Voz a texto, texto a voz, subtitulado y apoyo a la lectura

Autor: Zsolt Lengyel, iTStudy Hungary Training and Research Centre

Revisora: Mária Hartyányi, iTStudy Hungary Training and Research Centre

Objetivo de aprendizaje

La conversión de voz a texto y de texto a voz es una de las aplicaciones de la IA más maduras y útiles en educación. El objetivo de este material es mostrar cómo las herramientas de voz a texto, texto a voz, subtitulado y apoyo a la lectura pueden hacer que los materiales didácticos sean más accesibles y utilizables en la enseñanza diaria. Al terminar este material, los docentes serán capaces de elegir y utilizar estas herramientas para estudiantes con diferentes necesidades sensoriales, lingüísticas, lectoras o de concentración; crear subtítulos para vídeos, versiones en audio de textos y transcripciones de explicaciones; preparar apoyos de lectura simplificados, y comprobar la calidad de los resultados.

Introducción: un contenido, varios canales

La mayor parte del contenido didáctico llega al alumnado por un único canal: un texto impreso, una explicación oral o un vídeo. Cada canal único deja fuera a alguien: al estudiante que no puede oír la explicación, que no consigue leer un texto denso con suficiente rapidez o que pierde la concentración a mitad de camino. Las herramientas de voz y lectura basadas en IA facilitan ofrecer el mismo contenido por varios canales a la vez: un texto que también se puede escuchar, una explicación oral que también se puede leer, un vídeo que también tiene subtítulos. Es la forma más sencilla de llevar a la práctica el principio de "**proporcionar múltiples formas de representación**".

DUA son las siglas de Diseño Universal para el Aprendizaje (UDL, en inglés). Es un marco pedagógico basado en la idea de que no existe un "estudiante medio", por lo que la enseñanza debe diseñarse desde el principio para que sea accesible al mayor número posible de personas, en lugar de adaptarla a cada estudiante a posteriori.

Voz a texto

El reconocimiento de voz ya es lo bastante preciso –en la mayoría de los idiomas– como para usarlo a diario:

- El **dictado** está integrado en Windows (Win+H), macOS, los teclados de Android e iOS, Google Docs y Word, así como en asistentes de IA como ChatGPT. Los estudiantes con dificultades de escritura o discapacidad motora pueden expresar sus conocimientos hablando.
- **Transcripción** de grabaciones o de la voz en directo: Transcripción instantánea (Live Transcribe) en Android, la función Transcribir de Word y las transcripciones de reuniones en Teams o Google Meet. Los

servicios en línea convierten una hora de grabación en texto en pocos minutos y a bajo coste.

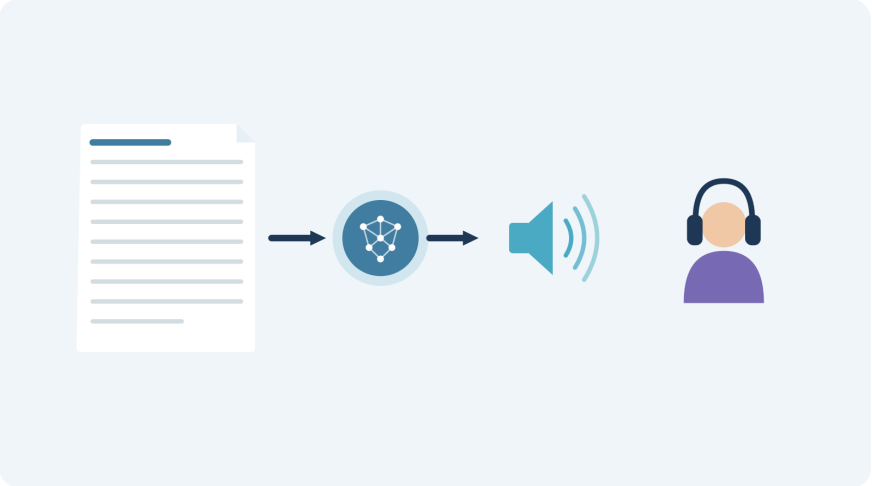
- Las **transcripciones de las explicaciones** ofrecen al alumnado un texto que puede releer, buscar y traducir, lo que resulta útil para estudiantes con pérdida auditiva, que no estuvieron presentes o que no estudian en su lengua materna.



La conversión de voz a texto transforma las explicaciones orales en texto editable

La precisión disminuye con el ruido de fondo, las voces superpuestas, los dialectos y la terminología especializada. Hable con claridad, utilice un micrófono externo siempre que sea posible y corrija los términos clave antes de compartir una transcripción. Para grabar a otras personas se necesita su consentimiento.

Texto a voz



La conversión de texto a voz transforma cualquier texto en una versión en audio de sonido natural

Las voces neuronales modernas suenan naturales y están disponibles en muchos idiomas, incluido el español:

- La función **Leer en voz alta** de Microsoft Edge y el Lector inmersivo de Word y OneNote leen documentos y páginas web en voz alta y resaltan las palabras a medida que se leen.

- Las **funciones de accesibilidad del móvil** (Seleccionar para pronunciar en Android, Leer pantalla en iOS) leen todo lo que aparece en la pantalla.
- Los **servicios de generación de voz**, como ElevenLabs, crean locuciones de alta calidad para las versiones en audio de los materiales didácticos.

Las versiones en audio ayudan a estudiantes con dislexia o discapacidad visual, a quienes leen en una segunda lengua y a cualquiera que prefiera escuchar, por ejemplo, de camino a las prácticas en empresa. Escuchar mientras se lee (lectura bimodal) puede mejorar tanto la comprensión como la capacidad de mantener la atención durante la lectura.

Subtítulos accesibles y traducidos

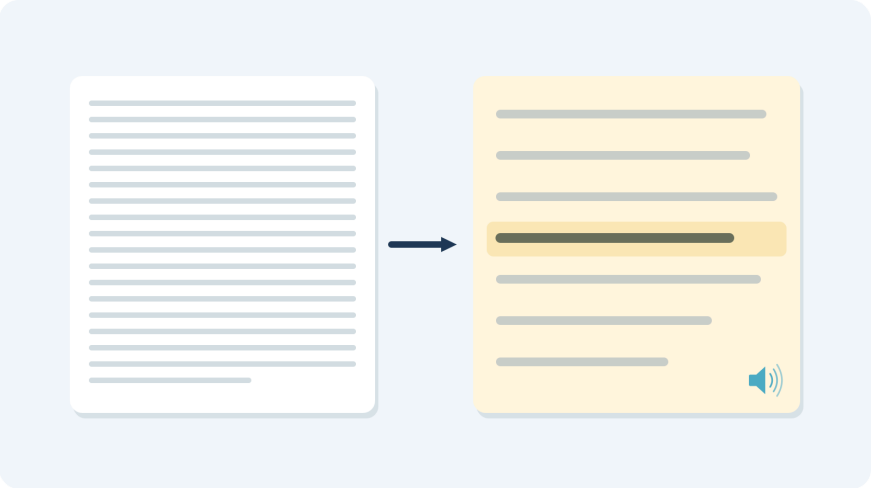


Los subtítulos hacen accesibles las explicaciones en vídeo, y resultan útiles en un taller ruidoso

Los subtítulos son imprescindibles para el alumnado sordo o con pérdida auditiva, y las **Pautas de Accesibilidad para el Contenido Web** (WCAG) los exigen en los vídeos pregrabados. Además, ayudan a quienes ven el vídeo en un entorno ruidoso, a quienes no estudian en su lengua materna o a quienes necesitan ver cómo se escribe un término técnico.

- YouTube, Microsoft Stream y la mayoría de las herramientas de vídeo generan **subtítulos automáticos**; los sistemas operativos y navegadores ofrecen **subtítulos en directo** para cualquier audio.
- Los subtítulos automáticos deben **revisarse y corregirse**, sobre todo en lo que respecta a términos técnicos, cifras y advertencias de seguridad.
- Los subtítulos pueden **traducirse** automáticamente, lo que ayuda a los estudiantes con otra lengua materna.
- Es preferible usar subtítulos opcionales, que el alumnado puede activar, desactivar y ampliar, en lugar de subtítulos incrustados en la imagen.

Apoyo a la lectura



Los modos de lectura reducen la saturación visual, resaltan la línea actual y pueden leer el texto en voz alta

Las herramientas de apoyo a la lectura reducen el esfuerzo de descodificación y ayudan al alumnado a centrarse en el significado:

- Los **modos de lectura** (Lector inmersivo, vista de lectura del navegador) eliminan los elementos superfluos y permiten cambiar el tamaño de letra, el espaciado, el color de fondo y el enfoque de línea; algunos pueden dividir las palabras en sílabas o mostrar imágenes para las palabras más comunes.
- **Simplificación y resúmenes con IA:** un modelo de lenguaje puede reescribir un texto con un nivel de lectura más bajo, dividirlo en secciones breves y añadir títulos y un resumen, manteniendo los términos técnicos y explicándolos.
- Los **glosarios** con los términos clave de una unidad, acompañados de explicaciones sencillas e imágenes, ayudan a todo el alumnado, en especial a quienes leen en una segunda lengua.

Receta: una ficha de trabajo, varios canales (unos 20 minutos)
1. Revise la estructura: títulos claros, párrafos breves y listas en lugar de frases largas.
2. Cree una versión en audio con Leer en voz alta o con una herramienta de texto a voz.
3. Pida a la IA una versión simplificada y un glosario de 8–10 términos clave, y revise ambos.
4. Añada subtítulos al vídeo de demostración y corrija los términos técnicos.
5. Ofrezca todas las versiones a todo el grupo, para que cada estudiante pueda elegir.

Límites y buenas prácticas

- Las transcripciones, los subtítulos y las simplificaciones generados por IA contienen errores: el docente sigue siendo responsable de la exactitud de lo que comparte.

- Las grabaciones de voz y los documentos pueden procesarse en la nube: siga las normas de protección de datos y no suba datos personales. Para material sensible, las herramientas que funcionan en local o las cuentas institucionales son opciones más seguras.
- Las herramientas de apoyo no sustituyen la enseñanza de la lectura y la escritura; dan acceso al contenido mientras esas destrezas se desarrollan.
- Ofrezca opciones en lugar de imponerlas: el alumnado sabe mejor que nadie qué formato le ayuda.

Ideas clave para docentes

- Ofrezca el mismo contenido por varios canales: texto, audio y vídeo subtitulado.
- El dictado y la transcripción ayudan a quienes tienen dificultades para escribir o escuchar.
- Las voces sintéticas de sonido natural permiten generar rápidamente versiones en audio de cualquier texto.
- Revise siempre los subtítulos y las transcripciones automáticos, sobre todo los términos técnicos y la información de seguridad.
- Los modos de lectura, las versiones simplificadas y los glosarios reducen el esfuerzo de lectura y favorecen la comprensión.

Referencias

- W3C (2023): Web Content Accessibility Guidelines (WCAG) 2.2 – Success Criterion 1.2.2 Captions (Prerecorded).
- CAST (2024): Universal Design for Learning Guidelines version 3.0.
- Soporte de Microsoft: Immersive Reader and Read Aloud – languages and voices.
- Lengyel, Zs. (2026): Concrete applications for content generation. Presentación de formación VITA.

Enlaces

- Idiomas y voces del Lector inmersivo: support.microsoft.com
- Live Transcribe & Sound Notifications: play.google.com
- Subtítulos – W3C: w3.org/WAI/media/av/captions
- Pautas del DUA: udlguidelines.cast.org
- W3C Web Accessibility Initiative: w3.org/WAI

Derechos de autor © 2026 Consorcio VITA

Esta obra puede compartirse y adaptarse bajo los términos de la Licencia Internacional Creative Commons Atribución/Reconocimiento-CompartirIgual 4.0 (CC BY-SA 4.0).

Las imágenes sin fuente indicada se han generado con Claude para Mac.

Financiado por la Unión Europea. Las opiniones expresadas son únicamente las de los autores y no reflejan necesariamente las de la Unión Europea ni las de la Agencia Nacional.