

Codificación y anotación preliminar de un corpus oral multilingüe de conversaciones telefónicas interpretadas para el estudio de los ataques a la imagen¹

MARCELO-YUJI HIMORO

Universidad Nacional de Educación a Distancia (UNED)
mhimoro1@alumno.uned.es
<https://orcid.org/0000-0002-7591-0354>

ANTONIO PAREJA-LORA

Universidad de Alcalá
antonio.parejal@uah.es
<https://orcid.org/0000-0001-5804-4119>

CODIFICACIÓN Y ANOTACIÓN PRELIMINAR DE UN CORPUS ORAL MULTILINGÜE DE CONVERSACIONES TELEFÓNICAS INTERPRETADAS PARA EL ESTUDIO DE LOS ATAQUES A LA IMAGEN

CONSTRUCTION AND PRELIMINARY ANNOTATION OF A TELEPHONE INTERPRETATION MULTILINGUAL SPOKEN CORPUS FOR THE STUDY OF FACE-THREATENING ACTS

CONSTITUTION ET ANNOTATION PRÉLIMINAIRE D'UN CORPUS ORAL MULTILINGUE D'ENTRETIENS TÉLÉPHONIQUES INTERPRÉTÉS POUR L'ÉTUDE DES ACTES DE LANGAGE MENAÇANTS LA FACE

RESUMEN: El crecimiento y la consolidación de la demanda de servicios de interpretación telefónica ha traído consigo un mayor estudio de los mismos en el entorno académico. El objetivo de este trabajo ha sido crear un corpus oral de interacciones telefónicas mediadas por intérpretes, orientado en concreto al estudio de los ataques contra la imagen. Estas interacciones, que siempre incluyen el español, se realizan también en alemán, chino, francés, inglés o ruso. Primeramente se describen brevemente los hitos alcanzados en este sentido en trabajos anteriores: la recopilación de las grabaciones anonimizadas de las conversaciones, su procesamiento inicial, su transcripción y su traducción. En segundo lugar, se detalla el proceso de conversión de las transcripciones al formato EXMARaLDA y su posterior sincronización con las grabaciones. Para terminar, se discuten las limitaciones y dificultades encontradas en estos procesos de conversión y sincronización.

ABSTRACT: The growth and consolidation of the demand for telephone interpreting services has been attracting an increasing attention of the academia to its study lately. The primary goal of this research has been to build a spoken corpus of interpreter-mediated telephone conversations geared towards the study of face-threatening acts (FTAs). These conversations involve Spanish and an additional language such as German, Chinese, French, English or Russian. Thus, firstly, we briefly outline the milestones reached in previous research, namely the compilation of anonymized conversation audio recordings, as well as their initial processing, transcription and translation. Secondly, we describe in detail the conversion of the transcripts to the EXMARaLDA format and its synchronization with the corresponding recordings. Finally, we discuss the restrictions and the challenges faced in these conversion and synchronization processes.

RÉSUMÉ: La croissance et la consolidation de la demande d'interprétation téléphonique attirent davantage l'intérêt des chercheurs sur son étude. Le but principal de cette recherche est de constituer un corpus oral de conversations téléphoniques médiées par interprètes, et orienté vers l'étude des actes de langage menaçants la face (FTA). Ces entretiens comprennent l'espagnol et une deuxième langue, c'est-à-dire, l'allemand, le chinois, le français, l'anglais ou le russe. Tout d'abord, on revoit brièvement les jalons atteints dans les travaux antérieurs, à savoir : la compilation des conversations enregistrées et anonymisées, leur traitement initial, et aussi leur transcription et traduction. Ensuite, on décrit en détail la conversion des transcriptions au format EXMARaLDA et sa synchronisation avec les enregistrements. Enfin, on discute des restrictions et des défis rencontrés dans ces processus de conversion et de synchronisation.

PALABRAS CLAVE: interpretación telefónica; corpus orales; anotación pragmática; herramientas de anotación; EXMARaLDA.

KEYWORDS: telephone interpreting; spoken corpus; pragmatic annotation; annotation tools; EXMARaLDA.

MOTS-CLÉS: interprétation téléphonique; corpus oraux; annotation pragmatique; outils d'annotation; EXMARaLDA.

Fecha de recepción: 15/12/2021

Fecha de revisión: 26/02/2022

Fecha de aceptación: 09/05/2022

Fecha de publicación: 01/12/2022

¹ Este estudio forma parte de los proyectos "Análisis de ataques contra la imagen en interpretación telefónica" (CM/JIN/2019-040), financiado por la Comunidad de Madrid; "Pragmática de corpus e interpretación telefónica: análisis de ataques contra la imagen en interpretación telefónica: análisis de ataques contra la imagen (PRAGMACOR)", financiado por el Ministerio de Ciencia e Innovación, programa de Generación del Conocimiento (PID2021-127196NA-I00); y LITHME (Language in the Human-Machine Era, COST Action CA19102), financiado por el Programa Marco Horizonte 2020. Los autores también desean agradecer al grupo de investigación FITISPos (Formación e Investigación en Traducción e Interpretación en los Servicios Públicos, UAH) su continuo soporte en el desarrollo de esta investigación.

SUMARIO: 1. Introducción. 2. Antecedentes y estado de la cuestión. 2.1. Corpus orales. 2.2. Anotación pragmática. 2.3. Actos del habla y ataques contra la imagen. 2.4. Herramientas de construcción de corpus orales. 2.5. Descripción del proyecto y de las tareas anteriores. 2.5.1. Paso 1: Obtención y codificación de los datos. 2.5.2. Pasos 2, 3 y 4: Creación de archivos digitales, selección de los datos y transcripción. 2.5.3. Paso 5: Traducción. 3. Objetivos y metodología. 4. Desarrollo. 4.1. Definición de los parámetros de configuración de las transcripciones. 4.2. Conversión de las transcripciones. 4.3. Sincronización de las transcripciones con sus ficheros de audios respectivos. 5. Limitaciones y trabajos futuros. 6. Conclusiones. 7. Referencias.

SUMMARY: 1. Introduction. 2. Background and state of the art. 2.1. Spoken corpora. 2.2. Pragmatic annotation. 2.3. Speech acts and face-threatening acts. 2.4. Spoken corpus construction tools. 2.5. Project description and previous work. 2.5.1. Step 1: Data compilation and encoding. 2.5.2. Steps 2, 3 and 4: Digital archives creation, data selection and transcription. 2.5.3. Steps 5: Translation. 3. Goals and methodology. 4. Development. 4.1. Transcript configuration settings. 4.2. Transcripts conversion. 4.3. Transcripts synchronization with the corresponding audio files. 5. Restrictions and future work. 6. Conclusions. 7. References.

SOMMAIRE: 1. Introduction. 2. Contexte et état de l'art. 2.1. Corpus oraux. 2.2. Annotation pragmatique. 2.3. Actes de langage et actes menaçants. 2.4. Outils de construction de corpus oraux. 2.5. Description du projet et travaux antérieurs. 2.5.1. Tape 1: Compilation des données et codification. 2.5.2. Tapes 2, 3 et 4: Création d'archives numériques, sélection des données et transcription. 2.5.3. Tape 5: Traduction. 3. Objectifs et méthodologie. 4. Développement. 4.1. Définition des paramètres de configuration des transcriptions. 4.2. Conversion des transcriptions. 4.3. Synchronisation des transcriptions avec leurs fichiers audio correspondants. 5. Restrictions et travaux futurs. 6. Conclusions. 7. Références.

1. Introducción

Estamos inmersos en un mundo globalizado en el que, sobre todo gracias a la tecnología, cada vez se difuminan más las fronteras y las distancias parecen acortarse enormemente. De esta forma, muchos servicios pasan a ser ofrecidos de manera descentralizada y telemática, lo cual permite no solo reducir costes, sino también ampliar enormemente su oferta y alcance, que ya no dependen de la existencia de oficinas físicas en diferentes localidades. Es en un contexto así que la interpretación telefónica va conquistando cada vez más espacio en la industria de los servicios lingüísticos, ya que posibilita ofrecer interpretación en gran número de lenguas de manera remota, sin necesidad de que las partes (los interlocutores y el intérprete) compartan un mismo espacio físico.

A pesar de que se usa desde los años 1970 y 1980 en países como Australia, Japón y EE.UU., su implantación es relativamente reciente en Europa, adonde no llega hasta los años 90, para extenderse a comienzos de los años 2000 (Ozolins, 1998; Phelan, 2001; Lázaro Gutiérrez, 2019). Probablemente, las claves principales de su éxito sean la mejora, omnipresencia y abaratamiento de las infraestructuras y servicios de telecomunicación, además de su utilidad en situaciones de emergencia, o bien como se ha observado en tiempos recientes con la pandemia del COVID-19, la posibilidad de seguir garantizando la oferta de servicios de interpretación aún en situaciones de contacto restringido. Por esta razón, pese a que la investigación en este campo es todavía incipiente, es más que posible que esta situación cambie radicalmente en los próximos años.

De hecho, uno de los aspectos que más atraen la atención de los investigadores de este campo es el estudio de los elementos pragmáticos que intervienen en el discurso de las interpretaciones telefónicas y cómo afectan a este tipo de comunicación especializada. En esta línea, la investigación que se describe en este artículo se ha centrado en la preparación de un corpus oral de interpretaciones telefónicas para el estudio de un conjunto particular de elementos pragmáticos, que son los ataques a la imagen (FTA, por sus siglas en inglés, *face-threatening acts*). Cómo facilitar este estudio mediante el uso del corpus en preparación es materia de investigaciones en curso que se detallan en otros artículos de este volumen (Lázaro Gutiérrez y Alcalde Peñalver, 2022; Pena Díaz, 2022; Valero Garcés y Li, 2022, y Vitalaru, 2022), por lo que queda fuera del ámbito de este trabajo.

Así, el resto del artículo está organizado de la siguiente manera. El apartado 2 describe los antecedentes relevantes para esta investigación, el proyecto en el cual está encuadrada y qué hitos se han alcanzado en trabajos anteriores. En el apartado 3 se presentan los objetivos y la metodología de investigación. El apartado 4 proporciona una descripción de la investigación realizada en este trabajo. En el apartado 5 se analizan las limitaciones del mismo y algunas de las dificultades encontradas y da cuenta de posibles trabajos futuros. Y, finalmente, el apartado 6 aborda las conclusiones finales de esta investigación.

2. Antecedentes y estado de la cuestión

En este apartado se presentan algunos conceptos fundamentales y trabajos anteriores relevantes para esta investigación. El subapartado 2.1 trata de los corpus orales. El subapartado 2.2 resume brevemente el estado de la cuestión en el campo de la anotación pragmática. El subapartado 2.3 define los actos del habla y los ataques contra la imagen. En el subapartado 2.4 se presentan algunas de las herramientas frecuentemente utilizadas en la construcción de corpus orales. Y, finalmente, el subapartado 2.5 describe el proyecto del que esta investigación forma parte y los objetivos alcanzados en etapas anteriores del mismo.

2.1. CORPUS ORALES

Los corpus orales son menos numerosos que los escritos, ya sea por los requisitos técnicos necesarios para su compilación (dificultades de obtención de los datos inclusive), ya sea por el tiempo necesario para su transcripción, además del tiempo habitual invertido en anotación. Constituyen un tipo especial de corpus, y así como los corpus escritos, pueden ser de ámbito general o específico, o bien suelen ser

una de las componentes de los corpus de referencia. Por un lado, son ejemplos notables de los primeros el *London-Lund Corpus of Spoken English* (LLC) (Svartvik, 1990), el *Michigan Corpus of Academic Spoken English* (MICASE) (Simpson *et al.*, 2002) y el CORLEC (Corpus Oral de Referencia de la Lengua Española Contemporánea) (Marcos-Marín, 1992). Por otro lado, son ejemplos de los segundos las componentes habladas del *British National Corpus* (BNC) (Crowdy, 1993; Love *et al.*, 2017) y del Corpus de Referencia del Español Actual (CREA) (Sánchez, 2005).

Por lo que respecta a los corpus orales relativos a la interpretación, son escasos y, salvo excepciones, aunque están en formato digital, no han sido procesados informáticamente, es decir, carecen de una codificación normalizada² y de cualquier forma de anotación. Entre estas excepciones cabría citar el *European Parliamentary Interpreting Corpus* (EPIC) (Monti *et al.*, 2005), que es el corpus de interpretaciones simultáneas del parlamento europeo; el *ComInDat Pilot Corpus* (Angermeyer *et al.*, 2012), perteneciente al ámbito de la interpretación de interacciones entre médicos y pacientes y de juicios interpretados (Lázaro Gutiérrez, 2018). A diferencia de otros campos, sin embargo, muchos de los fenómenos observados en el campo de la pragmática requieren un análisis del contexto en el que se produce la interacción, así como de las señales emitidas por los interlocutores. La dificultad de captar estadísticamente estos matices hace que la lingüística de corpus tenga un papel hasta cierto punto secundario, pues dicho análisis implica el estudio de una componente cualitativa importante.

2.2. ANOTACIÓN PRAGMÁTICA

Aunque la pragmática es ya una ciencia con una larga trayectoria, la pragmática computacional, como subrama compartida con la lingüística computacional, es un campo de investigación todavía poco explorado. Por tanto, se necesita armonizar las clasificaciones de los fenómenos de este campo y, asimismo, construir corpus específicos, debidamente anotados, para entrenar modelos de anotación automática de estos fenómenos. En este sentido, en el área de la anotación pragmática, la mayoría de los trabajos se concentra, sobre todo, en la

² Se refiere, en un sentido amplio, al uso de algún formato de codificación basado en lenguajes de marcado normalizados como XML (Extensible Markup Language) o RDF (Resource Description Framework), o, en un sentido más estricto, al uso de formatos normalizados para la representación de corpus que utilizan estos lenguajes, como puede ser el caso de TEI-XML (Text Encoding Initiative, 2009). Además, en la codificación de los caracteres, se suele utilizar un subconjunto del estándar Unicode para garantizar la interoperabilidad de los documentos, independientemente de la configuración del lenguaje en cada entorno informático. Normalmente, se utiliza el juego de caracteres UTF-8, el cual codifica con una longitud variable (1 a 4 bytes) todos los caracteres del documento (Yergeau, 2003).

anotación de los actos de habla o discurso y sus fenómenos lingüísticos relacionados (Sayers *et al.*, 2021).

En las últimas décadas, con el propósito de avanzar en estos aspectos, se han desarrollado diferentes normas ISO relativas a distintas subáreas: estructura del discurso (*International Organization for Standardization*, 2014), relaciones semánticas en el discurso (*International Organization for Standardization*, 2016), anotación de los actos del habla en diálogos (*International Organization for Standardization*, 2020). Además del desafío de lograr que las máquinas sean capaces de detectar e identificar adecuadamente los fenómenos pragmáticos, se debe conseguir que estos sean interpretados correctamente (Sayers *et al.*, 2021).

2.3. ACTOS DEL HABLA Y ATAQUES CONTRA LA IMAGEN

Los actos de habla, tal y como son definidos por Austin (1962), nacen de la idea de que las palabras no solo describen cosas, sino que también tienen el poder de realizar acciones (Rühlemann, 2018). De acuerdo con la teoría de la cortesía de Brown y Levinson (1987), los ataques contra la imagen (*face-threatening acts* o FTA) son un tipo particular de actos de habla que se suele clasificar de acuerdo con la imagen a la que afecten (positiva o negativa). La imagen positiva es atacada cuando el orador no tiene en cuenta los deseos de su interlocutor o su necesidad de independencia. A su vez, la imagen negativa es atacada cuando el orador limita la libertad de acción de su interlocutor. Según dicha teoría, la cortesía sería así la forma de evitar atacar la imagen de nuestros interlocutores. Aunque los FTA han sido ampliamente estudiados en trabajos previos (Lachenicht, 1980; Culpeper *et al.*, 2003; Bravo y Briz Gómez, 2004; O'Driscoll, 2007), incluso en el campo de la interpretación (Pöllabauer, 2004; Cambridge, 1997), estos se centran en lenguas particulares, distintas del español (el inglés, principalmente), por lo que se echan en falta estudios más particularizados para el castellano, y también con un enfoque eminentemente multilingüe.

2.4. HERRAMIENTAS DE CONSTRUCCIÓN DE CORPUS ORALES

Las herramientas más frecuentemente utilizadas en la creación de corpus orales son ANVIL (Kipp, 2001), Praat (Boersma y Van Heuven, 2001), ELAN (Brugman y Russel, 2004) y EXMARaLDA (Schmidt, 2004).

ANVIL fue originalmente desarrollada para la creación y anotación de corpus multimodales, pero actualmente se utiliza también en la creación de otros tipos de corpus. Su limitación principal, que la diferencia del resto de herramientas aquí analizadas, es que solo permite trabajar con vídeos y no con ficheros exclusivamente de audio.

Praat, por su parte, es una herramienta ampliamente usada en el análisis fonético y en la anotación de audio. En este sentido, su mayor fortaleza es su gran versatilidad, pero su principal debilidad es que carece de una función nativa de extracción de coocurrencias (*concordancer*), lo que requiere el uso adicional de otras herramientas externas para realizar este tipo de análisis.

A su vez, la herramienta ELAN se utiliza frecuentemente en la documentación de lenguas. Con una versatilidad similar a la de Praat, destaca, sobre todo, por ofrecer una amplia gama de recursos para el estudio de las lenguas de signos y de la multimodalidad. Además, aunque dispone de sus propias funciones básicas para la extracción de coocurrencias, sus datos pueden ser también analizados mediante otras herramientas externas para este mismo propósito.

Finalmente, EXMARaLDA es un conjunto de herramientas para la creación y análisis de corpus orales. Además de su versatilidad, sus puntos fuertes son su interoperabilidad con otras herramientas, proporcionada por sus funciones de importación y exportación de ficheros generados por otros sistemas, y sus funciones de extracción de coocurrencias y de gestión de corpus, que hacen de esta herramienta la más completa de las aquí mencionadas. Por dicha razón, fue la elegida para la realización de las tareas de este proyecto.

2.5. DESCRIPCIÓN DEL PROYECTO Y DE LAS TAREAS ANTERIORES

Por un lado, como ya ha sido corroborado por investigaciones previas que analizan las dificultades afrontadas por intérpretes telefónicos (Lázaro Gutiérrez y Cabrera Méndez, 2018; 2019), se produce una gran cantidad de ataques contra la imagen (FTA) en este tipo de conversaciones en los dominios estudiados.

Por otro lado, las interpretaciones telefónicas se consideran interacciones asíncronas, ya que la interpretación no es simultánea y el discurso de la persona interpretada es reformulado – es decir, no siempre se guarda completa fidelidad a lo que fue dicho – antes de llegar al otro interlocutor. En algunos casos, esta reformulación minimiza el impacto de los FTA en la interlocución.

El objetivo final de esta investigación es construir un corpus que permita detectar y clasificar los FTA, y analizar su impacto en el discurso y en la actuación de los intérpretes. Para construir tal corpus, se ha juzgado necesario seguir los siguientes pasos:

1. Obtención y codificación de los datos;
2. Creación de archivos digitales;
3. Selección de las conversaciones;
4. Transcripción de las grabaciones;
5. Traducción del discurso;
6. Conversión de las transcripciones al formato EXMARaLDA;

7. Sincronización de las transcripciones con las grabaciones;
8. Análisis del discurso del corpus y anotación de los fenómenos discursivos y pragmáticos de interés (FTA, etc.);
9. Evaluación y mantenimiento.

Mientras que los pasos 6 y 7 son el objeto de la investigación que aquí se presenta, los pasos 1 a 5 fueron completados en el proyecto “Diseño, compilación y análisis de un corpus multilingüe de interacciones mediadas sobre asistencia en carretera” (Ref. CCGP2017-HUM002), posible gracias a la cooperación de la Universidad de Alcalá (UAH) con la empresa Migralingua³ (Lázaro Gutiérrez, 2019), por lo que se describen de manera breve en los subapartados siguientes. Los pasos 8 y 9, que mostrarán la verdadera relación de este corpus y/o de su aplicación al estudio de los FTA, forman parte de trabajos e investigaciones en curso. Por ello, no serán abordados en este artículo.

2.5.1. Paso 1: Obtención y codificación de los datos

La mayor dificultad en la construcción de un corpus oral con las características descritas anteriormente es la obtención de grabaciones. Esto requiere el establecimiento de acuerdos con empresas que ofrecen servicios de interpretación telefónica, que, a su vez, necesitan acordar con sus clientes la posibilidad de almacenar y analizar grabaciones de conversaciones con fines de investigación y mejora de calidad.

La obtención de las grabaciones que conforman el corpus ha sido posible gracias a convenios establecidos entre la Universidad de Alcalá (UAH) y las principales empresas de interpretación telefónica de España, cumpliendo una serie de condiciones respecto a su almacenamiento y procesamiento, entre las cuales se incluye la eliminación de cualquier tipo de dato personal. Así, datos como nombre y apellido, matrícula, teléfono de contacto, número de contrato, dirección o cualquier otro dato que pudiera identificar al cliente son sustituidos por un pitido (R. Lázaro Gutiérrez, 2019).

El ámbito de las conversaciones son servicios de asistencia en carretera y de seguros de coche y hogar. Los participantes de una llamada son típicamente tres: el operador, hablante de español, el cliente, hablante de una lengua extranjera, y el intérprete, hablante de ambas lenguas que intermedia la conversa. En ciertos específicos, puede haber otros participantes, como un mecánico, el empleado de un establecimiento o otras personas que acompañan al cliente.

Las grabaciones suelen empezar con la llamada del operador de la compañía a la agencia de interpretación. Una vez conectado con un intérprete de la lengua deseada, el operador se identifica y le explica

³ <http://voze.es/>

brevemente las informaciones que necesita. A continuación, el operador pone en conferencia al intérprete y al cliente, el intérprete recoge los datos necesarios del cliente, y por fin los repasa al operador. Este ciclo se repite múltiples veces durante toda la interacción.

2.5.2. Pasos 2, 3 y 4: Creación de archivos digitales, selección de los datos y transcripción

Una vez obtenidas las grabaciones, los ficheros de audio de las conversaciones se incluyeron en archivos digitales distintos en función de la lengua interpretada. A tal efecto, se creó un archivo distinto para cada una de las lenguas consideradas en el corpus, es decir, alemán, chino, francés, inglés y ruso.

De los archivos digitales así recopilados, se eliminaron las conversaciones repetidas, las que se cortaban al inicio de la llamada, cuando el cliente llamaba a un número incorrecto, o cuando el intérprete hablaba una lengua diferente de la que hablaba el cliente.

Cada conversación fue transcrita fiel y literalmente (salvo error u omisión) y se incluyó en un fichero Word por separado. En la primera línea del fichero, se indican los participantes, que suelen ser el intérprete (I), el operador (O) y el cliente (C)⁴. En los casos excepcionales en que participan más personas, se indica de manera semejante en la misma línea, junto a su respectiva abreviatura.

En las líneas siguientes, se indica la abreviatura de la persona que habla, seguida de un punto y la respectiva transcripción de su intervención. Es decir, cada línea representa un turno. Las partes censuradas por un pitido (véase el Subapartado 2.5.1) se representan mediante una secuencia de "X" o de guiones ("-"). En los casos en los que el transcriptor tuvo dificultades para entender lo que decía alguno de los participantes de la interacción, ya fuera porque las voces se superponían, ya fuera por la mala calidad de la conexión o de la grabación, o por cualquier otro motivo, esto viene indicado por la etiqueta "[IN-AUDIBLE]".

El enlace de estos ficheros de transcripciones individuales con los ficheros de audio se hizo posteriormente y forma parte de la investigación que se detalla en el apartado 4.

2.5.3. Paso 5: Traducción

Cuando la lengua hablada por el cliente era el ruso o el chino, los diferentes sistemas de escritura de las transcripciones las hacía menos transparentes para su posterior análisis. En consecuencia, se

⁴ En todo el artículo, por economía lingüística y de espacio, y de acuerdo con la RAE, se adopta el masculino como género no marcado y para referirse a ambos géneros, masculino y femenino (como ocurre en este caso con "intérprete", "operador" o "cliente").

optó por traducir asimismo las intervenciones del cliente y sus interacciones con el intérprete al castellano. Esto facilitó el trabajo de análisis del corpus, aun contando con menos especialistas en esas lenguas que para las restantes, como fue finalmente el caso.

Además, se tradujo también al chino todo el discurso en castellano. Esto se debió a que, al principio del proyecto, el grupo de investigación incluía hablantes nativos de chino sin un dominio adecuado del castellano, por lo que no podían realizar análisis del corpus en ambas lenguas. No obstante, su aportación era claramente necesaria y relevante para el proyecto, por lo que se optó por esta solución mediante la subcontratación con cargo al proyecto de traductores especializados.

3. Objetivos y metodología

Como ya se ha indicado, este trabajo pretende complementar las tareas descritas en el apartado 2.4. Estas tareas realizadas anteriormente, junto con las aquí detalladas, tienen como fin la construcción de un corpus oral de interacciones telefónicas.

El objetivo de este trabajo, por lo tanto, ha sido:

1. Convertir las transcripciones de las grabaciones al formato EX-MARaLDA.
2. Enlazar cada una de las líneas de las transcripciones con su instante correspondiente en los ficheros de audio.

Para ello, el punto de partida fueron 342 grabaciones telefónicas de servicios para seguro de asistencia en carretera, de coche y de hogar en castellano interpretadas en cinco idiomas (es decir, alemán, chino, francés, inglés y ruso), con un total de más de 60 horas de audio (véase la Tabla 1).

IDIOMA	N.º DE LLAMADAS	DURACIÓN TOTAL
alemán	37	6 h 45 min 56 s
chino	27	4 h 43 min 21 s
francés	94	18 h 5 min 57 s
inglés	168	27 h 33 min 34 s
ruso	16	2 h 55 min 11 s
TOTAL	342	60 h 4 min

Tabla 1: Distribución de las grabaciones por idioma, número de llamadas y duración total

Los pasos y procesos de este desarrollo se detallan a continuación.

4. Desarrollo

El propósito de este desarrollo, por tanto, ha sido convertir las transcripciones originalmente hechas en formato Word al formato de transcripción de EXMARaLDA utilizando la herramienta Partitur-Editor, además de enlazar al mismo tiempo cada línea de las transcripciones con su instante correspondiente en los ficheros de audio. El interfaz de Partitur-Editor (véase la Figura 1) permite crear, editar y exportar transcripciones de intervenciones, tanto simultáneas como secuenciadas, y también sincronizar o anotar el discurso de los interlocutores con sus instantes asociados en un fichero multimedia (audio o vídeo).

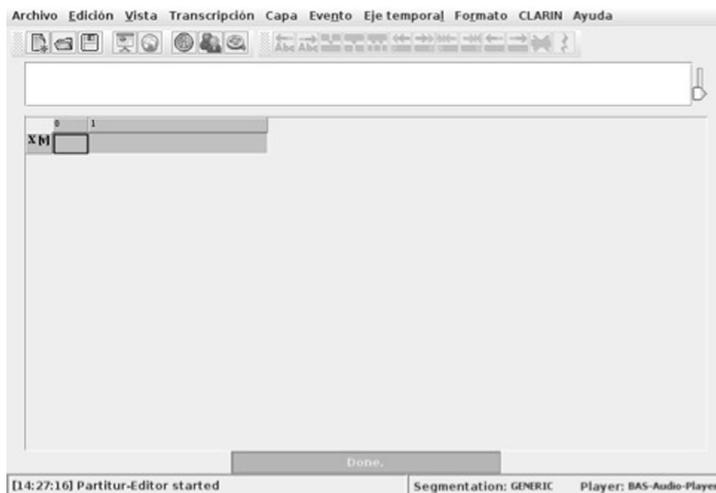


Figura 1: Interfaz de Partitur-Editor de EXMARaLDA

4.1. DEFINICIÓN DE LOS PARÁMETROS DE CONFIGURACIÓN DE LAS TRANSCRIPCIONES

Para cada una de las transcripciones, hay que definir un cierto número de configuraciones. En primer lugar, se pueden definir algunos metadatos (opcionales), como el nombre del proyecto y el nombre de la transcripción (véase la Figura 2). Se deben especificar los interlocutores que serán transcritos y, opcionalmente, los idiomas que cada uno de ellos habla en la interacción (véase la Figura 3). En este proyecto, se ha obviado el género de los interlocutores, ya que las grabaciones no venían etiquetadas con esta información y la deducción de la misma podría resultar dudosa en algunas ocasiones, por lo que no sería posible adjuntarla de forma general. Finalmente, se debe asociar un fichero multimedia (audio en nuestro caso, aunque la herramienta también permite la asociación con ficheros de vídeo) que se corresponde con la transcripción (véase la Figura 4).

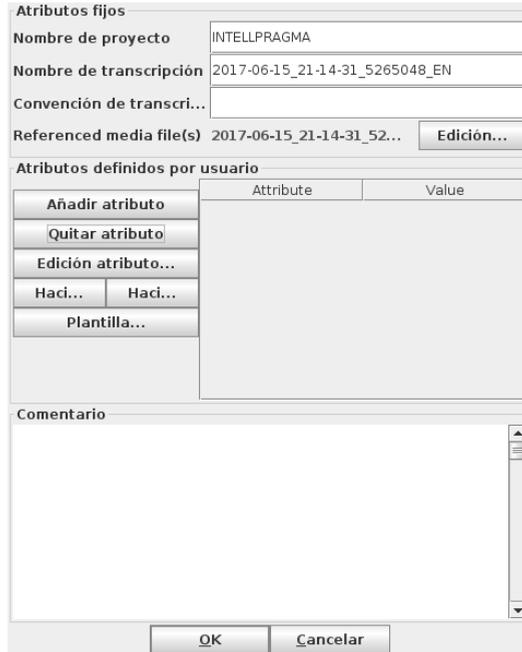


Figura 2: Interfaz de configuración de los metadatos de las transcripciones en la herramienta Partitur-Editor



Figura 3: Interfaz de caracterización de los interlocutores de la herramienta Partitur-Editor

4.2. CONVERSIÓN DE LAS TRANSCRIPCIONES

Una vez definidos estos parámetros de configuración, es posible empezar a trabajar en la conversión y sincronización de las conversaciones con Partitur-Editor. Como se ve en la Figura 5, se incluye por defecto una capa (*tier* en inglés) para cada interlocutor, y cada una de ellas es representada por una banda horizontal. Además, se muestra la señal de audio del fichero multimedia asociado a la transcripción, en forma de onda, sobre la cual se debe marcar el inicio y el fin de cada línea de la transcripción.

Pese a que los ficheros del proyecto originalmente se encontraban en formato MP3, debido a problemas de compatibilidad con los reproductores internos de Partitur-Editor, los ficheros tuvieron que convertirse al formato WAVE. La reproducción de los ficheros MP3 es posible, pero, al menos en nuestro entorno informático, la herramienta no fue capaz de mostrar la señal del audio en forma de ondas, lo que causaba dificultades a la hora de sincronizar las transcripciones con sus correspondientes instantes de las grabaciones (véase la Figura 6).

A partir de este punto, es posible empezar a crear la transcripción. Pese a que es posible importar ficheros de texto con la transcripción, de forma que se cree automáticamente un evento asociado a cada línea de la transcripción importada, posteriormente se hace necesario insertar vacíos entre cada uno de ellos para representar las pausas que separan las intervenciones, lo cual resulta en un proceso mucho más costoso (y tedioso). Por esta razón, se ha optado por crear cada evento manualmente y sincronizarlos después con el instante correspondiente del audio al mismo tiempo.



Figura 4: Ejemplo de grabación asociada a una transcripción en Partitur-Editor

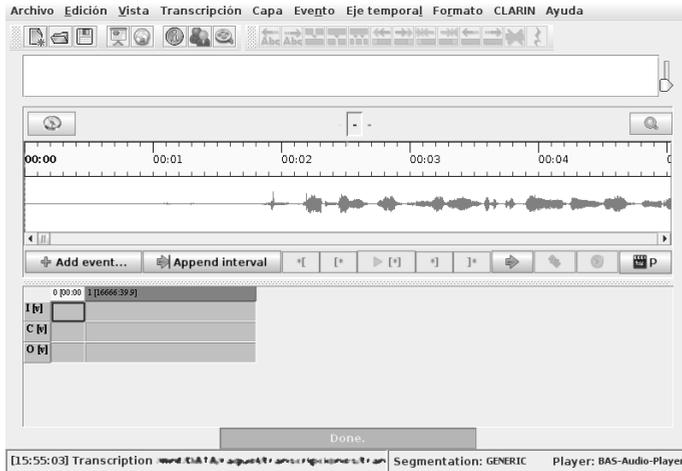


Figura 5: Captura de pantalla de la herramienta Partitur-Editor tras definir los parámetros básicos de la transcripción

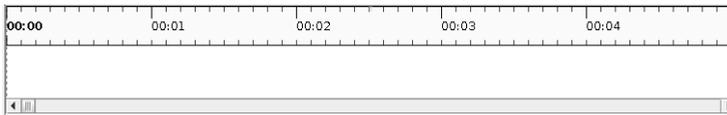


Figura 6: Problema de presentación de la señal de audio de un fichero MP3 en Partitur-Editor (campo en blanco)

Como se ve en la Figura 7, cada evento representa una intervención o turno de un interlocutor dado. Los eventos secuenciales se suceden en sentido horizontal, mientras que los que ocurren de manera simultánea se representan en paralelo en sentido vertical, todos ellos en la fila del interlocutor que lo realiza.

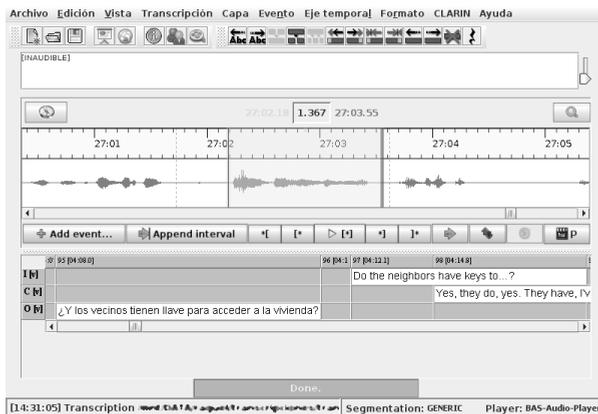


Figura 7: Ejemplo de transcripción en Partitur-Editor

Para cada interlocutor, es posible definir diferentes capas, que permiten anotar diferentes aspectos de su intervención de manera organizada e independiente. Un ejemplo de su uso en el proyecto ha sido la inclusión de la traducción de las intervenciones en ruso y chino al español (véase las Figuras 8 y 9). En ambos casos, tanto para el intérprete (I) como para el cliente (C), hay una capa adicional, indicada por “[ru-es]” o por “[zh-es]”, para el ruso y el chino, respectivamente, con la traducción de su discurso al castellano. Para el caso específico del chino se adjunta, además, la traducción al chino de la interacción en español entre el intérprete (I) y el operador (O), indicada por “[es-zh]”.

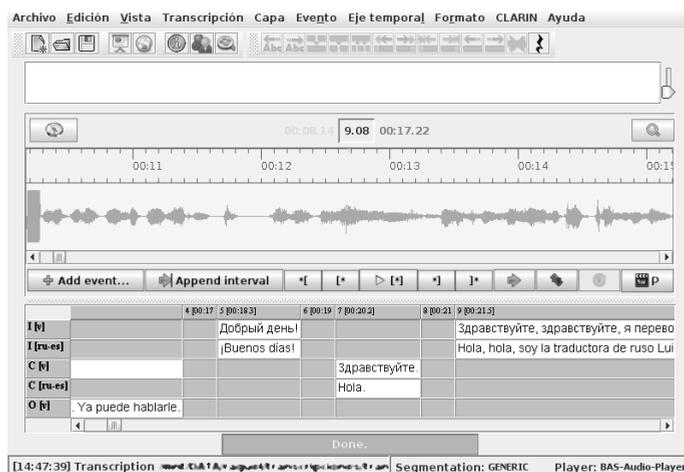


Figura 8: Ejemplo de transcripción de una conversación telefónica bilingüe español-ruso

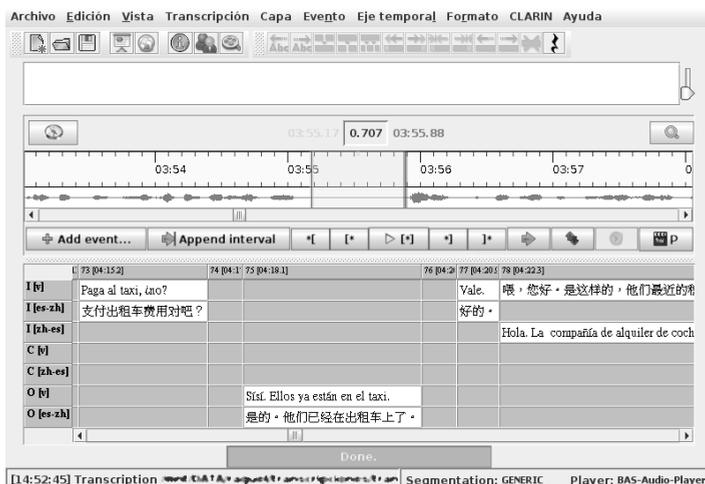


Figura 9: Ejemplo de transcripción de una conversación telefónica bilingüe español-chino

4.3. SINCRONIZACIÓN DE LAS TRANSCRIPCIONES CON SUS FICHEROS DE AUDIO RESPECTIVOS

Al pulsar el botón “*Add event...*”, se crea un nuevo evento (véase la Figura 7). Además de definir su texto, es posible marcar su momento de inicio y fin en la grabación arrastrando el ratón sobre la representación del audio en forma de onda. El botón “[*]” reproduce el fragmento seleccionado, mientras que los botones “[*” y “]” permiten escuchar los dos segundos anteriores y posteriores a la selección, respectivamente. El botón “*Append interval*” asocia los instantes del fragmento de audio seleccionado al evento en cuestión. De manera análoga, se puede editar y resincronizar el momento de un evento ya existente pinchando sobre él. La resincronización, cuando es necesaria, se realiza seleccionando el evento correspondiente y ejecutando sobre él estos mismos pasos.

5. Limitaciones y trabajos futuros

Las grabaciones de conversaciones telefónicas a tres bandas suelen presentar un gran número de inconvenientes, como, por ejemplo, la suma de los ruidos de fondo de las tres partes, ecos, problemas de conexión, etc. A esto se añade el fenómeno de lejanía, ya que, inevitablemente, las partes conectadas (cliente e intérprete) se escucharán con un volumen inferior o de manera más atenuada que la parte que realiza la conexión (operador). Asimismo, cuando se producen superposiciones (es decir, más de una persona hablando a la vez), no siempre resulta fácil escuchar a todos los interlocutores con claridad. Tal situación podría haber sido evitada, o al menos mitigada, mediante la captura de las intervenciones de cada una de las partes en canales de audio separados. Sin embargo, debido a limitaciones tecnológicas, las grabaciones telefónicas recogidas en el corpus incluyen las tres bandas en un mismo canal. En estas condiciones, la calidad del audio no siempre es la ideal, por lo que se hace imposible la transcripción automática del mismo. Aun realizando la transcripción de manera manual, en virtud de las dificultades de comprensión, se produjeron en las transcripciones un número considerable de omisiones. Éstas vienen indicadas por la etiqueta “[INAUDIBLE]” (véanse el subapartado 2.5.4 y la Figura 10).

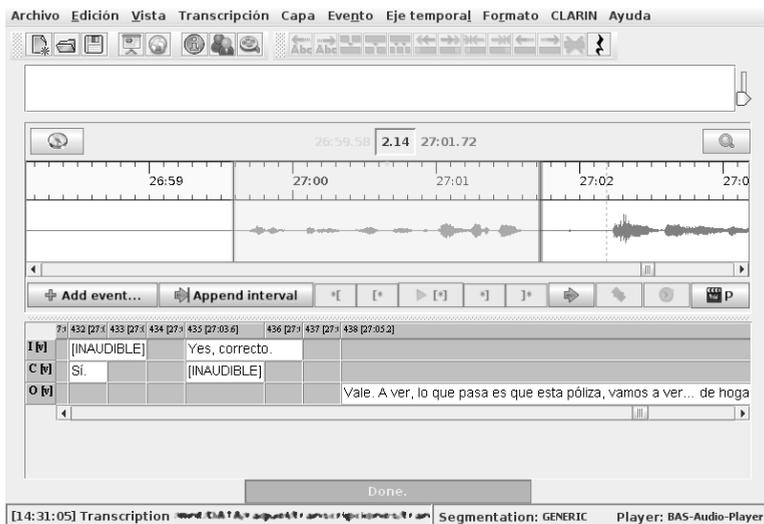


Figura 10: Ejemplo de fragmento omitido en la transcripción por dificultad de comprensión

Aunque se ha procurado no modificar las transcripciones originales, siempre que se ha detectado algún tipo de omisión en la transcripción sin razón aparente, se ha tomado la decisión de marcarlas por la etiqueta “[INAUDIBLE]” – la misma utilizada para indicar un fragmento de difícil comprensión – y sincronizarlas con la parte correspondiente del audio. Este también fue el caso de elementos extralingüísticos que no siempre aparecían en las transcripciones, como interjecciones, risas y suspiros, ya que en algún caso podrían ser de utilidad en la desambiguación de las emociones de su emisor. Esto posibilita acceder más fácilmente a estas ocurrencias e incluso anotarlas más tarde en caso de ser pertinente.

Tampoco se han anotado dentro del texto de las transcripciones la duración de las pausas que se producen en el discurso, ya que se han juzgado no relevantes para el estudio en cuestión (Lázaro Gutiérrez, 2018). Sin embargo, si fuera necesario, Partitur-Editor tiene mecanismos que facilitan su anotación sin necesidad de calcular los tiempos manualmente. Por último, aunque en esta fase no se ha utilizado, otro recurso de la herramienta es la posibilidad de definir estilos diferentes, como color y tipo de fuente, para diferenciar cada una de las capas y/o interlocutores, lo cual facilita su identificación.

En trabajos futuros, se deberá avanzar en el análisis y clasificación de los FTA, así como en la anotación de dichos fenómenos en las transcripciones. Una vez finalizado este proceso, sería posible integrar estas transcripciones dentro del corpus utilizando la herramienta Corpus Manager, que también forma parte del paquete EXMARaLDA.

6. Conclusiones

En este artículo, se han presentado los pasos dados hasta la fecha para la creación y anotación preliminar de un corpus oral de conversaciones mediadas por intérpretes, más específicamente de llamadas telefónicas sobre seguro de asistencia en carretera, de coche y hogar entre español y alemán, chino, francés, inglés y ruso. Dadas las características de este tipo de interacción asincrónica, se sabe de antemano que se produce un número elevado de ataques contra la imagen (*face-threatening acts* o FTA). Por tal motivo, el objetivo final de este corpus será su uso en el análisis, la detección y la clasificación de los FTA para su posterior anotación, así como la investigación del impacto de estos actos de habla en el discurso y en la actuación de los intérpretes.

Así, se han resumido los objetivos de la investigación y la descripción del proyecto en el que se encuadra, el proceso de obtención y de selección de las grabaciones relevantes, las características de los datos que conforman el corpus y el proceso de transcripción del audio. A continuación, se otorga especial importancia a los aspectos técnicos vinculados a la conversión de las transcripciones al formato EXMARaLDA por medio de la herramienta Partitur-Editor y la sincronización de cada una de las intervenciones de los interlocutores con sus instantes correspondientes en las grabaciones. Esto constituye un paso esencial en el marco del proyecto, ya que estas actividades iniciales, una vez completadas, posibilitarán múltiples trabajos en fases posteriores. Estos trabajos incluirán el análisis y clasificación de los FTA, además de la anotación de las transcripciones de las grabaciones. Un corpus así anotado permitiría incluso entrenar modelos automáticos de detección y/o clasificación de FTA en español y otras lenguas menos estudiadas que el inglés en este aspecto, lo cual es un reto aún no superado a fecha de hoy.

Como trabajos futuros adicionales quedarían asimismo (1) la realización de un análisis profundo del discurso del corpus y (2) la anotación de los fenómenos discursivos y pragmáticos de interés, así como (3) una evaluación adecuada de las anotaciones incluidas en el mismo de cara al análisis y estudio de estos fenómenos. A priori, es más que posible que los resultados del análisis y estudio de estos fenómenos mediante el corpus tengan un gran impacto para la comunidad científica que trabaja en este campo. No obstante, este impacto solo podrá ser confirmado mediante la aplicación de las anotaciones que se han de realizar, y que servirán como evaluación final de la utilidad del corpus y su interés para el estudio de FTA, entre otros fenómenos discursivos y/o pragmáticos.

7. Referencias

- ANGERMEYER, P., MEYER, B. y SCHMIDT, T. (2012): "Sharing Community Interpreting Corpora: A pilot study", Schmidt, T. y Wörner, K. (coords.), *Multilingual Corpora and Multilingual Corpus Analysis*, Amsterdam: John Benjamins Publishing Company, pp. 275-294. <https://doi.org/10.1075/hsm.14.19ang>
- AUSTIN, J. L. (1962): *How to Do Things with Words*, Oxford: University Press.
- BOERSMA, P. y VAN HEUVEN, V. (2001): "Praat, a system for doing phonetics by computer", *Glott International*, 5:9/10, pp. 341-345. Disponible en: https://www.fon.hum.uva.nl/paul/papers/speakUnspeakPraat_glott2001.pdf (Fecha de consulta: 03/12/2021).
- BRAVO, D. y BRIZ GÓMEZ, A. (2004): *Pragmática sociocultural: estudios sobre el discurso de cortesía en español*, Barcelona: Ariel.
- BROWN, P. y LEVINSON, S. C. (1987): *Politeness: Some universals in language usage*, Cambridge: Cambridge University Press.
- BRUGMAN, H. y RUSSEL, A. (2004): "Annotating Multimedia/Multi-modal resources with ELAN", *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*, Lisbon: European Language Resources Association, pp. 2065-2068. Disponible en: <https://aclanthology.org/L04-1285/> (Fecha de consulta: 03/12/2021).
- CAMBRIDGE, J. (1997): "Information exchange in bilingual medical interviews", Trabajo de fin de máster, University of Manchester, Manchester, Inglaterra.
- CROWDY, S. (1993): "Spoken Corpus Design", *Literary and Linguistic Computing*, 8(4), pp. 259-265. <https://doi.org/10.1093/llc/8.4.259>
- CULPEPER, J., BOUSFIELD, D. y WICHMANN A. (2003): "Impoliteness revisited: With special reference to dynamic and prosodic aspects", *Journal of Pragmatics*, 35, pp. 1545-1579. [https://doi.org/10.1016/S0378-2166\(02\)00118-2](https://doi.org/10.1016/S0378-2166(02)00118-2)
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (2014): *Language resource management — Semantic annotation framework (SemAF) — Part 5: Discourse structure (SemAF-DS)*, (ISO/TS Standard No. 24617-5). Disponible en: <https://www.iso.org/standard/57083.html> (Fecha de consulta: 03/12/2021).
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (2016): *Language resource management — Semantic annotation framework (SemAF) — Part 8: Semantic relations in discourse, core annotation schema (DR-core)*, (ISO Standard No. 24617-8). Disponible en: <https://www.iso.org/standard/60780.html> (Fecha de consulta: 03/12/2021).
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (2020): "Language resource management — Semantic annotation framework (SemAF) — Part 2: Dialogue acts (ISO Standard No. 24617-2). Disponible en: <https://www.iso.org/standard/76443.html> (Fecha de consulta: 03/12/2021).
- KIPP, M. (2001): "Anvil - A Generic Annotation Tool for Multimodal Dialogue", *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech 2001)*, Aalborg: International Speech Communication Association, pp. 1367-1370. <https://www.isca-speech.org/archive/pdfs/>

- eurospeech_2001/kipp01_eurospeech.pdf
- LACHENICHT, L. G. (1980): "Aggravating language: A study of abusive and insulting language". *International Journal of Human Communication*, 13(4), pp. 607-688. <https://doi.org/10.1080/08351818009370513>
- LÁZARO GUTIÉRREZ, R. (2018): "Design and Compilation of a Multilingual Corpus of Mediated Interactions about Roadside Assistance", Ruiz Mezcuca, A. (ed.), *Approaches to Telephone Interpretation. Research, Innovation, Teaching and Transference*, Bern: Peter Lang.
- LÁZARO GUTIÉRREZ, R. (2019): "Telephone interpreting and roadside assistance", *Translation and Translanguaging in Multilingual Contexts*, (5)3, pp. 215-240.
- LÁZARO GUTIÉRREZ, R. y ALCALDE PEÑALVER, E. (2022): "El cliente siempre lleva la razón: problemas de comunicación y soluciones en la interpretación telefónica para asistencia en carretera", *Pragmalingüística*, 30, pp. 433-446. <https://doi.org/10.25267/Pragmalingüística.2022.i30.20>
- LÁZARO GUTIÉRREZ, R. y CABRERA MÉNDEZ, G. (2018): "Pragmática e interpretación telefónica: un estudio sobre ataques contra la imagen de los intérpretes (FTA, Face threatening acts)", Curado Fuentes, A. (coord.), *LSP in Multi-disciplinary contexts of Teaching and Research. Papers from the 16th International AELFE Conference*, Manchester: EasyChair, pp. 85-90.
- LÁZARO GUTIÉRREZ, R. y CABRERA MÉNDEZ, G. (2019): "Context and pragmatic meaning in telephone interpreting", Garcés-Conejos Blitvich, P., Fernández Amaya, L. y Hernández-López, M. O. (coords.), *Technology Mediated Service Encounters*, Amsterdam: John Benjamins Publishing Company, pp. 45-67.
- LOVE, R., DEMBRY, C., HARDIE, A., BREZINA, V. y MCENERY, T. (2017): "The Spoken BNC2014: designing and building a spoken corpus of everyday conversations". *International Journal of Corpus Linguistics*, 22(3), pp. 319-344. <https://doi.org/10.1075/ijcl.22.3.02lov>
- MARCOS-MARÍN, F. (1992): "Corpus de referencia de la lengua española contemporánea: Corpus oral peninsular", *Laboratorio de Lingüística Informática*. Disponible en: <http://www.lllf.uam.es/ESP/Corlec.html> (Fecha de consulta: 03/12/2021).
- MONTI, C., BENDAZZOLI, C., SANDRELLI, A. y RUSSO, M. (2005): "Studying Directionality in Simultaneous Interpreting through an Electronic Corpus: EPIC (European Parliament Interpreting Corpus)", *Meta*, 50(4). <https://doi.org/10.7202/019850ar>
- O'DRISCOLL, J. (2007): "What's in an FTA? Reflections on a chance meeting with Claudine", *Journal of Politeness Research*, 3(2), pp. 243-268. <https://doi.org/10.1515/PR.2007.011>
- OZOLINS, U. (1998): *Interpreting & Translating in Australia: Current Issues and International Comparisons*, Melbourne: The National Language and Literacy Institute of Australia. Disponible en: <https://eric.ed.gov/?id=ED426597>
- PENA DÍAZ, C. (2022): "El uso de atenuantes retóricos en la interpretación telefónica en la asistencia en carretera", *Pragmalingüística*, 30, pp. 447-462. <https://doi.org/10.25267/Pragmalingüística.2022.i30.21>
- PHELAN, M. (2001): *The Interpreter's Resource*, Clevedon: Multilingual Matters.
- PÖLLABAUER, S. (2004): "Interpreting in asylum hearings. Issues

- of role, responsibility and power”, *Interpreting (International Journal of Research and Practice in Interpreting)*, 6(2), pp. 143-180. <https://doi.org/10.1075/intp.6.2.03pol>
- RÜHLEMANN, C. (2018): “CL and speech acts”, *Corpus Linguistics for Pragmatics: A Guide for Research*, Abingdon: Routledge, pp. 16-47.
- SÁNCHEZ, M. S. (2005): “El corpus de referencia del español actual (CREA): el CREA oral”. *Oralia: Análisis del discurso oral*, 8, pp. 37-56.
- SAYERS, D., SOUSA-SILVA, R., HÖHN, S. et al. (2021): “*The Dawn of the Human-Machine Era: A forecast of new and emerging language technologies*”, EU COST Action CA19102 ‘Language In The Human-Machine Era’. <https://doi.org/10.17011/jyx/reports/20210518/1>
- SCHMIDT, T. (2004): “Transcribing and annotating spoken language with EXMARaLDA”, *Proceedings of the LREC-Workshop on XML based richly annotated corpora*, Paris: European Language Resources Association. Disponible en: <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/2317> (Fecha de consulta: 03/12/2021).
- SIMPSON, R. C., BRIGGS, S. L., OVENS, J. y SWALES, J. M. (2002): *The Michigan Corpus of Academic Spoken English*, Ann Arbor: The Regents of the University of Michigan.
- SVARTVIK, J. (ed.) (1990): *The London-Lund corpus of spoken English: Description and research*, Lund: Lund University Press. Disponible en: <https://portal.research.lu.se/sv/publications/the-londonlund-corpus-of-spoken-english-description-and-research> (Fecha de consulta: 03/12/2021).
- TEXT ENCODING INITIATIVE (2019): “The TEI Guidelines”, *TEI: Text Encoding Initiative*. Disponible en: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (Fecha de consulta: 03/12/2021).
- VALERO GARCÉS, C. y LI, J. (2022): “La interpretación telefónica y presencial chino-español. Estudio de caso”, *Pragmalingüística*, 30, pp. 463-482. <https://doi.org/10.25267/Pragmalingüística.2022.i30.22>
- VITALARU, B. (2022): “Mitigación y estrategias atenuadoras en interpretación telefónica: estudio de caso sobre la combinación español-ruso”, *Pragmalingüística*, 30, pp. 483-514. <https://doi.org/10.25267/Pragmalingüística.2022.i30.23>
- YERGEAU, F. (2003): “RFC 2279: UTF-8, a transformation format of ISO 10646” (RFC Standard No. 2279). <https://doi.org/10.17487/RFC3629>