Análisis

Recepción: 11-10-2022
Aceptación: 28-10-2022

Adaptación de libros hablados digitales mediante síntesis de voz en el Servicio Bibliográfico de la ONCE

Adaptation of digital talking books by means of voice synthesis in the ONCE Bibliographic Service

Á. Sierra Berrocal

Resumen

En 1962, la ONCE empezó a producir libros hablados para cubrir las necesidades educativas, laborales y culturales de sus afiliados. Con la llegada de la era digital, estos libros evolucionaron gracias a un revolucionario sistema de producción que incrementó enormemente la accesibilidad a la información: el formato Daisy. Todo ello se vio aún más potenciado con la inclusión de voces sintéticas capaces de transformar un texto convenientemente etiquetado en un libro hablado digital (LHD) accesible y navegable, estableciendo una sincronización entre texto y audio. En este artículo, analizaremos la trayectoria del Servicio Bibliográfico de la ONCE (SBO) en su afán por alcanzar la eficacia y excelencia que exigen sus usuarios. También se analizarán las diferentes voces sintéticas que el servicio lleva utilizando desde el año 2002 y cómo han ido mejorando hasta llegar a las voces actuales generadas por inteligencia artificial, como es el caso de la nueva voz «neuronal» de Lucía integrada en el servicio en la nube Amazon Polly, con la que ya se están produciendo libros hablados digitales en formato Daisy con el texto completo sincronizado con el audio.

Palabras clave

Accesibilidad. Adaptación. Síntesis de voz. Tiflotecnología. Innovación.

Sierra, Á. (2022). Adaptación de libros hablados digitales mediante síntesis de voz en el Servicio Bibliográfico de la ONCE. *RED Visual: Revista Especializada en Discapacidad Visual, 80*, 106-126. https://doi.org/10.53094/IBOA4928.

Abstract

In 1962, the ONCE began producing talking books to meet the educational, occupational and cultural needs of its members. With the advent of the digital age, these books evolved thanks to a revolutionary production system that greatly increased accessibility to information: the Daisy format. This was further enhanced by the inclusion of synthetic voices capable of transforming a suitably labelled text into an accessible and navigable digital talking book (DTBs), achieving synchronisation between text and audio. In this article, we will analyse the path followed by the ONCE Bibliographic Service (SBO) in its quest to achieve the efficiency and excellence demanded by its users. The various synthetic voices that the service has been using since 2002 will also be analysed, as well as how they have improved to arrive at the current voices generated by artificial intelligence, such as the new «neural» voice of Lucia integrated in the Amazon Polly cloud service, which is already producing digital talking books in Daisy format in which the full text is synchronised with the audio.

Key words

Accessibility. Adaptation. Voice synthesis. Tiflotechnology. Innovation.

1. Introducción: antecedentes históricos

Desde comienzos de los años sesenta del siglo pasado, la Organización Nacional de Ciegos Españoles (ONCE) lleva realizando adaptaciones en braille y sonido para cubrir las necesidades de sus afiliados. Dentro de estos textos se incluyen libros, apuntes, exámenes, revistas, mapas, planos, cuentos infantiles, guías, manuales, cursos, oposiciones, etc., con una finalidad educativa, laboral, cultural, de ocio e institucional.

Centrándonos en la producción sonora, es necesario recordar que en 1962 se empezaron a producir los primeros libros hablados. Se trataba de grabaciones analógicas realizadas por locutores profesionales en estudios de grabación utilizando cintas magnéticas (Figura 1).

Al principio, los registros se realizaban en dos pistas (A y B), pero poco después, se empezaron a utilizar grabadores de cuatro pistas (A, B, C y D) que multiplicaban por dos el tiempo de cada cinta.

Sierra, Á. (2022). Adaptación de libros hablados digitales mediante síntesis de voz en el Servicio Bibliográfico de la ONCE. *RED Visual: Revista Especializada en Discapacidad Visual, 80*, 106-126. https://doi.org/10.53094/IBOA4928.

Figura 1. Magnetófono para grabación de libros

Con el propósito de mejorar el acceso a la información y la navegación por el libro, se integraron señales acústicas entre los diferentes capítulos o apartados del libro. Dichas señales consistían en tonos sinusoidales de baja frecuencia que, en avance rápido o rebobinado, podíamos percibir acústicamente. Algo así como un «pitido» que nos advertía dónde empezaba o finalizaba un epígrafe o apartado del libro.

Con el paso del tiempo, se tendió a normalizar y articular este proceso, y se publicaron unas normas para que los lectores tuvieran claro cómo se tenían que leer, adaptar y describir cualquier elemento que pudiese contener el libro, como notas, mapas, figuras, gráficos, diagramas, esquemas, tablas, cuadros, fórmulas, etc. También se parametrizó el sistema de «señalizado» para que cualquier usuario pudiese comprender la estructura del libro. Ejemplo: si un libro tenía partes y capítulos, se utilizaban señales largas («pitidos largos») para los capítulos y señales dobles («dos pitidos») para las partes. Por último, al final de la última pista de la última cinta se generaba un índice de señales acústicas con los diferentes apartados señalizados en la obra y su página correspondiente en el libro en «tinta».

Los usuarios utilizaban este dispositivo (Figura 2) para escuchar los libros hablados en cintas de casete.

Durante los posteriores años ochenta y noventa, la ONCE disponía de una infraestructura muy potente para producir libros hablados de todo tipo. Se seleccionaron y formaron lectores-adaptadores especializados en distintas disciplinas, como las ciencias

puras, idiomas, fisioterapia, ajedrez, psicología, derecho, música, etc. Este inmenso patrimonio analógico todavía se sigue digitalizando, es decir, se siguen volcando las pistas de estas cintas a formato digital.

Figura 2. Reproductor y grabador de libros hablados en casete

Pero la «revolución» del servicio llegaría con la era digital y la creación del Consorcio Daisy[1] en mayo de 1996, en el que la ONCE, junto con otras organizaciones como la *Asociación Japonesa de Bibliotecas para Ciegos*, el *Royal National Institute of Blind People* (RNIB) del Reino Unido y algunas bibliotecas de Suecia, Suiza y los Países Bajos publicaron un formato estándar para libros hablados digitales.

A grandes rasgos, un LHD Daisy dispone de opciones de navegación por frases,[2] secciones, páginas u otros elementos; utiliza tablas de contenido, permite búsquedas —ya que permite sincronizar el texto con el audio— y la creación de marcas.

El sistema Daisy permite generar seis tipos básicos de libros: libros con audio y texto completo, libros con audio y parte del texto, libros solo con audio, libros solo con texto, y libros con audio y estructura (NCC). En definitiva, un sistema que permite una *usabilidad* y accesibilidad realmente poderosas. No vamos a profundizar más sobre

1 Originalmente, acrónimo de *Digital Audio Information System*, que se amplió años después al nombre actual usando *Accessible* en lugar de *Audio* por incluirse más formatos (texto, imágenes, etc.), quedando establecido de la siguiente manera: *Digital Accessible Information System*.

2 Una frase en Daisy es una pausa en el audio que el estándar utiliza luego para moverse por el texto, pero puede estar en medio de una frase gramatical o puede incluir un par de ellas, dependiendo del modo de grabación.

este estándar ya que se requiere conocer cómo ha evolucionado durante todos estos años, pero sí que es importante saber que hoy en día se trabaja con las versiones Daisy 2.02 y Daisy 3. Las versiones anteriores Daisy 1 y Daisy 2.0 quedaron obsoletas, y no ha quedado más remedio que reconvertir estos fondos a Daisy 2.02.

Es posible reproducir libros Daisy con dispositivos específicos (algunos llevan el logo *Daisy* OK; Figuras 3, 4 y 5) o con aplicaciones específicas para ordenadores o dispositivos móviles (Android/iOS).

Figura 3. Reproductor PlexTalk

Figura 4. Reproductor Victor

Figura 5. Reproductor/grabador PlexTalk

Los dispositivos Daisy nos permiten funciones de lectura y reproducción, como:

- **Ir a página:** navegar a una página. La mayoría de estos reproductores cuentan con un teclado numérico (como un teléfono actual: 0-9, asterisco y almohadilla)

que facilitan esta función. Es de vital importancia disponer de un índice con la paginación para que el usuario pueda localizar cualquier apartado o sección del libro con precisión.

- **Ir a cabecera:** ir a un epígrafe concreto, siempre y cuando conozcamos su numeración relativa en la estructura, ya que también se utiliza el teclado numérico.

- **Avance/retroceso rápido:** ofrece varias velocidades de avance o «rebobinado».

- **Frase anterior/siguiente:** la frase supone el elemento navegable de más bajo nivel, ya que cualquier texto es un conjunto de frases.

- **Avanzar/retroceder en el tiempo:** es posible seleccionar el intervalo de salto de tiempo en minutos.

- **Incrementar o disminuir la velocidad de reproducción:** los «oídos entrenados» tienen la capacidad de escuchar un libro a gran velocidad, o todo lo contrario. También dependerá del ritmo de lectura que haya empleado el narrador.

- **Detener/Pausa/Continuar:** Play/Pause/Stop.

- **Hacer marca, Ir a marca:** puede entenderse como un «subrayado» de una parte del texto para acceder en cualquier momento. Algunos dispositivos permiten marcas de voz.

- **Dónde estoy:** función que me indicará en qué sección estoy, en qué página, tiempo transcurrido y restante.

- **Información del libro:** que nos recordará el título del libro, su duración, el número de páginas, etc.

- **Tono y Volumen**.

- **Seleccionar elemento navegable:** una opción realmente imprescindible que permite seleccionar con qué tipo de elemento (saltable) vamos a navegar. Si seleccionamos el elemento «cabecera» —que ofrece una profundidad de anidamiento de 1 a 6—, podremos navegar por el árbol o estructura del libro. Supongamos que

un libro se compone de partes (nivel 1), capítulos (nivel 2), apartados (nivel 3) y subapartados (nivel 4); si seleccionamos el nivel 1, podremos navegar por partes sin acceder al resto de niveles inferiores.

- **Elementos opcionales:** también es posible indicar qué elementos navegables serán opcionales; por ejemplo, podremos decidir si las notas al pie se leerán o no durante la escucha.

Los primeros modelos se centraban en reproducir discos compactos (CD), pero, posteriormente, se fueron añadiendo otros medios, como tarjetas de memoria o puertos USB.

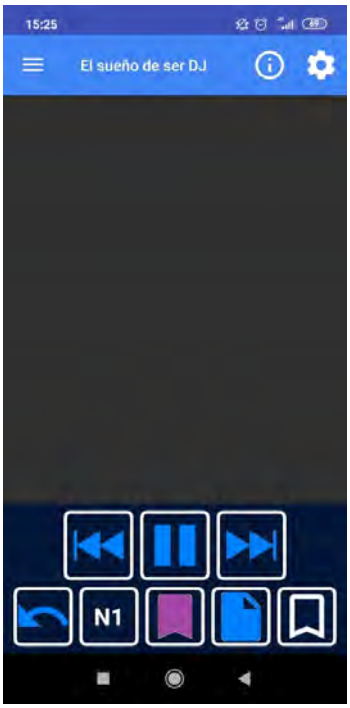
El *software* permite aún más funcionalidades a la hora de navegar por el libro; por ejemplo, será posible teclear una secuencia de texto para que el programa lo localice y lo reproduzca, imprescindible si el libro dispone del texto completo. También será posible visualizar imágenes, cambiar fuente, tamaño, contraste, espaciado, márgenes, seleccionar voz sintética del sistema (para libros que no dispongan de audio) y acceder a la Biblioteca Digital ONCE (BDO),[3] que hoy cuenta con cerca de 75000 obras.

Ahora mostramos dos ejemplos de reproductores Daisy por *software*: Easy Reader y GOLD (Figuras 6 y 7).

Figura 6. EasyReader

3 La BDO es un servicio restringido a través de Internet en el que los afiliados a la ONCE pueden acceder al catálogo de libros braille/sonido, revistas, fonoteca, teatro y audiodescripción.

Figura 7. GOLD



Por otro lado, la ONCE también estaba buscando otros recursos, como la voz sintética para convertir texto a voz (TTS, por sus siglas en inglés; *text to speech*) y poder así producir libros hablados digitales. Desde el lanzamiento del lector de pantalla JAWS (*Job Access With Speech*), creado por la corporación norteamericana Freedom Scientific para el sistema operativo Microsoft Windows 3.1/Windows 3.11 en 1995, los usuarios se dieron cuenta de la importancia que suponía disponer de una voz que pudiese leer todo lo que aparece en pantalla con la ayuda del teclado. Además, durante estos años fueron apareciendo otros lectores de pantalla, tanto para ordenadores (PC/Mac) como para dispositivos móviles. Destacaremos el lector de pantalla gratuito NVDA[4] para Windows y VoiceOver, lector de pantalla integrado en el sistema operativo de Mac.

En 2002, el SBO comenzó a producir libros Daisy en voz sintética con las voces de la corporación italiana Loquendo. Disponía de voces masculinas y femeninas en castellano además de un surtido paquete de voces en otros idiomas.

Esto fue solo el principio. La voz sintética había llegado para quedarse.

4 NVDA (Non Visual Desktop Access), acceso de escritorio no visual, desarrollado por NV Access.

2. Evolución y desarrollo

Las primeras voces de Loquendo adquiridas por la ONCE tenían nombres como Jorge, Juan, Carmen y Leonor. Todavía hoy podemos escuchar a Jorge en transportes públicos anunciando estaciones. Estas voces fueron mejorando y evolucionando paulatinamente. Cuando hablamos de la mejora de una voz sintética nos referimos a algo muy sencillo; es decir, se busca que la voz sintética sea muy parecida a la voz humana: natural e inteligible. Se requiere una correcta vocalización: articulación, modulación, acentuación; un timbre agradable; una entonación y expresividad que se ajusten a los diferentes tipos de textos, y, además, un ritmo de lectura que sea ajustable y que no distorsione a velocidad rápida o lenta. Una meta que parecía imposible en 2002. Sin embargo, ya podemos decir que estamos muy cerca de conseguirlo.

Las voces de Loquendo se instalaron en los ordenadores del Servicio Bibliográfico, de tal forma que podían invocarse a través de SAPI (*Speech Application Programming Interface*), una interfaz de programación de aplicaciones de voz de Microsoft. Esta interfaz se utiliza tanto para «reconocimiento automático del habla» (RAH) como para síntesis de voz (TTS, por sus siglas en inglés). SAPI ofrece subrutinas, funciones, procedimientos y una biblioteca para ser utilizada por otro *software*. Ese otro *software* fue desarrollado por un equipo de técnicos del SBO para generar libros con las voces o idiomas seleccionados. Esto supuso el pistoletazo de salida para la voz sintética. Se empezaron a producir libros en formato Daisy 2.02 con las voces de Jorge y Carmen, y también se generó audio para bibliografías en otros idiomas.

En 2011, Loquendo fue adquirida por la multinacional norteamericana Nuance Communications. Las voces siguieron evolucionando y aparecieron nuevas herramientas con las que se podían realizar ajustes para dar más naturalidad y expresividad. Un *partner* de Nuance en Cataluña, llamado Code Factory Global, fue el encargado de otorgar al SBO cierta formación y soporte para el uso de una herramienta llamada Nuance Vocalizer Expressive para «customizar» pronunciaciones, puntuación, diccionarios (lexicones)... y también se añadieron nuevos idiomas.

La producción seguía su curso, pero, mientras tanto, la ONCE apostó por otras opciones. En 2014 se añadió una nueva voz: Conchita. Fue desarrollada por una empresa del entorno de Google, denominada Ivona, para dispositivos Android. Se consiguió hacerla funcionar en Windows bajo la interfaz SAPI.

La llegada de Windows 10 en julio de 2015 también fue un momento decisivo. Este sistema operativo tiene integradas varias voces: Laura, Pablo y Helena. Tras horas y horas de pruebas, la ONCE se decantó por Helena, con la aprobación de Microsoft. A partir de este momento, se comenzaron a producir libros hablados con la voz Helena.

Más tarde, los proveedores de voces sintéticas fueron dejando de ofrecer voces «instalables», dando paso a los servicios TTS «en la nube». Aparecieron nuevas voces con una calidad realmente sorprendente. Destaquemos algunas pruebas realizadas:

- Enrique y Laura (IBM), a través de su servicio Watson Texto to Speech Voices integrado en IBM Cloud.

- Amazon, a través de su servicio Amazon Polly de la plataforma AWS (Amazon Web Services). De aquí surgió la voz de Lucía.

- Google desde Google Cloud.

- Microsoft, desde su servicio en la nube Microsoft Azure. Como dato adicional, decir que Microsoft adquirió Nuance Communications en abril de 2021, por lo que ahora es parte de la división Azure.

El «problema» de que las voces estén ubicadas en la nube creó la necesidad de establecer una nueva forma de producir, ya que, al invocar una voz determinada, se necesitaba algún tipo de aplicación capaz de conectarse a Internet para poder obtener los ficheros de audio. De esta forma, la ONCE se puso en contacto con empresas externas para desarrollar una «nueva» herramienta.

3. Proceso productivo

Para empezar, es importante recordar que, en 2011, el SBO obtuvo una certificación ISO-9001 de su Sistema de Gestión de la Calidad (SGC), emitido por la British Standards Institution. Este sistema garantiza y certifica los procesos y procedimientos del servicio para cubrir las necesidades de sus afiliados. Esto trae consigo el cumplimiento de objetivos, iniciativas de mejora, seguimiento, análisis y detección de las incidencias que puedan surgir. Este sistema se aplica en todas las adaptaciones que se realizan en el SBO.

Sierra, Á. (2022). Adaptación de libros hablados digitales mediante síntesis de voz en el Servicio Bibliográfico de la ONCE. *RED Visual: Revista Especializada en Discapacidad Visual, 80*, 106-126. https://doi.org/10.53094/IBOA4928.

Partiendo de esta premisa, y para comprender mejor cómo funciona este servicio, vamos a resumir las diferentes fases por las que pasa un documento para su adaptación sonora en formato Daisy.

3.1. Fases del proceso productivo

Dentro del proceso productivo, se pueden distinguir las siguientes fases:

- **Solicitudes**. Se recogen pedidos con cuatro finalidades: educativa, laboral, institucional y de ocio y cultura. Todas las solicitudes serán revisadas y aprobadas por una comisión de autorización. Posteriormente, se procederá a la adquisición de cada libro o documento y a su catalogación bibliográfica.

- **Planificación**. Las tareas de adaptación se organizan estableciendo prioridades en función del tipo y finalidad, dando prioridad absoluta a los libros educativos y laborales. Se establecerán plazos para las entregas y, además —para cubrir las necesidades de estudiantes y trabajadores—, se realizarán entregas parciales para que los usuarios dispongan de material para ir leyendo o estudiando. También se establecerá el grado de dificultad de la adaptación: mínima, sencilla, media, compleja y máxima. Esta complejidad tiene relación con el tipo de materia, ya que necesitan un perfil especializado. Un libro de ajedrez o un libro de matemáticas requiere amplios conocimientos en estas materias, y, además, contienen elementos cuya descripción es muy laboriosa, como un diagrama de ajedrez o una expresión matemática. Finalmente, antes de pasar a la adaptación propiamente dicha, se dictarán las pautas básicas para afrontar la adaptación; esto se denomina *análisis de adaptación*.

- **Adaptación**. Se entiende por *adaptación* al proceso o procesos necesarios para hacer accesible un documento; es decir, se trata de modificar un documento bajo unas normas concretas para que su contenido pueda ser leído por el mayor número de personas posible (incluidas, lógicamente, las que tienen algún tipo de discapacidad o dificultad para la lectura y/o comprensión).

- **Control de calidad**. Todas las adaptaciones tendrán que ser revisadas, aportando los informes y documentos necesarios que acrediten cómo se ha realizado dicha revisión o control. Esta «revisión» incluye aspectos tan importantes como el cotejo con el documento original, la revisión de la estructura, verificación de metadatos,

comprobación de adaptaciones y descripciones, la evaluación de la lectura y la calidad del sonido.

- **Posproducción**. Una vez comprobada que la adaptación es válida y cumple con las normas establecidas, y si no se requieren acciones correctivas, se genera un máster por cada libro o documento y se almacenará en los servidores del Servicio Bibliográfico.

- **Distribución**. Después del archivado y registro de cada máster, se podrán realizar copias en CD —si el usuario así lo requiere— y, además, se subirán a la BDO para que todos los afiliados puedan descargarlos a través de Internet en sus ordenadores o dispositivos móviles. Esta biblioteca ofrece la posibilidad de descargar libros en formato Daisy y en formato TLO[5] (braille).

3.2. Adaptación mediante síntesis de voz

Nos vamos a centrar en los procesos de adaptación. La primera pregunta que debemos hacernos es la siguiente: ¿qué libros se adaptarán mediante voz sintética?

Los libros que realmente funcionan en síntesis de voz suelen ser ensayos, libros de historia, derecho, autorrealización, psicología, antropología, biografías, oposiciones, guías, manuales, gastronomía, viajes y reportajes, divulgación científica, discapacidad, sociología y diccionarios. Quedan fuera novelas, cuentos o relatos, libros infantiles y juveniles, poesía y teatro, ya que las voces no son lo suficientemente expresivas. Tampoco se adaptan la musicografía o libros de maquetación puramente visual, como cómics o novelas gráficas, ya que casi habría que redactarlos de nuevo.

Una vez aprobadas las propuestas, las obras se catalogan y se crea una «Ruta de tareas» para su producción, que veremos a continuación.

3.2.1. Análisis previo de adaptación

Los técnicos del Servicio Bibliográfico indicarán cómo y en qué plazos se debe realizar la adaptación. En estas indicaciones se hará constar, por ejemplo, cómo se adaptarán las notas, gráficos, imágenes, cómo se confeccionará la jerarquía de niveles para

5 TLO: formato de libro electrónico de libros transcritos al braille.

crear la estructura del libro, cómo procesar idiomas, qué voz se va a utilizar o si es necesario crear un índice o tabla de contenidos. En general, se trata de orientar sobre la adaptación y/o descripción de los elementos que forman parte del documento.

3.2.2. Valoración de la procedencia del documento fuente

El origen del texto puede ser un libro en tinta (impreso) o un libro electrónico. En el caso de libros impresos se añade una tarea más, la digitalización. Para ello se utilizan escáneres y programas de reconocimiento óptico de caracteres (OCR, por sus siglas en inglés). Si el documento es un *e-book*, se pasará a la siguiente tarea.

3.2.3. Conversión de los ficheros de texto

Existen numerosos tipos de archivos de texto que requieren conversión. Por ejemplo, los libros electrónicos procedentes de Amazon tienen formato MOBI, otros utilizan PDF (Adobe), EPUB, ODT (Open Document) o DOC (MS Word). Además, existen otros formatos más específicos, como, por ejemplo, archivos BRA (producción braille), que también requieren conversión.

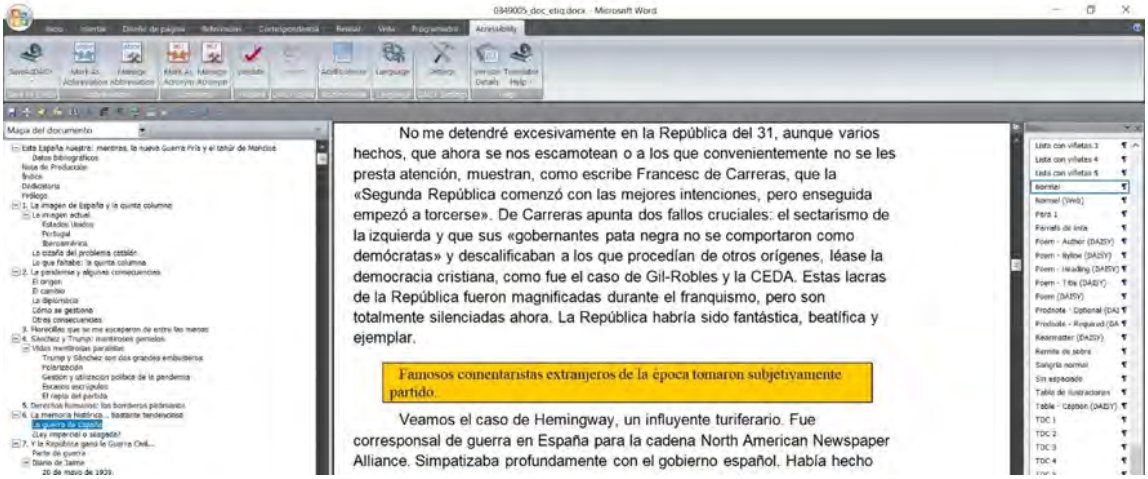
Nuestro objetivo es que el archivo de texto se convierta en un documento XML DTBook (Daisy Digital Talking Book), también conocido como Daisy XML. Este formato fue desarrollado por el Consorcio Daisy y consiste en un documento accesible que contiene elementos navegables con un etiquetado estandarizado creado por el Consorcio, en el que se define una DTD[6] como parte del estándar NISO Z39.86-2005.

Es evidente que generar XML desde, por ejemplo, un archivo PDF será algo muy diferente a crear un documento XML desde un archivo EPUB, ya que ambos formatos son intrínsecamente muy diferentes. Se barajaron varias opciones y soluciones al respecto, pero, al final, se optó por transformar a DOCX, formato Microsoft Word «basado» en XML.

Microsoft y el Consorcio Daisy idearon un complemento (*add-in*) gratuito para Word capaz de transformar un fichero DOCX en un fichero Daisy XML llamado: *Save as Daisy – MS Word Add-In*. Para utilizarlo, solamente se requiere su instalación (Figura 8).

6 DTD: *Document Type Definition*. Es un documento que define la estructura e indica cómo es el lenguaje en que se debe «escribir» (etiquetar) el documento XML.

Figura 8. Captura de pantalla de Microsoft Word con el complemento *Save as Daisy*



3.2.4. Maquetación o etiquetado del documento fuente

Para que esta conversión a DAISY XML sea válida, se exige «maquetar» (etiquetar) el documento DOCX aplicando determinados estilos reprogramados por el complemento *Save as Daisy* para MS Word.

Los estilos aplicados son: títulos (que definen las secciones con una jerarquía anidada de niveles 1 a 6 que dan estructura al texto), páginas (Daisy), listas (ordenadas/no ordenadas), notas a pie de página (*footnotes*), notas del productor (*noteprod*), anotaciones, siglas/acrónimos, referencias/enlaces, tablas u otros estilos, como etiquetas de énfasis (cursivas/negritas). Si el documento fuente contiene imágenes (no decorativas), será necesario hacer una descripción en función de su relevancia y utilidad para comprender y/o enriquecer el contexto. Esta descripción debe quedar integrada en un atributo de imagen concreto (*alt*) para que un lector de pantalla lea su descripción cuando se pase por ella.

Otro factor para tener en cuenta a la hora de etiquetar nuestro documento es la creación obligatoria de un índice o tabla de contenidos, indicando siempre la página en que se ubica cada apartado.

También es indispensable establecer idioma o idiomas (para bloques en otros idiomas) y revisar la ortografía, ya que, en ocasiones, algunos reconocimientos de texto dan problemas con algunos caracteres.

Por último, teniendo en cuenta que se trabaja con documentos accesibles, el documento DOCX generado con este sistema también debe ser accesible. Para ello, existen herramientas de comprobación y validación. Una de ellas está integrada actualmente en Word 365 y se denomina «Comprobar accesibilidad». Con un solo clic, el programa nos indicará si se han encontrado problemas de accesibilidad y cómo corregirlos. *Save as Daisy* para MS Word también dispone de validador, aunque no se basa en su accesibilidad, sino en que el documento sea válido y coherente con las directrices del estándar Daisy.

Una vez finalizado este proceso, el complemento nos permitirá exportar el fichero DOCX a Daisy XML, que servirá de entrada para generar nuestro LHD en formato con voz sintética.

3.2.5. Otras alternativas

Existen otras alternativas al complemento *Save as Daisy* para MS Word:

- Si no trabajamos en un entorno Windows, existe un *plug-in* llamado *Open XML to Daisy XML Translator*.

- Complemento *WordToEPUB* con MS Word. Una opción muy interesante con la que se está investigando en este momento.

- Usar cualquier editor XML.

3.2.6. Generación del proyecto con voz sintética: Daisy Pipeline

Es el momento de generar el libro Daisy con voz sintética. Para ello, utilizaremos Daisy Pipeline 1 y/o Daisy Pipeline 2.

Daisy Pipeline es una valiosa herramienta de código abierto desarrollada por el Consorcio Daisy en 2006 capaz de transformar documentos en formatos accesibles para personas con dificultades para acceder al texto impreso. Trabaja con módulos (*scripts*) con funciones como:

- Herramientas de validación: Daisy 2.02, Daisy 3, EPUB o Daisy XML.
- Transformar Daisy 2.02 a Daisy 3, y viceversa.

- Transformar EPUB a Daisy 2.02/Daisy 3.
- Transformar DTBook a Daisy, EPUB, XHTML, ODT, RTF, LaTeX o PEF.[7]
- Unir, dividir, recodificar, renombrar.
- Generar libros con síntesis de voz con el texto completo en formato Daisy

El método de trabajo tiene mucho que ver con la voz inglesa *pipeline*, «tubería». Para atravesar la tubería, se requiere una entrada y una salida (*Input/Output*). En nuestro caso, la entrada es el documento DTBook (Daisy XML) y la salida es el libro Daisy.

La «tubería» no funcionará correctamente si la entrada no es válida o no está bien formada; por tanto, antes de generar el libro hay que validar el documento XML de entrada.

Por último, tenemos que entender que el «funcionamiento» de cada voz sintética es diferente. No todas interpretan el texto de la misma forma. Por ejemplo, la voz de Helena (Microsoft) no lee correctamente algunas palabras. Cuando aparece «sale», la voz lee /séil/. Para corregirlo, basta con añadir una tilde: «sále». También es posible crear «diccionarios» (lexicones) con palabras que se desean modificar. Actualmente, se dispone de una lista de más de 300 entradas con palabras que esta voz no lee o pronuncia correctamente. Esto implica que el documento de entrada debe ser «falseado» para que dichas palabras se lean correctamente, sin olvidar de revertir posteriormente los cambios al final del proceso.

3.3. Diferencias entre Pipeline 1 y Pipeline 2

Existen dos versiones que funcionan de forma diferente.

3.3.1. Daisy Pipeline 1

Como se indicó en la introducción, en el SBO de Madrid se utilizaba una herramienta propia para generar libros en síntesis de voz en formato Daisy. Pero, desde finales de 2006, se optó por la utilización de un *script* de Pipeline 1 para generar libros con cualquier voz instalada en nuestro sistema operativo a través de SAPI4, y, posteriormente, SAPI5. Solamente teníamos que dar instrucciones a Pipeline sobre qué voz instalada queríamos usar, como, por ejemplo, Jorge (Loquendo) o Conchita (Ivona). Tras añadir

7 PEF: *Portable Embosser Format*. Un tipo de archivo para mostrar braille de forma digital. Daisy Pipeline 2 también es capaz de generar archivos de braille en formato BRF (*Braille Ready Format*).

como entrada el documento XML, el *software* nos daba como salida dos carpetas, una con el proyecto Daisy 2.02 y otra con el mismo proyecto en formato Daisy 3.

Con el paso del tiempo, algunas empresas que producían y/o vendían voces sintéticas dejaron de proporcionar voces instalables, y, por otro lado, el Consorcio Daisy publicó Pipeline 2 en 2010 con versiones para Windows, MacOs y Linux. Se dio paso a que los servicios que ofertaban los proveedores de TTS se adquiriesen vía *online*; por tanto, el SBO necesitaba «reescribir» Pipeline 1 (o migrar a Pipeline 2). Por otro lado, el «mantenimiento» de Pipeline 1 dejó de estar realmente actualizado a partir de 2015 y, además, da problemas con las nuevas versiones de Java. Esto no significa que ya no se esté utilizando Pipeline 1, pero es cierto que requiere realizar ajustes en los sistemas operativos actuales.

3.3.2. Daisy Pipeline 2: la nueva herramienta

El lanzamiento de Daisy Pipeline 2, también de código abierto, abrió nuevas perspectivas y añadió nuevas funcionalidades para los nuevos estándares como EPUB 3, Daisy 3 o PEF. Un avance muy significativo es que, además de ser una aplicación de escritorio, se puede ejecutar como aplicación web para que uno o varios usuarios puedan trabajar juntos en la misma aplicación desde su navegador web, de cara a un servidor central (aplicación cliente-servidor).

Daisy Pipeline 2 ha pasado a ser la herramienta actual para generar libros Daisy 3 mediante síntesis de voz, sobre todo desde que, en 2020, la ONCE seleccionara la voz de Lucía de Amazon. Esta voz no se «instala» en ningún equipo: se accede a ella a través del servicio Amazon Polly, integrado en Amazon Web Services (AWS), una enorme colección de servicios de computación en la nube.

La elección de Lucía como voz sintética ha sido consensuada por unanimidad dentro de nuestra institución, y con respecto a otras voces dispone de las siguientes ventajas:

- Timbre, vocalización, entonación y puntuación adecuadas.
- Óptimo ritmo de lectura que además es personalizable.
- Buena pronunciación en otros idiomas.
- Acentuación correcta.
- Voz generada por Inteligencia Artificial.
- Facilidad para crear diccionarios o lexicones para diferentes materias.

Como única desventaja, podemos decir que la generación TTS no es gratuita, ya que se factura por carácter.

Pipeline 2 viene preconfigurado para utilizar varios servicios o procesos generadores de voz sintética, como Google Cloud TTS, SAPI u otros proveedores, como el Grupo Acapela, el sintetizador de voz gratuito eSpeak, MacOS Speech (Apple) o los escoceses CereProc.

Como Pipeline 2 no está preconfigurado para trabajar con Lucía (Amazon), se tomó la decisión de reprogramar una nueva versión de Daisy Pipeline 2 capaz de conectarse con la interfaz de Amazon Polly. Esta tarea fue encomendada a la consultora tecnológica madrileña NTT Data España, conocida como Everis hasta octubre de 2021.

El proceso es el mismo de siempre, una entrada (XML) y una salida (Daisy 3). Pipeline 2, a diferencia de Pipeline 1, que ofrecía dos salidas (Daisy 2.02 y Daisy 3), solo da como salida una carpeta con el libro en formato Daisy 3. Por tanto, si se quería seguir con el estándar Daisy 2.02, estaríamos obligados a un proceso más: transformar Daisy 3 a Daisy 2.02. Esta conversión que *a priori* parece sencilla y se puede realizar con Pipeline, todavía no está muy pulida y no es capaz de transformar libros Daisy con texto completo. Este aspecto fue decisivo para que el SBO tomase la decisión de migrar del formato Daisy 2.02 al Daisy 3. Por tanto, desde noviembre de 2022, los libros hablados digitales publicados por el Servicio Bibliográfico tienen el estándar Daisy 3.

A continuación, pasamos a ver algunas ventajas e inconvenientes de Daisy 3 vs. Daisy 2.02. Respecto a las ventajas, las más reseñables:

- Estándar basado en XML: más potente, más versátil, más sólido, sin ambigüedades y personalizable.
- Es el estándar más actualizado.
- Acceso a la información más eficiente.
- Nuevas opciones de etiquetado y/o navegación que pueden ser opcionales o no.
- Es posible generar pausas en la reproducción esperando a que el usuario tome una decisión. Ejemplos: opciones para continuar la lectura en una página u otra, o que el texto nos dé un tiempo determinado para reflexionar.

- Es posible omitir y/o desordenar partes del libro.
- Mejoras para la creación y manipulación de marcas o textos resaltados.
- Es un estándar que comparte elementos y funcionalidades con el estándar de libros electrónicos EPUB 3.
- Los libros generados con Daisy Pipeline 2 con síntesis de Lucía dispondrán del texto completo.

En cuanto a las desventajas, las más reseñables son:

- Algunos reproductores Daisy antiguos no podrán reproducir libros en formato Daisy 3, como, por ejemplo, el Victor antiguo, aunque el número de dispositivos es casi inexistente.
- En otros reproductores, como el PlexTalk PTN1, el acceso al disco es un poco más lento.
- Admite dos formas diferentes de «escribir» tiempos en los archivos de sincronización (SMIL): «00:00:00» o «npt="0.000s"», a tener cuenta a la hora de crear aplicaciones de lectura Daisy, como el Gestor ONCE de Libros Digitales (GOLD), desarrollado por el Centro de Tiflotecnología e Innovación de la ONCE (CTI).

El paso a Daisy 3 no implica convertir todos los títulos en Daisy 2.02, ya que ambos estándares están vigentes hoy en día.

4. Conclusiones. Futuro

Este artículo ha intentado dar a conocer cómo ha sido la trayectoria y evolución en la adaptación de libros hablados en la ONCE. Cómo se han llegado a buscar nuevas soluciones, como la voz sintética, no en detrimento de la voz humana, sino como complemento al servicio. Es evidente que, actualmente, la voz sintética se parece mucho a la voz humana, y permite una producción rápida y versátil, pero eso no significa que se vayan a dejar de producir libros con voz humana.

Actualmente, el avance tecnológico para la generación de voces sintéticas ha dado pasos de gigante. Al principio, las voces se sintetizaban por métodos como la conca-

tenación, es decir, algo parecido a unir fragmentos de voces pregrabadas. Pero, en los últimos años, el método es mucho más eficaz y complejo. Con la Inteligencia Artificial, la síntesis se basa en el uso de redes neuronales, algoritmos matemáticos que imitan la conexión de las neuronas en nuestro cerebro. El resultado se traduce en una enorme mejora de la voz sintética, ya que se parecen mucho más a la voz humana, no solo por tener una pronunciación mejor, sino porque resultan más «sentimentales», más próximas y expresivas.

Además de Lucía, el SBO tiene previsto añadir nuevas voces inteligentes de otros servicios en la nube como Microsoft Azure, Google Cloud e IBM Cloud.

De forma paralela, la Fundación ONCE está desarrollando nuevos proyectos con voces sintéticas generadas con Inteligencia Artificial gracias a su colaboración con Vicomtech, centro tecnológico especializado en «Artificial Intelligence, Visual Computing & Interaction». Aunque este proyecto todavía está en desarrollo, la Fundación ONCE ha tenido la amabilidad de arrojar algo de información sobre el Proyecto IANA (Inteligencia Artificial para Narrativa Accesible) que surge con el objetivo de permitir a todo tipo de entidades públicas, privadas y particulares generar audio a partir de texto con voces reales, no robóticas, generadas mediante Inteligencia Artificial para un fin social (AI4SG, por sus siglas en inglés; *AI for Social Good*) con voces de locutores profesionales. IANA cuenta con tecnología relacionada con la clonación y síntesis neuronal de voz, y dispone de técnicas de Aprendizaje Profundo a través de redes neuronales que aportan inteligibilidad y naturalidad a la voz.

Por último, y como conclusión, vamos a resumir las expectativas que tiene la ONCE:

- Seguir buscando y experimentado con nuevas voces que puedan surgir en el mercado de otros proveedores en la nube como Microsoft Azure, IBM Cloud o Google Cloud.

- Buscar otros métodos de maquetación de textos. Ya se han realizado con éxito diferentes pruebas desde EPUB a Daisy 3 con el complemento *WordToEPUB*. Las ventajas más relevantes son la rapidez en la generación del documento EPUB y la posibilidad de generar síntesis con editores o distribuidores que trabajen en este formato estandarizado y libre.

- Incrementar y optimizar la producción de libros con síntesis de voz en los centros del Servicio Bibliográfico de Madrid y Barcelona.

Sierra, Á. (2022). Adaptación de libros hablados digitales mediante síntesis de voz en el Servicio Bibliográfico de la ONCE. *RED Visual: Revista Especializada en Discapacidad Visual, 80*, 106-126. https://doi.org/10.53094/IBOA4928.

- La posibilidad de generar nuestras propias voces para que puedan ser utilizadas en un futuro próximo.

- Ampliar el uso de las voces sintéticas para otros usos, como centralitas, audio corporativo o mensajes.

- Formación continua del equipo de producción, tanto trabajadores internos como externos (proveedores).

- Informar al mundo editorial sobre la necesidad de publicar libros electrónicos accesibles de cara a la normativa europea que entrará en vigor en 2025.

Ángel Sierra Berrocal. Técnico del Departamento de Adaptación, Producción y Tiflotecas; Sonido. Servicio Bibliográfico de la ONCE en Madrid. Calle de La Coruña, 18; 28020 Madrid (España). Correo electrónico: ansb@once.es.

Sierra, Á. (2022). Adaptación de libros hablados digitales mediante síntesis de voz en el Servicio Bibliográfico de la ONCE. *RED Visual: Revista Especializada en Discapacidad Visual, 80*, 106-126. https://doi.org/10.53094/IBOA4928.