Ameresco. Dicho corpus está compuesto por un conjunto de microcorpus de conversaciones coloquiales obtenidas en las distintas ciudades que se integran en el proyecto. Con el objetivo de analizar estas muestras de habla, en este proyecto se recogen conversaciones informales espontáneas (coloquiales) reales grabadas en lugares cotidianos para los hablantes, en una situación de familiaridad o amistad para su posterior análisis lingüístico.

La Universitat de València se compromete a que cualquier divulgación pública de los resultados obtenidos con motivo de la investigación se realizará anonimizando debidamente los datos utilizados, de modo que los sujetos de la investigación no resultarán identificados o identificables.

La base jurídica del tratamiento es el consentimiento del afectado/a y se prevé la conservación de los datos personales durante cinco años. Transcurrido ese periodo, los datos se conservarán debidamente disociados para garantizar el anonimato.

2. Registro de imagen o sonido

En el marco del desarrollo de la actividad se obtendrán registros de audio. Ud. Autoriza a la Universitat de València al uso, edición, difusión y explotación de estos registros exclusivamente para fines de investigación. En caso de utilización, se asegurará que el afectado/a nunca sea identificado por su nombre ni mediante información alguna que le haga identificable.

Todo ello con la única salvedad y limitación de aquellas utilizaciones o aplicaciones que pudieran atentar a los derechos garantizados en la Ley Orgánica 1/1982, de 5 de mayo, de Protección Civil al Derecho al Honor, la Intimidad Personal y familiar y a la Propia Imagen, así como del pleno respeto de las previsiones específicas del art. 4 de la Ley Orgánica 1/1996, de 15 de enero, de protección jurídica del menor.

3. Publicación

Los resultados del proyecto son susceptibles de publicación. En caso de tal utilización, se asegurará que Ud. nunca sea identificado/a por su nombre apellidos, ni mediante información alguna que le haga identificable.

4. Ejercicio de derechos

Las autorizaciones concedidas en este documento podrán ser revocadas mediante la presentación del oportuno escrito. La revocación comportará la retirada de la información de los sistemas de la Universitat de València en un plazo prudencial de tiempo en función de la disponibilidad de recursos.

Puede obtener más información acerca de sus derechos en: links.uv.es/lopd/derechos

Y en prueba de conformidad, firmo el presente documento en el lugar y la fecha indicados en el encabezamiento.

Nombre y apellidos Firma	Nombre y apellidos Firma PADRE / MADRE / TUTOR *Rellenar solo en caso de menores o personas con incapacidad legal.*

FICHA TÉCNICA

a) Investigador:

b) Datos identificadores de la grabación:

- Fecha de la grabación:

- Tiempo de la grabación

Duración del audio	
Momento de inicio de la transcripción (minuto:segundo)	
Momento de finalización de la transcripción (minuto:segundo)	

- Lugar de grabación**:**

Municipio	
Espacio concreto (casa, bar, aula…)	

c) Situación comunicativa:

- Temas:

- Propósito o tenor funcional predominante (marcar con una **X** una de las dos opciones):

Transaccional	
Interpersonal	

- Tono:

- Modo o canal:

d) Tipo de discurso (conversación, debate…):

e) Técnica de grabación (seleccionar una de las dos opciones de cada fila):

Conversación libre / semidirigida	
Investigador participante / no participante	
Grabación secreta / no secreta	

f) Descripción de los participantes:

- Número de participantes:

- Tipo de relación que los une (familia, amigos, hermanos, compañeros de piso…):

- Cuadro de rasgos sociolingüísticos:

Clave hablante	**Sexo (V/M)**	**Edad (especificar edad exacta dentro de una de las tres franjas etarias)**			**Nivel de instrucción**			**Activo/Pasivo**	**Monolingüe cast./Bilingüe**	**Profesión**	**Residencia habitual (municipio)**
		18-25	**26-55**	**≥56**	**Bajo**	**Medio**	**Alto**				
A											
B											
C											
D											
E											
F											
G											

g) Grado de prototipicidad coloquial (marcar con una **X** una de las dos opciones)**:**

Conversación coloquial prototípica	
Conversación coloquial periférica	