



Universidad de Valladolid

FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN
MÁSTER EN TRADUCCIÓN EN ENTORNOS DIGITALES
MULTLINGÜES

TRABAJO DE FIN DE MÁSTER

Aproximación a dos sistemas de traducción de
voz a voz inglés-español:
un estudio de caso

Presentado por D.^a Sofía Calvo del Barrio

Tutelado por la Dra. María Teresa Ortego Antón

Soria, junio de 2025

ÍNDICE

RESUMEN	1
ABSTRACT.....	1
1. Introducción.....	2
2. Competencias.....	4
3. Objetivos	6
3.1. Objetivo principal	6
3.2. Objetivos secundarios	6
4. Marco teórico.....	7
4.1. Tecnologías de la traducción y de la interpretación	7
4.2. Clasificación de las tecnologías de interpretación.....	8
4.2.1. Sistemas de interpretación asistida por ordenador.....	9
4.2.2. <i>Smart pen</i>	13
4.2.3. <i>Tablet</i>	13
4.2.4. Sistemas de gestión de servicios lingüísticos.....	14
4.2.5. Sistemas de interpretación remota	14
4.2.6. Traducción de voz a voz	15
4.3. Sector agroalimentario	20
5. Metodología	23
5.1. Sistemas de traducción de voz a voz.....	23
5.1.1. Interprefy.....	23
5.1.2. KUDO	24
5.2. Muestras de análisis.....	25
5.2.1. Muestra 1	25
5.2.2. Muestra 2	27

5.3. Modelo de evaluación	28
5.4. Procedimiento de recogida y análisis de datos	34
6. Análisis y resultados.....	36
6.1. Muestra 1	36
6.1.1. Interprefy.....	36
6.1.2. KUDO	39
6.2. Muestra 2	41
6.2.1. Interprefy.....	41
6.2.2. KUDO	44
6.3. Contraste de resultados	47
6.3.1. Interprefy.....	47
6.3.1.1. Muestra 1	47
6.3.1.2. Muestra 2	48
6.3.1.3. Comparativa de muestras en Interprefy	49
6.3.2. KUDO	50
6.3.2.1. Muestra 1	50
6.3.2.2. Muestra 2	51
6.3.2.3. Comparativa de muestras en KUDO.....	52
6.3.3. Contraste de resultados entre sistemas (Interprefy y KUDO)	53
7. Conclusiones	55
8. Bibliografía	58

ÍNDICE DE FIGURAS

Figura 1. Propuesta de revisión de modelo: parámetros lingüísticos.....	30
Figura 2. Propuesta de revisión de modelo: parámetros paralingüísticos.....	31
Figura 3. Gráfico de resultados de la muestra 1 (Interprefy).....	47
Figura 4. Gráfico de resultados de la muestra 2 (Interprefy).....	48
Figura 5. Gráfico para el contraste de resultados de M1 y M2 (Interprefy).	49
Figura 6. Gráfico de resultados de la muestra 1 (KUDO).	50
Figura 7. Gráfico de resultados de la muestra 2 (KUDO).	51
Figura 8. Gráfico para el contraste de resultados de M1 y M2 (KUDO).....	52

ÍNDICE DE IMÁGENES

Imagen 1. Interfaz del Sistema Integrado Voz-Texto para Intérpretes (VIP).....	12
Imagen 2. Secuencias del sistema de traducción de voz a voz en cascada (Fantinuoli, 2023d)	17
Imagen 3. Taxonomía de las tipologías textuales del sector agroalimentario (Policastro Ponce, 2017).	22
Imagen 4. Interfaz del sistema Interprefy.	23
Imagen 5. Interfaz del sistema KUDO.....	24

ÍNDICE DE TABLAS

Tabla 1. Transcripción del discurso original de la muestra 1.	26
Tabla 2. Transcripción del discurso original de la muestra 2.	27
Tabla 3. Propuesta de modelo de análisis de partida (Calvo del Barrio, 2024).	28
Tabla 4. Identificación de errores en la muestra 1 (Interprefy).	36
Tabla 5. Identificación de errores en la muestra 1 (KUDO).	39
Tabla 6. Identificación de errores en la muestra 2 (Interprefy). Fehler! Textmarke nicht definiert.	
Tabla 7. Identificación de errores en la muestra 2 (KUDO).	44

RESUMEN

Los últimos avances en materia de inteligencia artificial (IA) y redes neuronales han tenido una repercusión considerable en los servicios lingüísticos, tanto en la disciplina de traducción, como de interpretación. En el presente Trabajo de Fin de Máster, nos centraremos en la traducción de voz a voz, uno de los avances que suscita un interesante debate sobre la supervivencia de los profesionales humanos y sobre la convivencia entre estos y la tecnología. El objetivo es evaluar la calidad de dos sistemas de traducción de voz a voz mediante el análisis de la traducción al español en forma de producto oral de dos muestras originales en inglés relacionadas con la industria agroalimentaria y de distinto grado de especialización. Para llevar a cabo dicho análisis, revisaremos el baremo propuesto en un estudio previo por Calvo del Barrio y Ortego Antón (en prensa) con el fin de proponer un nuevo modelo de análisis adaptado a las características de los sistemas de traducción de voz a voz. La identificación de errores según dicho modelo, además de proporcionarnos información sobre la prevalencia de cada parámetro, nos permitirá conocer aquellos aspectos sobre los que se esperan mejoras y reflexionar sobre la posibilidad de emplear estos sistemas en situaciones reales.

Palabras clave: inteligencia artificial, reconocimiento de voz automático, traducción automática, conversión de texto a voz, error.

ABSTRACT

Recent developments in the field of Artificial Intelligence (AI) and neuronal networking have had a profound impact on the language service industry, both concerning translation and interpreting disciplines. This research project will focus on speech-to-speech translation, which arises an interesting debate on the survival of human professionals and on the interaction between humans and technology. The aim of the study is to assess the quality of two speech-to-speech translation systems by analysing the oral translation into Spanish of two pieces of audio with a different level of specialization. Both are original in English and deal with the agri-food industry. A model proposed by Calvo del Barrio and Ortego Antón (in press) in a previous project will be reviewed in order to present a new yardstick that will take into consideration the special features of speech-to-speech translation systems. The results will provide us not only with information about the frequency of parameters, but they will also serve as food for thought about the improvements that are required and about the possibilities of employing these systems in real contexts.

Key words: artificial intelligence, automatic speech recognition, machine translation, text-to-speech synthesis, mistake.

1. Introducción

En un mundo en el que la tecnología avanza a una velocidad sin precedentes y juega un papel cada vez más determinante en numerosos campos, la industria de los servicios lingüísticos y los Estudios de Traducción e Interpretación no permanecen en absoluto ajenos a dichos avances, cuyo mayor exponente es, sin lugar a dudas, la inteligencia artificial (IA), siempre dependiente del desarrollo que se produzca en materia de aprendizaje profundo (AP), aprendizaje automático (AA) y procesamiento del lenguaje natural (PLN). Es posible afirmar que el mundo ha sido consciente en mayor medida de los avances en la disciplina de traducción, en particular con los sistemas de traducción automática, que han evolucionado desde sistemas basados en reglas (RBMT, por sus siglas en inglés), pasando por los sistemas de traducción automática estadística (SMT, por sus siglas en inglés) hasta llegar a sistemas neuronales (NMT, por sus siglas en inglés), que funcionan fundamentalmente con base en el aprendizaje automático (AA). Sin embargo, a pesar de que la tecnología comenzó a ocupar un lugar importante en la disciplina de interpretación en el siglo pasado, con el hito del empleo de sistemas alámbricos de transmisión de voz, ha sido recientemente cuando los avances tecnológicos en interpretación han adquirido un mayor interés en la práctica profesional y en la literatura académica. De hecho, la IA ha multiplicado las posibilidades para este tipo de comunicación multilingüe y contamos con diversos dispositivos que ya existían para otros usos, pero que se han aplicado a la disciplina, como el teléfono móvil o los sistemas de videoconferencia, así como otras soluciones lingüísticas creadas de forma específica para cubrir alguna de las necesidades profesionales (Corpas Pastor, 2021, p. 91-97). Ante la velocidad a la que tienen lugar este tipo de avances, resulta difícil tanto para los profesionales y las empresas del sector, como para los investigadores adaptarse a dichos avances al mismo ritmo en el contexto de un mundo globalizado en constante cambio (Davitti *et al.*, 2025, p. 1-3). Una de las propuestas tecnológicas que aúna algunos de los últimos y más significativos avances es la traducción de voz a voz, que es la tecnología de la que nos ocuparemos en el presente trabajo. A pesar de lo lejana que pueda parecer esta tecnología que nos proporciona un producto oral en una lengua distinta a la de la lengua del material oral origen, contamos con ejemplos reales del empleo de la IA para este fin. Uno de los modelos de *smartphone* actuales que ofrece funcionalidades de IA es la línea Samsung Galaxy S24. Entre las funcionalidades se encuentra Chat Assist, que permite traducir en tiempo real todo lo que escribimos e incluso la interpretación simultánea de una llamada telefónica. Asimismo, Google lanzó al mercado unos auriculares inteligentes, Pixel Buds, que cuentan con una función inteligente de intérprete en tiempo real en 40 idiomas (Holgado, 2024). Estos ejemplos, como veremos más adelante, se basan en tecnología de reconocimiento automático de voz, de traducción automática y de conversión de texto a voz, los tres procesos que conforman el modelo de traducción de voz a voz en cascada. En un contexto en el que estas aplicaciones de la IA a la comunicación humana multilingüe están a disposición de los usuarios, consideramos que resulta de vital importancia evaluar la calidad del servicio que proporcionan, para garantizar el correcto desarrollo de los actos de

comunicación en los que se emplean. Desde esta premisa, el objetivo principal de este Trabajo de Fin de Máster es obtener una aproximación de los productos resultantes en dos muestras de diferente grado de especialización en dos sistemas de traducción de voz a voz. Para ello, en primer lugar, en el Capítulo 4, titulado *Marco teórico*, definiremos los conceptos que vertebran este trabajo, con especial atención a la traducción de voz a voz y una breve referencia al sector agroalimentario, dado que la temática de las muestras de análisis guarda relación con esta industria. Posteriormente, en el Capítulo 5, dedicado a la metodología, describiremos las dos muestras originales en inglés seleccionadas, así como los dos sistemas de traducción de voz a voz que emplearemos para obtener los productos resultantes para el posterior análisis: Interprefy y KUDO. En este mismo capítulo, presentaremos una propuesta de modelo de evaluación, en la que revisaremos el baremo propuesto en un estudio previo. Continuaremos con la descripción del análisis y de los resultados (Capítulo 6), que irá seguido de las conclusiones (Capítulo 7) y, por último, de la bibliografía (Capítulo 8).

2. Competencias

En el presente Trabajo de Fin de Máster se han aplicado diferentes conocimientos teóricos y prácticos adquiridos a lo largo del Máster en Traducción en Entornos Digitales Multilingües. Paralelamente, se han puesto en práctica una serie de competencias específicas, todas ellas recogidas en la guía docente para el Trabajo de Fin de Máster del curso 2024-25.

En primer lugar, las competencias de carácter general y transversal que se han puesto en práctica en la realización de este trabajo son:

- G1. Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.
- G2. Aplicar los conocimientos adquiridos y la capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.
- G3. Integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios
- G4. Comunicar conclusiones –y los conocimientos y razones últimas que las sustentan– a públicos especializados y no especializados de un modo claro y sin ambigüedades.
- T1. Planificar y gestionar el tiempo, el estrés y la carga de trabajo.
- T2. Cumplir plazos, ceñirse a instrucciones y especificaciones.

En lo que se refiere a competencias específicas, estas se han ido adquiriendo en las diferentes asignaturas de la titulación y las que se han aplicado a este trabajo en concreto se corresponden con:

- E1. Analizar un documento origen, identificando posibles dificultades textuales y cognitivas y evaluar las estrategias y recursos necesarios para reformular apropiadamente en función de las necesidades comunicativas.
- E2. Resumir, reformular, reestructurar, adaptar, transcribir y acortar rápidamente y con precisión en español y en inglés por escrito.
- E3. Evaluar la relevancia y fiabilidad de fuentes de información de cara a las necesidades de traducción y comunicación en entornos multilingües.
- E4. Adquirir, desarrollar y usar conocimientos específicos relevantes necesarios para la traducción y la comunicación en entornos multilingües (dominar sistemas de conceptos, métodos de razonamiento, estándares de presentación, terminología y fraseología, fuentes especializadas, etc.).

- E5. Aplicar instrucciones, guías de estilo o convenciones relevantes en situaciones de traducción y comunicación en entornos multilingües.
- E7. Analizar y justificar soluciones de traducción y opciones de redacción, usando el metalenguaje apropiado y aplicando propuestas teóricas adecuadas.
- E8. Comprobar y revisar el propio trabajo y el de otros sobre la base de estándares o objetivos de calidad específicos del encargo de traducción o redacción multilingüe.
- E11. Usar las principales herramientas informáticas —incluida toda la gama de programas de ofimática— y adaptarse rápidamente a nuevas herramientas y recursos informáticos
- E12. Hacer un uso efectivo de motores de búsqueda, herramientas basadas en corpus, herramientas de análisis lingüístico y herramientas de traducción asistida por ordenador.
- E13. Manejar tecnologías web.

3. Objetivos

En este capítulo, procedemos a describir el objetivo principal de este Trabajo de Fin de Máster en el primero de los apartados, seguido de los objetivos secundarios, los cuales quedan explicados en el segundo apartado.

3.1. Objetivo principal

- Verificar la calidad de los productos resultantes de los sistemas de traducción de voz a voz del inglés al español.

3.2. Objetivos secundarios

- Elaborar una propuesta de modelo de evaluación lo más precisa y adaptada posible a las características de los sistemas de traducción de voz a voz.
- Detectar errores en los productos resultantes de los sistemas de traducción de voz a voz.
- Obtener una aproximación de la evolución de la aplicación de la inteligencia artificial (IA), el aprendizaje profundo (AP) y el procesamiento del lenguaje natural (PLN) a los servicios lingüísticos.

4. Marco teórico

4.1. Tecnologías de la traducción y de la interpretación

En primer lugar, procedemos a ofrecer una definición para tecnologías de la traducción y una definición para tecnologías de la interpretación. En el sentido más amplio, entendemos tecnologías de la traducción como el gran conjunto de herramientas informáticas que ayudan al traductor en su labor, tales como correctores gramaticales y de escritura, procesadores de texto, herramientas de compilación de corpus, de gestión terminológica, de memorias de traducción o sistemas de traducción automática, entre otras. Es posible establecer un continuo desde las herramientas menos automatizadas que requieren mayor intervención humana, hasta las más automatizadas, como es la traducción automática (O'Brien y Rodríguez Vázquez, 2020, p. 264-265). En este sentido, es fundamental hacer una distinción entre sistemas de traducción asistida por ordenador (TAO) (en inglés, *computer-assisted translation* o CAT) y traducción automática (TA) (en inglés, *machine translation* o MT). Por su parte, la traducción automática no requiere intervención humana y permite procesar un texto original en una determinada lengua y ofrecer una versión en otra lengua. Por otra parte, los sistemas de traducción asistida por ordenador sí necesitan intervención humana, ya que no permiten obtener la traducción de un texto propiamente, sino que asisten al profesional para sumar eficacia a su trabajo, al permitirle crear memorias de traducción o llevar a cabo el proceso de gestión terminológica (Kent State University, s.f.).

Por otra parte, por tecnologías de la interpretación nos referimos a las tecnologías que emplean los intérpretes en cualquiera de las etapas de su labor, ya sea para la memorización de terminología, para posibilitar el propio servicio de interpretación o para la descodificación de información, entre otras funcionalidades. En esta línea, como ocurre con las tecnologías de la traducción, es necesario diferenciar el tipo de funcionalidad que presentan, ya que, como veremos en el siguiente apartado, pueden asistir al intérprete, actuar como soporte para realizar su trabajo o generar directamente el producto de interpretación (Prandi, 2023, p. 27-31). Como consecuencia de la integración de la tecnología, los procesos cognitivos propios de la práctica de interpretación se complementan cada vez más con herramientas tecnológicas, que, a su vez, plantean nuevos desafíos cognitivos, de forma que resulta interesante analizar y afrontar estos retos para mejorar el rendimiento de los profesionales y la formación de los mismos (Chen y Doherty, 2024, p. 411-412).

Tras definir el concepto de tecnologías de la traducción y de la interpretación, se hace necesario puntualizar que, a continuación, atenderemos a la clasificación de las tecnologías de la disciplina de interpretación, por ser la más estrechamente relacionada con el concepto central de este trabajo, la traducción de voz a voz, que se ocupa de trasladar material oral de una lengua a otra, la característica principal de la interpretación.

4.2. Clasificación de las tecnologías de interpretación

En cuanto a la clasificación de las tecnologías de interpretación, es imprescindible aclarar que no existe consenso entre los investigadores respecto a la clasificación y a la definición de las tipologías. Ante las múltiples clasificaciones existentes, en este trabajo nos hemos decantado por la que propone Prandi (2023, p. 25-38) y que procedemos a describir a continuación.

En primer lugar, atenderemos a las herramientas tecnológicas que median en la interpretación (*technology mediating interpreting*) y que ofrecen grandes posibilidades para el intérprete, el orador y los oyentes. Se trata de herramientas que han cambiado los escenarios tradicionales de interpretación o que ha aportado alguna solución dentro de la disciplina. La siguiente categoría corresponde a las tecnologías que dan lugar al propio resultado de la interpretación (*technologies generating interpreting*), marcadas por los avances en materia de inteligencia artificial (IA) y redes neuronales. El ejemplo más claro es la traducción de voz a voz, en la que nos centraremos más adelante. Por último, distinguimos las tecnologías que ofrecen apoyo durante el proceso y en cualquiera de sus etapas (*technologies supporting interpreting*), ya sean *software* o *hardware* (Prandi, 2023, p. 27-31).

Antes de pasar a comentar algunos ejemplos de estas tecnologías en los siguientes apartados, es necesario indicar que, a pesar del modesto impacto que ha tenido hasta ahora el desarrollo tecnológico en la interpretación, se han producido dos avances significativos a lo largo de la historia. En primer lugar, la transmisión de voz por sistemas alámbricos, que se dio a conocer oficialmente en los juicios de Núremberg (1945), cambió para siempre la forma de realizar este trabajo, puesto que, aunque el factor de la simultaneidad no era nuevo y ya venía estando presente en el *chuchotage*, estos sistemas supusieron un valor añadido para el prestigio y el estatus de los profesionales. Sin embargo, resulta interesante mencionar que en aquel momento los intérpretes también entendieron ese cambio tecnológico y el hecho de tener que trasladarse del escenario, que solían compartir con diplomáticos, a las cabinas de interpretación como una pérdida de calidad en su trabajo. Lejos de esta percepción, el cambio supuso la profesionalización de la disciplina. El segundo hito tecnológico que influyó en los servicios lingüísticos fue la irrupción de Internet en los años 90 del siglo pasado. Este gran repositorio multilingüe cambió completamente la forma en la que los profesionales obtenían el conocimiento necesario en la fase de preparación, hasta convertirse en la herramienta más habitual para traductores e intérpretes. Podríamos afirmar que la profesión se encuentra ya en el tercer cambio tecnológico, de una envergadura y un alcance aún por explorar (Fantinuoli, 2023a, p. 46-48). En las últimas dos décadas, hemos sido testigos de una tendencia creciente de digitalización y de tecnologización de los diferentes sectores, incluido el de la interpretación (Stengers *et al.*, 2023, p. 111). A pesar de que esta disciplina no ha sido ajena a los últimos avances en inteligencia artificial (IA), aprendizaje profundo (AP), aprendizaje automático (AA) y procesamiento del lenguaje natural (PLN), la acogida de las herramientas que vienen a facilitar el trabajo de los profesionales y a permitirles adaptarse a un mercado cambiante no se ha producido a la misma velocidad que en la disciplina de traducción y, por

consiguiente, los intérpretes no se han beneficiado de igual forma de estos avances. Las razones principales por las que la tecnología no ha tenido una gran acogida entre los profesionales de la interpretación son el desconocimiento respecto a estos avances por parte de los propios intérpretes, las complejidades que entraña la comunicación humana, la actitud de reticencia de los profesionales con respecto a la tecnología, así como la ausencia de sistemas y herramientas que se ajusten específicamente a las necesidades de los usuarios. Pese a esta realidad y al hecho de que las tecnologías de la interpretación no son un tema central en la literatura académica, la transformación tecnológica se hace evidente en cómo ha evolucionado la forma de trabajar de muchos de los profesionales, como, por ejemplo, con la transición a la modalidad remota o con el empleo de diversos dispositivos. Por otra parte, en los últimos años se ha observado un creciente interés en estos recursos para intérpretes entre los profesionales y los académicos, por ejemplo, dado el aumento en el número de conferencias y cursos (Corpas Pastor, 2021, p. 91-97). Los propios profesionales van adoptando una actitud más positiva, al colaborar incluso con técnicos para desarrollar soluciones tecnológicas para la práctica profesional (Corpas Pastor y Gaber, 2024, p. 125). Asimismo, desde una perspectiva pedagógica, en los años 90 del siglo pasado el ámbito de la docencia se dio cuenta de que los estudiantes de interpretación debían ser conscientes del potencial de la tecnología y desde entonces existen evidencias, aunque no muy frecuentes todavía en la actualidad, de la integración de estos conocimientos en la formación y de la posibilidad de acceso para los estudiantes a este tipo de materiales tecnológicos. Es posible enumerar algunas herramientas que están a disposición de los intérpretes en formación, como son Interpreter's Resources Information System (IRIS), un sistema que permite el acceso a materiales multilingües en diversos formatos (orales y escritos), o Speech Repository, la base de datos de la Comisión Europea en la que se pueden encontrar discursos de distintos organismos internacionales seleccionados por intérpretes (Chen y Doherty, 2024, p. 408-409). En esta línea, resulta interesante mencionar que algunas instituciones también se han interesado recientemente por este tipo de herramientas, como en el caso de la Dirección General de Interpretación de la Comisión Europea, que lanzó una plataforma, The Interpreter's Digital Toolbox Project, que fomenta el trabajo colaborativo de los intérpretes que trabajan para la institución y la automatización del proceso de gestión documental y de consulta terminológica (Corpas Pastor, 2021, p. 91-97). En definitiva, es posible afirmar que nos encontramos en plena transición gradual a lo digital en el campo de los servicios lingüísticos.

A continuación, introduciremos algunos ejemplos que se encuadran en la categoría de *technology mediating interpreting*.

4.2.1. Sistemas de interpretación asistida por ordenador

Se trata de herramientas diseñadas específicamente para apoyar al intérprete en al menos uno de los procesos de su actividad, ya sea en la gestión documental o en el acceso a recursos lexicográficos en tiempo real, entre otros. Del mismo modo que se crearon herramientas que asistían al traductor en los diferentes procesos, el objetivo ha sido crear sistemas similares para intérpretes, que permitan

mantener la calidad del servicio, al mismo tiempo que restan carga cognitiva a los profesionales (Fantinuoli, 2023a, p. 46-48). La forma de trabajar de los traductores previa a la aparición de estos sistemas de traducción asistida por ordenador (TAO) (en inglés, *computer-assisted translation* o CAT) solía llevar más tiempo, implicar mayor cantidad de ciertos errores y dificultar el trabajo colaborativo. Con la llegada de estas herramientas, los traductores pudieron acceder a diversas funcionalidades que tenían por objetivo aumentar la eficacia de su labor, tales como las memorias de traducción, las bases de datos terminológicas o el control de calidad (Falempin y Ranadireska, 2024, p. 178). En el caso de los sistemas de interpretación asistida por ordenador (IAO) (en inglés, *computer-assisted interpreting* o CAI), las funcionalidades y las estructuras también varían considerablemente, sobre todo, con los últimos avances en IA, AP, AA y PLN que se han incorporado en funciones, como el reconocimiento automático de voz, la traducción automática o la modelación del lenguaje (Fantinuoli, 2023a, p. 46-48).

La literatura inicial con respecto a estos sistemas se remonta a la primera década de los 2000 de la mano de los académicos alemanes. Rütten afirmaba que un sistema ideal permitiría al usuario llevar a cabo de forma *online* u *offline* una labor de documentación y gestión de la misma, así como de extracción, gestión y memorización de términos. Por su parte, Stoll destacaba la importancia de que la máquina y el humano pudieran acoplarse adecuadamente. A partir de lo expuesto por estos autores, Fantinuoli presenta en 2011 una nueva generación de IAO que incorpora avances de PLN al introducir la creación de corpus o la extracción terminológica automática. Por último, cabe mencionar, que la tendencia actual es crear plataformas que integren diferentes módulos, entre ellos un módulo de IAO o incluso de interpretación remota (Fantinuoli, 2023a, p. 50-54).

Aunque el empleo de estas herramientas es limitado, conviene destacar aquellas funcionalidades que ofrecen en cada uno de los procesos de la interpretación:

- Preparación: por norma general, existe una brecha de conocimiento entre el público meta, el orador principal y el intérprete, que este último debe subsanar mediante una exhaustiva fase de preparación, que le permita combinar el conocimiento lingüístico y extralingüístico. En este sentido, los sistemas de IAO son una herramienta útil en la gestión de la documentación y en la creación de glosarios (Fantinuoli, 2023a, p. 48-49). La fase de preparación para un encargo es fundamental tanto en traducción, como en interpretación. Sin embargo, esta labor de documentación implica mayores dificultades técnicas para la interpretación. En el caso de los intérpretes, es frecuente que estos recurran a material audiovisual en la fase de preparación para familiarizarse con el acento del orador o conocer sus expresiones más recurrentes, entre otros aspectos. La dificultad para obtener las transcripciones de estos materiales y construir con ellas un corpus, así como la ausencia de herramientas que asistieran a los profesionales en esta labor, complicaban el propio proceso de preparación. No obstante, los sistemas de

reconocimiento automático de voz permiten solucionar estas dificultades técnicas y ayudar al profesional en la práctica. La aplicación de estos sistemas en este ámbito se divide en dos líneas fundamentalmente. Por un lado, como se detallará más adelante, el desarrollo de estos sistemas está encaminado a la traducción de voz a voz. La segunda línea se centra en aquellas soluciones para la práctica profesional o formativa basadas en estos sistemas e integradas en herramientas de IAO. Estas soluciones se refieren tanto al proceso de preparación, a la hora de construir corpus, como al propio proceso de interpretación (Corpas Pastor y Gaber, 2024, p. 123-131).

- Durante la interpretación: considerada como una estrategia de respaldo, estas herramientas permiten consultar la terminología en los descansos, mientras se está llevando a cabo la interpretación o para ayudar al compañero de cabina. Se trata de una función importante, teniendo en cuenta que cuando los plazos son limitados, los intérpretes no tienen tiempo de memorizar toda la terminología que han incluido en sus bases. Por tanto, podemos concluir que el objetivo es reducir la carga cognitiva en una actividad que de por sí supone una gran carga cognitiva. Sin embargo, se valora la posibilidad de que la búsqueda de términos mientras se está interpretando añada y no reduzca carga cognitiva. Por ello, se busca automatizar este proceso por medio de herramientas, como el reconocimiento automático de voz (Fantinuoli, 2023a, p. 49). Como se ha adelantado en el apartado de la fase de preparación, los sistemas de reconocimiento automático de voz integrados en herramientas de IAO pueden ser de gran utilidad a la hora de restar carga cognitiva al profesional cuando este debe lidiar con la traducción de nombres de entidades, de números o de términos complejos (Corpas Pastor y Gaber, 2024, p. 130)
- Después de la interpretación: una vez ha finalizado el ejercicio de interpretación, lo más común es que los intérpretes actualicen sus bases de datos terminológicas, sus repositorios documentales, tomen notas o corrijan traducciones, entre otras tareas. Estas herramientas resultan útiles para automatizar estos procesos o al menos reducir el tiempo que se invierte en ellos. Asimismo, se hace necesario realizar una evaluación de la calidad del servicio, una actividad que suelen desempeñar los propios profesionales. Por el momento, no se conoce ninguna herramienta que realice esta actividad de evaluación de forma automática, precisamente por la ambigüedad que envuelve al concepto de calidad en interpretación (Fantinuoli, 2023a, p. 50).

Uno de los sistemas de IAO de referencia es el Sistema Integrado Voz-Texto para Intérpretes (VIP)¹, desarrollado en la Universidad de Málaga (España). Se trata de un sistema gratuito y de acceso abierto disponible que ofrece al intérprete una serie de herramientas para los distintos procesos de su labor organizadas en un catálogo. Por otra parte, cuenta con cuatro módulos especializados: módulo de corpus, módulo de glosarios, módulo complementario y módulo de entrenamiento. Además de la gestión de corpus monolingües y paralelos y de la posibilidad de importar un corpus de forma manual y extraer la terminología, el módulo de corpus permite seleccionar una lengua y una serie de palabras clave para compilar un corpus de forma (semi)automática. El usuario podrá acceder a los enlaces que le devuelve la herramienta para escoger aquellos que sean de utilidad y poder almacenar el corpus final. Por su parte, el módulo de glosarios permite la consulta, la gestión y la creación de glosarios desde cero o desde la consulta de corpus. Las herramientas de traducción automática, generación de resúmenes, asistente para toma de notas y extracción de entidades nombradas se encuentran en el módulo complementario. Por último, diferenciamos el módulo de entrenamiento, en el que tanto intérpretes profesionales, como en formación pueden realizar ejercicios de traducción a vista, de cifras o de anticipación. Con el desarrollo de las nuevas tecnologías, estos sistemas de IAO se adaptan en mayor medida a las necesidades del sector y de la práctica profesional, al tiempo que observamos que asistentes conversacionales o *chatbots*, como ChatGPT o Gemini, se han convertido en aliados de los profesionales en la fase de preparación. Sin embargo, a pesar de la rapidez y la comodidad que ofrecen estos asistentes, no proporcionan todas las funcionalidades, la precisión, la plenitud ni la adecuación en los resultados que sí ofrecen sistemas como VIP (Lima Florido y Corpas Pastor, 2024, p. 7-29).



Imagen 1. Interfaz del Sistema Integrado Voz-Texto para Intérpretes (VIP).

¹ El Sistema Integrado Voz-Texto para Intérpretes (VIP) está disponible en el siguiente enlace: <https://www.lexytrad.es/VIP/site/?ln=en&ln=es>.

4.2.2. *Smart pen*

El empleo de bolígrafos que permiten escribir a mano y recoger la información de forma digital se remonta a principios de este siglo en el ámbito de la formación en diseño o la formación de profesorado. En ese momento, también se produjo el avance de incorporar la grabación de audio en estos dispositivos. En 2008, llegaría el modelo Pulse, un *smart pen* creado por Livescribe, que presenta una gran capacidad de memoria, así como una pantalla y un micrófono (Orlando, 2023, p. 6-9).

En el ámbito de la interpretación, se han investigado estos dispositivos con el objetivo de evaluar actividades de entrenamiento de toma de notas, como parte de una perspectiva más pedagógica, o para explorar diferentes áreas de la interpretación, como métodos de toma de notas, métodos híbridos o el nivel de carga cognitiva. En 2009, se dotó a este dispositivo de la capacidad de recoger de forma sincronizada lo que escribe el usuario y de grabar lo que se dice en la sala, de forma que, aunque en un principio esta herramienta estaba pensada para aplicarse en ámbitos como la ingeniería, la medicina o la educación, se observó que podría tener cabida en la disciplina de interpretación. De esta manera, se contribuyó a que se pusiera el foco en el proceso de toma de notas y no tanto en las notas del intérprete como producto (Orlando, 2023, p. 6-9).

4.2.3. *Tablet*

Con el aumento de la compra de *tablets*, sobre todo después de que Apple lanzara su primer iPad en 2010, los intérpretes se plantearon el empleo de este dispositivo como herramienta de ayuda en uno o varios aspectos del proceso de interpretación. Entre los usos para la interpretación que tienen las *tablets*, consideradas como un subconjunto dentro de las herramientas IAO, destaca la toma de notas, la posibilidad de grabación de voz o las herramientas de creación de glosarios o de gestión de la documentación. A pesar de que la investigación en relación a estas herramientas aún está en sus inicios y existen muchas incógnitas sobre sus consecuencias en la calidad y en la productividad, entre otros aspectos, contamos con algunos estudios que nos han permitido conocer algunas de las aplicaciones de estos dispositivos y sus ventajas. Además, sabemos que instituciones como las Naciones Unidas, el Parlamento Europeo o la Comisión Europea han incorporado estas herramientas. En la fase de preparación, proceso para el que se suele disponer de un tiempo limitado y que es fundamental para adquirir el conocimiento necesario, las *tablets* son útiles para gestionar los recursos documentales de forma intuitiva, construir glosarios y poder consultarlos y repasarlos cómodamente antes del ejercicio de interpretación. En el caso de la interpretación consecutiva, se pueden emplear para llevar a cabo la toma de notas, al mismo tiempo que se pueden consultar recursos, tener acceso a glosarios, emplear diferentes colores o grosores para las notas y moverse en el documento de forma vertical. Se afirma que, además de ser un recurso ligero, duradero y cómodo, es respetuoso con el medio ambiente y más silencioso que el papel. En los últimos años, con el foco en reducir la carga cognitiva, se ha tratado de automatizar ciertos procesos por medio de la IA, como el de consulta de términos durante la interpretación (Goldsmith, 2023, p. 27-32).

4.2.4. Sistemas de gestión de servicios lingüísticos

Se trata de programas que facilitan el proceso de organizar el trabajo de los intérpretes y de asignar el profesional más adecuado a cada encargo, en función de una serie de factores, como la combinación lingüística, la disponibilidad o el grado de conocimiento del tema. En un primer momento, era el gestor de proyectos quien se encargaba de ello principalmente, mientras que en la actualidad existen sistemas que funcionan con IA y que llevan a cabo esta labor de organización de forma automática. Se presenta como una solución viable cuando nos encontramos ante grandes proyectos de interpretación en los que el número de encargos y de profesionales es considerablemente alto (Fantinuoli, 2023a, p. 65).

Una vez han quedado explicados algunos ejemplos de la categoría de *technologies supporting interpreting*, nos gustaría prestar atención a la interpretación remota, dado que desencadenó el desarrollo de la traducción de voz a voz.

4.2.5. Sistemas de interpretación remota

El concepto de interpretación remota (RI, por sus siglas en inglés), como se ha mencionado anteriormente, corresponde a la categoría de herramientas tecnológicas que intervienen en el proceso de interpretación (*technology mediating interpreting*). Esta tecnología cambia por completo el escenario tradicional de un servicio de interpretación, puesto que elimina la necesidad de que las partes implicadas (intérprete, orador y público) deban estar presentes en el mismo lugar, lo que lo hace más flexible y permite el ahorro de costes y de equipos. Un orador emite su discurso y su imagen y su voz llegan al intérprete a través de una plataforma en la nube diseñada para ello. El intérprete necesita simplemente un ordenador con cámara, un acceso fiable a internet y unos cascos con micrófono para poder emitir la interpretación a los oyentes, que generalmente estarán también conectados a la plataforma (Interprefy, 2022).

Antes de la pandemia del COVID-19, que tuvo lugar desde marzo de 2020, se pensaba que la interpretación de conferencias en su modalidad presencial era algo estable y duradero. A pesar de que la interpretación remota ya existía, fue a raíz de esta emergencia mundial cuando esta nueva forma de trabajar se incorporó rápidamente a la realidad, más rápido incluso de lo que la comunidad de intérpretes se pudo amoldar a la nueva modalidad, que en verano de ese año ya suponía el 100 % de la actividad de interpretación. Los intérpretes tuvieron que trabajar nuevas habilidades a gran velocidad para poder mantener su competitividad en el mercado y no sufrir pérdidas. Por otra parte, no nos gustaría dejar de mencionar que estos acontecimientos tan recientes nos enseñan que es fundamental la capacidad de adaptación y la adquisición de conocimiento y de habilidades respecto a los avances tecnológicos antes de que estos terminen por establecerse. La velocidad a la que avanzó esta tecnología fue tal que los proveedores de estos sistemas también aumentaron considerablemente e incluso la industria se diversificó para ofrecer diferentes servicios a los profesionales (Corpas Pastor y Defranq, 2023, p. 1-3).

Dentro de la disciplina de interpretación, los profesionales de la modalidad de interpretación para los servicios públicos (PSI, por sus siglas en inglés) tuvieron que pasar a llevar a cabo una parte importante de sus encargos de forma remota. Los primeros indicios de PSI en modalidad remota tuvieron lugar por medio del teléfono y algunos de los inconvenientes que surgían eran la calidad del sonido, la falta de contexto visual o la dificultad a la hora de coordinar los turnos de habla. En principio, la interpretación remota en vídeo (VRI, por sus siglas en inglés, *video remote interpreting*) se presenta como una solución para los inconvenientes mencionados. La literatura también nos indica que es necesario que tanto los intérpretes, como los proveedores del servicio de interpretación reciban la formación adecuada para el empleo de estos sistemas. Por otra parte, la literatura se ha encargado de estudiar cómo afecta el cambio a modalidad remota a la percepción de los usuarios con respecto al servicio y a las propias impresiones de los profesionales (Stengers *et al.*, 2023, p. 109-113). En este sentido, es preciso señalar que, desde sus inicios, se han llevado a cabo numerosos experimentos para evaluar esta modalidad y, en muchas ocasiones, los profesionales han coincidido en que el hecho de no estar presentes en la situación comunicativa como tal les provoca una sensación de pérdida de control de la situación, así como estrés, dolor de cabeza, dificultad de concentración y fatiga visual (Ziegler y Gigliobianco, 2018, p. 123-139).

4.2.6. Traducción de voz a voz

Una vez revisado el concepto de RI, nos ocuparemos ahora del aspecto central del presente trabajo: la traducción de voz a voz (S2ST, por sus siglas en inglés). Como se ha explicado anteriormente, pertenece a la categoría de herramientas tecnológicas que generan interpretación (*technologies generating interpreting*), es decir, sistemas que ofrecen el producto final (Prandi, 2023, p. 28-31). Antes de atender a la definición que nos ofrece Fantinuoli para este concepto, cabe indicar que el término que emplearemos de ahora en adelante será «traducción de voz a voz». Tal y como explica Downie (2020, p. 37-38) en *How computers “interpret”*, lo único que el sistema toma en consideración en el modelo de traducción de voz a voz en cascada, el cual explicaremos en este apartado, son las palabras, mientras que el resto de los elementos, como el contexto social, las particularidades de la lengua y otros aspectos paralingüísticos que rodean a las palabras, no se tienen en cuenta. Por tanto, es posible afirmar que los avances y mejoras en estos sistemas dependen en gran medida de los que se produzcan en el campo de la traducción automática. De esta forma, no sería adecuado hablar de interpretación, ya que la manera de proceder descrita no coincide con lo que los humanos consideramos «interpretar» (Downie, 2020, p. 36-38). Así, emplearemos el término «traducción de voz a voz», definida como: «...an automated language translation process that converts spoken content from one language to another in the form of speech» (Fantinuoli, 2023b, p. 10). Esta área emergente de la IA que permite convertir un mensaje oral en otro mensaje oral en otra lengua en tiempo real tiene como objetivo derribar barreras lingüísticas. Este proceso puede tener lugar de forma simultánea al discurso original que se quiere traducir, ya sea en situaciones informales, como una conversación entre amigos que no hablan la misma lengua, o en

contextos más formales, como el de un discurso político. Por otra parte, se puede emplear la traducción de voz a voz para vídeos o material audiovisual que ya ha sido grabado previamente (Fantinuoli, 2023d).

Es posible distinguir al menos dos tendencias en traducción de voz a voz, basándonos en Fantinuoli (2023d):

1. Sistemas en cascada: reciben este nombre porque en esta tendencia tienen lugar tres procesos automáticos en secuencia, que pueden modelarse en programas informáticos:
 - Reconocimiento automático de voz (ASR, por sus siglas en inglés): transcribe el mensaje oral a escrito. Es imprescindible alimentar a estos sistemas de forma continua con datos, puesto que suelen presentar problemas cuando escuchan un elemento que nunca antes han detectado, como puede ser un determinado acento o una expresión idiomática. En este sentido, no hay que perder de vista que la lengua cambia constantemente. Otro de los aspectos cruciales en la mejora del rendimiento de estos sistemas es el empleo de algoritmos predictivos (Downie, 2020, p. 38-39).
 - Traducción automática (MT, por sus siglas en inglés): traduce el mensaje escrito de una lengua a otra. Antes de la llegada de la traducción automática neuronal (NMT, por sus siglas en inglés), diferenciamos dos modelos. Los sistemas de traducción automática basados en reglas (RBMT, por sus siglas en inglés), que a pesar de estar fundados sobre el conocimiento experto de los lingüistas, fallan cuando aparecen ambigüedades o cuando los hablantes prescinden de esas reglas. Por otra parte, los sistemas de traducción automática estadística (SMT, por sus siglas en inglés) funcionan a partir del análisis de patrones lingüísticos y de estadística. La mayor dificultad a la que se enfrentan son la creatividad y la variedad que va implícita en la lengua. Los sistemas NMT actuales funcionan a partir de redes neuronales con neuronas artificiales que realizan cálculos y, en base al AA, el sistema aprende a hacer mejor una tarea determinada, al combinar patrones lingüísticos en función de los resultados obtenidos previamente. Como hemos explicado al principio de este apartado, el inconveniente de estos sistemas es que solo consideran el cotexto de las palabras y no el contexto de las mismas. Por otra parte, dado que funcionan con algoritmos matemáticos, resulta difícil que puedan aprender en el marco de la traducción de voz a voz (Downie, 2020, p. 39-46). Además de los sistemas NMT, los asistentes conversacionales de IA, como ChatGPT, son capaces de resolver encargos de traducción en segundos con una calidad sorprendente, a partir de un entrenamiento en dos fases y el acceso a gigantes modelos de lenguaje. Sin embargo, sobre todo, en el caso de los textos más especializados o más complejos, estos sistemas basados en IA no son capaces

de resolver ciertos aspectos relacionados con las culturas, los matices o las expresiones idiomáticas (Falempin y Ranadireska, 2024, p. 179).

- Conversión de texto a voz (TTS, por sus siglas en inglés): ofrece una versión oral de la traducción escrita. A partir de lo que la traducción automática genera como resultado en forma de texto escrito, el sistema de conversión de texto a voz asocia ese resultado a grabaciones de palabras, fonemas o incluso frases enteras y lo combina de la mejor forma posible y con la mejor entonación posible para obtener un producto final oral. El objetivo es que este producto final sea claro y natural y que se asemeje a la voz humana. Sin embargo, uno de los principales obstáculos para que estos sistemas alcancen dicho objetivo es que, al funcionar a partir de material escrito y, como consecuencia, perder rasgos de oralidad que albergan una gran carga de significado, se pueden producir gran cantidad de ambigüedades (Downie, 2020, p. 48-49).

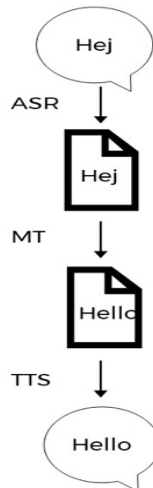


Imagen 2. Secuencias del sistema de traducción de voz a voz en cascada (Fantinuoli, 2023d).

No podemos dejar de destacar que estos sistemas funcionan en base a las tecnologías disponibles, de forma que los nuevos avances se van incorporando a la secuencia. Sin embargo, implican algunas deficiencias, ya que, de producirse algún error en alguno de los procesos, este se arrastra en los siguientes hasta el propio producto final (Fantinuoli, 2023b, p. 10). Por ejemplo, en caso de que el sistema de reconocimiento automático de voz cometa un error al no identificar correctamente los homófonos, la transcripción incorrecta se alargará hasta el proceso de traducción automática y, por tanto, se verá reflejado en el producto final. Por otra parte, el empleo de grandes modelos de traducción basados en recursos escritos significa que tienen dificultad a la hora de identificar y resolver los fenómenos propios del lenguaje oral (Fantinuoli, 2023d).

2. *End-to-end approach*: a diferencia del modelo anterior (en cascada), este traslada el mensaje oral en una lengua a un mensaje oral en otra lengua sin necesidad de transcripción, al aplicarse las mismas técnicas de AA propias de la traducción automática a los datos orales bilingües. Por el momento, se trata de un método experimental en el que los resultados obtenidos hasta ahora son inferiores a los del modelo anterior (Fantinuoli, 2023d). Esta tendencia consiste en una gran red neuronal unificada que elimina la necesidad de trabajar con elementos independientes que, a su vez, necesitan un entrenamiento determinado e independiente. De esta forma, se logra un proceso más rápido con un menor riesgo de errores acumulativos y una mejor gestión de los fenómenos interdependientes del lenguaje humano. No obstante, se requiere una gran potencia de cálculo para estos sistemas y grandes corpus de entrenamiento (Cattoni *et al.*, 2021, p. 1). Los avances de estos sistemas directos de traducción de voz a voz son los que encontramos en el desarrollado por Google en 2019, Translatotron, y en la investigación publicada por Meta AI en 2022 sobre su propio sistema (Corpas Pastor y Gaber, 2024, p. 130).

Por último, cabe mencionar la existencia de un modelo híbrido en el que se combina un módulo de traducción de material oral a texto y otro de conversión de texto a voz. En cualquier caso, a pesar de los rápidos avances, la traducción de voz a voz se enfrenta a los infinitos retos de la comunicación oral multilingüe y por el momento estos sistemas permanecen en el plano lingüístico, es decir, trabajan en la superficie de la comunicación (Fantinuoli, 2023d).

No sería posible abordar de forma completa estos sistemas sin explicar brevemente los siguientes conceptos, ya mencionados previamente en este trabajo: procesamiento del lenguaje natural (PLN) o *Natural Language Processing* (NLP, por sus siglas en inglés) y aprendizaje automático (AA) o *Machine Learning* (ML, por sus siglas en inglés). Con PLN nos referimos a la disciplina de las ciencias computacionales que automatiza el procesamiento del lenguaje natural con el objetivo de que la máquina comprenda el lenguaje humano en textos escritos y orales para diversos fines. A pesar de que estas herramientas existen desde hace unos 50 años, en aplicaciones como los correctores automáticos, actualmente se nutren de grandes cantidades de datos lingüísticos que dan lugar a modelos de lenguaje generales. Estos modelos representan el lenguaje mediante operaciones matemáticas y permiten llevar a cabo tareas más complejas, como el reconocimiento automático de voz o la traducción automática (Fantinuoli, 2023f). El PLN se basa en los avances en AA, un área perteneciente a la IA que permite que una máquina lleve a cabo una determinada tarea a partir de una serie de datos y algoritmos. En lugar de funcionar de acuerdo con reglas, la máquina es capaz de aprender a generar resultados mediante un entrenamiento en el que está expuesto a gran cantidad de datos, por ejemplo, con materiales de audio y sus correspondientes transcripciones. Se trata de un elemento fundamental para los modelos de traducción automática actuales (Fantinuoli, 2023e). El entrenamiento con datos es, precisamente, uno de los aspectos que contribuye al denominado *machine bias*, que se refiere al sesgo que puede

observarse en estos sistemas que funcionan con AA. En el caso de los humanos, con sesgo nos referimos a la preferencia, consciente o inconsciente, que mostramos hacia determinados individuos o grupos, influidos por estereotipos y prejuicios que condicionan nuestras decisiones. Este sesgo humano se traslada a los sistemas de IA y, en ocasiones, puede desembocar en discriminaciones y desigualdades. Como se ha adelantado anteriormente, el sesgo viene dado tanto por los algoritmos, como por los datos con los que se entrenan estos sistemas, que cuentan con el sesgo de quienes los generan y carecen, además, de la propia diversidad de los datos, al no contar con la inmensa totalidad de los mismos (Fantinuoli, 2023c).

Ante el surgimiento de la traducción de voz a voz, es inevitable que tenga lugar un debate sobre la supervivencia de los intérpretes y los traductores humanos. En la actualidad, es posible diferenciar diversas posturas sobre el empleo de la traducción de voz a voz, así como de diferentes recursos tecnológicos. Una de las posturas más fatalistas asume que el empleo de la tecnología en general y de la traducción de voz a voz en particular supone una amenaza que, además de desprofesionalizar la disciplina y afectar negativamente a la calidad del servicio, podría terminar por sustituir a los profesionales humanos (Ahmed Ahmed, 2022, p. 327-329). Sin embargo, hay quien es capaz de identificar otro tipo de implicaciones de la traducción de voz a voz más allá de la posible pérdida de empleos. El hecho de que los sistemas automáticos tomen el relevo de la interpretación humana significará que todos los datos que brinda la comunicación interlingüística estarán en manos de los gigantes de los macrodatos y del AA (Downie, 2020, p. 71-75). Los profesionales humanos se rigen por un código deontológico que, entre otros aspectos, establece que deben garantizar la confidencialidad de los datos a los que tienen acceso al prestar sus servicios. Sin embargo, la IA y el hecho de que estos sistemas se entrenen con gran cantidad de datos implica un cierto riesgo para la privacidad y la seguridad de esos datos (Eby *et al.*, 2024, p. 2). En este sentido, hoy en día ya es posible hablar del primer marco jurídico sobre IA, la Ley de IA, que entró en vigor en la Unión Europea el 1 de agosto de 2024 y que establece cuatro niveles de riesgo diferentes (Comisión Europea, 2024). Por otro lado, diferenciamos la postura de quienes defienden la colaboración entre este tipo de herramientas y los profesionales humanos como una ventaja (Ahmed Ahmed, 2022, p. 327-329). En esta línea, surge un enfoque conocido como *human-in-the-loop* (HITL, por sus siglas en inglés), que combina la eficacia que puede brindar el empleo de la IA con el conocimiento y las competencias indispensables del profesional humano. Esta combinación estratégica consta de cinco procesos: el humano recopila datos e información y los prepara para la siguiente fase de entrenamiento del sistema, posteriormente analiza su rendimiento y proporciona información para ajustar el sistema y una vez ha vuelto a observar su rendimiento, el humano realiza los ajustes necesarios. Por tanto, la IA proporciona una primera versión de la tarea que el humano debe revisar, corregir y refinar, para que el resultado final tenga la calidad esperada. Este enfoque colaborativo cobra gran sentido en el ámbito de los servicios lingüísticos, pues es fundamental que los profesionales humanos intervengan con su conocimiento y con sus competencias

para proporcionar retroalimentación al sistema y pulir aquellos aspectos que la IA no es capaz de identificar o solucionar, como los relacionados con la cultura o con la idiosincrasia de cada lengua (Falempin y Ranadireska, 2024, p. 180-181).

En cualquiera de los casos, citando las palabras de Downie (2020, p. 7), nos parece inteligente adoptar la siguiente actitud: «predicting the future is a fool's game; preparing for the future certainly is not».

4.3. Sector agroalimentario

El presente estudio está estrechamente relacionado con el sector agroalimentario, puesto que la temática de las muestras de análisis, en las que nos centraremos más adelante, se encuadra en esta industria, que se caracteriza por ser verdaderamente amplia, al englobar tanto los sectores sobre los que se fundamenta, como los que genera como consecuencia de su actividad. Por un lado, está formado por dos macrosectores, como son el sector primario, entendido como el conjunto de actividades que forman la agricultura, la ganadería, la silvicultura y la pesca, y el sector secundario, que consiste en la transformación de las materias primas. Por otro lado, el sector agroalimentario está fuertemente vinculado a otras áreas, como el medioambiente, la salud, la sociología, la cultura, la gastronomía o el turismo, entre otras (Rodríguez Rodríguez, 2013, p. 156-158). De hecho, en el caso de este trabajo, es posible observar cómo se relaciona el sector agroalimentario con temas de salud, un asunto muy recurrente en los medios de comunicación hoy en día.

La relevancia y el volumen de este sector en el escenario internacional es tal que existen diversos organismos reguladores asociados a él, como la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, por sus siglas en inglés) o la Organización Mundial de la Salud (OMS), que surgieron ya en el siglo pasado. A nivel europeo, nació en 2002 la Autoridad Europea de Seguridad Alimentaria (EFSA) con la finalidad de proporcionar asesoramiento para las legislaciones y las políticas europeas sobre los riesgos alimentarios. En España, contamos desde 2001 con la Agencia Española de Seguridad Alimentaria y Nutrición (AESAN), además de ministerios específicos para la materia, como es el caso del actual Ministerio de Agricultura, Pesca y Alimentación (Rodríguez Rodríguez, 2013, p. 156-158). Centrados ya en el territorio nacional, el sector agroalimentario, formado en su mayoría por pequeñas y medianas empresas, es clave en la economía española, al suponer el 17 % del sector industrial del país y estar conformado por casi medio millón de trabajadores en unas 30 000 empresas, que cada vez más apuestan por la digitalización y la sostenibilidad. Además, la cuarta posición de España como potencia agroalimentaria en Europa dota a este sector de gran prestigio internacional. El pilar fundamental que fomenta el crecimiento del sector y hace que sea fuerte y sólido es la internacionalización (ICEX, s.f.). En 2023, el valor de las exportaciones fue de 72, 431 millones de euros, un 3 % más que en 2022. El 67,4 % de estas exportaciones se dirigió a países de la UE, como Francia o Alemania, mientras que el 32,6 % correspondió a países extracomunitarios, como Reino

Unido o EE. UU (Ministerio de Agricultura, Pesca y Alimentación, 2024). Este carácter internacional del sector agroalimentario genera una gran variedad de documentos que, en muchas ocasiones, es necesario traducir, así como situaciones en las que se requiere algún otro servicio lingüístico, que resulta fundamental para hacer posible la internacionalización de la industria (Rodríguez Rodríguez, 2013, p. 157).

Como ocurre con otras áreas de traducción especializada, en el caso del sector agroalimentario, podemos considerar como un lenguaje de especialidad, dado que está marcado por la presencia de un campo temático y, por ende, de una terminología para denominar conceptos de un determinado campo del saber. En consecuencia, resulta indispensable que el traductor y el intérprete conozcan las convenciones textuales de los distintos géneros en cada lengua, cuenten con conocimientos temáticos y terminológicos del campo en cuestión y, por supuesto, presenten la competencia necesaria para documentarse. A su vez, la capacidad de documentarse le permitirá al profesional cubrir aquellas lagunas de conocimiento a nivel de convenciones textuales, de campo temático y de terminología. De hecho, traductores e intérpretes tienen a su disposición una serie de bases de datos encuadradas en este sector y creadas por algunas de las instituciones mencionadas en el párrafo anterior, como AGROVOC², el tesoro de la FAO que cuenta con más de 40 000 conceptos en 21 idiomas diferentes. La FAO también ofrece un Depósito de Documentos³ que es de gran utilidad a nivel terminológico, por ejemplo, a la hora de extraer concordancias monolingües o bilingües con la ayuda de herramientas, como Sketch Engine. Este repositorio documental también resulta útil para documentarse a nivel de convenciones textuales (Rodríguez Rodríguez, 2013, p. 159-163). Precisamente, dada la gran diversidad de textos que podemos encontrar en esta industria, es fundamental que los profesionales sepan identificar las características de los distintos géneros textuales en cada lengua y sistematizar de forma adecuada los diferentes textos para poder tomar las decisiones lingüísticas y traductológicas adecuadas (Policastro Ponce, 2017, p. 83). En este punto, procedemos a mostrar la taxonomía que propone Policastro Ponce (2017, p. 85-86), ya que nos parece que recoge la gran variedad de textos relacionados con este sector de forma muy clara:

Tipo de texto	Grados y campo de especialización (Hoffmann 1998: 56-69)	Ejemplos documentales
Patentes	Nivel de abstracción: más elevado Campo: científico, técnico, físico, químico, etc.	Patentes para la protección de nuevos productos del sector agroalimentario, nuevas composiciones.

² AGROVOC está disponible en el siguiente enlace: <https://www.fao.org/agrovoc/es>.

³ El Depósito de Documentos de la FAO está disponible en <https://openknowledge.fao.org/home>

Documentación mercantil y comercial	Nivel de abstracción: elevado Campo: económico, comercial, aduanero, tributario, etc. (ICEX, 2017)	Documentos comerciales (facturas, certificados de origen, cuadernos ATA y CPD); documentos de transporte (seguro de transporte, conocimiento de embarque marítimo/aéreo), cuaderno TIR, carta de porte por carretera CMR, carta de porte ferroviario); documentos aduaneros y tributarios (partida arancelaria, censo VIES, certificado de circulación, DUA)
Normativa y legislación	Nivel de abstracción: muy elevado Campo: Legislación en materia alimentaria, económica, comercial, aduanera, etiquetado, etc.	Codex Alimentarius, normativa europea, normativa nacional, organismos o instituciones reguladoras, etc.
Etiquetado	Nivel de abstracción: muy bajo Campo: científico, técnico, nutricional, publicitario, etc.	Información nutricional, declaraciones saludables, propiedades del producto, información persuasiva, publicidad de la empresa distribuidora o productora, elementos multimodales.
Hostelería y restauración	Nivel de abstracción: muy bajo Campo: gastronómico, sector servicios, etc.	Cartas de restaurantes o menús, folletos informativos.
Documentación divulgativa	Nivel de abstracción: muy bajo Campo: publicitario, periodístico, interés general	Revistas de divulgación, reportajes, anuncios publicitarios, publicaciones virtuales periódicas (blogs, newsletter, etc.)

Imagen 3. Taxonomía de las tipologías textuales del sector agroalimentario (Policastro Ponce, 2017).

A pesar de que años atrás la traducción del sector agroalimentario no gozaba de la misma popularidad que otros sectores, como el de la medicina o la economía, seguramente por verse encuadrada como un subcampo del sector de la hostelería y la restauración, hemos podido observar un reciente interés en este sector en los Estudios de Traducción e Interpretación. En el caso concreto de la Traducción, es posible destacar diversos grupos de investigación, que incluso se centran en productos concretos dentro del sector agroalimentario, como es el GIR TRADUVINO de la Universidad de Valladolid, centrado en el sector vitivinícola, o los trabajos centrados en el aceite de oliva y su traducción a diferentes lenguas de la Universidad de Jaén, cuyo resultado más notable es el *Diccionario de Términos de Aceite de Oliva (DTAO) (español, inglés, chino)* de Mercedes Roldán Vendrell. Asimismo, es necesario destacar las aportaciones del grupo de investigación Ciencia y Traducción de la Universidad de Córdoba. Por su parte, el grupo interuniversitario ACTRES ha desarrollado una serie de programas semiautomáticos basados en PLN para el trasvase lingüístico de textos agroalimentarios de inglés a español, como son BiTexCook o GDQ (Ortego Antón, 2019, p. 25-28). Por el contrario, en el caso de la interpretación, son más escasos los estudios que se centran en este sector tan amplio, de ahí el interés de abordarlo en el presente trabajo.

Una vez abordados los conceptos que vertebran este Trabajo de Fin de Máster: la traducción de voz a voz y las particularidades de la traducción en el ámbito agroalimentario, expondremos la metodología en el siguiente capítulo.

5. Metodología

En este capítulo realizaremos una descripción de los dos sistemas de traducción de voz a voz seleccionados y de las muestras que serán objeto de análisis, así como de los criterios de evaluación de la interpretación del inglés al español realizada por los sistemas. Asimismo, detallaremos el procedimiento de recogida y análisis de datos.

5.1. Sistemas de traducción de voz a voz

Para este estudio, hemos recurrido a dos sistemas de traducción de voz a voz, que nos servirán para extraer el producto resultante que someteremos a análisis: Interprefy y Kudo.

5.1.1. Interprefy

Uno de los *softwares* seleccionados es Interprefy, una plataforma que tiene por objetivo derribar barreras lingüísticas, al ofrecer al usuario un servicio de interpretación remota, de subtitulado en tiempo real en más de 80 idiomas y de traducción de voz a voz con IA. Desde su creación en 2015, Interprefy ha prestado sus servicios a empresas como la gigante alemana Siemens o la biofarmacéutica británica GSK Glaxosmithkline, a instituciones como la Casa Blanca y en eventos como la UEFA Euro 2020. El módulo que emplearemos en este caso es el de traducción de voz a voz, que se presenta como un servicio de calidad y puntero a nivel tecnológico que ofrece traducciones multilingües en tiempo real y que está disponible para cualquier urgencia y para cualquier contexto (Interprefy, s.f.). Por último, es necesario indicar que, dado que se trata del único sistema de estas características del que la Facultad de Traducción e Interpretación tiene licencias, nos hemos decantado por este *software*.

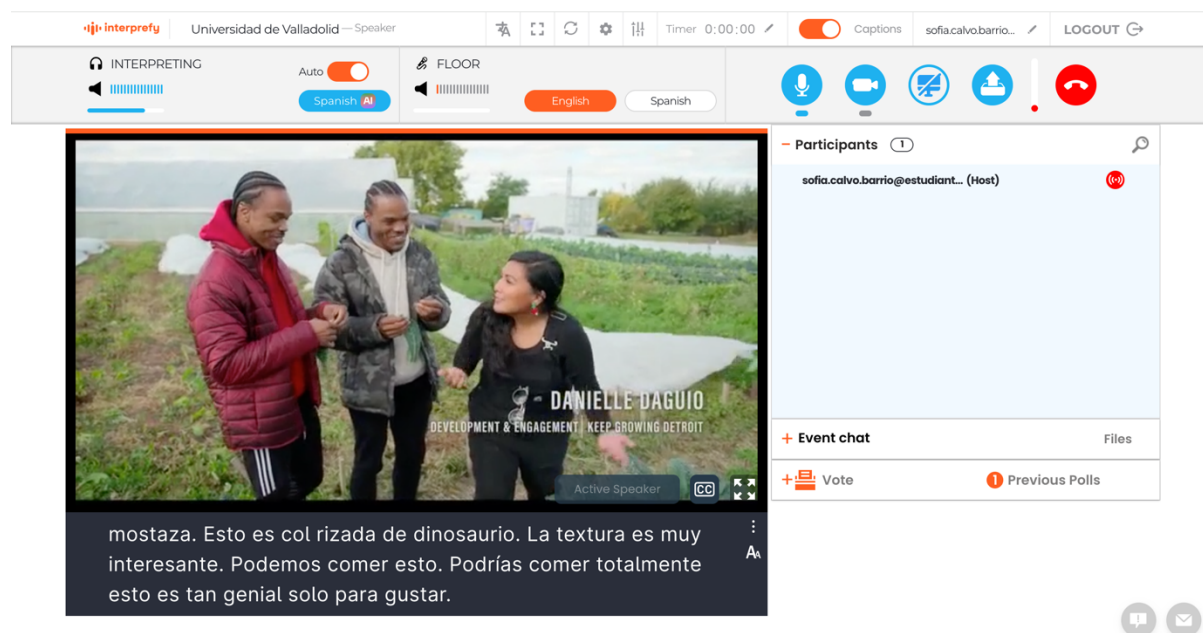


Imagen 4. Interfaz del sistema Interprefy.

5.1.2. KUDO

El siguiente *software* del que nos serviremos para extraer la interpretación de las muestras de análisis es KUDO, que permite emplear el servicio de traducción de voz a voz con IA en diferentes dispositivos en más de 60 idiomas, incluidas algunas variantes regionales. Además de contar con un equipo de más de 12 000 intérpretes profesionales que trabajan en más de 200 lenguas, ofrece un servicio de subtítulo automático con IA en diferentes idiomas, así como la posibilidad de descargar y conservar la grabación de la traducción de voz a voz. Disponible 24 horas los 7 días de la semana para cualquier contexto, desde KUDO se reconoce la importancia y la necesidad de este tipo de servicios de interpretación, al afirmar que el 80 % de los trabajadores presentan un mayor rendimiento cuando tienen acceso a su lengua materna en el trabajo y que el 59 % de las personas que trabajan con una lengua distinta a la suya muestran preocupación por estar perdiendo parte de la información (KUDO, s.f.). Con el objetivo de obtener una visión más global y unas conclusiones lo más sólidas posible, hemos decidido contar con este segundo sistema.

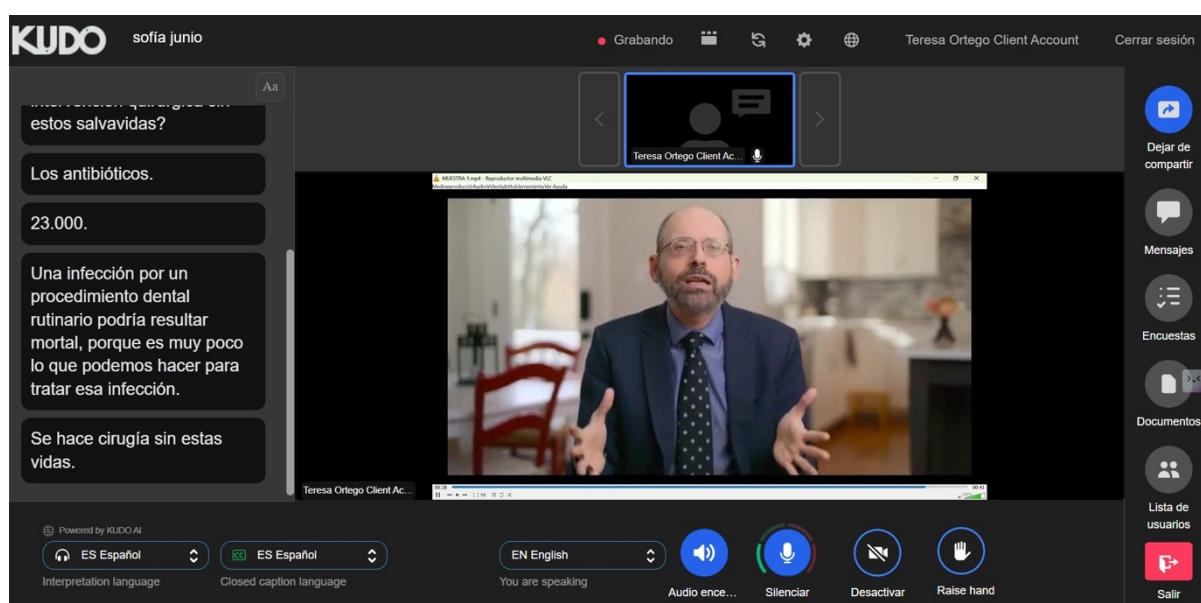


Imagen 5. Interfaz del sistema KUDO.

5.2. Muestras de análisis

Las muestras de análisis están constituidas por dos fragmentos de audio en lengua inglesa (versión original) que han sido extraídas del episodio número tres de la serie documental de cuatro episodios *Somos lo que comemos*. Emitida desde 2023 en la conocida plataforma de *streaming* Netflix, esta serie documental da a conocer varias parejas de gemelos idénticos que cambian su dieta y su estilo de vida para participar en un estudio científico que pretende examinar cómo influye la alimentación en la salud. Desde las premisas de que cada cuerpo puede reaccionar de forma diferente a un mismo patrón dietético y de que los gemelos comparten las mismas características genéticas, cada una de las 22 parejas de gemelos idénticos deberá seguir durante ocho semanas una de las dos dietas saludables propuestas. Además, esta miniserie invita a reflexionar sobre los modos actuales de producción de alimentos y cómo estos afectan a nuestra salud y al medio ambiente (Netflix, 2024). Tal y como hemos explicado en el apartado 4.3. *Sector agroalimentario*, este estudio está relacionado con el sector agroalimentario, una industria verdaderamente amplia y vinculada a numerosas otras áreas. En este caso, la temática de las muestras de análisis, que presentan grados diferentes de especialización, nos hace centrarnos más concretamente en el sector primario (agricultura y ganadería) y en la alimentación humana relacionada con la salud.

5.2.1. Muestra 1

Como muestra de audio con mayor grado de especialización (Muestra 1), se ha seleccionado el fragmento desde 8'02'' hasta 8'42'' (40 segundos de duración). Esta primera muestra está precedida por una serie de intervenciones sobre prácticas ganaderas actuales, con especial hincapié en el empleo de antibióticos en la alimentación de los animales para conseguir un crecimiento más rápido. En la propia muestra, se puede escuchar un reportaje de noticias de ABC News sobre las consecuencias en humanos de la resistencia a los antibióticos. Esta primera parte de la muestra con voz en off sirve como introducción para las palabras de una abogada miembro de Natural Resources Defense Council (NRDC), una ONG ubicada en EE. UU. que lucha por la protección del medio ambiente, y de un médico y escritor especializado en nutrición, seguridad alimentaria y salud pública. Ambos explican el peligro de la resistencia a los antibióticos a la hora de tratar determinadas enfermedades, que aparentemente no son demasiado graves. La presencia de términos especializados y el tono ligeramente formal nos han hecho decantarnos por este fragmento del episodio.

EN
<i>They now have strong evidence of a dangerous strain of E. Coli, that is now highly resistant to antibiotic treatment.</i>
<i>At least 2 million Americans get antibiotic resistant infections each year, with 23,000 deaths.</i>
<i>We are continuing to barrel towards a future where a cut on your hand or an infection from a routine dental procedure could prove to be deadly because there is very little that we can do to treat that infection.</i>
<i>How can you do surgery without these life saving critical drugs that are being squandered just to make cheaper meat?</i>

Tabla 1. Transcripción del discurso original de la muestra 1.

5.2.2. Muestra 2

La segunda de las muestras (Muestra 2) se corresponde con el fragmento desde 20'10'' hasta 20'52'' (42 segundos de duración) y presenta un menor grado de especialización. Esta parte del episodio se centra en una de las parejas de gemelos de las 22 que protagonizan esta miniserie. Tras aprender a elaborar recetas veganas para una dieta basada en vegetales, en este fragmento podemos ver cómo realizan una visita a un huerto en Detroit y mantienen una conversación distendida con una persona que les va mostrando los distintos cultivos. Con el objetivo de contagiarles los beneficios de cultivar sus propios alimentos, esta responsable de desarrollo y participación del huerto les permite probar algunos de los cultivos recién recogidos. Se trata de una conversación coloquial con vocabulario cotidiano, en su mayoría relacionado con el mundo de los alimentos.

EN
“The first time I've ever tried a tomato that I grew myself I was like ‘wait, this is a whole other tomato, like it tastes different’.” “Right.”
“This is all poblano peppers.”
“Let's try some beans.” “How do you eat these?” “It's just like eat it.” “The whole thing, you could eat it?” “Yes, you could eat it, except for like you know the end.” “It's just like fresh, you know.” “Yes, it's very fresh.”
“Those are all mustard greens. This is dinosaur kale.” “The texture is so interesting.” “We could eat this?” “You could totally eat this.” “This is so cool just to like, just take.” “It's like, this is like you're shopping through the grocery store.”

Tabla 2. Tabla 2. Transcripción del discurso original de la muestra 2.

5.3. Modelo de evaluación

A la hora de elaborar el modelo de evaluación de sistemas de traducción de voz a voz que emplearemos para analizar la interpretación al español de las dos muestras de análisis descritas en el apartado anterior, hemos revisado y refinado el modelo presentado en Calvo del Barrio y Ortego Antón (en prensa). Para dicha tarea, nos fijaremos, asimismo, en la revisión del mismo modelo propuesta por Ortego Antón (2025). Dichos baremos toman en consideración los modelos propuestos por Ortiz Boix (2016) y por Collado Aís (2007). En primer lugar, procederemos a mostrar el modelo de partida propuesto en un estudio previo (Calvo del Barrio, 2024) y explicaremos brevemente los parámetros que lo conforman, para finalmente presentar y justificar con detalle la propuesta de revisión de dicho modelo.

Aspectos lingüísticos		Aspectos paralingüísticos
Terminología		Entonación
Traducción incorrecta	Demasiado literal	Fluidez
	Calco	Pronunciación
	Falso sentido	Latencia de habla
	Cifra	
	Nombre propio	
	Traducción innecesaria	
Omisión		
Adición innecesaria		
Elemento no traducido		
Uso incorrecto de la gramática		
Registro o estilo inadecuado a la situación comunicativa		

Tabla 3. Propuesta de modelo de análisis de partida (Calvo del Barrio, 2024).

La propuesta de modelo en la que nos basaremos para llevar a cabo la revisión recoge un total de 17 parámetros, que quedan divididos en aspectos lingüísticos y paralingüísticos y organizados según un código de colores, que facilitó la labor de identificación de errores y su posterior análisis. Como hemos puntualizado en el párrafo anterior, se tomaron dos baremos de distinta naturaleza para diseñar este modelo. Por su parte, Ortiz Boix (2016) sugiere en su tesis doctoral *Implementing machine translation and post-editing to the translation of wildlife documentaries through voice-over and off-screen dubbing* una serie de parámetros orientados a la evaluación de traducciones realizadas con traducción automática que ha adaptado de la clasificación de errores propuesta por MQM⁴. Dado que

⁴ Dicha clasificación ha sido revisada y actualmente contamos con una nueva métrica del modelo de MQM de tipología de errores disponible en <https://themqm.org/error-types-2/typology/>.

la traducción automática es uno de los tres escenarios presentes en los modelos en cascada, resulta lógico tomar en consideración algunos aspectos de este baremo. La totalidad de los parámetros que se tomaron o que se adaptaron a partir de la propuesta de Ortiz Boix (2016, p 63-64) quedaron ubicados en el apartado de aspectos lingüísticos. Una de las adaptaciones más significativas tiene que ver con el parámetro de calco, ya que en el modelo original recibe del nombre de *false friend*, un anglicismo que sirve para referirse a aquellas palabras a las que, por su parecido ortográfico a otra palabra de una lengua diferente, se les atribuye un significado erróneo. Sin embargo, se decidió ampliar esta categoría a calco con el objetivo de abarcar un mayor número de elementos más allá de los heterosemánticos, como son los calcos de estructuras sintácticas. Por otra parte, se optó por mantener los parámetros de cifra y nombre propio, por ser aspectos que presentan gran dificultad a la hora de interpretar, y se eliminó el parámetro de conversión de unidades de medida, dado el rasgo de inmediatez que hace verdaderamente complicado que un intérprete humano o un sistema de traducción de voz a voz sean capaces de adaptar las unidades de medida a la situación comunicativa, pues esto añade una carga cognitiva excesiva. Por otra parte, el modelo de Collados Aís (2007), que aparece recogido en su obra *La evaluación de la calidad en interpretación simultánea. La importancia de la comunicación no verbal*, está diseñado para evaluar el trabajo de un intérprete profesional. Como hemos mencionado en otras ocasiones, este estudio guarda cierta relación con la disciplina de interpretación y, teniendo en cuenta que en los sistemas de traducción de voz a voz los rasgos de oralidad son fundamentales, decidimos considerar este modelo e incluir y adaptar algunos de sus parámetros, para recogerlos en el apartado de aspectos paralingüísticos. La fluidez es uno de los aspectos más importantes que se tomaron de la propuesta de Collados Aís (2007), puesto que representa la doble faceta que tiene un intérprete, en este caso el sistema de traducción de voz a voz: por un lado, la función fundamental de comunicador y, por otro lado, la labor de generar un discurso agradable y de entretener al público, para lo que es imprescindible que haya fluidez en la emisión del discurso. Sin embargo, hay una serie de aspectos de los que prescindimos en primer momento, por considerarlos relacionados exclusivamente con las características humanas: acento nativo, impresión de fiabilidad y seguridad y agradabilidad de la voz.

Una vez comentada la propuesta de modelo en la que nos hemos basado para este trabajo, presentaremos y explicaremos, a continuación, la propuesta de revisión de dicho modelo. En primer lugar, podemos observar el esquema con los distintos parámetros que conforman el baremo, que hemos dividido en parámetros lingüísticos (*linguistic criteria*), relacionados fundamentalmente con el propio proceso de traducción, y parámetros paralingüísticos (*paralinguistic criteria*), que representan el factor oral que está presente en la traducción de voz a voz. A cada uno de los parámetros se le ha asignado un

color diferente, lo cual facilita el proceso de identificación de errores, al permitir señalarlos de forma visual sobre la propia transcripción del resultado de las muestras de análisis

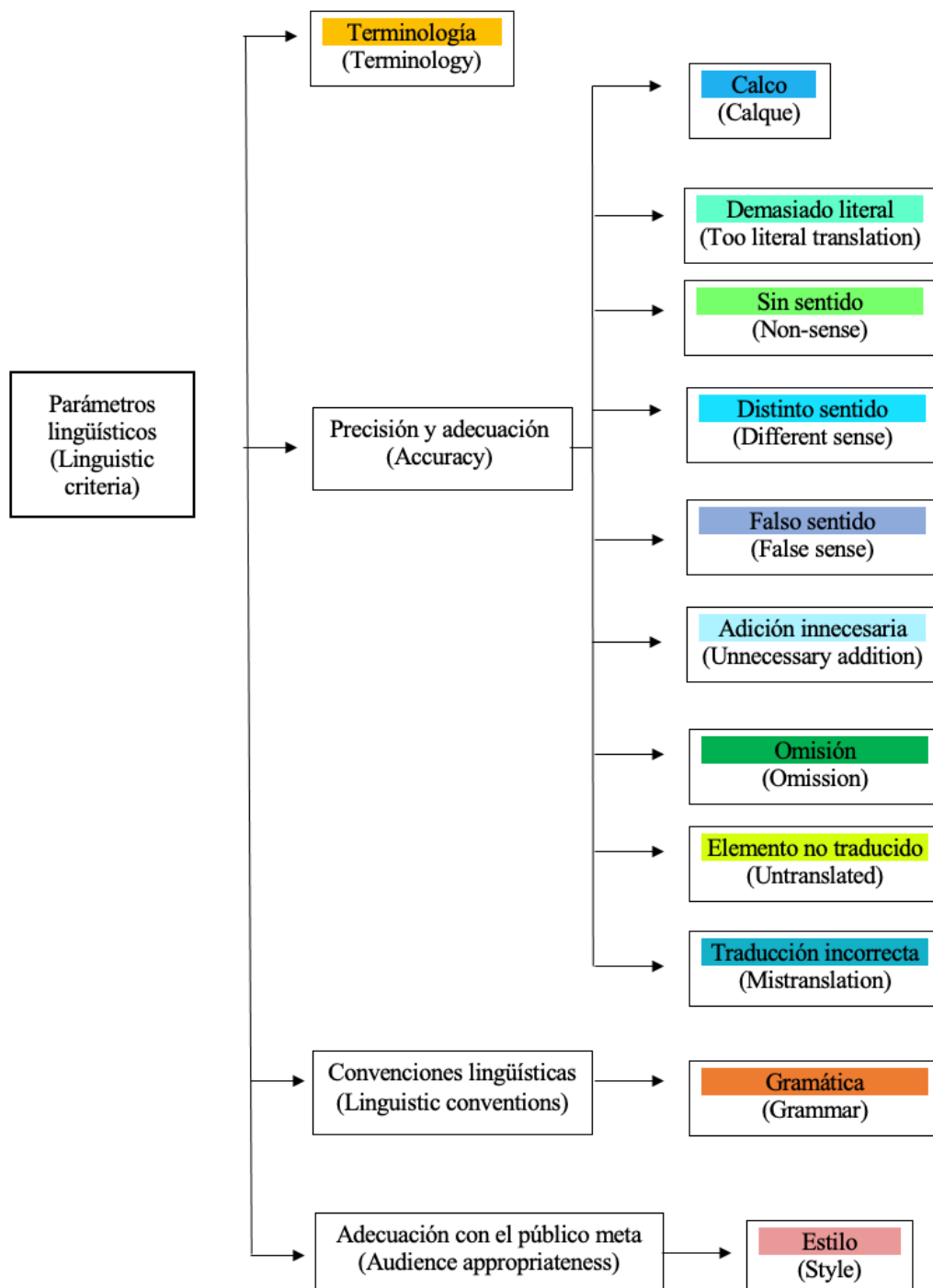


Figura 1. Propuesta de revisión de modelo: parámetros lingüísticos.

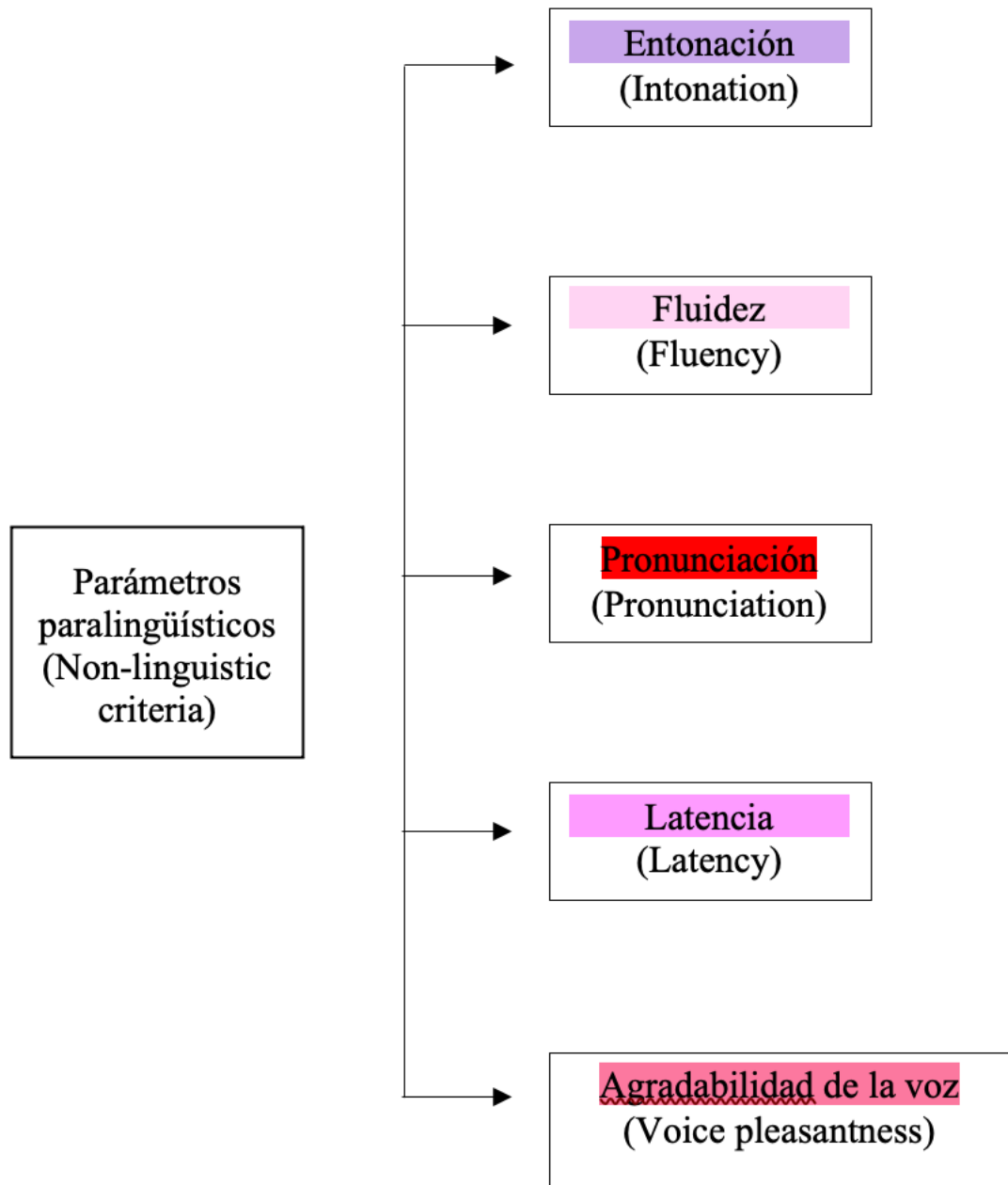


Figura 2. Propuesta de revisión de modelo: parámetros paralingüísticos.

La primera parte del baremo corresponde a los parámetros lingüísticos (*linguistic criteria*), un total de once aspectos que están, en su mayoría, relacionados con el proceso de traducción automática, el tercero de los pasos que conforman el modelo en cascada. Basándonos en la nueva clasificación de errores de traducción automática propuesta por MQM (2025), hemos dividido los once parámetros en las categorías de terminología (*terminology*), precisión y adecuación (*accuracy*), convenciones lingüísticas (*linguistic conventions*) y adecuación al público meta (*audience appropriateness*). La revisión del modelo propuesta por Ortego Antón (2025) nos ha servido para establecer esta categorización de los parámetros más detallada y precisa, en comparación con el modelo de partida para la presente revisión propuesto por Calvo del Barrio y Ortego Antón (en prensa). En primer lugar, el parámetro de terminología, presente en el modelo anterior, ha sido tomado del modelo de Ortiz Boix (2016), orientado a traducción automática. En caso de seleccionarse un término que no sea adecuado por no ajustarse al campo, a la situación comunicativa o a las especificaciones del cliente, si las hubiera, lo catalogaremos como error de terminología. Posteriormente, hemos sustituido la categoría de traducción incorrecta por precisión y adecuación (*accuracy*). A pesar de que traducción incorrecta cumplía una función de categoría general en la que tenían cabida aquellos errores de traducción que no pudieran encuadrarse en otras categorías y que tuvieran, efectivamente, un carácter más general, consideramos que los conceptos de precisión y de adecuación representan de forma más clara lo que esperamos del rendimiento del sistema de traducción automática que interviene en ambos sistemas. Sin perder el carácter general de una mala decisión de traducción que puede estar relacionada con diversos aspectos, sin llegar a ser de una naturaleza específica, como un falso sentido, hemos decidido mantener este parámetro de traducción incorrecta (*mistranslation*), en el que tienen cabida, por ejemplo, los errores a la hora de enfrentarse a cifras o nombres propios. Estos dos últimos elementos daban nombre a dos parámetros independientes en el modelo anterior. Sin embargo, por tratarse de aspectos demasiado específicos, hemos decidido incluirlos en el parámetro de traducción incorrecta (*mistranslation*), junto con otros errores relacionados con un desajuste de contenido entre el texto original y el texto meta que, por las características del propio desajuste, no tengan cabida en el resto de aspectos. El siguiente parámetro que se encuentra dentro de la categoría de precisión y adecuación (*accuracy*) es el de calco (*calque*), en el que se incluyen, por ejemplo, errores por calco de unidades léxicas. Este aspecto está estrechamente relacionado con el de traducción demasiado literal (*too literal translation*), destinado a errores en los que la traducción de una oración recuerde a una traducción palabra por palabra sin prestar atención al sentido. Posteriormente, diferenciamos el parámetro de falso sentido (*false sense*), en el que incluiremos todos aquellos errores en los que se transmita un significado contrario al original y, por tanto, se cometa un error importante relacionado con el significado. En esta línea, hemos considerado añadir el parámetro de sin sentido (*non-sense*). El proceso de identificación y análisis de errores llevado a cabo en el estudio anterior nos permitió detectar que, en muchas ocasiones, los errores tenían que ver con una formulación totalmente carente de sentido, que no permitía al receptor obtener la información adecuadamente. En esos casos, clasificábamos este tipo de errores en la categoría general de traducción

incorrecta, explicada anteriormente. No obstante, dada la frecuencia de aparición de este tipo de errores y, sobre todo, su relevancia a la hora de llevar a cabo el acto de comunicación con éxito, la decisión ha sido crear un parámetro independiente. Con el objetivo de centrar la atención en la importancia de la reproducción del sentido y del contenido original de forma fidedigna, así como de clasificar de forma precisa aquellos casos en los que esto no se cumpla, hemos considerado oportuno incorporar el parámetro de distinto sentido (*different sense*). En este aspecto, tendrán cabida aquellas decisiones de traducción que den a entender al receptor un significado ligeramente distinto al original, sin llegar a ser opuesto ni carecer de sentido completamente. Otros de los parámetros que se tomaron del modelo de Ortiz Boix (2007) y que se han mantenido en esta propuesta de modelo son el de adición innecesaria (*unnecessary addition*), el de omisión (*omission*) y el de elemento no traducido (*untranslated*). Por una parte, consideraremos como error por adición innecesaria aquellas informaciones que se incluyan en el discurso meta, pero que no se mencionen en el texto origen y que no sea necesario que se añadan. Por otra parte, se clasificarán como errores de omisión aquellos en los que se prescinda en el discurso meta de información presente en el discurso origen, lo que hace que no se transmita el mensaje de forma completa. Por último, hablaremos de error de elemento no traducido en aquellas partes que permanezcan en lengua origen en el discurso meta, a pesar de que deberían haberse traducido. A continuación, pasamos a la categoría de convenciones lingüísticas (*linguistic conventions*), integrada por el parámetro de gramática (*grammar*). No es de dudar que la mayoría de los receptores esperan de un texto que cumpla con las normas gramaticales de la lengua en cuestión. De hecho, la corrección gramatical contribuye en gran medida a la transmisión correcta del mensaje, puesto que en ocasiones los errores gramaticales pueden derivar en errores a nivel de contenido. De ahí, la importancia de señalar aquellos elementos que no respeten las normas gramaticales. El siguiente parámetro, que es el único integrante de la categoría de adecuación con el público meta (*audience appropriateness*), también está vinculado a los receptores y a sus expectativas, parte fundamental de los procesos que se llevan a cabo en las distintas modalidades de servicios lingüísticos. Catalogaremos como error de estilo (*style*) aquellos aspectos que no se adapten a las características y a los requisitos de la situación comunicativa. Cabe mencionar, no obstante, que las cuestiones relacionadas con el estilo suelen resultar altamente subjetivas y depender, en algunos casos, de las preferencias del traductor o del intérprete.

La segunda parte del esquema recoge los cinco parámetros paralingüísticos (*paralinguistic criteria*), que subrayan la importancia del factor oral en la traducción de voz a voz y para los que se ha tomado en consideración el modelo propuesto por Collados Aís (2007), orientado a interpretación simultánea realizada por humanos. Los cuatro primeros ya estaban presentes en la propuesta anterior y, aunque se optó por prescindir del quinto, hemos considerado oportuno incluirlo en esta revisión. En primer lugar, se hace necesario puntualizar que la entonación puede transmitir casi tanto como las palabras, de modo que resulta fundamental replicar la entonación del discurso original para reproducir la intención del orador. Por tanto, de cara al éxito del acto de comunicación, hemos considerado esencial

mantener el parámetro de entonación (*intonation*). Como hemos adelantado al principio de este apartado 2.3. *Modelo de evaluación*, la fluidez en un discurso permite que este sea agradable y entretenga al público. Uno de los elementos que puede restar fluidez al discurso y que consideraremos como error de fluidez (*fluency*) son las pausas innecesarias, sobre todo, si su duración es considerable. En el modelo anterior, decidimos sustituir el parámetro de acento nativo, por resultar muy abstracto y más bien propio de un intérprete humano y no de un sistema automático, por pronunciación (*pronunciation*), ya que una correcta pronunciación y dicción son de gran relevancia a la hora de transmitir el contenido de forma clara y agradable. Con respecto a la velocidad de habla, en la propuesta anterior incorporamos el parámetro de latencia (*latency*) y, dado que resulta interesante analizar cómo se adecúa la velocidad de transmisión de la interpretación con respecto a la del original, hemos decidido seguir manteniéndolo. Además, nos permitirá conocer más acerca de la gestión del tiempo y de las pausas por parte de los sistemas. Por último, a pesar de que en el estudio anterior indicamos que la agradabilidad de la voz se relacionaba más fácilmente con las actitudes humanas y no con sistemas de IA, que es lo que ocupa ambos estudios, hemos considerado oportuno incorporar este parámetro en esta nueva propuesta, teniendo en cuenta que uno de los aspectos que se han destacado en el marco teórico es que resulta importante que la voz del sistema se asemeje lo máximo posible a la voz humana y resulte, por lo tanto, natural. De este modo, mostramos cierto interés en la experiencia del receptor, lo cual debería tenerse en cuenta a la hora de tratar con este tipo de sistemas.

Tras explicar y justificar la presencia de cada uno de los parámetros que forman parte de nuestra propuesta, es necesario indicar cuál será el proceso de análisis según dicho modelo.

5.4. Procedimiento de recogida y análisis de datos

Una vez seleccionados los dos fragmentos del episodio, hemos recurrido a Interprefy y a Kudo para obtener la interpretación al español de las dos muestras de análisis. En el caso de Interprefy, hemos seleccionado primeramente la lengua de partida (*English*) y la lengua de llegada (*Spanish AI*) y posteriormente hemos ajustado el volumen del discurso en cada lengua, para garantizar que la interpretación en español se escucha correctamente, mientras que puede escucharse de fondo el discurso original en inglés. A continuación, hemos subido el fragmento de vídeo y con solo pulsar al botón «Play» el vídeo empieza a reproducirse y comenzamos a escuchar la interpretación al español, con el discurso original en inglés de fondo. Para guardar ambas muestras, hemos recurrido a una grabación de pantalla. A la hora de obtener la transcripción del discurso original, hemos activado la opción de «Live Captions» y a partir de ello hemos llevado a cabo una revisión manual. En el caso de la transcripción del discurso meta, hemos optado por un proceso manual, por tratarse de un fragmento muy breve en lengua materna.

En el caso de KUDO, tras crear una sesión de AI Speech Translation y determinar los idiomas de trabajo para dicha sesión, hemos seleccionado el archivo de vídeo de las muestras de análisis en la

opción de «Compartir Pantalla» para comenzar a escuchar la traducción a español. Posteriormente, hemos finalizado la sesión y hemos podido descargar la grabación de la misma. A la hora de obtener la transcripción del discurso original, hemos recurrido a la opción de «Live Captions» y posteriormente hemos revisado el resultado. Para la transcripción del discurso meta, hemos llevado a cabo un proceso manual. Una vez contamos con las transcripciones del discurso origen y del discurso meta, las hemos dispuesto en segmentos alineados en un documento Excel de Microsoft, a partir del cual procederemos al proceso de identificación de errores, en el señalaremos sobre la propia transcripción aquellos errores de carácter lingüístico y paralingüístico que se detecten. Para ello, trabajaremos de forma paralela con las muestras de análisis en audio en lengua origen y meta. Como hemos observado en el esquema, contamos con un código de colores, de forma que cada color corresponderá a un parámetro diferente y podremos reconocer en la propia tabla los distintos errores y la categoría a la que pertenecen de forma más rápida. De este modo, llevaremos a cabo un análisis cuantitativo, además de cualitativo, ya que nos detendremos a comentar cada error.

6. Análisis y resultados

En este capítulo, mostraremos el análisis de los errores que se han detectado en la interpretación al español de ambas muestras. En primer lugar, nos centraremos en la muestra 1, con un apartado independiente para cada uno de los sistemas de traducción de voz a voz empleados, y posteriormente abordaremos del mismo modo la muestra 2. Para ello, presentaremos una tabla con la transcripción del discurso original y la interpretación al español sobre la que hemos señalado los diferentes errores según el código de colores que figura en el modelo de evaluación del capítulo anterior. Además del código de colores, que hace posible identificar con mayor rapidez y de forma más visual el parámetro al que pertenece cada error, los errores aparecen acompañados de un número, que nos permite localizar el comentario correspondiente, que aparecerá tras la tabla.

6.1. Muestra 1

6.1.1. Interprefy

EN	ES
<i>They now have strong evidence of a dangerous strain of E. Coli, that is now highly resistant to antibiotic treatment.</i>	⁽¹⁾ Ahora tienen pruebas contundentes ⁽²⁾ de una cepa peligrosa de coli ⁽³⁾ que ⁽⁴⁾ ahora ⁽⁵⁾ es altamente resistente al tratamiento con antibióticos.
<i>At least 2 million Americans get antibiotic resistant infections each year, with 23,000 deaths.</i>	Al menos dos millones de estadounidenses contraen enfermedades resistentes a los antibióticos cada año, con 23 000 muertes.
<i>We are continuing to barrel towards a future where a cut on your hand or an infection from a routine dental procedure could prove to be deadly because there is very little that we can do to treat that infection.</i>	Seguimos avanzando hacia un futuro en el que un corte en la mano o una infección de ⁽⁶⁾ un procedimiento dental de rutina ⁽⁷⁾ podrían resultar mortales, porque hay muy poco que podamos ⁽⁸⁾ hacer para tratar esa ⁽⁹⁾ infección.
<i>How can you do surgery without these life saving critical drugs that are being squandered just to make cheaper meat?</i>	¿Cómo se puede hacer ⁽¹⁰⁾ una cirugía sin estos medicamentos ⁽¹¹⁾ críticos ⁽¹²⁾ que salvan vidas y que se están desperdiciando solo ⁽¹³⁾ para hacer ⁽¹⁴⁾ carne más barata?

Tabla 4. Identificación de errores en la muestra 1 (Interprefy).

⁽¹⁾ Este primer error se corresponde con el parámetro de latencia de habla, dentro de los parámetros paralingüísticos. El tiempo de desfase, esto es, el tiempo que transcurre desde que el orador principal comienza su discurso hasta que la interpretación comienza, es incluso mayor que una oración, de modo

que la coordinación entre el discurso original y la interpretación no es la adecuada y este tiempo de desfase excesivo se va arrastrando hasta el final del fragmento.

⁽²⁾ Nos encontramos ante un error relacionado con una cuestión tan subjetiva, como es el estilo. En este caso, consideramos que se hace necesario buscar una alternativa que resulte más idiomática y que rescate alguna de las estructuras que suelen emplearse en el lenguaje científico a la hora de informar sobre nuevos hallazgos. En este sentido, proponemos: «Actualmente contamos con evidencias sólidas de...».

⁽³⁾ Detectamos que no se ha seleccionado el equivalente adecuado: «E. Coli», una bacteria que se transmite al ser humano por ingesta de alimentos contaminados.

⁽⁴⁾ Este error está relacionado con la fluidez, al alargarse la pronunciación de la preposición «de» en una pausa innecesaria entre la preposición y el sustantivo. Posteriormente sucede todo lo contrario: no se realiza apenas pausa entre el sustantivo antecedente y el enlace de la proposición de relativo.

⁽⁵⁾ Nuevamente hemos identificado un error de estilo, al encontrar repetido de forma muy seguida el adverbio «ahora». A pesar de que el adverbio *now* aparece en dos ocasiones bastante próximo en la misma oración en el discurso origen, hemos determinado que sería conveniente sustituirlo en una de las ocasiones, con el objetivo de sumar riqueza lingüística y fluidez al discurso en español. Dado que se trata de una circunstancia nueva, es posible emplear el verbo «volverse» en la siguiente estructura: «...una cepa peligrosa de E. Coli que se ha vuelto altamente resistente al tratamiento con antibióticos...» o bien «...una cepa peligrosa de E. Coli que ha desarrollado una alta resistencia al tratamiento con antibióticos...».

⁽⁶⁾ Estamos ante un error de traducción incorrecta por la elección errónea de la preposición. Para mayor precisión y fiabilidad con el contenido original, proponemos emplear la preposición «durante», dado que el tratamiento dental no es la causa directa de la infección, como podría darse a entender si mantenemos la preposición «de», sino que se produce en el contexto de la misma.

⁽⁷⁾ Tras consultar la página web de diversas clínicas dentales ubicadas en España, hemos asumido este sintagma como un calco de sus tres elementos. En este sentido, cabe mencionar que es importante prestar atención al término *procedure*, ya que el equivalente «procedimiento» no es el adecuado en la inmensa mayoría de los casos. En este ejemplo, podríamos optar, entre otras opciones, por «tratamiento de odontología general». Al contrastar los servicios que se incluyen en «tratamiento de odontología general» y en *routine dental procedure*, observamos una clara equivalencia de significado, con tratamientos, como la limpieza bucal o el empaste.

⁽⁸⁾ A pesar de no ser un texto excesivamente especializado, encontramos conveniente mantener un tono formal y de despersonalización que cumpla con las convenciones propias del lenguaje científico, debido a las propias características de los profesionales que intervienen. Por ello, dentro del parámetro de estilo,

proponemos el empleo de una oración impersonal refleja para sustituir el uso de la primera persona de plural: «se pueda hacer».

⁽⁹⁾ Teniendo en cuenta que no se refiere a una infección en concreto, sino a una serie de infecciones semejantes a los ejemplos previos, es decir, infecciones o lesiones que *a priori* no deberían presentar mayor complicación, sugerimos eliminar el calco de *that* por «esa» y sumar precisión con el sintagma «este tipo de infecciones».

⁽¹⁰⁾ Nos encontramos una vez más ante una cuestión de estilo, estrechamente relacionada con el error ⁽⁸⁾. Proponemos apostar nuevamente por un tono más formal, por la propia naturaleza del tema que se está tratando, y emplear el verbo «llevar a cabo», en lugar de «hacer».

⁽¹¹⁾ El término *drug* suele resultar un tanto problemático en la disciplina de traducción e interpretación, puesto que en español puede equivaler a tres términos diferentes: «medicamento», «fármaco» o «droga». En este caso, al no referirnos a ninguna marca comercial en concreto, no sería adecuado emplear el término «medicamento», sino «fármaco».

⁽¹²⁾ Al consultar el Diccionario de la Real Academia de la Lengua Española (2025), hemos comprobado que no hay ninguna acepción que recoja el adjetivo «crítico/a» con el significado de alta importancia o relevancia. Por ello, hemos catalogado este error como calco y proponemos otros adjetivos, como «determinante» o «indispensable». Por otra parte, dado que inmediatamente después contamos con una proposición de relativo tan enfática, como es «que salvan vidas», consideramos que sería aceptable prescindir del adjetivo para evitar la redundancia.

⁽¹³⁾ De igual forma que hemos procedido con los errores anteriores que pertenecen a la categoría de estilo, proponemos emplear un tono más formal y sustituir «solo» por «únicamente» o «con el único pretexto de».

⁽¹⁴⁾ Además de considerar que en este caso también es necesario recurrir a un tono más formal, nos parece que la elección del verbo «hacer» puede dar lugar a ambigüedad y que sumaríamos precisión si optamos por el verbo «producir».

6.1.2. KUDO

EN	ES
<i>They now have strong evidence of a dangerous strain of E. Coli, that is now highly resistant to antibiotic treatment.</i>	⁽¹⁾ Ahora tienen pruebas fehacientes ⁽²⁾ de la existencia de una cepa peligrosa de Coli ⁽³⁾ que ⁽⁴⁾ ahora ⁽⁵⁾ es altamente resistente al tratamiento antibiótico.
<i>At least 2 million Americans get antibiotic resistant infections each year, with 23,000 deaths.</i>	Al menos dos millones de estadounidenses contraen infecciones resistentes a los antibióticos cada año, con 23 000 muertes.
<i>We are continuing to barrel towards a future where a cut on your hand or an infection from a ⁽⁷⁾ routine dental procedure could prove to be deadly because there is very little that we can do to treat that infection.</i>	Seguimos ... ⁽⁶⁾ y no pueden rechazarlos a la acción de un procedimiento dental rutinario ⁽⁸⁾ podría resultar mortal ⁽⁹⁾ porque es muy poco lo que podemos ⁽¹⁰⁾ hacer para tratar esa ⁽¹¹⁾ infección.
<i>How can you do surgery without these life saving critical drugs that are being squandered just to make cheaper meat ⁽¹³⁾?</i>	¿Cómo se puede operar ⁽¹²⁾ sin ellos ⁽¹⁴⁾ ⁽¹⁵⁾ barato ⁽¹⁶⁾ ?

Tabla 5. Identificación de errores en la muestra 1 (KUDO).

⁽¹⁾ Dado que el tiempo de desfase es demasiado extenso, la coordinación entre el discurso original y la interpretación no es la adecuada. Este desajuste se reproduce hasta el final del fragmento con las constantes diferencias en la velocidad de habla, lo que genera un error de latencia.

⁽²⁾ Nos encontramos ante un error relacionado con el parámetro de estilo, pues la formulación no resulta idiomática y se hace necesario buscar otra alternativa que, además, cumpla con las estructuras que tienden a emplearse en el lenguaje científico para informar sobre nuevos hallazgos. Proponemos sustituir el sustantivo «prueba» por otro sustantivo más propio del lenguaje científico: «evidencia». Con el objetivo de no resultar redundante, ya que «evidente» y «fehaciente» son sinónimos, sugerimos emplear el adjetivo «sólido», en lugar de «fehaciente». De esta forma, obtendríamos el siguiente resultado: «Actualmente contamos con evidencias sólidas de una cepa peligrosa...» o «Actualmente existen evidencias sólidas de una cepa peligrosa...».

⁽³⁾ En este caso, el equivalente adecuado es «E. Coli», una bacteria que se transmite al ser humano por ingesta de alimentos contaminados.

⁽⁴⁾ Dado que se da una pausa entre la preposición y el sustantivo antecedente, al alargarse innecesariamente la pronunciación de la preposición «de» y posteriormente no existe apenas pausa entre el sustantivo y el enlace de la proposición de relativo, encuadramos este error en el parámetro de fluidez.

⁽⁵⁾ Se trata de un error de estilo en el que el adverbio «ahora» aparece repetido en dos ocasiones de forma muy seguida. Si bien es cierto que el adverbio *now* aparece en dos ocasiones bastante próximo en la misma oración, consideramos oportuno sustituirlo en una de las ocasiones con el objetivo de sumar riqueza lingüística y fluidez al discurso en español. En este sentido, dado que se trata de una circunstancia nueva, es posible emplear el verbo «volverse» en la siguiente estructura: «...una cepa peligrosa de E. Coli que se ha vuelto altamente resistente al tratamiento con antibióticos...» o bien «...una cepa peligrosa de E. Coli que ha desarrollado una alta resistencia al tratamiento con antibióticos...».

⁽⁶⁾ Tiene lugar una pausa innecesaria demasiado extensa que resta fluidez al discurso.

⁽⁷⁾ y ⁽⁹⁾ Hemos decidido tratar los errores ⁽⁷⁾ y ⁽⁹⁾ en conjunto, puesto que están muy relacionados entre sí. En primer lugar, diferenciamos un error por omisión considerablemente extenso. Como consecuencia de esta omisión tan larga, se produce un error de la tipología de sin sentido, ya que la formulación de esta oración, además de no corresponder con la información del discurso original, carece de sentido completamente. Para subsanar ambos errores, proponemos lo siguiente: «Avanzamos hacia un futuro en el que un corte en la mano o una infección durante un tratamiento de odontología general podría resultar mortal...».

⁽⁸⁾ Conviene señalar que el término *procedure* tiende a generar confusión, pues el equivalente «procedimiento» no es el adecuado en la inmensa mayoría de los casos. Para seleccionar el equivalente adecuado y evitar este error por calco de este sintagma nominal, podríamos optar por «tratamiento de odontología general». Al contrastar los servicios que se incluyen en «tratamiento de odontología general» y en *routine dental procedure*, observamos una clara equivalencia de significado, con tratamientos como la limpieza bucal o el empaste.

⁽¹⁰⁾ Encuadramos este error en el parámetro estilo, puesto que consideramos que, dado el tema de la intervención y el propio interlocutor, convendría adoptar un tono un tanto más formal. Para ello, proponemos recurrir a un estilo impersonal, en lugar el uso de la primera persona del plural: «...las posibilidades de tratamiento para este tipo de infecciones son muy limitadas».

⁽¹¹⁾ Dado que no se trata de una infección en concreto, sino a una serie de infecciones parecidas a los ejemplos previos, es decir, infecciones que *a priori* no deberían presentar mayor complicación, el empleo del adjetivo demostrativo «esa» resulta restrictivo a nivel de significado. Por tanto, proponemos eliminar el calco de *that* por «esa» y sumar precisión con el sintagma «este tipo de infecciones».

⁽¹²⁾ Como en el resto de los errores de estilo previos, hemos determinado que se hace necesario sustituir el verbo «operar» por un equivalente ampliado más formal: «llevar a cabo una cirugía».

⁽¹³⁾ y ⁽¹⁶⁾ Hemos estimado conveniente analizar estos dos errores en conjunto, dado que están conectados. La mayor parte del contenido original ha sido omitido y, como consecuencia, el texto en español carece

de sentido por completo, lo cual impide al receptor obtener la información del original. Por tanto, estamos ante un error por omisión y un error de sin sentido, para los que realizamos la siguiente propuesta: «¿Cómo se puede llevar a cabo una cirugía sin estos fármacos que realmente salvan vidas y que se están desperdiciando con el único pretexto de producir carne más barata?».

⁽¹⁴⁾ y ⁽¹⁵⁾ Además de encontrarnos ante una pausa innecesaria entre dos elementos de la oración, que definimos como error de fluidez, hemos detectado una entonación errónea que hace al receptor deducir que la oración finaliza en el lugar equivocado.

6.2. Muestra 2

6.2.1. Interpretfy

EN	ES
“The first time I've ever tried a tomato that I grew myself I was like ‘wait, this is a whole other tomato, like it tastes different’.” “Right.” ⁽⁶⁾	— ⁽¹⁾ La primera vez que probé un tomate que cultivé yo ⁽²⁾ mismo ⁽³⁾⁽⁴⁾ pensé «espera, este es un tomate completamente diferente, como si tuviera un sabor diferente ⁽⁵⁾ .”
“This is all poblano peppers.”	— Esto es todo ⁽⁷⁾ chiles poblanos.
“Let's try some beans.” “How do you eat these?” “It's just like eat it.” “The whole thing, you could eat it?” “Yes, you could eat it, except for like you know the end.” “It's just like fresh, you know.” “Yes, it's very fresh.”	— Probemos algunos frijoles ⁽⁸⁾ . — ¿Cómo comemos estos ? ⁽⁹⁾ — Simplemente comérselo . ⁽¹⁰⁾ — ¿Todo, te lo podías comer ⁽¹¹⁾ ? ⁽¹²⁾ — ⁽¹³⁾ Excepto por el final. Es como fresco ⁽¹⁴⁾ . — Estaba muy fresco ⁽¹⁵⁾ .
“Those are all mustard greens. This is dinosaur kale.” “The texture is so interesting.” “We could eat this?” “You could totally eat this.” “This is so cool just to like, just take ⁽¹⁹⁾ .” “It's like, this is like you're shopping through the grocery store.”	— Esas son todas hojas de mostaza. Esto es col rizada de dinosaurio ⁽¹⁶⁾ . — La textura es muy interesante. — ¿Podemos comer esto ? ⁽¹⁷⁾ — Podrías comer totalmente esto . ⁽¹⁸⁾ — ⁽²⁰⁾ Es tan genial. ⁽²¹⁾ — Como si fuera lo correcto ⁽²²⁾ , esto como si esto fuera como ⁽²³⁾ ir de compras ⁽²⁴⁾ al supermercado.

Tabla 6. Identificación de errores en la muestra 2 (Interprefy).

⁽¹⁾ El tiempo de desfase se extiende hasta casi una oración, de modo que la coordinación del discurso original y la interpretación no es la adecuada y este tiempo de desfase excesivo se va arrastrando hasta el final del fragmento por las constantes diferencias en la velocidad de habla, lo que genera un error de latencia.

⁽²⁾ Consideramos que el empleo del pretérito perfecto simple de indicativo («cultivé») no suma fluidez discursiva. Por ello, sugerimos emplear el pretérito pluscuamperfecto de indicativo («había cultivado») o incluso prescindir de la forma verbal y sustituirla por un adjetivo («un tomate cultivado por mí misma»).

⁽³⁾ Se trata de un error relacionado con el parámetro de gramática, puesto que no existe concordancia entre el género del adjetivo y la persona que interviene. Dado que la interlocutora pertenece al sexo femenino, debe emplearse el género femenino: «misma».

⁽⁴⁾ La entonación entre el complemento de tiempo previo y la forma verbal nos da a entender que la oración va a finalizar, cuando en realidad, tras el complemento de tiempo, encontramos la forma verbal. Por tanto, diferenciamos un error de entonación.

⁽⁵⁾ La formulación de esta oración no resulta idiomática en la lengua meta y no presenta rasgos de oralidad propios del español. Con el objetivo de aportar fluidez discursiva a esta intervención de carácter coloquial, realizamos la siguiente propuesta: «Este tomate es completamente diferente, tiene un sabor bien distinto».

⁽⁶⁾ Nos encontramos ante la omisión de la respuesta de uno de los gemelos a la intervención de la persona que les muestra los diferentes cultivos («*Right*»), que proponemos traducir como «Entiendo».

⁽⁷⁾ Además de no resultar idiomático en español, esta breve intervención contiene un error de gramática, puesto que no hay concordancia de número entre el sujeto («chiles poblanos»), que tiene número plural, y el verbo («es»), de número singular. Más allá de corregir el error gramatical, realizamos dos propuestas con un estilo más idiomático: «Aquí tenemos chiles poblanos» o «Todo esto es chile poblano».

⁽⁸⁾ La forma alargada y cilíndrica, junto con el color verde que podemos ver en imagen nos lleva a pensar que lo que los gemelos tienen en las manos es una judía verde. Al hablar de frijoles nos referimos a un tipo de legumbre que en la Península Ibérica suele asociarse con la cocina latinoamericana. El apoyo visual del episodio nos ha permitido detectar este error de terminología y determinar que el equivalente adecuado es «judía verde».

⁽⁹⁾ Como ocurría en el error de estilo anterior, nos encontramos ante una formulación que en absoluto resulta idiomática en español. En su lugar, proponemos lo siguiente: «¿Cómo se comen?».

⁽¹⁰⁾ Volvemos a diferenciar un error de estilo, por no resultar idiomático en la lengua meta. Con el objetivo de mantener la idea del mensaje original y lograr una formulación más cercana a los rasgos de oralidad en español, realizamos la siguiente propuesta: «Se comen así, tal cual». Dado que contamos con la ayuda visual del vídeo, en el que aparece la responsable de desarrollo y participación del huerto comiendo una judía, podemos permitirnos emplear el adverbio «así» y la locución adverbial «tal cual».

⁽¹¹⁾ Estamos ante un error de traducción incorrecta en el que se ha tomado *could* desde la perspectiva de capacidad o posibilidad en el pasado y no con el significado de solicitar permiso de forma cortés en el presente. Por tanto, sería adecuado unificar ambas partes de la pregunta para transmitir el sentido del original: «¿Se puede comer todo entero?».

⁽¹²⁾ Distinguimos un error de entonación, ya que no se aprecia una variación del tono para indicar que se trata de una pregunta.

⁽¹³⁾ El hecho de que no tenga lugar una breve pausa entre la pregunta y la respuesta no permite establecer una diferencia clara entre las dos intervenciones, lo cual catalogamos como error de fluidez. Cabe mencionar que en este punto se ha omitido la primera parte de la respuesta. Sin embargo, hemos determinado que no se trata de un error, ya que la segunda parte de la respuesta («Excepto por el final») lleva implícito el hecho de que el resto de la verdura sí se puede comer, de modo que no afecta a la comprensión del mensaje.

⁽¹⁴⁾ Tal y como está formulada esta oración, no estaríamos ante una afirmación, como sí la encontramos en el discurso original. Podemos deducir que las fórmulas coloquiales que aparecen en el discurso en inglés, tales como *just* o *like*, han provocado que la traducción al español haya perdido firmeza en la afirmación. Por ello, se da a entender un distinto sentido, que podría corregirse sin restar rasgos de coloquialidad de la siguiente manera: «Está muy fresco».

⁽¹⁵⁾ Catalogamos este error como traducción incorrecta, puesto que, al trasladar de forma errónea el tiempo de la forma verbal, se ha dado a entender una información incorrecta. Proponemos corregirlo de la siguiente manera: «Muy fresco, sí».

⁽¹⁶⁾ El equivalente adecuado para *dinosaur kale* es «col rizada».

⁽¹⁷⁾ Como ocurría en el caso de otras preguntas en este fragmento, no presenta un estilo idiomático que se asemeje a los rasgos de coloquialidad en español. Por ello, sugerimos la siguiente alternativa: «¿Esto se puede comer?».

⁽¹⁸⁾ Nuevamente realizamos una propuesta de corrección con el fin de sumar idiomática: «Por supuesto, se puede comer».

⁽¹⁹⁾ y ⁽²¹⁾ En primer lugar, identificamos un error por omisión que *a priori* pasa un tanto inadvertido. En la siguiente propuesta incorporamos la información que faltaba y tratamos de sumar idiomatidad en la formulación: «Es fantástico, lo coges y ya está».

⁽²⁰⁾ Estamos ante un error de fluidez, ya que no tiene lugar una breve pausa entre la pregunta y la respuesta, lo cual no permite establecer una diferencia clara entre las dos intervenciones.

⁽²²⁾ La información que figura en el original no se corresponde con ninguno de los segmentos del texto original, por lo que proponemos eliminarlo y catalogarlo como error de adición innecesaria.

⁽²³⁾ Dado que esta oración carece de sentido y sus elementos no están ordenados correctamente, consideramos este caso como un error de sin sentido, para el que realizamos la siguiente corrección: «Esto es como...».

6.2.2. KUDO

EN	ES
“The first time I’ve ever tried a tomato that I grew myself I was like ‘wait, this is a whole other tomato, like it tastes different’.” “Right.” ⁽⁹⁾	— ⁽¹⁾ La primera vez. Se fue ⁽²⁾ . La primera vez que probé un tomate ⁽³⁾ que cultivé yo mismo ⁽⁴⁾ ⁽⁵⁾ ⁽⁶⁾ yo ⁽⁷⁾ estaba como espera esto es como...es diferente ⁽⁸⁾ .
“This is all poblano peppers.”	— Esto es todo ⁽¹⁰⁾ pimientos poblanos.
“Let’s try some beans.” “How do you eat these?” “It’s just like eat it.” “The whole thing, you could eat it?” “Yes, you could eat it, except for like you know the end.” ⁽¹¹⁾ “It’s just like fresh, you know.” “Yes, it’s very fresh.”	— Al igual que sí, estoy para como ya sabes el final ⁽¹²⁾ . Es solo como fresco, ¿sabes? ⁽¹³⁾ — Es muy fresco.
“Those are all mustard greens. This is dinosaur kale.” “The texture is so interesting.” “We could eat this?” “You could totally eat this.” “This is so cool just to like, just take.” “It’s like this is like you’re shopping through the grocery store.”	— Los dinosaurios son tan interesantes ⁽¹⁴⁾ . — ¿Usted podría comer esto? ⁽¹⁵⁾ — Podrías comerte esto ⁽¹⁶⁾ , esto, ⁽¹⁷⁾ solo ⁽¹⁸⁾ . — Es como ir de compras ⁽¹⁹⁾ a la tienda de comestibles ⁽²⁰⁾ .

Tabla 7. Tabla 6. Identificación de errores en la muestra 2 (KUDO).

⁽¹⁾ Consideramos que el tiempo de desfase es demasiado extenso, de modo que la coordinación del discurso original y la interpretación no es la adecuada y este tiempo de desfase excesivo se va arrastrando hasta el final del fragmento por las constantes diferencias en la velocidad de habla, lo que genera un error de latencia.

⁽²⁾ Esta oración no corresponde a ningún fragmento del discurso original, por lo que la consideramos como un error de adición innecesaria y optamos por eliminarla.

⁽³⁾ Nos encontramos ante un error de entonación, puesto que la entonación después del sustantivo «tomate» da a entender que la oración finaliza en ese punto, cuando contamos con una proposición de relativo inmediatamente después, que da continuidad a esta oración.

⁽⁴⁾ Estamos ante un error de gramática, puesto que no existe concordancia entre el género del adjetivo y la persona que interviene. Dado que la interlocutora es una mujer, debe emplearse el género femenino: «misma».

⁽⁵⁾ y ⁽⁶⁾ En este caso, además de producirse una pausa innecesaria, que definimos como error de fluidez, la entonación vuelve a hacernos pensar que la oración finaliza en este punto, mientras que a continuación nos encontramos con el sintagma verbal principal.

⁽⁷⁾ Las particularidades propias de la lengua española hacen posible omitir el sujeto y, de hecho, esta práctica es la más común, de modo que, en la mayoría de las ocasiones, tanto en el lenguaje escrito, como oral, tendemos a omitir el sujeto en las oraciones. Por ello, consideramos que el haber incluido el pronombre personal «yo» es una adición innecesaria y optamos por eliminarlo.

⁽⁸⁾ En este fragmento en particular, consideramos que la voz de la interpretación resulta un tanto robótica y entrecortada, de forma que diferenciamos un error de agradabilidad de la voz.

⁽⁹⁾ Detectamos que se omite la respuesta de uno de los gemelos a la intervención de la persona que les muestra los diferentes cultivos («*Right*»), que proponemos traducir como «Entiendo».

⁽¹⁰⁾ Además de no resultar idiomático en español, esta breve intervención contiene un error de gramática, puesto que no hay concordancia de número entre el sujeto («chiles poblanos»), que tiene número plural, y el verbo («es»), de número singular. Después de corregir el error gramatical, tratamos de emplear un estilo más idiomático con estas dos propuestas: «Aquí tenemos chiles poblanos» o «Todo esto es chile poblano».

⁽¹¹⁾ y ⁽¹²⁾ Identificamos un error de omisión considerablemente extenso que provoca una formulación carente de sentido que impide al receptor obtener la información de forma correcta. Con el objetivo de transmitir el mensaje de forma completa, sugerimos la siguiente alternativa:

—Vamos a probar unas judías verdes.

—¿Cómo se comen?

—Se comen así, tal cual.

—¿Se puede comer todo entero?

—Sí, se come todo, menos el extremo.

—Es muy fresco, ¿veis?

—Muy fresco, sí.

⁽¹³⁾ Se ha llevado a cabo una traducción palabra por palabra y no se han adaptado las expresiones coloquiales propias de la lengua inglesa (*just y like*), lo cual ha resultado en un estilo poco idiomático. Consideramos oportuno alejarnos del texto origen y tomar en consideración los rasgos del lenguaje coloquial de la lengua española: «Es muy fresco, ¿veis?».

⁽¹⁴⁾ En esta parte de la intervención, la responsable de desarrollo y participación del huerto les muestra dos cultivos diferentes y les da a conocer el nombre de los mismos. Sin embargo, la traducción de estas dos intervenciones de apenas cuatro palabras cada una no se corresponde en absoluto, de modo que lo consideramos un error de traducción incorrecta. Proponemos la siguiente solución: «Aquí tenemos hojas de mostaza y esto es todo col rizada»

⁽¹⁵⁾ Además de no presentar un estilo demasiado idiomático, se emplea el pronombre personal como fórmula de tratamiento de cortesía o distanciamiento «usted», a pesar de que podemos observar que se trata de una conversación distendida con lenguaje familiar. Planteamos la siguiente alternativa, en la que hemos eliminado la fórmula de cortesía y hemos tratado de sumar idiomática a la intervención: «¿Esto se puede comer?».

⁽¹⁶⁾ Como ocurría en el caso anterior, hemos tratado de realizar una propuesta alternativa que presente mayor fluidez discursiva: «Por supuesto, se puede comer».

⁽¹⁷⁾ y ⁽¹⁸⁾ Identificamos, por un lado, un error por pausa innecesaria dentro de una formulación carente de sentido que no se corresponde en absoluto con la información del discurso original. En la siguiente propuesta, incluimos la información que figura en el texto origen: «Es fantástico, lo coges y ya está».

⁽¹⁹⁾ Nos encontramos ante un error de distinto sentido por la diferencia de significado que se da entre las expresiones «ir de compras» e «ir a hacer la compra» o «ir a la compra». En el caso de la primera, nos referimos una actividad que podría considerarse de ocio y en la que acudimos a grandes almacenes o a distintas tiendas en busca de productos, como ropa o complementos. Por otra parte, en el caso de las dos últimas expresiones, hacemos referencia a la tarea de comprar alimentos o productos necesarios para el hogar. Por consiguiente, consideramos conveniente emplear bien «ir a hacer la compra», bien «ir a la compra».

⁽²⁰⁾ A pesar de que el equivalente «tienda de comestibles» para *grocery store* es correcto, conviene reconocer que los hablantes nativos de español de España no suelen recurrir a este término para referirse a este tipo de establecimientos, sino que resulta más común emplear «supermercado».

6.3. Contraste de resultados

En este apartado, procedemos a reflexionar sobre el análisis anterior, para lo que comparemos los resultados obtenidos en cada sistema para ambas muestras de análisis y realizaremos un análisis comparativo de los dos sistemas de traducción de voz a voz.

6.3.1. Interpretfy

6.3.1.1. Muestra 1

A partir del siguiente gráfico circular, comentaremos la frecuencia de aparición de los distintos errores y el parámetro al que pertenece cada porcentaje, con el fin de obtener una visión general de la muestra.

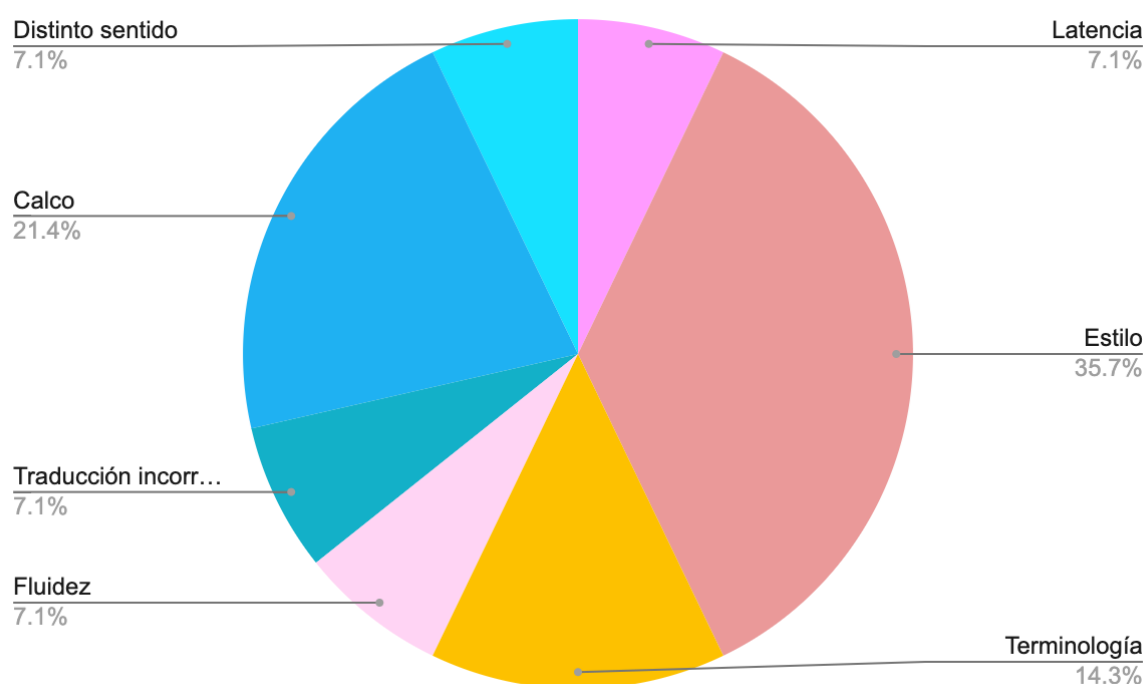


Figura 3. Gráfico de resultados de la muestra 1 (Interprefy).

Con una diferencia considerable, el parámetro que acumula la mayor parte de los errores es el aspecto lingüístico del estilo (35,7 %). En la mayoría de los casos, hemos optado por proponer alternativas con un estilo más formal, dada la temática de las intervenciones y los propios interlocutores. Por otra parte, en uno de los casos se ha tratado de evitar una repetición bastante próxima, para sumar riqueza léxica al discurso, y en otro de los casos el objetivo de la propuesta de reformulación ha sido adaptarla a las estructuras propias del lenguaje científico. En segunda posición, distinguimos el parámetro de calco, al que corresponden un total de tres errores (21,4 %), uno de ellos de un sintagma

nominal que ha resultado en la elección incorrecta de un término. Posteriormente, encontramos el parámetro de terminología (14,3 %), con dos elecciones erróneas de términos. Con el 7,1 % figuran los parámetros de distinto sentido, fluidez, latencia y traducción incorrecta. Conviene recordar que el error de latencia de habla está relacionado con el tiempo de desfase, que hemos considerado excesivo, al acabar provocando una descoordinación entre el audio de la interpretación al español y la imagen durante todo el discurso.

6.3.1.2. Muestra 2

En primer lugar, ofrecemos el siguiente gráfico con el objetivo de obtener una visión general y cuantitativa de los errores identificados.

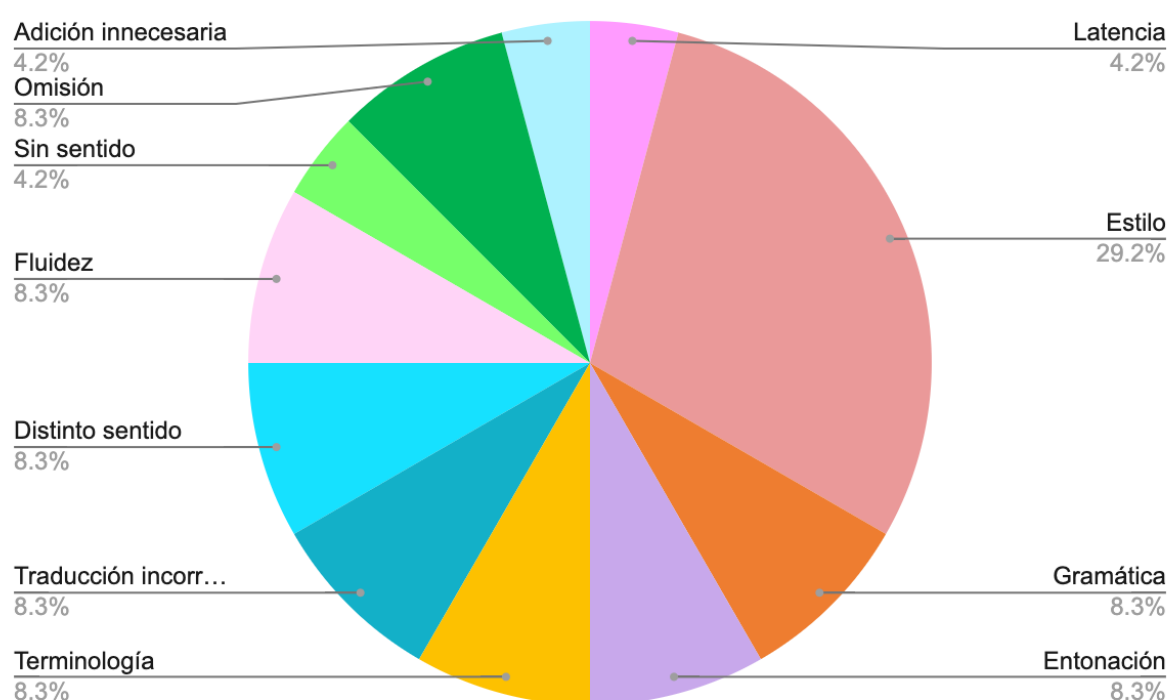


Figura 4. Gráfico de resultados de la muestra 2 (Interprefy).

La mayoría de los errores corresponden al parámetro de estilo, concretamente casi el 30 % de los errores totales. La totalidad de los errores de estilo están relacionados con el concepto de idiomática, esto es, los usos particulares y propios de una lengua. Un estilo adecuado, además de contribuir a la comprensión del mensaje, contribuye a lograr un discurso fluido y agradable para los receptores. Por tanto, en todos los casos hemos tratado de conseguir un estilo idiomático, desde la premisa de que las cuestiones de estilo dependen, en muchos casos, de las preferencias de los hablantes. En lo que se refiere a la frecuencia de aparición, son siete los parámetros que ocupan la segunda posición, con el 8,3 % cada uno: omisión, fluidez, distinto sentido, traducción incorrecta, terminología, gramática y entonación. Por un lado, diferenciamos aquellos parámetros pertenecientes a la categoría de aspectos paralingüísticos: fluidez y entonación. Más allá de no resultar agradables para los

receptores, las pausas innecesarias y, sobre todo, las entonaciones erróneas, por ejemplo, cuando no se aprecia la entonación de pregunta en los casos en los que corresponde, afectan a la propia transmisión del mensaje, al causar cierta confusión. Por otro lado, cabe mencionar, dentro de los aspectos lingüísticos, que los dos errores de omisión afectan únicamente a uno o dos elementos. En lo que se refiere a los errores de distinto sentido y traducción incorrecta, hemos detectado una falta de precisión y de coherencia con respecto a lo que establecía el texto origen. Por último, hemos detectado dos errores de concordancia de número y género que hemos asociado al parámetro de gramática, así como dos errores de terminología, por la elección errónea de las denominaciones de dos cultivos.

6.3.1.3. Comparativa de muestras en Interprefy

Una vez hemos abordado ambas muestras de forma independiente, tomaremos el siguiente gráfico para reflexionar sobre las similitudes y diferencias entre la interpretación al español de ambas muestras realizada por Interprefy.

MUESTRA 1 y MUESTRA 2

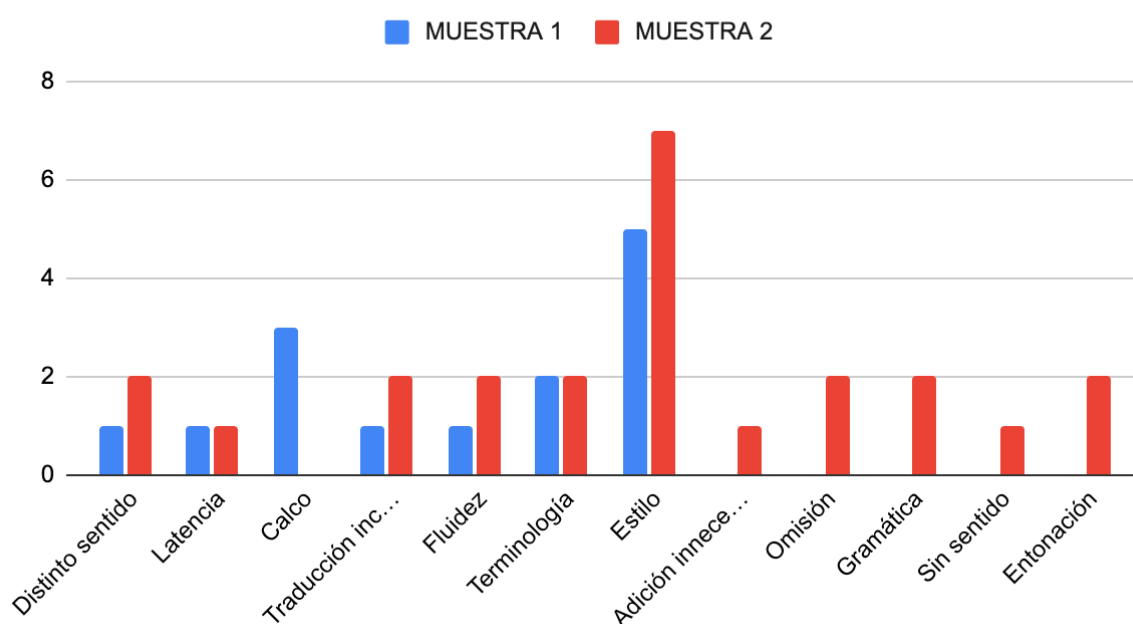


Figura 5. Gráfico para el contraste de resultados de M1 y M2 (Interprefy).

El primer aspecto que llama nuestra atención es el hecho de que los errores identificados en la muestra 2 pertenecen a una mayor variedad de parámetros, concretamente a once. Uno de los parámetros que merece la pena destacar es el parámetro de omisión, pues en la muestra 1 no se han producido omisiones, de modo que podemos afirmar que, pese a diversos desajustes en la traducción, se ha transmitido el mensaje de forma completa. Por otra parte, el único parámetro del gráfico sin presencia en la muestra 2 es el de calco, un parámetro con cierta relevancia en la muestra 1. En el caso de los parámetros que están presentes en ambas muestras, prestamos atención al estilo, dado que en ambos

casos se trata de la categoría que acumula el mayor número de errores. En este sentido, conviene tener en cuenta que en cuestiones de estilo los traductores humanos siguen siendo superiores a la traducción automática (Downie, 2020, p. 42). Hemos comprobado que este hecho se acrecienta cuando el tono del discurso tiende a la coloquialidad y al lenguaje familiar. Asimismo, en ambas muestras hemos detectado un error de latencia que ha tenido una gran relevancia en el conjunto de los discursos. Dado que el tiempo de desfase desde el inicio de los fragmentos ha sido demasiado extenso, el discurso origen y la interpretación en español han quedado descoordinados. Además, este fenómeno no solo ha provocado que la velocidad de habla se haya aumentado en demasía en algunos puntos y que no se hayan producido pausas según establecía el discurso origen, sino que también ha generado una descoordinación entre el audio y la imagen.

Una vez que hemos contrastado los resultados en Inteprefy, procedemos a replicarlos con el segundo sistema, Kudo.

6.3.2. KUDO

6.3.2.1. Muestra 1

El siguiente gráfico circular nos servirá para obtener una visión general de los errores identificados y de la frecuencia de aparición de los mismos.

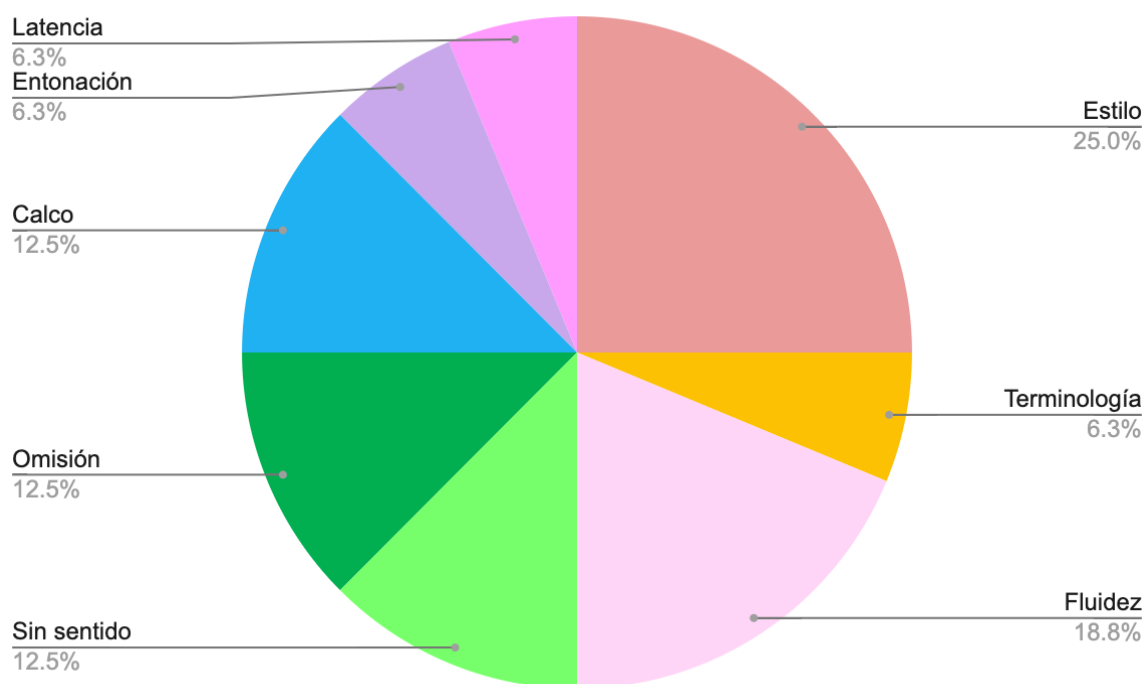


Figura 6. Gráfico de resultados de la muestra 1 (KUDO).

El 25 % de los errores pertenecen al parámetro de estilo. En la mayoría de los casos, hemos considerado que era necesario emplear un tono más formal, dada la temática de las intervenciones. Posteriormente, señalamos el parámetro de fluidez (18,8 %), al que corresponden un total de tres errores,

todos ellos debidos a una pausa innecesaria. En tercera posición, se encuentran los parámetros de calco, omisión y sin sentido (todos ellos con el 12,5 %). Resulta interesante destacar los errores de omisión y sin sentido, pues afectan en mayor medida a la transmisión del mensaje. Los dos errores de omisión abarcan una parte considerable del discurso, concretamente afectan a elementos fundamentales de las oraciones, de modo que el resultado de las mismas ha sido una formulación carente de sentido. En concreto, el segundo error de sin sentido imposibilita por completo que el receptor pueda obtener la más mínima idea que se pretende. Por último, los errores de entonación, latencia y terminología suponen cada uno el 6,7 % del total. En particular, los dos errores de entonación son de especial relevancia para la correcta transmisión del mensaje, ya que se infiere que la oración ha finalizado, cuando tiene continuación. Con respecto al error de latencia, a pesar de haber contabilizado un único error, supone una descoordinación entre el audio y la imagen durante la totalidad del fragmento.

6.3.2.2. Muestra 2

A partir del siguiente gráfico, procederemos a analizar en conjunto todos los errores identificados y la cantidad de errores correspondientes a cada parámetro.

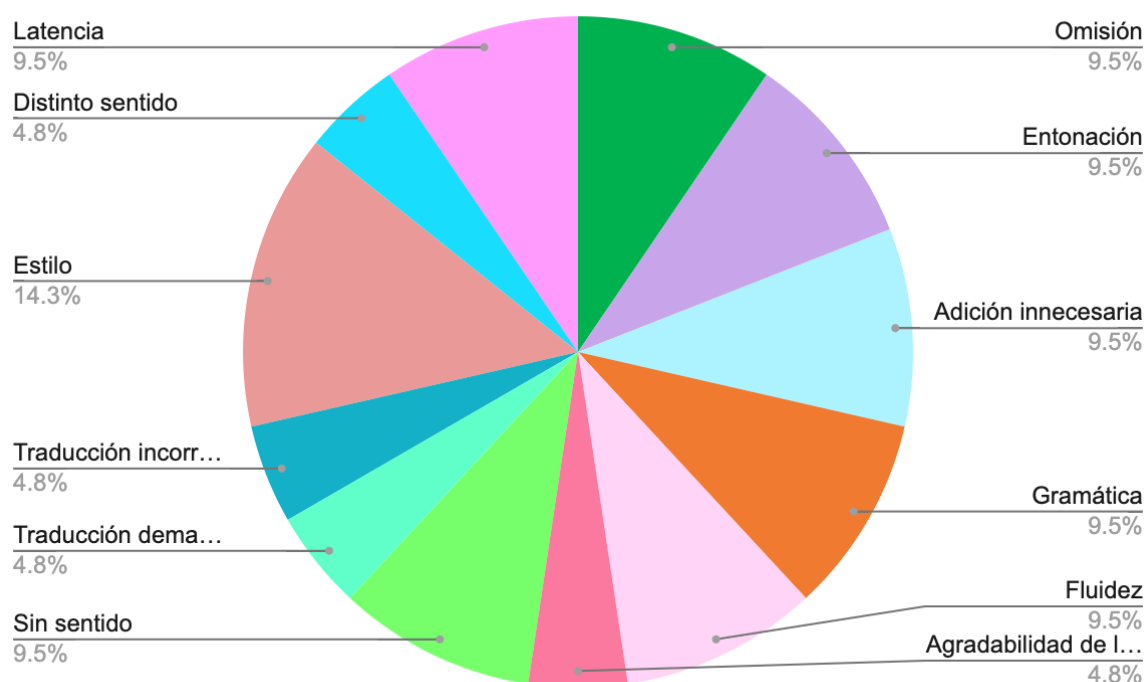


Figura 7. Gráfico de resultados de la muestra 2 (KUDO).

El porcentaje mayor corresponde al parámetro de estilo (14,3 %), más concretamente, a cuestiones relacionadas con la idiomaticidad. Teniendo en cuenta que los aspectos relacionados con el estilo pueden depender de las preferencias de cada hablante, hemos tratado de dotar a los fragmentos donde hemos localizado los errores de un estilo más idiomático y fluido a nivel discursivo. Posteriormente encontramos una serie de parámetros que comparten el 9,5 % de frecuencia de aparición: latencia, sin sentido, omisión, entonación, adición innecesaria, gramática y fluidez. Con

respecto a los errores relacionados con parámetros lingüísticos, destacamos el caso de las dos omisiones, parámetro que consideramos de gran importancia. Uno de ellos ha provocado una formulación carente de sentido, error que catalogamos como muy grave, dado que supone un obstáculo importante para la transmisión correcta y completa del mensaje. En la línea de los aspectos relacionados con el propio proceso de traducción, encontramos un error de traducción incorrecta especialmente grave, pues transmite una información incorrecta y totalmente inconexa con el resto del discurso, así como un error de traducción demasiado literal, que, además, resta idiomatización al discurso, y un error de distinto sentido, en el que hemos identificado una falta de precisión. Por otro lado, diferenciamos dos errores de gramática, relacionados con la concordancia de género y número, y dos errores de adición innecesaria, uno de ellos vinculado a la particularidad de la lengua española de no incluir el sujeto en forma de pronombre personal de forma explícita.

6.3.2.3. Comparativa de muestras en KUDO

Tras haber comentado de forma independiente los gráficos de cada una de las muestras, prestaremos atención al siguiente diagrama de barras para comparar los resultados de ambas muestras con la traducción voz a voz proporcionada por KUDO.

MUESTRA 1 y MUESTRA 2

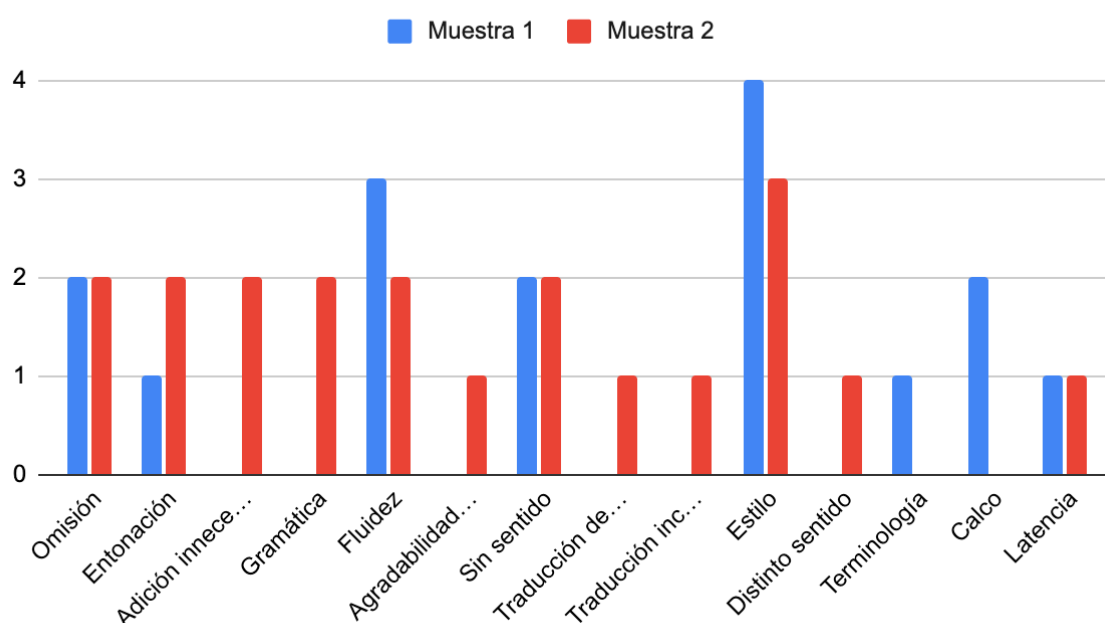


Figura 8. Gráfico para el contraste de resultados de M1 y M2 (KUDO).

En primer lugar, observamos que el parámetro que reúne el mayor número de errores en ambas muestras se corresponde con el estilo. A diferencia de lo que apreciamos en el caso de las muestras de Interprefy, la muestra 1, que es la de mayor grado de especialización, es la que presenta un mayor número de errores de estilo, en comparación con la muestra 2, que en la que el empleo del lenguaje familiar debería suponer una dificultad a la hora de conseguir un tono discursivo fluido e idiomático.

Por otra parte, observamos que existen omisiones en ambos casos y merece la pena destacar que en todos los casos se trata de omisiones de una buena parte del contenido, lo cual ha generado dos errores de sin sentido en ambos casos. Por tanto, podemos afirmar que en ambas muestras la transmisión completa y correcta del mensaje se ha visto negativamente afectada. Otros de los parámetros que están presentes en las dos muestras es el de entonación, con una diferencia de un error más en la muestra 2. En ambos casos, hemos diferenciado un error por dar tono de fin de oración en el momento equivocado. Este parámetro está estrechamente relacionado con el parámetro de fluidez, que también tiene presencia en ambas muestras en forma de pausas innecesarias o de ausencia de pausas necesarias. Por otra parte, los parámetros de gramática, adición innecesaria, traducción demasiado literal, agradabilidad de la voz y distinto sentido están presentes únicamente en la muestra 2, mientras que hemos detectado errores de terminología y calco solo en la muestra 1. Merece la pena destacar que la cuestión de agradabilidad de la voz es la más subjetiva dentro de los parámetros del modelo, junto con el estilo, y convendría analizarla según el conjunto de la intervención. Sin embargo, hemos estimado oportuno señalar este error en concreto, dado que nos parece que la voz en este fragmento no se asemeja en absoluto a la voz humana y no resulta natural, que es uno de los objetivos de estos sistemas de traducción de voz a voz.

6.3.3. Contraste de resultados entre sistemas (Interprefy y KUDO)

En el presente estudio, hemos recurrido a dos sistemas de traducción de voz a voz diferentes, con el objetivo de poder realizar una comparación de los resultados y obtener unas conclusiones lo más sólidas posible. En este contexto, procedemos en este punto a reflexionar de forma comparativa sobre los resultados obtenidos por parte de Interprefy y KUDO.

Si atendemos a un aspecto tan determinante para el éxito del acto de comunicación como es la transmisión completa del mensaje y fiel a nivel informativo al texto origen, debemos reconocer que los errores de omisión y de sin sentido que hemos observado en las dos interpretaciones aportadas por KUDO han provocado una importante falta de información, acompañada de unas formulaciones con elementos inconexos. Por ejemplo, si tomamos la muestra 2 de KUDO, comprobaremos que al receptor le resultaría imposible obtener una mínima idea de la información que se pretende transmitir. Por tanto, no podría considerarse que el acto de comunicación se desarrolla con éxito.

Dado que uno de los objetivos de este tipo de sistemas de IA es que el discurso que se obtiene resulte natural y agradable a nivel lingüístico y paralingüístico, las cuestiones relacionadas con el estilo se convierten en especialmente relevantes. No obstante, merece la pena recordar que todo lo relacionado con el estilo es profundamente subjetivo y puede dar lugar a debate. En cualquier caso, es posible afirmar que el estilo ha sido un asunto de gran peso en el análisis de los resultados de ambos sistemas. En el caso de la muestra 1, hemos tratado de dotar al discurso de un tono más formal y propio del lenguaje científico, a pesar de que el nivel de especialización del fragmento no era excesivamente alto. Por otra parte, en el caso de la muestra 2, se ha tratado de adaptar las propuestas a un estilo más

idiomático y próximo a los rasgos del lenguaje coloquial de la lengua española, por tratarse de una conversación distendida. En definitiva, podemos afirmar que tanto en el caso de Interprefy, como en el caso de KUDO los aspectos relacionados con el estilo requieren mejoras.

Otros de los aspectos en los que hemos fijado nuestra atención son los parámetros de traducción literal y calco. En el caso de los errores de calco, hemos observado que se producen en los mismos puntos y con las mismas soluciones en ambos sistemas. Asimismo, el error terminológico de la denominación de la bacteria que se menciona en la muestra 1 también se ha repetido en ambos sistemas.

Por lo que respecta a aspectos paralingüísticos, destacamos el error de latencia que se produce al principio de los cuatro fragmentos como consecuencia de un tiempo de desfase excesivamente largo y que hace que la velocidad de habla sufra desajustes constantes. Más allá de que la velocidad de habla aumente o se desacelere, sin que así suceda en el discurso original, el audio y la imagen no están coordinados, lo cual puede crear confusión al receptor, por ejemplo, a la hora de saber cuál de los interlocutores interviene en cada momento, teniendo en cuenta que se mantiene la misma voz durante toda la interpretación. Este inconveniente se acentúa en la muestra 2 en ambos casos, al tratarse de una conversación. Este problema que afecta a la totalidad de la duración de los fragmentos hace que se produzca un discurso menos agradable, a lo que se suman los errores de fluidez y de entonación. En particular, aunque encontramos un mayor número de errores de entonación en los resultados de KUDO, este tipo de errores son en todos los casos relevantes, puesto que la entonación conlleva una gran carga de significado y en la mayoría de los casos el tono del discurso original no se reproduce en el discurso meta.

Tras finalizar el proceso de análisis de los resultados, en el siguiente capítulo procederemos a exponer las conclusiones y a reflexionar sobre los resultados obtenidos en el presente estudio.

7. Conclusiones

Una vez finalizado el proceso de análisis, podemos afirmar que los resultados obtenidos de los dos sistemas de traducción de voz a voz son aceptables. De hecho, en el caso del sistema Interprefy, empleado también en un estudio previo en junio de 2024 (Calvo del Barrio, 2024; Calvo del Barrio y Ortego Antón, en prensa), hemos observado mejoras sustanciales en los resultados. No obstante, hay una serie de aspectos en ambos casos que están sujetos a posibles implementaciones. En consecuencia, podemos confirmar que hemos alcanzado tanto el objetivo principal propuesto (verificar la calidad de los productos resultantes de los sistemas de traducción de voz a voz), como los tres objetivos secundarios: elaborar un modelo de evaluación, que hemos puesto en práctica y nos ha servido para detectar los errores en los productos resultantes de los sistemas de traducción de voz a voz y, en consecuencia, obtener una aproximación a la evolución de la IA, AP y PLN aplicado a los servicios lingüísticos.

En primer lugar, centraremos nuestra atención en los aspectos relativos a los parámetros lingüísticos, que consideramos que están relacionados con el proceso de traducción automática, que es el segundo de los tres que tienen lugar en los sistemas de traducción de voz a voz en cascada. En el caso del previo realizado en 2024 (Calvo del Barrio, 2024; Calvo del Barrio y Ortego Antón, en prensa), concluimos que el fenómeno de las formulaciones carentes de sentido era de especial relevancia, sobre todo, por lo que respecta a la transmisión completa y adecuada del mensaje. Por consiguiente, en el baremo presentado en este trabajo decidimos incluir un parámetro independiente denominado «sin sentido», partiendo de la hipótesis de que iba a surgir un número importante de errores relacionados con este parámetro. Finalmente, ninguno de los productos ha presentado una cantidad considerable de errores de sin sentido. No obstante, en los casos en los que sí hemos identificado errores correspondientes a este parámetro, estos han supuesto un problema considerable para la transmisión del mensaje. Además de los problemas directamente relacionados con el motor de traducción automática, nos inclinamos a que la información no ha llegado, en algunos casos, de forma completa a este segundo proceso de traducción automática, puesto que el tiempo de desfase al comienzo de la traducción de voz a voz al español de ambas muestras ha sido excesivo. En consecuencia, sospechamos que se han originado desajustes a nivel de reconocimiento de voz. Por tanto, es posible confirmar el carácter acumulativo de los errores en los sistemas en cascada, de modo que destacamos la importancia de un buen rendimiento en el proceso de reconocimiento automático de voz, desde la idea de que si la calidad de lo que se proporciona al sistema de traducción automática desde el sistema de reconocimiento automático no es óptima, no se puede esperar que la calidad del producto final de la traducción sea aceptable (Downie, 2020, p. 38-49).

Por otro lado, en el estudio anterior (Calvo del Barrio, 2024; Calvo del Barrio y Ortego Antón, en prensa), buena parte del total de errores en ambas muestras correspondieron a parámetros

paralingüísticos. Reconocimos que, además de resultar fundamentales para lograr un discurso agradable para el receptor, los aspectos paralingüísticos albergan una gran carga de significado y son de gran relevancia a la hora de transmitir una idea. En caso de no detectar correctamente la idea o el significado que pretende transmitir una determinada entonación, un determinado tono o, incluso, los gestos o las expresiones faciales, pueden producirse ambigüedades. Los errores relacionados con estos parámetros han tenido una presencia y una relevancia similar en este estudio. Con el objetivo de que estos sistemas puedan detectar y trabajar correctamente con distintas entonaciones y tonos, convendría entrenarlos con una gran variedad de acentos y discursos en los que se aprecie la diversidad de entonaciones (Downie, 2020, p. 38-49). Si bien estos aspectos resultan esenciales a nivel informativo, su relevancia es igualmente destacable a la hora de producir un discurso agradable. El tono de voz robótico y las pausas innecesarias son precisamente los aspectos que restan agradabilidad en la voz, un aspecto que concierne directamente a la experiencia del receptor cuando recurre a un servicio de traducción de voz a voz con IA. En esta línea, respetar las pausas que tienen lugar entre intervenciones y emplear voces diferentes ayudaría a los receptores a seguir el discurso más fácilmente y a diferenciar claramente quién interviene en cada momento, lo cual mejoraría la experiencia del receptor.

Tras llevar a cabo el análisis, hemos confirmado la elección del término «traducción de voz a voz», dado que la manera de proceder de estos sistemas, que no toman en consideración aspectos fundamentales en un acto de comunicación, como el contexto social y cultural, no coincide con lo que los humanos consideramos «interpretar» (Downie, 2020, p. 37-38). Ciertos errores detectados en los productos resultantes nos llevan a deducir que, efectivamente, estos sistemas no llevan a cabo un proceso de interpretación propiamente. Es el caso del empleo del pronombre «usted», reservado para expresar cortesía en situaciones formales, en un contexto de una conversación distendida y coloquial. En definitiva, los grandes retos de la comunicación humana multilingüe hacen fundamental un proceso de interpretación en el que se tengan en cuenta una serie de aspectos fundamentales a la hora de comprender el discurso origen y de dar lugar a una traducción adecuada para el público y el contexto meta. Los grandes retos del lenguaje humano, más allá de suponer que estos sistemas puedan presentar un mejor rendimiento cuando se trata de oraciones directas y sencillas, ponen en juego una serie de limitaciones técnicas que la tecnología aún no ha sido capaz de resolver. En el caso concreto de este estudio, haber tomado en consideración la imagen hubiera servido de gran ayuda, por ejemplo, a la hora de seleccionar los términos correctos para los alimentos que se mostraban. Por tanto, podemos afirmar que los profesionales humanos son una pieza fundamental para la comunicación, dado que son capaces de proporcionar la calidad y la precisión que requieren los servicios de interpretación, sobre todo, en situaciones delicadas y complejas que requieren de decisiones responsables y éticas (Jiménez Serrano, 2025, p. 154-155).

The answer to the question: “will AI replace language interpreters?” is not a simple yes or no. While AI technology has made significant strides in recent years, it seems that

it is not yet capable of replacing human interpreters entirely. At least not in industries that place a high value on accuracy and cultural sensitivity. However, for business that are looking to expand globally and reduce interpretation costs, AI-powered solutions may be the answer (Zabetian, 2024).

Tras haber examinado el potencial de estos sistemas de traducción de voz a voz y haber reconocido la necesidad de controlar la calidad de los resultados que ofrecen para garantizar el correcto desarrollo de los actos de comunicación, consideraríamos muy interesante seguir muy de cerca la evolución de la IA y los avances que esta evolución genere en estos sistemas. Asimismo, nos gustaría volver a abordar y analizar el rendimiento de estos sistemas con el objetivo de seguir refinando el modelo de evaluación a partir de nuevas metodologías, nuevas muestras y un mayor número de sistemas que puedan sumar precisión y adecuación a las nuevas propuestas. En particular, despierta gran curiosidad conocer cómo abordan estos sistemas aspectos relacionados con el humor y la ironía.

Por último, nos gustaría finalizar este trabajo con la siguiente reflexión: «el binomio intérprete-tecnología será el artífice principal del gran reto y la gran oportunidad de transformación que supone para el sector la aplicación y la integración de la IA en todos sus procesos y flujos de trabajo. Y es que el futuro no ha hecho más que comenzar...» (Corpas Pastor, 2025, p. 127).

8. Bibliografía

- AHMED AHMED, S. (2022). «Technology and artificial intelligence in simultaneous interpreting: a multidisciplinary approach». *Occasional Papers*, 78. https://opde.journals.ekb.eg/article_249945_b9781cf0b9712533721caeeb6fe0ff2c.pdf (Fecha de consulta: 28/12/2024).
- CALVO DEL BARRIO, S. (2024). *La calidad en interpretación automática inglés-español en el ámbito agroalimentario: un estudio de caso*. Trabajo de Fin de Grado. Soria: Universidad de Valladolid. <https://uvadoc.uva.es/handle/10324/73158>. (Fecha de consulta: 25/05/2025).
- CALVO DEL BARRIO, S. Y ORTEGO ANTÓN, M.T. (En prensa). «La calidad en interpretación automática inglés-español en el ámbito agroalimentario: una propuesta de modelo de evaluación». En M.M. Rivas Carmona y M.C. Aguilar Camacho (Eds.), *Enfoques multidisciplinares de lengua, literatura y traducción especializada*. Berlín: Peter Lang.
- CATTONI, R., DI GANGI, M. A., BENTIVOGLI, L., NEGRI, M. Y TURCHI, M. (2021). «MuST-C: A multilingual corpus for end-to-end speech translation». *Computer Speech and Language*, 66, p. 1-14.
- CHEN, S., DOHERY, S. (2024). «Interpreting and technologies». En MELLINGER, C. D. (Ed.). *The Routledge Handbook of Interpreting and Cognition*. Abingdon: Routledge.
- COLLADOS AÍS, Á. (2007). *La evaluación de la calidad en interpretación simultánea. La importancia de la comunicación no verbal*. Granada: Editorial Comares.
- COMISIÓN EUROPEA. (2024). *Ley de IA*. Comisión Europea. <https://digital-strategy.ec.europa.eu/es/policies/regulatory-framework-ai> (Fecha de consulta: 20/02/2025).
- CORPAS PASTOR, G. (2021). «Technology solutions for interpreters: the VIP system». *Hermēneus. Revista de Traducción e Interpretación*, 23, 91-123.
- CORPAS PASTOR, G. Y DEFRANQ, B. (2023). *Interpreting Technologies — Current and Future Trends*. Ámsterdam / Filadelfia: John Benjamins.
- CORPAS PASTOR, G., Y GABER, M. (2024). «La tecnología del reconocimiento automático del habla: recurso de documentación y apoyo tecnológico para la docencia en interpretación». En CORPAS PASTOR, G. Y SEGHIRI, M. (Eds.). *Tecnología e Interpretación: nuevos horizontes didácticos y profesionales*. Granada: Editorial Comares, p. 120-131.

- CORPAS PASTOR, G. (2025). «Anexo: La voz de los expertos: el futuro inmediato en una página». En JIMÉNEZ SERRANO, O. *La interpretación (,) a distancia. El futuro inmediato de la profesión*. Granada: Comares.
- DAVITTI, E., KORYBSKI, T. Y BRAUN, S. (2025). «The Routledge Handbook of Interpreting, Technology and AI». Oxon / New York: Routledge, Taylor & Francis Group.
- DOWNIE, J. (2020). *Interpreters vs Machines. Can Interpreters Survive in an AI-Dominated World?* Londres: Routledge.
- EBY, H., GREEN, C. Y HYLAK, B. (2024). *Human vs. Machine Interpreting Today: A Professional Assessment*. https://www.ata-divisions.org/ID/wp-content/uploads/2024/09/20240906_machine_interpreting_today_v2.pdf (Fecha de consulta: 06/03/2025).
- FALEMPIN, A. Y RANADIRESKA, D. (2024). «Human vs. Machine: The Future of Translation in an AI-Driven World». En *Proceedings of the Widyatama International Conference on Engineering 2024 (WICOENG 2024)*. <https://www.atlantispress.com/proceedings/wicoeng-24/126007241> (Fecha de consulta: 06/03/2025).
- FANTINUOLI, C. (2018). «Interpreting and Technology». En FANTINUOLI, C. (2018). *Interpreting and Technology*, 11. Berlin: Language Science Press, p. 1-12.
- FANTINUOLI, C. (2023a). «Towards AI-enhanced computer assisted interpreting». En CORPAS PASTOR, G. Y DEFRANQ, B. (2023). *Interpreting Technologies — Current and Future Trends*. Ámsterdam / Filadelfia: John Benjamins, p. 47-71.
- FANTINUOLI, C. (2023b). «The Emergence of Machine Interpreting». *European Society for Translation Studies*, 62. https://www.claudiofantinuoli.org/docs/ESTNL_May_2023.pdf (Fecha de consulta: 29/12/2024).
- FANTINUOLI, C. (2023c). «Machine Bias». *EasyAI*. <https://easyai.uni-mainz.de/html/bias.html#id21> (Fecha de consulta: 02/03/2025)
- FANTINUOLI, C. (2023d). «Machine Interpreting». *EasyAI*. <https://easyai.uni-mainz.de/html/machine-interpreting.html> (Fecha de consulta: 01/03/2025)
- FANTINUOLI, C. (2023e). «Machine Learning». *EasyAI*. <https://easyai.uni-mainz.de/html/what-is-machine-learning.html> (Fecha de consulta: 01/03/2025)
- FANTINUOLI, C. (2023f). «What is NLP». *EasyAI*. <https://easyai.uni-mainz.de/html/what-is-nlp.html> (Fecha de consulta: 01/03/2025)

- GOLDSMITH, J. (2023). «Tablet interpreting. A decade of research and practice». En CORPAS PASTOR, G. Y DEFRANQ, B. (2023). *Interpreting Technologies — Current and Future Trends*. Ámsterdam / Filadelfia: John Benjamins, p-27-45.
- HOLGADO, R. (2024). «Dispositivos móviles y auriculares para traducir al instante otro idioma con la ayuda de la IA». *20 minutos*. <https://www.20minutos.es/tecnologia/moviles-dispositivos/dispositivos-moviles-para-traducir-instante-otro-idioma-5220170/> (Fecha de consulta: 22/05/2025).
- ICEX España Inversiones y Exportaciones. (s.f.). *Industria agroalimentaria*. <https://www.investinspain.org/es/sectores/industria-agroalimentaria> (Fecha de consulta: 01/03/2025)
- INTERPREFY. (s.f.). *Your next-generation AI speech translator*. <https://www.interprefy.com/solutions/event-ai-speech-translator> (Fecha de consulta: 08/04/2024)
- INTERPREFY. (29 de agosto de 2022). *How does remote simultaneous interpreting work?* <https://www.interprefy.com/resources/blog/how-does-remote-simultaneous-interpretation-work> (Fecha de consulta: 15/05/2024).
- JIMÉNEZ SERRANO, O. (2025). *La interpretación (,) a distancia*. Granada: Editorial Comares.
- KENT STATE UNIVERSITY. (s.f.). *The Difference Between Machine Translation and Computer-Assisted Translation*. Kent State University. <https://www.kent.edu/mcls/blog/machinetranslationvscat> (Fecha de consulta: 08/04/2024)
- KUDO. (s.f.). *KUDO AI Speech Translator*. <https://kudo.ai/> (Fecha de consulta: 08/04/2024)
- LIMA FLORIDO, F. J. Y CORPAS PASTOR, G. (2024). «Posibilidades (y limitaciones) de los asistentes conversacionales como recursos tecnológicos para la interpretación: ChatGPT y Gemini frente a VIP». En CORPAS PASTOR, G. Y VEREDAS NAVARRO, F. J. (Eds.). *Tecnologías lingüísticas multilingües. Desarrollos actuales y transición digital*. Granada: Editorial Comares.
- MINISTERIO DE AGRICULTURA, PESCA Y ALIMENTACIÓN. (2024). *Avance anual de comercio exterior 2023*. [Infografía]. <https://www.mapa.gob.es/es/prensa/ultimas-noticias/las-exportaciones-agroalimentarias-y-pesqueras-alcanzaron-en-2023-su-cifra-r%C3%A9cord-70.431-millones-de-euros/tcm:30-677513#prettyPhoto> (Fecha de consulta: 01/03/2025)
- MQM. (2025). *The MQM Error Typology*. <https://themqm.org/error-types-2/typology/> (Fecha de consulta: 28/04/2025)

- NETFLIX. (2024). *Somos lo que comemos*. <https://www.netflix.com/es/title/81133260> (Fecha de consulta: 29/04/2025)
- O'BRIEN, S. Y RODRÍGUEZ VÁZQUEZ, S. (2020). «Translation and technology». En LAVIOSA, S. Y GONZÁLEZ-DAVIES, M. *The Routledge Handbook of Translation and Education*. (Eds.). Oxon / New York: Routledge, p. 264-267.
- ORLANDO, M. (2023). «Using smartpens and digital pens in interpreter training and interpreting research. Taking stock and looking ahead». En CORPAS PASTOR, G. Y DEFRANQ, B. (2023). *Interpreting Technologies — Current and Future Trends*. Ámsterdam / Filadelfia: John Benjamins, p. 7-26.
- ORTEGO ANTÓN, M.T. (2019). *La terminología del sector agroalimentario (español-inglés) en los estudios contrastivos y de traducción especializada basados en corpus: los embutidos*. Berlín: Peter Lang Alemania, p. 25-28.
- ORTEGO ANTÓN, M.T. (2025). «Cooking up a storm: Assessing machine interpreting of marinated meat preparation». En *VIII International Congress Science and Translation*. Innsbruck, Innsbruck Universitaat, 26-28 March 2025.
- ORTIZ BOIX, C. (2016). *Implementing Machine Translation and Post-Editing to the Translation of Wildlife Documentaries through Voice-over and Off-screen Dubbing. A Research on Effort and Quality*. Tesis doctoral. Barcelona: Universidad Autónoma de Barcelona. <https://tesisenred.net/bitstream/handle/10803/400020/cob1de1.pdf?sequence=1&isAllowed=y> (Fecha de consulta: 28/04/2025).
- POLICASTRO PONCE, G. (2017). «La taxonomía textual de la industria alimentaria desde una perspectiva traductológica». *Futhark: revista de investigación y cultura*, 12, p. 79-88. <https://dialnet.unirioja.es/servlet/articulo?codigo=7915577> (Fecha de consulta: 01/03/2025)
- PRANDI, B. (2023). «Computer-assisted simultaneous interpreting: A cognitive-experimental study on terminology». Berlín: Language Science Press.
- RODRÍGUEZ RODRÍGUEZ, F. (2013). «La traducción en el sector agroalimentario: una simbiosis en auge». *Skopos: revista internacional de traducción e interpretación*, 2, p. 155-172. <https://doi.org/10.21071/skopos.v2i.4423> (Fecha de consulta: 06/03/2025).
- STENGERS, H., LÁZARO GUTIÉRREZ, R. Y KERREMANS, K. (2023). «Public service interpreters' perceptions and acceptance of remote interpreting technologies in times of a pandemic». En CORPAS PASTOR, G. Y DEFRANQ, B. (2023). *Interpreting Technologies — Current and Future Trends*. Ámsterdam / Filadelfia: John Benjamins, p. 109-136.

- ZABETIAN, F. (2024). «KUDO CEO Takes On the ‘AI vs Human Language Interpreters’ Debate». *KUDO*. <https://kudo.ai/blog/ai-vs-human-language-interpreters-a-side-by-side-video-comparison/> (Fecha de consulta: 31/05/2025)
- ZIEGLER, K. Y GIGLIOBIANCO, S. (2018). «Present? Remote? Remotely present! New technological approaches to remote simultaneous conference interpreting». En FANTINUOLI, C. (Ed.), *Interpreting and technology*. Berlín: Language Science Press, p. 119–139. DOI:10.5281/zenodo.1493299 (Fecha de consulta: 06/03/2025).