



Análisis de mercado de subtítulos traducidos por IA

Evaluación tecnológica independiente

Índice

- 3 Aspectos clave**
- 5 Introducción**
- 7 Calidad de traducción**
 - 7 Rendimiento general
 - 8 Rendimiento según la gravedad de los errores
 - 9 Rendimiento por idioma
 - 11 Preferencias de los lingüistas
- 12 Estabilidad de los subtítulos**
 - 13 Rendimiento general
 - 13 Rendimiento por idioma
- 16 Conclusión**
 - 16 Resultados generales
 - 17 Repercusiones para las empresas
- 18 Apéndice**
 - 18 Metodología

Aspectos clave

1. DeepL Voice destaca tanto por la calidad de traducción como por la estabilidad de los subtítulos

DeepL Voice obtuvo las puntuaciones más altas tanto en la evaluación lingüística humana como en la medición automatizada de la estabilidad de los subtítulos. DeepL Voice para Zoom logró un índice de calidad de **96,4/100**, mientras que DeepL Voice para Teams alcanzó un **96,3/100** frente a los **87-89** puntos de otras plataformas. Los productos de DeepL Voice también generaron subtítulos más estables, con índices de **88,6** y **85,8** respectivamente.

2. DeepL reduce significativamente los errores graves de traducción

En todas las combinaciones de idiomas, DeepL Voice redujo la tasa de errores graves o críticos en un **76 % de media** frente a las demás plataformas analizadas.

El **79 % de los segmentos analizados con DeepL Voice** superó la evaluación lingüística en su totalidad, en comparación con el **42 % registrado en otras herramientas** de la competencia.

En todas las combinaciones de idiomas, DeepL Voice **redujo** la media de errores de traducción por segmento en un **66 % frente a Microsoft Teams** y en un **64 % respecto a Zoom**.

3. La estabilidad de los subtítulos varía considerablemente según la plataforma

En todas las plataformas se observó cierta volatilidad en los subtítulos, entendida como el parpadeo o rectificación repetidos del texto en pantalla. Sin embargo, los productos de DeepL Voice mostraron la menor tasa de volatilidad.

En todas las combinaciones de idiomas, **DeepL Voice redujo este fenómeno en un 37,6 %** en comparación con **Microsoft Teams**, y en un **54,7 % de media** frente a **Zoom**.

4. La gran mayoría de lingüistas prefiere DeepL Voice

Tras completar las evaluaciones a ciegas, el 96 % de los lingüistas se decantó por un producto de DeepL Voice como su plataforma preferida para generar subtítulos traducidos.

5. Los productos de DeepL Voice, a la vanguardia de la subtitulación con IA

Con el fin de sintetizar los resultados de precisión lingüística y estabilidad de los subtítulos, Slator evaluó las plataformas basándose en dos criterios:

1. La calidad de traducción, analizada mediante una evaluación lingüística humana.
2. La estabilidad de los subtítulos, medida a partir de un análisis automatizado por fotograma.

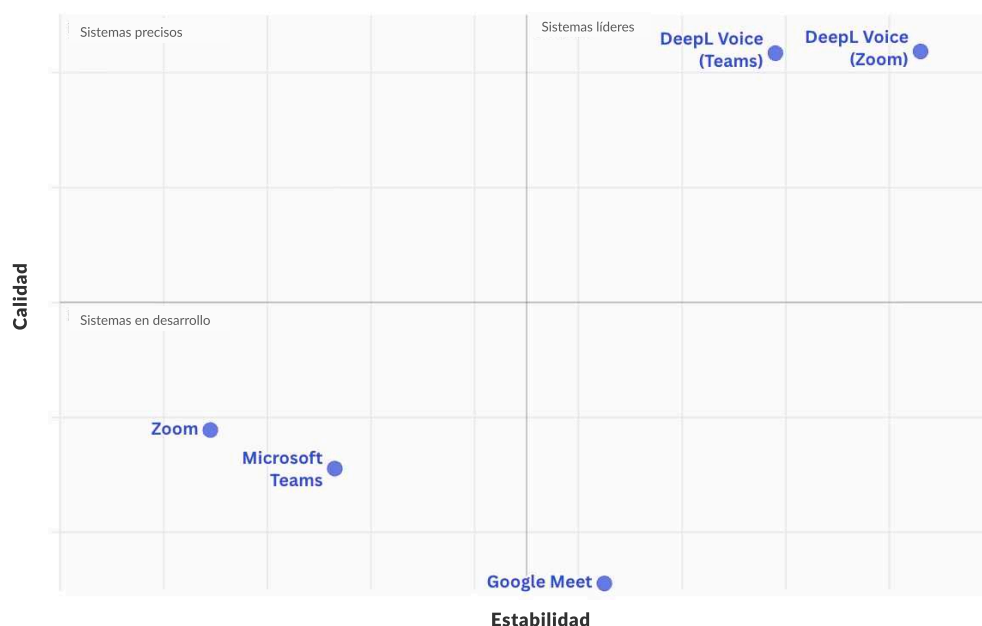
A partir de aquí, se identificaron cuatro posibles perfiles del sistema:

| Cuadrante | Descripción |
|---------------|---|
| Líderes | Buena calidad de traducción y estabilidad de los subtítulos |
| Precisos | Buena calidad, pero menor estabilidad |
| Eficientes | Menor calidad, aunque con subtítulos relativamente estables |
| En desarrollo | Menor rendimiento tanto en calidad como en estabilidad |

Los resultados sitúan a los productos de DeepL Voice en el cuadrante de los sistemas líderes dentro de este marco de evaluación:

Cuadrante de plataformas con traducción de subtítulos por IA

Comparativa de sistemas de traducción de subtítulos en tiempo real basada en la calidad lingüística y la estabilidad de la renderización.



Introducción

La traducción se está integrando cada vez más como una función clave en las herramientas corporativas; de hecho, las plataformas de reuniones globales ya incorporan subtítulos traducidos por IA. La traducción de subtítulos en tiempo real ya está disponible en plataformas como Google Meet, Microsoft Teams y Zoom para facilitar la comunicación entre diferentes idiomas.

No obstante, pese a que cada vez son más comunes, todavía faltan análisis comparativos independientes que evalúen su eficacia en reuniones reales. En la práctica, lo que define la experiencia de usuario con los subtítulos traducidos en tiempo real son dos aspectos fundamentales:

- Calidad de traducción: si los subtítulos reflejan fielmente el mensaje del orador y
- Estabilidad de los subtítulos: si los subtítulos se visualizan con fluidez en pantalla sin sufrir cambios bruscos ni parpadeos según se procesa el audio.

Los cambios constantes en el texto y las traducciones inestables [pueden dificultar la comprensión](#), por muy fiel que sea la traducción final. Por tanto, evaluar tanto la precisión lingüística como la estabilidad visual permite hacerse una idea mucho más clara de la experiencia que ofrecen estos sistemas de subtítulo al usuario final.

Para dar respuesta a esta falta de datos, Slator ha llevado a cabo un estudio independiente de la traducción de subtítulos por IA en cinco plataformas distintas:

1. Google Meet
2. Microsoft Teams
3. Zoom
4. DeepL Voice para Microsoft Teams
5. DeepL Voice para Zoom

El estudio examinó tanto la calidad de traducción como la estabilidad de los subtítulos en 14 combinaciones de idiomas: siete idiomas traducidos al inglés y otras siete con el inglés como lengua de origen: español, francés, alemán, italiano, portugués, coreano y japonés.

El estudio evalúa los subtítulos que el usuario ve directamente en pantalla, más allá de los procesos subyacentes de reconocimiento de voz o de la traducción que genera el sistema en segundo plano. Slator grabó y analizó reuniones en vídeo y extrajo los subtítulos de los propios fotogramas; esto permitió evaluar con precisión la experiencia real de usuario con respecto a los subtítulos visibles en pantalla.

28 lingüistas profesionales llevaron a cabo una evaluación a ciegas de subtítulos traducidos con el fin de contrastar el rendimiento de cada plataforma. No se les comunicó que los sistemas personalizados de DeepL Voice formaban parte del estudio, para garantizar de este modo que las valoraciones se centraran exclusivamente en la calidad del subtítulo y la facilidad de uso.

Las plataformas se evaluaron con la configuración estándar predeterminada para Google Meet, Microsoft Teams y Zoom. En la evaluación de DeepL Voice para Teams y DeepL Voice para Zoom se emplearon las funciones nativas disponibles para los usuarios finales, como los glosarios de traducción. Además, en el caso de DeepL Voice para Teams se incluyeron las funciones de reconocimiento de términos hablados, que facilitan la transcripción de nombres propios y tecnicismo.

El material de audio se obtuvo de grabaciones de podcast en los que dos personas conversaban sobre temas de negocios en un contexto coloquial. Cada grabación se editó para generar unos 12 minutos de conversación ininterrumpida por idioma.

Gracias a esta metodología, se logró que el estudio reflejara patrones de diálogo naturales, la terminología específica del sector y las características de la comunicación verbal propias de las reuniones en entornos profesionales. El objetivo de la metodología de evaluación era analizar la experiencia de uso real de los subtítulos en diferentes plataformas y en condiciones comparables.

Como la estructura lingüística de cada idioma es distinta, algunas traducciones requieren de forma natural más rectificaciones provisionales que otras. Por ejemplo, en idiomas en los que el orden de los elementos varía o el verbo va al final de la frase (como el japonés o el coreano), es habitual que los sistemas de traducción tengan que ajustar los subtítulos a medida que disponen de más contexto. Por esta razón, el análisis de los resultados se lleva a cabo tanto de forma conjunta como desglosada por idioma.

Además de la evaluación lingüística humana, Slator desarrolló un sistema de medición automatizado para cuantificar la estabilidad de los subtítulos mediante el análisis de los cambios del texto renderizado en cada fotograma.

El proceso fue el siguiente: Grabación de vídeo > Extracción de fotogramas (10 fps) > Encuadre del área de subtítulos > Reconocimiento óptico de caracteres (Tesseract) > Estandarización del texto > Comparación de fotogramas > Detección de cambios.

Estos análisis humanos y automatizados ofrecen conjuntamente una visión integral de la eficacia del subtítulo con IA durante reuniones multilingües en tiempo real.

En el apéndice se incluye una descripción completa de la metodología y el proceso.

Calidad de traducción

Slator contó con un equipo de 28 lingüistas nativos para analizar subtítulos traducidos por IA en 14 combinaciones de idiomas (siete con el inglés como idioma de llegada y otras siete con el inglés como idioma de partida) en una evaluación a ciegas. En el apéndice se puede consultar más información sobre la metodología.

DeepL Voice superó sistemáticamente a las herramientas de subtítulos nativas, tanto por su calidad general como por presentar menos errores de traducción graves.

Los resultados son los siguientes:

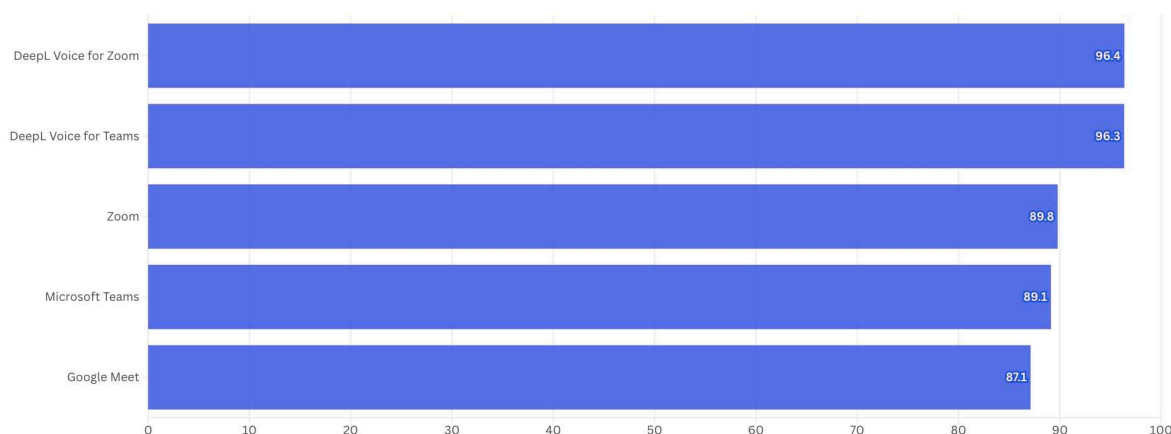
Rendimiento general

Slator consolidó los resultados en un único índice de calidad de 0 a 100 que refleja la gravedad global de los errores de traducción cometidos en todos los segmentos evaluados. De este modo, Slator pudo dar respuesta a la pregunta: «en general, ¿qué grado de calidad ofrecen los subtítulos traducidos por IA en las cinco plataformas?»

Los resultados de la evaluación son los siguientes:

DeepL Voice para Zoom obtuvo la puntuación más alta en la evaluación humana de la calidad de los subtítulos traducidos por IA

Índice de calidad general (0-100) de subtítulos traducidos por IA en todas las plataformas y combinaciones de idiomas analizadas.



Fuente: Slator • El índice de calidad general se calculó a partir de evaluaciones a ciegas de la precisión y fluidez de los subtítulos traducidos por IA, llevadas a cabo por dos lingüistas nativos en cada combinación de idiomas. El índice de calidad refleja el promedio de la gravedad de los errores detectados en todos los segmentos analizados, expresado en una escala de 0 a 100. Las combinaciones lingüísticas analizadas fueron del inglés al español, francés, alemán, italiano, portugués, coreano y japonés, y viceversa.

- DeepL Voice para Zoom logró la máxima puntuación de calidad en las evaluaciones humanas de subtítulos traducidos por IA, con un 96,4/100.
- A escasa distancia se situó DeepL Voice para Teams, con un 96,3/100.
- Google Meet obtuvo la puntuación más baja, con un 87/100.

Rendimiento según la gravedad de los errores

Slator desglosó los índices de calidad mencionados anteriormente para dar respuesta a la pregunta: «¿en qué medida influye la calidad de la traducción por IA en la comprensión de los subtítulos por parte del usuario final?» De este modo, podemos determinar hasta qué punto son graves los errores de traducción en cada plataforma.

Slator analizó el porcentaje de segmentos que presentaban errores de precisión críticos o graves. Se pidió a los lingüistas que calificaran al gravedad de los errores de traducción basándose en la precisión (traducciones incorrectas, omisiones y adiciones) y la fluidez (estilo, gramática, ortografía). De este modo, Slator pudo determinar la tasa de error general de los subtítulos traducidos por IA; es decir, el porcentaje de segmentos con errores de precisión críticos o graves, que provocaban una pérdida total del sentido original de la frase o una traducción errónea.

Este dato complementa el índice de calidad mencionado anteriormente, ya que la tasa de error muestra con qué frecuencia los subtítulos alteran de forma significativa el mensaje original o carecen de sentido. En contrapartida, la tasa de éxito refleja en qué medida cada plataforma transmite con precisión el significado del audio original y mantiene la fluidez en el idioma de llegada.

A continuación, se presentan ejemplos de cada categoría (traducciones de Google Meet):

Ejemplo 1 (no apto):

- Audio original (inglés): «Let me start by saying that I think when you sit in this position as CEO of a company like Merck, there are many stakeholders who have interests».
- Subtítulo traducido (español): «Entonces, permítanme comenzar diciendo que creo que cuando uno se sienta en esta posición, como director ejecutivo de una empresa como Killing, hay muchas partes interesadas que tienen intereses».
- Tipo de error: traducción incorrecta (grave), estilo (grave)
- Comentario del lingüista: una traducción literal da lugar a errores semánticos, concretamente en la frase «cuando uno se sienta en esta posición», además de un lenguaje forzado como «partes interesadas que tienen intereses». «Merck» se interpretó como «murder» (asesinato) en español.

Ejemplo 2 (apto con incidencias):

- Audio original (inglés): «If you have to pay that much money, all of the things that I just said are not going to make you pleased with the system».
- Subtítulo traducido (español): «Y si tienes que pagar tanto dinero, todas las cosas que acabo de decir no te harán sentir satisfecho con el sistema».
- Tipo de error: estilo (grave)
- Comentario del lingüista: la traducción por IA ha cambiado el registro a informal («tú»). Es una traducción poco natural, puesto que en español se emplearía una fórmula neutra o impersonal («Y si hay que pagar tanto dinero...»).

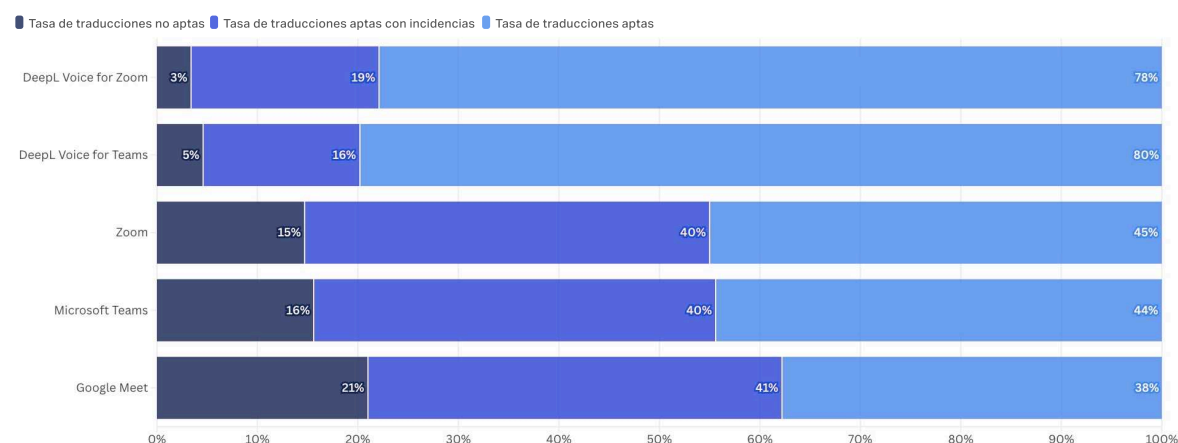
Ejemplo 3 (apto):

- Audio original (inglés): «I think it's my job to ensure that the company functions in a sustainable way to create long term value for all of its stakeholders, including its shareholders».
- Subtítulo traducido (español): «Creo que mi trabajo es garantizar que la empresa funcione de manera sostenible para crear valor a largo plazo para todas sus partes interesadas, incluidos los accionistas».
- Tipo de error: ninguno
- Comentario del lingüista: «La traducción al español de este segmento refleja el lenguaje oral en inglés».

Los resultados de este análisis son los siguientes:

DeepL Voice reduce los errores graves en un 13 % de media y alcanza una tasa de éxito promedio del 79 %

Tasa general de traducciones aptas con incidencias y no aptas de subtítulos traducidos por IA en todas las plataformas y combinaciones de idiomas analizadas.



Fuente: Slator • Las tasas se calcularon a partir de evaluaciones a ciegas de la precisión y fluidez de los subtítulos traducidos por IA, llevadas a cabo por dos lingüistas nativos en cada combinación de idiomas. Las tasas reflejan la gravedad de los errores de traducción detectados en todos los segmentos analizados. Las combinaciones lingüísticas analizadas fueron del inglés al español, francés, alemán, italiano, portugués, coreano y japonés, y viceversa.

- Los subtítulos traducidos por IA generados por los **productos de DeepL Voice** obtuvieron una **tasa media de error del 4 %**, en comparación con el **17 %** promedio del resto de herramientas. Esto supone una **reducción del 76 % de los errores de precisión críticos o graves gracias a los productos de DeepL Voice**.
- Los productos de DeepL Voice alcanzaron una **tasa media de traducciones aptas del 79 %** en todos los segmentos frente al **42 %** promedio del resto de herramientas. Esto supone un **aumento relativo del 88 %** en los segmentos aptos respecto a las demás herramientas comparadas.
- Aproximadamente, el **60 %** de los segmentos de Google Meet, así como la mitad de los de Zoom y Microsoft Teams, presentaron errores graves o problemas de traducción importantes.

Rendimiento por idioma

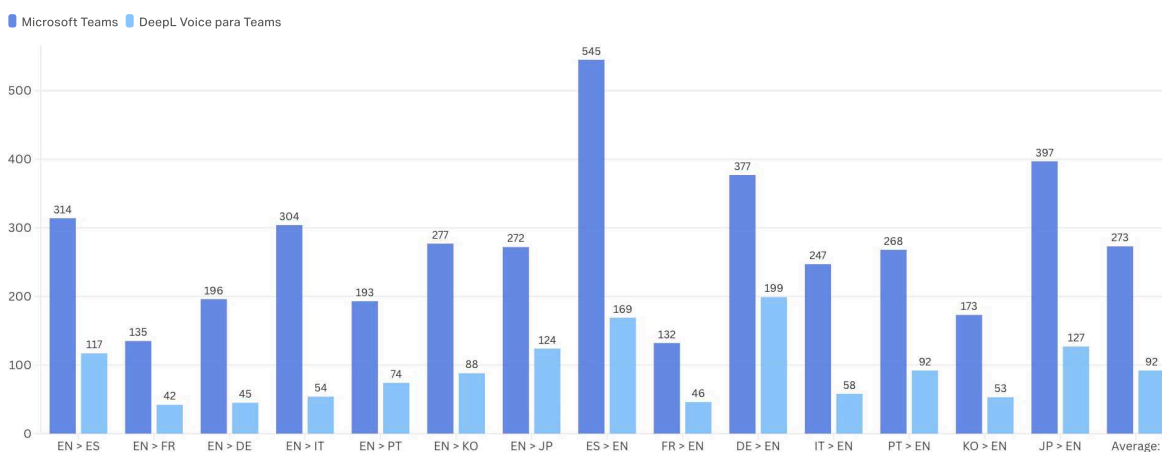
Slator profundizó aún más en el análisis para entender las diferencias lingüísticas entre las plataformas convencionales (Microsoft Teams, Zoom) y las personalizadas (DeepL Voice para Teams y DeepL Voice para Zoom) para dar respuesta a la pregunta: «¿en qué medida es DeepL Voice mejor o peor frente a las plataformas estándar?»

En todas las combinaciones de idiomas, DeepL Voice redujo la media de errores por segmento en un 66 % frente a Microsoft Teams, y un 64 % respecto a Zoom.

Al comparar Microsoft Teams y DeepL Voice para Microsoft Teams:

DeepL Voice para Teams reduce la tasa media de errores en los subtítulos en un 66 %

Análisis comparativo de las tasas de error en los subtítulos traducidos por IA en todas las combinaciones de idiomas entre Microsoft Teams y DeepL Voice para Teams.



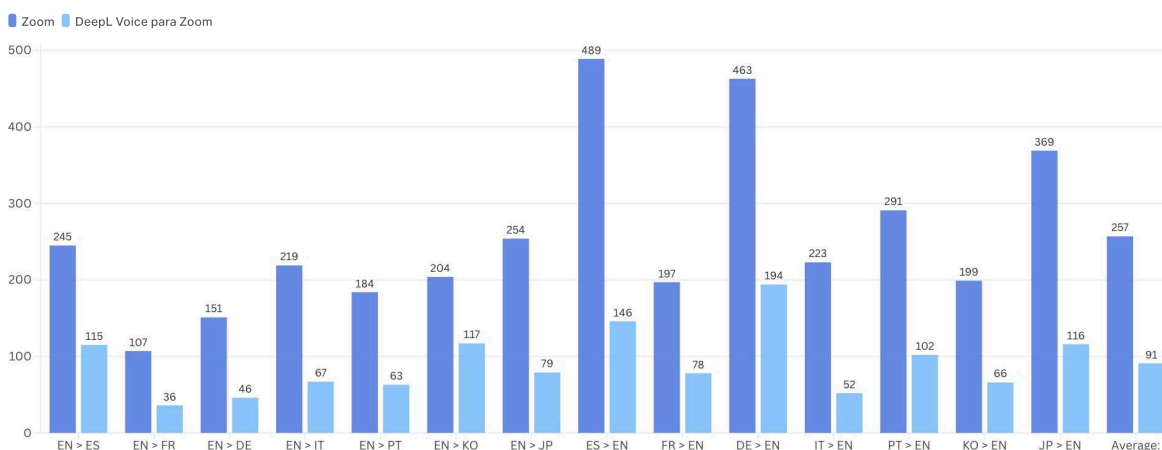
Fuente: Slator • Los valores más bajos indican una mayor calidad. Los subtítulos traducidos fueron revisados por dos lingüistas de cada combinación de idiomas y se puntuaron según la precisión y la fluidez. En el sistema de puntuación, los errores de fluidez se ponderaron con un 0,5.

- DeepL Voice para Teams logró las reducciones más significativas en la tasa media de errores por segmento en las siguientes combinaciones de idiomas:
 - De inglés a italiano (82 %)
 - De inglés a alemán (77 %)
 - De italiano a inglés (76 %)
- Solo una combinación de idiomas registró una mejora inferior al 50 % al emplear DeepL Voice para Teams (del alemán al inglés, un 47 %).

Al comparar Zoom y DeepL Voice para Zoom:

DeepL Voice para Zoom reduce la tasa media de errores en los subtítulos en un 65 %

Análisis comparativo de la tasa de error en los subtítulos traducidos por IA en todas las combinaciones de idiomas entre Zoom y DeepL Voice para Zoom.



Fuente: Slator • Los valores más bajos indican una mayor calidad. Los subtítulos traducidos fueron revisados por dos lingüistas de cada combinación de idiomas y se puntuaron según la precisión y la fluidez. En el sistema de puntuación, los errores de fluidez se ponderaron con un 0,5.

- **DeepL Voice para Zoom** logró las **reducciones más significativas** en la tasa media de errores por segmento en las siguientes combinaciones de idiomas:
 - Del italiano al inglés (77 %)
 - Del español al inglés (70 %)
 - Del inglés al alemán (70 %)
- Solo una combinación de idiomas registró una mejora inferior al 50 % al emplear DeepL Voice para Zoom (del inglés al coreano, un 43 %).

Preferencias de los lingüistas

Una vez finalizadas las evaluaciones de calidad, Slator solicitó a los 28 lingüistas participantes que clasificaran su plataforma preferida de subtítulos traducidos por IA basándose en la experiencia de usuario. Los lingüistas llevaron a cabo una evaluación a ciegas e indicaron su plataforma preferida utilizando los nombres en clave que identificaban a cada una.

Los resultados son los siguientes:

- El **52 %** de los lingüistas (15) eligió **DeepL Voice para Teams** como su plataforma preferida.
- El **44 %** de los lingüistas (12) eligió **DeepL Voice para Zoom** como su plataforma preferida.
- En total, el **96 %** (27) de los lingüistas situó a una de las dos versiones de DeepL Voice (para Teams o para Zoom) **entre sus dos opciones preferidas**.
- Zoom quedó por lo general en tercer lugar, si bien obtuvo un voto como la herramienta preferida. Microsoft Teams y Google Meet quedaron fuera de las dos opciones principales y ocuparon de forma sistemática el cuarto o quinto puesto.

Estabilidad de los subtítulos

Slator contabilizó automáticamente el número de cambios en todos los idiomas y plataformas. A modo de ejemplo, a continuación se presentan tres casos extraídos de los audios de muestra para ilustrar el procesamiento de los subtítulos antes de que la frase se estabilice por completo:

Ejemplo 1 (extensión gradual de las frases, del francés al inglés):

- Fotograma 1: «Hola Nathalie el tema de la»
- Fotograma 2: «Hola Nathalie el tema del día es la»
- Fotograma 3: «Hola Nathalie el tema del día es el asunto de la farmacia»
- Fotograma 4 (estable): «Hola Nathalie el tema del día es el asunto de los grupos farmacéuticos».

En este ejemplo, los subtítulos en tiempo real muestran traducciones fragmentadas antes de que la persona que habla haya completado la frase. Conforme se procesa el resto de palabras, los subtítulos se van extendiendo hasta que se estabiliza la oración completa.

Ejemplo 2 (reescritura de subtítulos, del español al inglés):

- Fotograma 1: «Cuando haces una demanda cuál es la cantidad económica mínima»
- Fotograma 2: «Cuando haces una reclamación cuál es la cantidad económica mínima»
- Fotograma 3: «Cuando pones una reclamación cuál es la cantidad económica mínima para que»
- Fotograma 4 (estable): «Cuando pones una reclamación, ¿cuál es la cuantía económica mínima para poder presentar una demanda?»

En este ejemplo, los subtítulos en tiempo real rectifican las palabras ya generadas a medida que se dispone de un mayor contexto lingüístico.

Ejemplo 3 (parpadeo de los subtítulos, del coreano al inglés):

- Fotograma 1: «el cuarto trimestre terminó unos diez días»
- Fotograma 2: «el cuarto trimestre es ahora de unos diez días»
- Fotograma 3 (estable): «el cuarto trimestre terminó hace unos diez días»

En este ejemplo, los subtítulos en tiempo real rectifican por completo las palabras ya generadas antes de mostrar de nuevo una versión final (siguiendo secuencias del tipo A>B>A, A>B>C o variantes similares).

Ejemplo 4 (gran estabilidad, del coreano al inglés):

- Fotograma 1: «Hablemos de los resultados de Samsung Electronics».
- Fotograma 2: «Hablemos de los resultados de Samsung Electronics».
- Fotograma 3: «Hablemos de los resultados de Samsung Electronics».

En este ejemplo, la versión final del subtítulo traducido se mostró en pantalla y permaneció sin cambios durante toda la secuencia.

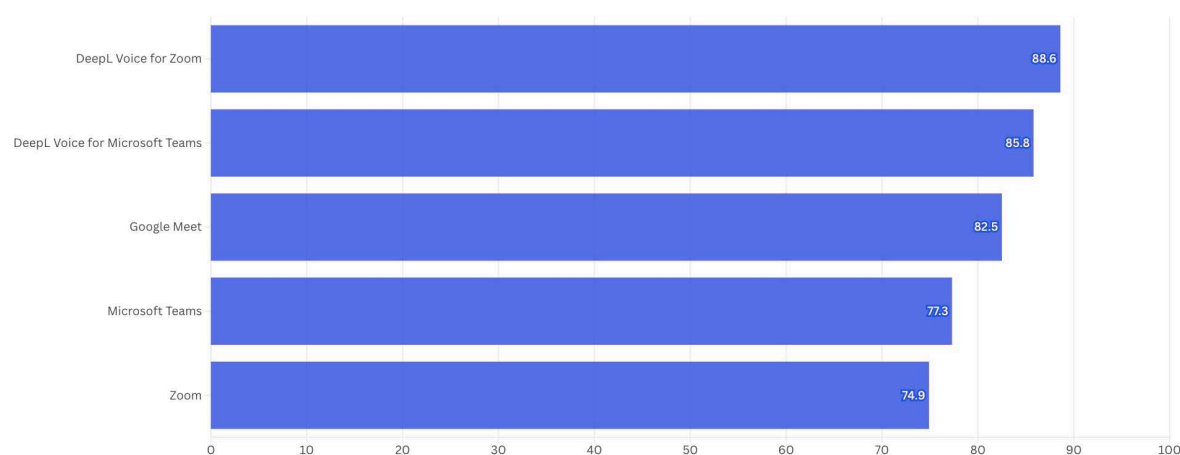
Rendimiento general

Slator consolidó los resultados en un único índice de calidad de 0 a 100 que refleja la estabilidad general de los subtítulos traducidos en todas las plataformas. De este modo, Slator pudo dar respuesta a la pregunta: «¿cuál es el grado de parpadeo o rectificación de las traducciones en cada plataforma hasta que los subtítulos se estabilizan del todo?»

Los resultados de la evaluación son los siguientes:

DeepL Voice para Zoom es la plataforma más estable para subtítulos traducidos

Índice de estabilidad general (0-100) de subtítulos traducidos en todas las plataformas y combinaciones de idiomas analizadas.



Fuente: Slator • La puntuación media de estabilidad se calculó mediante el análisis secuencial de los fotogramas de los subtítulos traducidos en cada herramienta y contabilizando el número total de cambios en cada combinación de idiomas. Estas puntuaciones se normalizaron en un único índice de estabilidad de 0 a 100.

- **DeepL Voice para Zoom** logró la **máxima puntuación** de calidad en las evaluaciones automáticas de la estabilidad de los subtítulos traducidos por IA, con un **88,6/100**.
- A escasa distancia se situó **DeepL Voice para Teams**, con un **85,8/100**.
- Zoom obtuvo la puntuación más baja, con un **74,9/100**.

Rendimiento por idioma

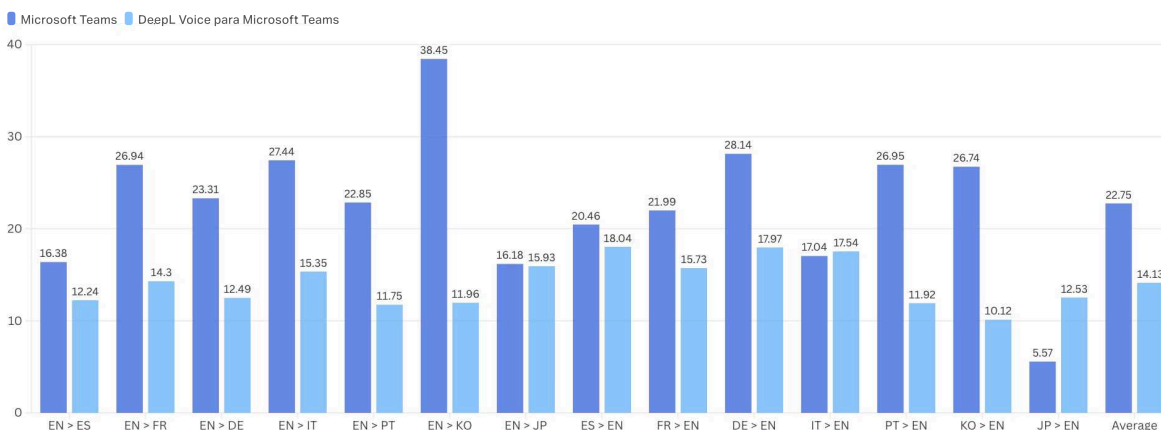
Slator profundizó aún más en el análisis para entender las diferencias lingüísticas entre las plataformas convencionales (Microsoft Teams, Zoom) y las personalizadas (DeepL Voice para Teams y DeepL Voice para Zoom) para dar respuesta a la pregunta: «¿los productos de DeepL Voice presentan una mayor o menor estabilidad en cada combinación de idiomas en comparación con las plataformas estándar?»

En todas las combinaciones de idiomas, **DeepL Voice redujo la volatilidad de los subtítulos en un 37,6 %** de media frente a **Microsoft Teams**, y un **54,7 %** con respecto a **Zoom**.

Al comparar Microsoft Teams y DeepL Voice para Microsoft Teams:

DeepL Voice para Teams mejora la estabilidad en un 38 % de media

Análisis comparativo de las tasas de volatilidad en los subtítulos traducidos por IA en todas las combinaciones de idiomas entre Microsoft Teams y DeepL Voice para Teams.



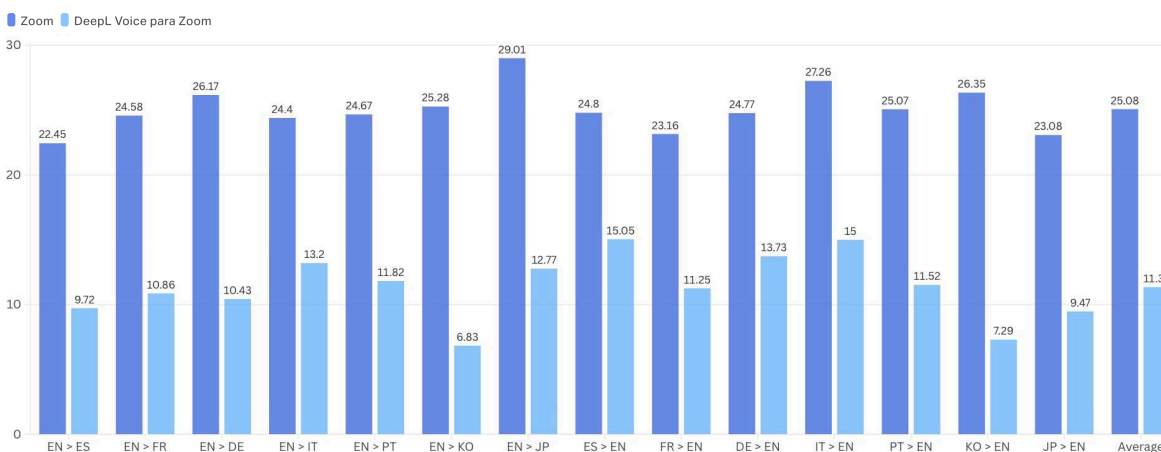
Fuente: Slator • La tasa de volatilidad de los subtítulos se define como el porcentaje de fotogramas en los que se produjo alguna modificación del texto generado. De esta forma, se cuantifica de manera directa cuántas veces aparecen cambios de texto o interrupciones en los subtítulos tal y como los ve el usuario. Los valores se calcularon automáticamente en Python.

- Las mejoras más significativas en la estabilidad de los subtítulos con DeepL Voice para Teams se registraron en las siguientes combinaciones de idiomas:
 - Del inglés al coreano (> 69 %)
 - Del coreano al inglés (> 62 %)
 - Del portugués al inglés (> 56 %)
- Solo dos combinaciones de idiomas mostraron más estabilidad en los subtítulos traducidos por IA de Microsoft Teams: del italiano al inglés (un 3 % más estables en Microsoft Teams) y del japonés al inglés (un 125 % más estables en Microsoft Teams).

Al comparar Zoom y DeepL Voice para Zoom:

DeepL Voice para Zoom mejora la estabilidad en un 55 % de media

Análisis comparativo de la tasa de volatilidad en los subtítulos traducidos por IA en todas las combinaciones de idiomas entre Zoom y DeepL Voice para Zoom.



Fuente: Slator • La tasa de volatilidad de los subtítulos se define como el porcentaje de fotogramas en los que se produjo alguna modificación del texto generado. De esta forma, se cuantifica de manera directa cuántas veces aparecen cambios de texto o interrupciones en los subtítulos tal y como los ve el usuario. Los valores se calcularon automáticamente en Python.

- Las mejoras más **significativas** en la estabilidad de los subtítulos con **DeepL Voice para Zoom** se registraron en las siguientes combinaciones de idiomas:
 - Del inglés al coreano (73 %)
 - Del coreano al inglés (72 %)
 - Del inglés al alemán (60 %)
- En ninguna de las combinaciones de idiomas la estabilidad de DeepL Voice para Zoom fue inferior a la de Zoom.

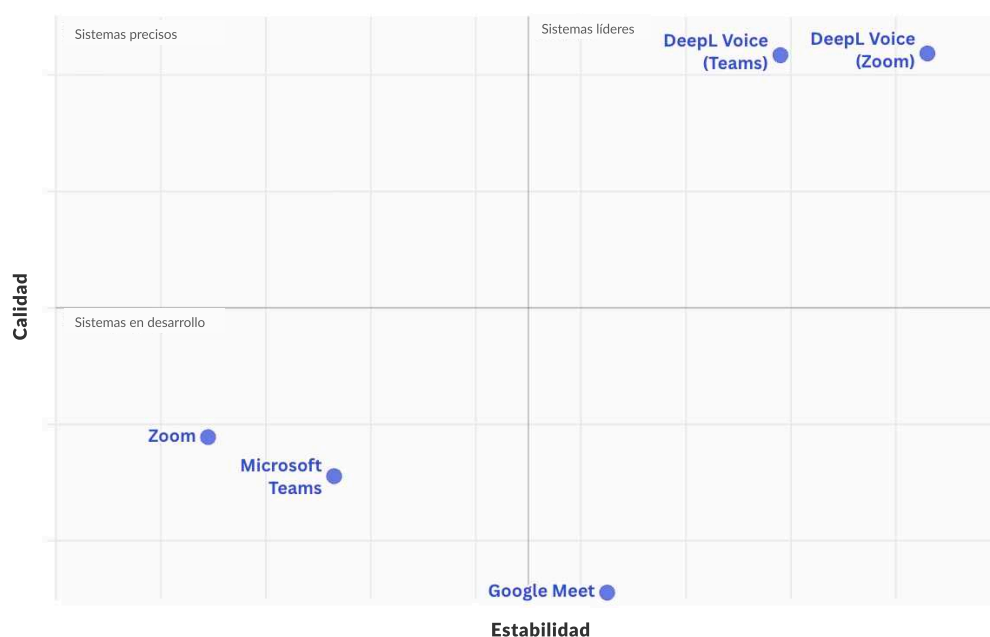
Conclusión

Resultados generales

A continuación, se detalla la clasificación general de las cinco plataformas con traducción de subtítulos por IA:

Cuadrante de plataformas con traducción de subtítulos por IA

Comparativa de sistemas de traducción de subtítulos en tiempo real basada en la calidad lingüística y la estabilidad de la renderización.



DeepL Voice para Teams y **DeepL Voice para Zoom** obtuvieron puntuaciones altas tanto en calidad como en estabilidad, por lo que se consolidaron como las soluciones **líderes** de **traducción de subtítulos con IA**. Los sistemas de subtítulos en tiempo real pueden verse en el compromiso de equilibrar la precisión de la traducción y la estabilidad de los subtítulos; la dificultad reside en aportar valor con traducciones fieles al mensaje original y minimizar, al mismo tiempo, la latencia al generarlas. No obstante, los resultados de este estudio sugieren que DeepL Voice mejora ambos aspectos de forma simultánea.

Tanto **Microsoft Teams** como **Zoom** se sitúan en la categoría de **sistemas en desarrollo**, al haber logrado un rendimiento inferior a la media tanto en la generación de subtítulos como en la calidad de las traducciones. Aunque Microsoft Teams obtuvo un rendimiento ligeramente superior en cuanto a estabilidad y Zoom respecto a la calidad, ambas herramientas requerirían optimizaciones en ambas categorías para mejorar su oferta actual.

Si bien **Google Meet** es relativamente estable en comparación con otras herramientas del mercado, obtuvo la **puntuación más baja** en el análisis que llevaron a cabo los expertos lingüísticos de Slator. Google Meet se sitúa así en la categoría de sistemas eficientes dentro de la clasificación general. Su salto a la categoría de líderes dependerá de si logra mejorar la calidad de las traducciones en todas las combinaciones de idiomas sin comprometer la estabilidad de los subtítulos.

Repercusiones para las empresas

Tras completar esta evaluación, Slator concluye las siguientes implicaciones para las empresas:

1. La calidad de los subtítulos generados por IA está mejorando, pero sigue siendo dispar dependiendo de la plataforma

Los resultados muestran una variación significativa en la calidad de la traducción en función de la herramienta elegida para las reuniones. Las organizaciones que dependen de las funciones de subtítulos integradas podrían encontrar diferencias sustanciales en la precisión de las traducciones según la herramienta elegida.

2. La estabilidad de los subtítulos afecta a la usabilidad

Aun cuando las traducciones son correctas, las rectificaciones constantes de los subtítulos pueden dificultar la comprensión. Medir la estabilidad de los subtítulos permite hacerse una idea muy clara de la experiencia real del usuario en reuniones multilingües.

3. Las plataformas especializadas con opciones de personalización pueden mejorar las traducciones

Los resultados sugieren que el uso de herramientas especializadas de traducción por IA, junto con los glosarios específicos y los sistemas de optimización de transcripción, mejoran significativamente la eficacia de los subtítulos traducidos por IA en tiempo real a distintos idiomas.

Apéndice

Metodología

Recopilación de muestras de audio

Slator compiló grabaciones de pódcast en todos los idiomas objeto de estudio. Se seleccionaron grabaciones de pódcast por ofrecer una calidad de audio homogénea en todos los idiomas, lo que facilitó una comparación controlada del rendimiento de las traducciones sin interferencias por el ruido de fondo o la calidad de los micrófonos.

Cada pódcast contaba con dos personas que mantenían una conversación fluida sobre temas de sectores regulados (finanzas, biociencia y derecho), donde el ruido ambiental era inexistente o inapreciable. Este enfoque permitió configurar el experimento para medir la calidad de la traducción de terminología especializada y analizar la capacidad de las plataformas para procesar el lenguaje espontáneo entre hablantes nativos de la lengua de partida.

Cada pódcast se editó para generar un fragmento de audio independiente de unos 12 minutos por idioma. Este archivo de audio original se procesó mediante una herramienta de transcripción automática para generar el texto original.

Configuración técnica

Slator utilizó un cable virtual para redirigir el audio original a cada plataforma (Google Meet, Microsoft Teams y Zoom). De este modo, se evitó añadir ruido ambiental durante el proceso de grabación. Slator utilizó una herramienta de captura de pantalla para grabar un vídeo en formato .mp4 de los subtítulos traducidos en cada plataforma y mostrarlos así en tiempo real, tal y como los vería cualquier usuario.

En Google Meet, Microsoft Teams y Zoom se mantuvo la configuración predeterminada para la traducción de subtítulos. En el caso concreto de Google Meet, la traducción del inglés al italiano se identificó como versión beta.

En las pruebas de DeepL Voice para Microsoft Teams y Zoom, se seleccionó un glosario terminológico y un registro informal en los ajustes de formalidad de DeepL, por el carácter distendido de las conversaciones del pódcast.

Además, se habilitó la función de términos mencionados de DeepL Voice, disponible únicamente en Microsoft Teams. Se trata de un listado monolingüe de términos mencionados (como acrónimos o nombres propios) que mejora la precisión de la herramienta a la hora de identificar el vocabulario propio del idioma de partida durante el proceso de transcripción de voz a texto.

Slator procesó todas las grabaciones mediante un motor automatizado (tal y como se describe en el apartado «Evaluaciones de calidad automatizadas») y las compartió con los lingüistas.

Selección de lingüistas

Slator seleccionó a dos lingüistas para cada combinación de idiomas, que debían ser hablantes nativos o contar con un dominio equivalente del idioma de llegada. Los lingüistas fueron elegidos minuciosamente por su experiencia en los temas relacionados con los audios y son profesionales cualificados.

El perfil de los lingüistas participantes destaca por su trayectoria en localización de contenidos multimedia, interpretación y control de calidad (incluido el entrenamiento y la evaluación de IA). Esta mezcla de perfiles subrayó la importancia de contrastar el audio con la calidad de los subtítulos que aparecían en pantalla en tiempo real.

Evaluaciones de calidad humanas

Slator creó una ficha de evaluación de calidad para cada uno de los archivos de audio originales. A continuación se muestra un ejemplo:

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---------------------|------------|----------------|--|-----------------------|----------|----------|-------|---------|----------|--------------------|
| 1 | Baseline Parameters | | | | Final Caption Quality | | | | | | |
| 2 | Tool Number | Segment ID | Start Time | Original Transcript | Mistranslation | Omission | Addition | Style | Grammar | Spelling | Evaluator Comments |
| 3 | 1 | 1 | [00:00:00.000] | Let me start by saying that I think when you sit in this position as CEO of a company like Merck, there are many stakeholders who have interests. | ▼ | ▼ | ▼ | ▼ | ▼ | ▼ | |
| 4 | 1 | 2 | [00:00:08.120] | I see my job, generally speaking, as to try to meet the needs of multiple stakeholders whose interests are often, if not opposed to one another, it's some dynamic tension with one another. | ▼ | ▼ | ▼ | ▼ | ▼ | ▼ | |
| 5 | 1 | 3 | [00:00:21.720] | That actually gets to the short term versus the long term issue. | ▼ | ▼ | ▼ | ▼ | ▼ | ▼ | |

La ficha de evaluación incluye:

- Parámetros iniciales:
 - Número de herramienta. Se trata de una referencia anónima a cada herramienta o plataforma. El objetivo es evitar que los lingüistas sepan qué plataforma están evaluando y si se ha personalizado la configuración o no. Los lingüistas desconocían que en la evaluación se habían incluido modelos personalizados de DeepL.
 - Identificador del segmento. Hace referencia al identificador del segmento correspondiente a la transcripción del audio original. Esto permitió a Slator basar la evaluación en fragmentos de texto segmentados, lo que facilitó un análisis más claro entre los distintos lingüistas de una misma combinación de idiomas, así como entre diferentes lenguas y plataformas.
 - Marca de tiempo inicial. Este dato se basa en la transcripción inicial y permitió a los lingüistas localizar con facilidad segmentos específicos en el archivo de audio original según fuese necesario.
 - Transcripción original. Se trata de la transcripción del archivo original. Se advirtió a los lingüistas de que el contenido podía contener errores. Slator mantuvo a propósito los errores en la transcripción

para evitar que los lingüistas contrastaran los subtítulos traducidos con la transcripción escrita y obligarlos así a centrarse en el audio original. De igual modo y por los mismos motivos, Slator optó por no proporcionar la transcripción de los subtítulos traducidos.

- Calidad final de los subtítulos:
 - Criterios de evaluación de la precisión:
 - Traducción incorrecta. Según la definición de Slator, esta categoría se aplica cuando «el significado del subtítulo traducido no es correcto o resulta engañoso». Se establecieron las siguientes categorías de errores:
 - Ninguno: sin errores apreciables.
 - Leve: el significado es correcto en general, solo se pierden algunos matices.
 - Grave: el significado está parcialmente distorsionado.
 - Crítico: el significado es sustancialmente incorrecto.
 - Omisión. Según la definición de Slator, esta categoría se aplica cuando «falta información importante en el subtítulo traducido». Se establecieron las siguientes categorías de errores:
 - Ninguno: sin errores apreciables.
 - Leve: el significado es correcto en general, solo se pierden algunos matices.
 - Grave: el significado está parcialmente distorsionado.
 - Crítico: el significado es sustancialmente incorrecto. Se pidió a los lingüistas que calificaran un segmento como crítico si la plataforma no lo traducía o si mostraba el idioma de partida en el subtítulo final.
 - Adición. Según la definición de Slator, esta categoría se aplica cuando «se añade información al subtítulo traducido que no aparece en el audio». Se establecieron las siguientes categorías de errores:
 - Ninguno: sin errores apreciables.
 - Leve: el significado es correcto en general, solo se pierden algunos matices.
 - Grave: el significado está parcialmente distorsionado.
 - Crítico: el significado es sustancialmente incorrecto.
 - Criterios de evaluación de la fluidez:
 - Estilo. Según la definición de Slator, esta categoría se aplica cuando «el estilo o el registro resulta inadecuado para subtítulos en tiempo real. Ejemplos: los subtítulos son demasiado extensos, formales o informales, el texto parece prosa escrita, etc.» Se establecieron las siguientes categorías de errores:
 - Ninguno: sin errores apreciables.
 - Leve: el error es perceptible, pero no dificulta la lectura.
 - Grave: distrae o dificulta la lectura.
 - Gramática. Según la definición de Slator, esta categoría se aplica cuando «la frase incumple las normas gramaticales del idioma de llegada.

Ejemplos: errores de concordancia, tiempos verbales erróneos o fallos de conjugación».

Se establecieron las siguientes categorías de errores:

- Ninguno: sin errores apreciables.
- Leve: el error es perceptible, pero no dificulta la lectura.
- Grave: distrae o dificulta la lectura.

_ Ortografía. Según la definición de Slator, esta categoría se aplica cuando «aparecen errores tipográficos u ortográficos que dificultan la lectura. Ejemplos: faltas de ortografía, tildes o signos diacríticos incorrectos, uso erróneo de las mayúsculas». Se establecieron las siguientes categorías de errores:

- Ninguno: sin errores apreciables.
 - Leve: el error es perceptible, pero no dificulta la lectura.
 - Grave: distrae o dificulta la lectura.
- Comentarios de evaluación. Slator pidió a los lingüistas que justificaran sus valoraciones en inglés. De este modo, Slator pudo cotejar las puntuaciones entre los distintos idiomas y plataformas para comprobar que las instrucciones se habían seguido correctamente y que las valoraciones eran coherentes. Asimismo, esta metodología fomentó que los lingüistas asignaran las puntuaciones con mayor criterio y que todo el proceso de evaluación fuera congruente. Además, permitió obtener valoraciones cualitativas sobre los subtítulos traducidos en los distintos idiomas y plataformas.

Se pidió a los lingüistas que rellenaran la ficha de evaluación de calidad basándose únicamente en los subtítulos traducidos finales o estabilizados.

Además, Slator les solicitó que clasificaran las plataformas por orden según la experiencia de usuario, su preferencia y la calidad percibida. A partir de estas puntuaciones, Slator extrajo conclusiones sobre las plataformas mejor valoradas por los lingüistas frente a las que ofrecen una peor experiencia de usuario.

Sistema de puntuación de las evaluaciones humanas

Tratamiento y ponderación de los datos

Dado que cada combinación de idiomas fue evaluada por dos lingüistas independientes, Slator revisó la coherencia de las valoraciones y descartó anomalías antes de consolidar los resultados. Se calculó el promedio de las puntuaciones para obtener una nota media por segmento en cada plataforma y combinación de idiomas. Si bien es normal que haya una ligera variación en la gravedad de los errores al tratarse de una evaluación humana, el cotejo de los datos ratificó que los evaluadores mantuvieron criterios muy similares a la hora de clasificarlos.

Slator asignó valores numéricos a las categorías de errores (ninguno [0 puntos], leve [1 punto], grave [3 puntos], crítico [5 puntos]).

A las categorías relacionadas con la precisión (es decir, traducciones incorrectas, omisiones y adiciones) se les asignó una puntuación total de precisión.

A las categorías relativas a la fluidez (como errores de estilo, ortografía y gramática) también se les asignó una puntuación total de fluidez.

Ambas puntuaciones de precisión y fluidez se combinaron para obtener una puntuación total por segmento. Las puntuaciones de los distintos lingüistas de una misma combinación de idiomas se consolidaron y promediaron para obtener la media de precisión, fluidez y del total por segmento.

Para el cálculo de la puntuación total por segmento, la de fluidez se ponderó con un 0,5. De este modo, se evitó que un segmento resultara «no apto» únicamente por un error grave de ortografía, gramática o estilo. Al mismo tiempo, se ponderó la puntuación total en favor de la precisión y se penalizaron más las traducciones incorrectas o las omisiones, que suelen percibirse como errores de mayor gravedad.

Clasificación de datos

La puntuación total por segmento también indicaba si cada uno de ellos se clasificaba como «apto», «apto con incidencias» o «no apto». Con este sistema, Slator pudo calcular las tasas globales de subtítulos no aptos por plataforma e idioma.

Dada la ponderación en favor de la precisión descrita anteriormente, cualquier segmento que presente alguna de las siguientes condiciones se considerará de forma automática como «no apto»:

- Presenta algún error de precisión (traducción incorrecta, omisión o adición) calificado como crítico;
- Presenta dos o más errores de precisión «graves» (p. ej., una omisión y una adición graves);
- Una traducción incorrecta se califica como un error grave (o crítico, de acuerdo con el punto 1).

Un segmento se califica como «apto con incidencias» si se cumplen las siguientes condiciones:

- Presenta algún error de fluidez (ortografía, gramática o estilo) calificado como grave;
- o bien presenta algún error de precisión calificado como grave (sin llegar a cumplir los criterios para considerarse no apto);
- o bien contiene más de tres errores leves entre todas las categorías.

Los segmentos con un máximo de dos errores leves se consideran «aptos».

Puntuación de los datos

El cálculo de la puntuación total por segmento para cada plataforma y combinación de idiomas permitió la comparación directa entre ellas.

A continuación se detallan las puntuaciones totales ponderadas (los valores más bajos indican una mayor calidad):

| Combinación de idiomas | Google Meet | Microsoft Teams | Zoom | DeepL Voice para Teams | DeepL Voice para Zoom |
|------------------------|-------------|-----------------|------|------------------------|-----------------------|
| EN-ES | 288 | 314 | 245 | 117 | 115 |
| EN-FR | 164 | 135 | 107 | 42 | 36 |
| EN-DE | 179 | 196 | 151 | 45 | 46 |
| EN-IT | 252 | 304 | 219 | 54 | 67 |
| EN-PT | 210 | 193 | 184 | 74 | 63 |
| EN-KO | 237 | 277 | 204 | 88 | 117 |
| EN-JP | 291 | 272 | 254 | 124 | 79 |
| ES-EN | 531 | 545 | 489 | 169 | 146 |
| FR-EN | 274 | 132 | 197 | 46 | 78 |
| DE-EN | 408 | 377 | 463 | 199 | 194 |
| IT-EN | 311 | 247 | 223 | 58 | 52 |
| PT-EN | 390 | 268 | 291 | 92 | 102 |
| KO-EN | 329 | 173 | 199 | 53 | 66 |
| JP-EN | 671 | 397 | 369 | 127 | 116 |

Las tasas de subtítulos aptos y no aptos se calcularon sumando el número total de segmentos de cada categoría («no aptos», «aptos» y «aptos con incidencias»), según se indica a continuación:

| Herramienta | Tasa de traducciones no aptas | Tasa de traducciones aptas con incidencias | Tasa de traducciones aptas |
|-----------------------|-------------------------------|--|----------------------------|
| Google Meet | 307 | 603 | 555 |
| Microsoft Teams | 229 | 586 | 650 |
| Zoom | 216 | 590 | 659 |
| DeepL Voice per Teams | 67 | 228 | 1170 |
| DeepL Voice per Zoom | 50 | 274 | 1141 |

Para obtener los porcentajes globales, se dividieron estas cifras por el total de segmentos analizados en todos los idiomas (1465):

| Herramienta | Tasa de traducciones no aptas | Tasa de traducciones aptas con incidencias | Tasa de traducciones aptas |
|-----------------------|-------------------------------|--|----------------------------|
| Google Meet | 21,0% | 41,2% | 37,9% |
| Microsoft Teams | 15,6% | 40,0% | 44,4% |
| Zoom | 14,7% | 40,3% | 45,0% |
| DeepL Voice per Teams | 4,6% | 15,6% | 79,9% |
| DeepL Voice per Zoom | 3,4% | 18,7% | 77,9% |

Procesamiento de los datos

Con el objetivo de facilitar una comparación clara entre las distintas herramientas y combinaciones de idiomas, las puntuaciones totales por segmento se convirtieron en una puntuación de calidad normalizada de 0 a 100. Se calculó la penalización media por segmento de cada herramienta, entendida como la suma de las puntuaciones totales de todos los segmentos evaluados (en todas las combinaciones de idiomas y en ambos sentidos) dividida por el número total de segmentos analizados.

La asignación de puntos por error siguió el baremo de puntuación establecido (categorías de precisión y fluidez con ponderaciones según la gravedad). Así, la penalización máxima teórica por segmento es de 24 puntos (15 correspondientes a las categorías de precisión y 9 a las de fluidez). A partir de ahí, la puntuación normalizada se calculó de la siguiente manera: $100 \times (1 - \text{penalización media por segmento} / 24)$.

Según esta fórmula, la puntuación de 100 corresponde a un sistema sin errores observados. Por el contrario, una puntuación de cero equivaldría al supuesto teórico de que todos los segmentos hubieran recibido la máxima penalización posible.

Los datos son los siguientes:

| Herramienta | Puntuación total ponderada | Total de segmentos | Penalización media | Puntuación normalizada (0-100) |
|------------------------|----------------------------|--------------------|--------------------|--------------------------------|
| Google Meet | 4534 | 1456 | 3,09 | 87,10537543 |
| Microsoft Teams | 3828 | 1456 | 2,61 | 89,11333902 |
| Zoom | 3592 | 1456 | 2,45 | 89,78313424 |
| DeepL Voice para Teams | 1286 | 1456 | 0,88 | 96,34172355 |
| DeepL Voice para Zoom | 1277 | 1456 | 0,87 | 96,36732082 |

Los resultados observados (87-96) sitúan el rendimiento de las herramientas en relación con este límite teórico de error, lo que garantiza la comparabilidad entre distintas herramientas, idiomas y futuras evaluaciones.

Evaluaciones de calidad automatizadas

Hemos desarrollado un flujo de medición automatizado que cuantifica la estabilidad de los subtítulos en tiempo real según aparecen en pantalla. El análisis comprende Google Meet, Microsoft Teams, Zoom, DeepL Voice para Microsoft Teams y DeepL Voice para Zoom, y analiza múltiples idiomas mediante un reconocimiento óptico de caracteres adaptado a cada lengua.

En resumen, el sistema procesa todas las grabaciones de las reuniones (una por cada idioma y plataforma), aísla el área de los subtítulos de cada fotograma y emplea el reconocimiento óptico de caracteres para reconstruir el texto visible en pantalla. A continuación, calcula los indicadores de estabilidad y genera un informe resumido.

El objetivo es evaluar la experiencia real de visionado, más allá de la transcripción interna del sistema de reconocimiento de voz. Si los subtítulos parpadean, se reescriben o presentan oscilaciones, Slator lo detecta directamente a partir de los fotogramas del vídeo renderizado.

En concreto, Slator extrajo aproximadamente diez fotogramas por segundo de cada grabación, almacenados como imágenes independientes. Este método facilitó un análisis fotograma a fotograma de lo que se mostraba en pantalla en todos los idiomas y plataformas, que permitió contabilizar incluso las rectificaciones rápidas que se producen en milésimas de segundo.

Por cada fotograma extraído, el sistema de Slator aisló un rectángulo que contenía únicamente el área específica de los subtítulos, con el fin de eliminar el contenido irrelevante de la pantalla.

Slator utilizó Tesseract, un motor de reconocimiento óptico de caracteres capaz de detectar los idiomas de cada subtítulo, para procesar todas las imágenes aisladas. De este modo, extrajo el texto visible y generó un registro por fotograma.

Los resultados sin procesar del reconocimiento óptico de caracteres pueden presentar variaciones debido al espaciado, las variantes de Unicode y los caracteres invisibles. Antes de contrastar los fotogramas, el sistema normaliza el texto de forma controlada, de modo que las comparaciones reflejen cambios visuales reales y no cualquier distorsión propia del reconocimiento óptico de caracteres.

Este proceso de normalización comprende:

- Normalización de Unicode
- Eliminación de marcas de tiempo
- Eliminación de los nombres de los interlocutores
- Eliminación del texto de la interfaz (p. ej., «Los subtítulos traducidos están activados»)
- Simplificación de espacios en blanco
- Recorte de espacios iniciales y finales
- Eliminación de caracteres de ancho cero
- Tratamiento específico por grupos lingüísticos (alfabeto latino frente a caracteres CJK)

De este modo, se garantiza que los «cambios» representen, en la medida de lo posible, rectificaciones reales visibles para el usuario, manteniendo al mismo tiempo la capacidad de identificar variaciones significativas.

El sistema procesa los resultados del reconocimiento óptico de caracteres fotograma a fotograma y analiza la evolución del texto para detectar cuándo se actualizan los subtítulos, con qué frecuencia se rectifican y si presentan parpadeos u oscilaciones.

El sistema genera entonces un informe resumido en formato CSV con los indicadores clave de cada vídeo; por ejemplo, el número de fotogramas activos y el número de cambios detectados (véase el apartado Sistema de puntuación de las evaluaciones automatizadas a continuación). Estos resultados están pensados para facilitar una comparación directa entre plataformas e idiomas, así como entre las distintas condiciones de análisis (con y sin DeepL).

Sistema de puntuación de las evaluaciones automatizadas

Para medir la estabilidad de los subtítulos de forma objetiva, Slator analizó la frecuencia con la que variaban entre un fotograma y el siguiente. Cualquier modificación visible, ya fuera por adición de palabras, cambios en la segmentación o leves parpadeos y oscilaciones, se contabilizó como un cambio. En concreto, contaron como cambios los siguientes fenómenos:

- Adiciones o eliminaciones de caracteres
- Autocompletado de palabras
- Sustituciones leves
- Ajustes de puntuación
- Cambios de formato
- Oscilaciones regresivas de texto
- Inestabilidad del texto y
- Parpadeos perceptibles y disruptivos

La proporción de fotogramas en los que no se produjo ningún cambio visible se convirtió en un índice de estabilidad de 0 a 100, donde 100 corresponde a subtítulos totalmente estables (sin cambios entre fotogramas) y las puntuaciones más bajas reflejan actualizaciones en pantalla más frecuentes.

Al mismo tiempo, se calculó la volatilidad de los subtítulos, entendida como el porcentaje de fotogramas en los que se produjo alguna variación del subtítulo en pantalla. Mientras que el índice de estabilidad destaca la proporción de fotogramas visualmente estables, la volatilidad cuantifica directamente la frecuencia con la que los usuarios perciben rectificaciones o interrupciones en el subtitulado.

En conjunto, el índice de estabilidad y la volatilidad proporcionan una metodología transparente, basada en el comportamiento del sistema, para contrastar el rendimiento de los subtítulos en las distintas plataformas. No miden la precisión lingüística, la calidad de la traducción ni la corrección gramatical o semántica (estos aspectos se evaluaron de forma independiente mediante una revisión lingüística llevada a cabo por expertos).

Datos de estabilidad sin procesar

Los datos sin procesar son los siguientes:

Clasificación de herramientas por estabilidad

Estos datos permitieron a Slator determinar la variación porcentual que supone el uso de DeepL Voice en comparación con la versión estándar de la misma plataforma.

| Herramienta | Índice de estabilidad media (0-100) | Volatilidad |
|----------------------------------|-------------------------------------|-------------|
| DeepL Voice para Zoom | 88,6 | 11,35 % |
| DeepL Voice para Microsoft Teams | 85,8 | 14,20 % |
| Google Meet | 82,5 | 17,50 % |
| Microsoft Teams | 77,3 | 22,75 % |
| Zoom | 74,9 | 25,08 % |

Las mejoras porcentuales se calcularon a partir de la reducción de la volatilidad de los subtítulos al comparar DeepL Voice con las funciones nativas de cada plataforma.

Clasificación de estabilidad por combinación de idiomas

Los índices de estabilidad media por idioma son los siguientes:

| Combinación de idiomas | DeepL Voice para Zoom | DeepL Voice para Microsoft Teams | Google Meet | Microsoft Teams | Zoom |
|------------------------|-----------------------|----------------------------------|-------------|-----------------|------|
| EN-ES | 90,3 | 87,8 | 87,5 | 83,6 | 77,6 |
| EN-FR | 89,1 | 85,7 | 70,5 | 73,1 | 75,4 |
| EN-DE | 86,8 | 87,5 | 77,5 | 76,7 | 73,8 |
| EN-IT | 88,2 | 84,7 | 85,4 | 72,6 | 75,6 |
| EN-PT | 93,2 | 88,2 | 91,4 | 77,2 | 75,3 |
| EN-KO | 93,2 | 88,0 | 80,2 | 61,5 | 74,7 |
| EN-JP | 87,2 | 84,1 | 67,2 | 83,8 | 71,0 |
| ES-EN | 84,9 | 82,0 | 88,7 | 79,5 | 75,2 |
| FR-EN | 88,8 | 84,3 | 89,4 | 78,0 | 76,8 |
| DE-EN | 86,3 | 82,0 | 75,5 | 71,9 | 75,2 |
| IT-EN | 85,0 | 82,5 | 88,4 | 83,0 | 72,7 |
| PT-EN | 88,5 | 88,1 | 78,6 | 73,1 | 74,9 |
| KO-EN | 92,7 | 89,9 | 87,9 | 73,3 | 73,7 |
| JP-EN | 90,5 | 87,5 | 86,7 | 94,4 | 76,9 |
| Media | 88,6 | 85,9 | 82,5 | 77,3 | 74,9 |

La volatilidad –el porcentaje de fotogramas en los que el subtítulo mostrado sufre variaciones, calculado a partir de la comparación de los productos de DeepL Voice y las plataformas habituales – es la siguiente:

| Combinación de idiomas | DeepL Voice para Zoom | DeepL Voice para Microsoft Teams | Google Meet | Microsoft Teams | Zoom | Diferencia entre DeepL Voice para Teams y Teams | Diferencia entre DeepL Voice para Zoom y Zoom |
|------------------------|-----------------------|----------------------------------|-------------|-----------------|---------|---|---|
| EN-ES | 9,72 % | 12,24 % | 12,53 % | 16,38 % | 22,45 % | 25 % | 57 % |
| EN-FR | 10,86 | 14,30 % | 29,45 % | 26,94 % | 24,58 % | 47 % | 56 % |
| EN-DE | 10,43 % | 12,49 % | 22,54 % | 23,31 % | 26,17 % | 46 % | 60 % |
| EN-IT | 13,20 % | 15,35 % | 14,63 % | 27,44 % | 24,40 % | 44 % | 46 % |

| | | | | | | | |
|-------|---------|---------|---------|---------|---------|--------|------|
| EN-PT | 11,82 % | 11,75 % | 8,57 % | 22,85 % | 24,67 % | 49 % | 52 % |
| EN-KO | 6,83 % | 11,96 % | 19,76 % | 38,45 % | 25,28 % | 69 % | 73 % |
| EN-JP | 12,77 % | 15,93 % | 32,78 % | 16,18 % | 29,01 % | 2 % | 56 % |
| ES-EN | 15,05 % | 18,04 % | 11,32 % | 20,46 % | 24,80 % | 12 % | 39 % |
| FR-EN | 11,25 % | 15,73 % | 10,57 % | 21,99 % | 23,16 % | 28 % | 51 % |
| DE-EN | 13,73 % | 17,97 % | 24,53 % | 28,14 % | 24,77 % | 36 % | 45 % |
| IT-EN | 15,00 % | 17,54 % | 11,58 % | 17,04 % | 27,26 % | -3 % | 45 % |
| PT-EN | 11,52 % | 11,92 % | 21,39 % | 26,95 % | 25,07 % | 56 % | 54 % |
| KO-EN | 7,29 % | 10,12 % | 12,11 % | 26,74 % | 26,35 % | 62 % | 72 % |
| JP-EN | 9,47 % | 12,53 % | 13,30 % | 5,57 % | 23,08 % | -125 % | 59 % |
| Media | 11,35 % | 14,13 % | 17,50 % | 22,75 % | 25,08 % | 38 % | 55 % |

En las dos últimas columnas, los porcentajes positivos reflejan la mejora al utilizar productos de DeepL Voice en comparación con las funciones predeterminadas de la plataforma. Los porcentajes negativos representan una mayor estabilidad de serie con respecto a los productos de DeepL Voice.

Conjunto de datos completo

A continuación se detallan los datos íntegros de todas las herramientas e idiomas.

| Herramienta | Combinación de idiomas | Fotogramas activos | Fotogramas por segundo | Total de cambios |
|------------------------|------------------------|--------------------|------------------------|------------------|
| DeepL Voice para Teams | DE-EN | 8987 | 11 | 1615 |
| DeepL Voice para Teams | EN-DE | 7212 | 11 | 901 |
| DeepL Voice para Teams | EN-ES | 7167 | 11 | 877 |
| DeepL Voice para Teams | EN-FR | 7099 | 10 | 1015 |
| DeepL Voice para Teams | EN-IT | 7005 | 10 | 1075 |
| DeepL Voice para Teams | EN-JP | 6957 | 10 | 1108 |
| DeepL Voice para Teams | EN-KO | 7021 | 10 | 840 |
| DeepL Voice para Teams | EN-PT | 7249 | 11 | 852 |
| DeepL Voice para Teams | ES-EN | 7057 | 10 | 1273 |
| DeepL Voice para Teams | FR-EN | 8123 | 11 | 1278 |
| DeepL Voice para Teams | IT-EN | 7214 | 10 | 1265 |
| DeepL Voice para Teams | JP-EN | 5921 | 10 | 742 |
| DeepL Voice para Teams | KO-EN | 5285 | 10 | 535 |
| DeepL Voice para Teams | PT-EN | 6360 | 10 | 758 |
| DeepL Voice para Zoom | DE-EN | 8766 | 11 | 1204 |
| DeepL Voice para Zoom | EN-DE | 7112 | 10 | 742 |
| DeepL Voice para Zoom | EN-ES | 7246 | 11 | 704 |
| DeepL Voice para Zoom | EN-FR | 7192 | 11 | 781 |
| DeepL Voice para Zoom | EN-IT | 7157 | 11 | 945 |
| DeepL Voice para Zoom | EN-JP | 7241 | 11 | 925 |
| DeepL Voice para Zoom | EN-KO | 7272 | 11 | 497 |
| DeepL Voice para Zoom | EN-PT | 7239 | 11 | 856 |

| | | | | |
|-----------------------|-------|------|----|------|
| DeepL Voice para Zoom | ES-EN | 7268 | 11 | 1094 |
| DeepL Voice para Zoom | FR-EN | 7923 | 10 | 891 |
| DeepL Voice para Zoom | IT-EN | 7299 | 10 | 1095 |
| DeepL Voice para Zoom | JP-EN | 6117 | 11 | 579 |
| DeepL Voice para Zoom | KO-EN | 5380 | 11 | 392 |
| DeepL Voice para Zoom | PT-EN | 6258 | 10 | 721 |
| Google Meet | DE-EN | 8512 | 10 | 2088 |
| Google Meet | EN-DE | 6185 | 9 | 1394 |
| Google Meet | EN-ES | 6712 | 10 | 841 |
| Google Meet | EN-FR | 6183 | 9 | 1821 |
| Google Meet | EN-IT | 7141 | 11 | 1045 |
| Google Meet | EN-JP | 6977 | 10 | 2287 |
| Google Meet | EN-KO | 7139 | 11 | 1411 |
| Google Meet | EN-PT | 7097 | 10 | 608 |
| Google Meet | ES-EN | 7237 | 11 | 819 |
| Google Meet | FR-EN | 7783 | 10 | 823 |
| Google Meet | IT-EN | 6917 | 9 | 801 |
| Google Meet | JP-EN | 6022 | 11 | 801 |
| Google Meet | KO-EN | 5045 | 10 | 611 |
| Google Meet | PT-EN | 6260 | 10 | 1339 |
| Microsoft Teams | DE-EN | 8443 | 10 | 2376 |
| Microsoft Teams | EN-DE | 6957 | 10 | 1622 |
| Microsoft Teams | EN-ES | 6793 | 10 | 1113 |
| Microsoft Teams | EN-FR | 7035 | 10 | 1895 |
| Microsoft Teams | EN-IT | 6796 | 10 | 1865 |
| Microsoft Teams | EN-JP | 6731 | 10 | 1089 |
| Microsoft Teams | EN-KO | 7053 | 10 | 2712 |
| Microsoft Teams | EN-PT | 7170 | 11 | 1638 |
| Microsoft Teams | ES-EN | 7141 | 10 | 1461 |
| Microsoft Teams | FR-EN | 7748 | 10 | 1704 |
| Microsoft Teams | IT-EN | 7117 | 10 | 1213 |
| Microsoft Teams | JP-EN | 5923 | 10 | 330 |
| Microsoft Teams | KO-EN | 5071 | 10 | 1356 |
| Microsoft Teams | PT-EN | 6160 | 10 | 1660 |
| Zoom | DE-EN | 7528 | 9 | 1865 |
| Zoom | EN-DE | 6780 | 10 | 1774 |
| Zoom | EN-ES | 7176 | 11 | 1611 |
| Zoom | EN-FR | 7023 | 10 | 1726 |
| Zoom | EN-IT | 7099 | 10 | 1732 |
| Zoom | EN-JP | 7066 | 10 | 2050 |
| Zoom | EN-KO | 6787 | 10 | 1716 |
| Zoom | EN-PT | 6805 | 10 | 1679 |
| Zoom | ES-EN | 6257 | 9 | 1552 |
| Zoom | FR-EN | 6903 | 9 | 1599 |
| Zoom | IT-EN | 6328 | 8 | 1725 |

| | | | | |
|------|-------|------|---|------|
| Zoom | JP-EN | 5207 | 9 | 1202 |
| Zoom | KO-EN | 4619 | 9 | 1217 |
| Zoom | PT-EN | 5504 | 9 | 1380 |

Conclusiones finales

Las conclusiones finales de este informe se han elaborado de forma ajena a DeepL o a cualquier entidad externa. En el marco de este análisis, Slator no ha favorecido a ninguna plataforma de manera artificial o intencionada, con el firme compromiso de ofrecer un estudio completamente objetivo y neutral. Slator desarrolló la metodología de forma independiente y mantuvo en todo momento el pleno control editorial del análisis y los resultados obtenidos.



Acerca de Slator

Slator es la principal fuente de investigación e inteligencia de mercado para los sectores de traducción, localización, interpretación e IA lingüística. El área de asesoramiento de Slator actúa como un socio estratégico para clientes interesados en servicios de fusiones y adquisiciones, además de análisis independientes. Con sede en Zúrich y oficinas en Londres, Slator dispone también de analistas en Asia, Europa y EE. UU.

Equipo del proyecto

FLORIAN FAES

Director general
Slator

Correo electrónico:
florian@slator.com

ALEX EDWARDS

Director de
Consultoría
Slator

Correo electrónico:
alex@slator.com

ROCIO TXABARRIAGA

Analista sénior de
Investigación
Slator

Correo electrónico:
rocio@slator.com