



Dando voz a la pragmática: creación de un sistema de anotación de actos de habla en el corpus ESLORA

Giving voice to Pragmatics: Creating an annotation system for speech acts in the ESLORA corpus

María Sampedro Mella 

Université catholique de Louvain

maria.sampedro@uclouvain.be

Resumen

Este artículo, enmarcado en la pragmática de corpus, tiene como objetivo presentar el sistema de anotación pragmática desarrollado e implementado en el subcorpus de entrevistas semidirigidas del Corpus para el Estudio del Español (ESLORA) de la Universidad de Santiago de Compostela. Este sistema permite recuperar de manera automática una selección de actos de habla: saludos y presentaciones, actos directivos, ofrecimientos, preguntas de distinta tipología y citas. Si bien la mayoría de los corpus anotados del español (y de otras lenguas) disponen de sistemas de etiquetado morfosintáctico que permiten la obtención de unidades lingüísticas simples y complejas, la anotación pragmática supone todavía una novedad en la lingüística de corpus. Por tanto, se busca contribuir a la investigación pragmática del español con la aportación de una nueva funcionalidad que permite realizar varios tipos de búsqueda y obtener de manera automática actos de habla, o bien combinarlos con otras unidades lingüísticas. El objetivo de este artículo es, pues, dar a conocer este sistema de anotación pragmática para que pueda ser utilizado en la investigación pragmático-discursiva y en la didáctica del español como lengua extranjera. También se espera que, tanto el diseño lingüístico como el procedimiento de anotación, puedan ser replicados e implementados en otros corpus.

Palabras clave: actos de habla, anotación pragmática, pragmática de corpus, lingüística de corpus, corpus oral

Abstract

Within the field of Corpus Pragmatics, this paper presents the pragmatic annotation system developed and implemented in the sociolinguistic interview subcorpus of the Corpus for the Study of Spanish (ESLORA). The system enables the automatic retrieval of a selection of speech acts, including greetings and introductions, directive acts, offers, different types of questions, and reported speech. Although most annotated corpora of Spanish and other languages incorporate morphosyntactic tagging systems that facilitate the retrieval of simple and complex linguistic units, pragmatic annotation remains relatively underdeveloped in corpus linguistics. This study therefore seeks to contribute to research on Spanish pragmatics by introducing a new annotation layer that supports different types of searches and allows for the automatic retrieval of speech acts, either independently or in combination with other linguistic units. The aim of the paper is thus to describe this pragmatic annotation system and to show its potential applications in pragmatic-discursive research and in the teaching of Spanish as a foreign language. In addition, it is hoped that both the linguistic design and the annotation procedures presented here may be replicated and adapted for implementation in other corpora.

Keywords: speech acts, pragmatic annotation, corpus pragmatics, corpus linguistics, spoken corpus



1. INTRODUCCION¹

En las últimas décadas, resulta innegable la importancia que han adquirido los corpus en la investigación pragmática, con el creciente número de publicaciones específicas (ej., Aijmer y Rühlemann, 2014; Jucker, 2018; Weisser, 2018; Rühlemann, 2019; Landert *et al.*, 2023; Aijmer, 2026), el desarrollo de corpus con sistemas de anotación que permiten recuperar contenidos pragmáticos de manera automática (ej., COREMAH, CLANES; §2.2.) y hasta la creación de una revista específica (*Corpus Pragmatics*). Todo ello se enmarca en una nueva disciplina denominada “pragmática de corpus”, que ha experimentado un considerable crecimiento.

Ahora bien, todavía en la actualidad, la mayoría de estudios fundamentados en datos de corpus continúan concentrándose en niveles lingüísticos formales o estructurales, es decir, en aspectos morfosintácticos y léxicos. Además de la influencia de la tradición gramatical, este desequilibrio obedece también al sistema de anotación morfosintáctica con que se enriquece la mayoría de los corpus, que facilita la recuperación de palabras y sintagmas, categorías gramaticales o lemas; pero no de contenidos pragmáticos. Por tanto, para llevar a cabo investigaciones sobre cuestiones pragmáticas con datos de corpus, como el estudio de las estrategias de cortesía o de los actos de habla, suele ser preciso realizar búsquedas de formas lingüísticas específicas (O’Keffe, 2018), o bien revisar manualmente los materiales por los que está constituido cada corpus (Sampedro Mella, 2021a). La anotación pragmática de los corpus constituye un avance necesario para ampliar el alcance de la investigación basada en corpus dentro de la lingüística, pues “la anotación morfosintáctica es solo el primer paso, absolutamente necesario, para llegar a corpus anotados con informaciones sintácticas, semánticas o pragmáticas” (Rojo, 2021, p. 3). En este sentido, la anotación pragmática no solo incrementa las posibilidades de explotación de los corpus, sino que también contribuye a ampliar el foco de la investigación lingüística hacia una perspectiva más abarcadora centrada en el uso lingüístico en todas sus dimensiones.

En este contexto, el objetivo de este trabajo es presentar el sistema de anotación pragmática implementado en el Corpus para el Estudio del Español (ESLORA) de la Universidade de Santiago de Compostela (USC). El sistema se diseñó con la intención de enriquecer un corpus oral ya existente, alineado, lematizado y que incluía anotación morfosintáctica, con un nuevo sistema que permite recuperar de manera automática una selección de actos de habla. Con la presente publicación se pretende, por un lado, dar a conocer este sistema para que pueda ser explotado con fines de investigación pragmático-discursiva y también de didáctica en el ámbito del español como lengua extranjera. Por otro lado, se busca poner a disposición pública esta información para que el diseño lingüístico y la metodología de anotación puedan ser replicados e implementados en otros corpus.

Para contextualizar y describir el sistema de etiquetado pragmático del corpus ESLORA, la sección 2. presenta la pragmática de corpus como disciplina emergente, así como las características de la anotación pragmática (2.1.) y algunos corpus del español que incorporan este tipo de etiquetado (2.2.). La sección 3. se dedica a la metodología: primero se presenta brevemente el corpus ESLORA (3.1.), a continuación se describen los actos de habla que se

¹ El sistema de anotación pragmática descrito en este trabajo se diseñó e implementó en el corpus ESLORA en el marco del proyecto *Explotación de corpus generales y de aprendices para la investigación y la enseñanza del discurso en el español como lengua extranjera* (ELE), financiado por la Xunta de Galicia en concurrencia competitiva (ED481D-2022-016) y desarrollado en el GI Gramática del Español de la Universidade de Santiago de Compostela (2022-2024). La autora agradece a los evaluadores anónimos sus comentarios y sugerencias, que han permitido mejorar y enriquecer la versión final de este artículo.



han incorporado (3.2.), el procedimiento de anotación (3.3.) y los desafíos lingüísticos y técnicos encontrados durante el proceso de etiquetado (3.4.), que afectan, principalmente, a estructuras incompletas (3.4.1.), al encadenamiento de varios actos (3.4.2.), y a la delimitación y los problemas de identificación de algunos de ellos (3.4.3.). En 4. se muestran las distintas opciones de búsqueda que ofrece este sistema en la interfaz del corpus: búsquedas generales de actos de habla (4.1.), o bien combinados con otras unidades lingüísticas (4.2.). Finalmente, las conclusiones se presentan en la sección 5.

2. PRAGMÁTICA DE CORPUS

La “pragmática de corpus” es una disciplina emergente que aúna los objetivos y metodologías de la lingüística de corpus y de la pragmática (Aijmer y Rühlemann, 2015; Weisser 2018; Albelda Marco, 2022; Aijmer, 2026). En ella se combinan el análisis cuantitativo de la lingüística de corpus con el enfoque cualitativo, característico de la pragmática: “[...] corpus-pragmatic research is more than just pragmatic research and it is more than just corpus-linguistic analysis in that it integrates the horizontal (qualitative) methodology typical of pragmatics with the vertical (quantitative) methodology predominant in corpus linguistics” (Aijmer y Rühlemann, 2015, p. 12). El término “pragmática de corpus”, para designar a esta disciplina, se generaliza a partir de 2015, con la publicación de Aijmer y Rühlemann (2015). No obstante, el uso de los corpus en los estudios pragmáticos es anterior, por ejemplo, para el análisis de los marcadores discursivos, las estrategias de cortesía o la conversación (véanse, como muestra en español, los estudios del Grupo Val.Es.Co.). Del mismo modo, tanto el desarrollo de herramientas computacionales específicas como los intentos de identificación y anotación de fenómenos pragmáticos, como actos de habla, en los corpus, son también anteriores (ej., Leech y Weisser, 2003; Weisser, 2003; Maynard y Leicher, 2007; Archer, Culpeper y Davies, 2008; O'Donnell, 2008).

A pesar de su potencial para la investigación pragmático-discursiva y la enseñanza de lenguas extranjeras mediante el enfoque comunicativo por tareas, la anotación pragmática de los corpus no ha sido un asunto priorizado ni en español ni en otras lenguas cercanas, como el inglés (Jucker, 2018), el francés o el portugués. Los primeros corpus con anotación pragmática se crearon para el estudio del inglés en los años 90: el Switchboard Dialogue Act Corpus (Jurafsky, Shriberg y Biasca, 1997), que recoge conversaciones telefónicas y contiene actos de diálogo, o el HRC Map Task (Anderson *et al.*, 1991) de grabaciones de audio y vídeo de estudiantes universitarios. Entre otros proyectos posteriores, destaca el corpus SPAADIA (Leech y Weisser, 2013), que anota actos de habla en muestras conversacionales, y el corpus SPICE-Ireland (Kirk, 2013), que incorpora la anotación pragmática dentro de un marco más amplio de lingüística de corpus y prosodia. En francés, cabe señalar el CID (Bertrand *et al.* 2008), con distintas capas de anotación: prosódica, sintáctica, morfosintáctica, discursiva y mímico-gestual.

Las causas por las que este tipo de anotación no ha tenido la misma popularidad que la morfosintáctica obedecen principalmente a las dificultades lingüísticas y técnicas que plantea (§2.1.). Como resultado, solo un reducido número de corpus cuenta con sistemas de estas características (§2.2.).



2.1. Anotación pragmática

La creación de sistemas computacionales que favorezcan el enriquecimiento y la explotación pragmática de los corpus resulta de vital importancia, tanto en el ámbito de la investigación empírica como en el de la didáctica de las lenguas extranjeras (Sampedro Mella y Rodríguez Domínguez, 2026). Para ello, una de las líneas de trabajo que requiere un mayor desarrollo actualmente es la que afecta al diseño de sistemas de anotación que permitan el reconocimiento y la recuperación automática de información pragmática.

Los corpus han experimentado considerables avances en lo que atañe a la anotación morfosintáctica, como se evidencia, por ejemplo, en la creación de distintos etiquetadores, como XIADA o FreeLing (*vid.* Vázquez Rozas *et al.*, 2020), así como en la aplicación de lenguajes de programación, como Python y la librería *spaCy*, que permiten etiquetar cantidades ingentes de textos. No ha sido este el caso de la anotación pragmática, debido a la complejidad que implica su implementación: por un lado, por la interpretación que requiere el proceso de etiquetado (Weisser, 2015; Milà-García, 2018) y, por otro, por la falta de correspondencias unívocas entre formas lingüísticas y funciones pragmáticas, es decir “form-to-function” y “function-to-form” (Landert *et al.*, 2023, p. 6). Como consecuencia de todo ello, “pragmatic annotation is still much less advanced than other linguistic annotation types” (Rabadán, Ramón y Sanjurjo-González, 2021, p. 210).

Para realizar estudios pragmáticos fundamentados en datos de corpus, se suele recurrir a búsquedas de formas lingüísticas específicas asociadas a esos componentes (O’Keffe, 2018). En el caso de los actos de habla, como señalaba hace varias décadas Kohnen (2000, p. 184; citado por Landert *et al.*, 2023, p. 6), “[p]erhaps a corpus-based study of speech acts will have to focus on the patterns representing the most typical and common manifestations of a speech act and will not seek to cover all the possible manifestations of that speech act”. El problema persiste hoy en día: ese tipo de procedimiento de búsqueda reduce el extenso campo de investigación de actos de habla a un limitado elenco de unidades que responden a un esquema lingüístico prototípico, por lo que inevitablemente se excluyen los actos indirectos o aquellos que no cuenten con la expresión anticipada por el investigador. Para evitar este inconveniente, se pueden llevar a cabo revisiones completas de los corpus e identificar de forma manual los actos de habla. Como muestra, en una investigación anterior sobre la petición en contextos transaccionales con datos de corpus, se revisaron siete corpus orales², ya que “[...] las peticiones, como otros actos de habla, no aparecen identificadas en las etiquetas de los corpus ni en los buscadores, lo que exige leer de manera completa las transcripciones de los corpus para poder reconocer y seleccionar cada caso” (Sampedro Mella, 2021a, p. 383). El resultado de esta ingente tarea fue un total de 34 enunciados válidos, muchos de ellos seleccionados entre muestras de discurso directo, lo que evidencia que el esfuerzo invertido no siempre redundaba en el producto obtenido.

La anotación pragmática consiste en identificar, segmentar y etiquetar elementos lingüísticos y paralingüísticos relacionados con el contexto, las intenciones comunicativas o la interacción entre los participantes, típicamente actos de habla, pero también estrategias de cortesía, de

² En concreto, los siguientes: Corpus de Habla Culta de Salamanca (CHCS) (Fernández Juncal, 2005); Corpus Oral de Lenguaje Coloquial (COLA) (Jørgensen, no disponible); Corpus para el Estudio del Español Oral (ESLORA) de Santiago de Compostela; Corpus Sociolingüístico de Castellón de la Plana y su Área Metropolitana (CSCP) (Blas Arroyo, Navarro Morales y Casañ Núñez (2009); Proyecto para el Estudio del Español de España y América (PRESEEA) Alcalá de Henares (Moreno Fernández *et al.*, 2002); Proyecto para el Estudio del Español de España y América (PRESEEA). Valencia (Gómez Molina, 2007); Valencia Español Coloquial (Briz y Grupo Val.Es.Co., 2002).



atenuación e intensificación; aspectos prosódicos y entonativos relevantes en la interacción, emociones y elementos relacionados kinésicos y proxémicos. De este modo, este tipo de anotación va más allá del significado literal y de la forma lingüística. Entre sus principales ventajas se encuentra la posibilidad de recuperar automáticamente información pragmática, como actos de habla, para poder estudiarlos en su contexto de uso, es decir, teniendo en cuenta variables como el perfil sociolingüístico de los participantes, las intenciones comunicativas y las características de la situación. Esto hace que los corpus que contienen estos sistemas de anotación sean especialmente valiosos en la investigación en el ámbito de la pragmática, el análisis del discurso y el análisis de la conversación (Milà-Garcia, 2018), así como en la enseñanza de lenguas extranjeras, a partir de métodos de aprendizaje por descubrimiento, como el “*data-driven learning*” (Leech, 1997).

Ahora bien, una de las principales limitaciones de la anotación pragmática es que se lleva a cabo de forma manual. Como señala Milà-Garcia (2018, p. 1): “Although (semi)automatic speech act annotation keeps improving, manual annotation is still needed in order to carry out in-depth analyses of speech acts that aim to go beyond the search for speech act verbs with specific linguistic structures”. Esta necesidad de contar con la intervención humana obedece a la segmentación del discurso oral y a los límites lingüísticos de cada unidad pragmática, además de la variabilidad del habla espontánea, en la que se incluyen fenómenos como solapamientos, interrupciones o reformulaciones. A todo ello se suma también la variabilidad prosódica y las diferentes interpretaciones que pueden derivar de un mismo enunciado, como es el caso de los imperativos, condicionales o actos de habla indirectos, como peticiones indirectas. Todo ello dificulta la delimitación de las unidades de análisis y la posibilidad de establecer correspondencias entre las unidades lingüísticas y sus funciones discursivas, implicando un alto grado de interpretación por parte de los anotadores en el proceso. Para minimizarlo, es frecuente desarrollar manuales de anotación minuciosos, como el estándar ISO 24617-2 (Bunt *et al.*, 2020); realizar revisiones periódicas y calcular métricas de acuerdo, como el coeficiente kappa de Cohen³. Una vez más, todo este procedimiento exige invertir una gran cantidad de tiempo y recursos, lo que no está al alcance de todos los equipos: “Any type of linguistic annotation is a highly complex and interpretive process, but none more so than pragmatic annotation” (Weisser, 2015, p. 84).

En síntesis, la anotación pragmática de corpus orales presenta aplicaciones relevantes, si bien su desarrollo e implementación dependen de medios computacionales, y de contar con altos presupuestos, tiempo y equipos con rigor y precisión metodológicas, además de encontrar un equilibrio entre la variabilidad del uso lingüístico real y la restricción de los sistemas computacionales⁴.

³ Se trata de una medida estadística que permite evaluar el grado de acuerdo entre dos investigadores o anotadores (*inter-rater agreement*) cuando clasifican los mismos datos en categorías (*vid.* Rau y Shih, 2021, para más información).

⁴ A este respecto, puede consultarse <https://www.palabrasbinarias.com/articles/2025-03-14-informacion-pragmatica-corpus-eslora> (2025). Se trata de una entrevista entre M. Sampedro Mella, responsable del diseño lingüístico del sistema de anotación de actos de habla del corpus ESLORA, y M. Barcala, ingeniero informático y director de la empresa NLPgo Technologies S.L., encargado del desarrollo de un entorno de anotación específico para el corpus. En ella se analizan los desafíos de articular la naturaleza dinámica de la lengua oral con los requisitos estructurales de los sistemas computacionales de procesamiento automático.



2.2. Corpus con anotación pragmática del español

En la actualidad, el número de corpus orales en español que incorpora sistemas de anotación pragmática es todavía reducido. Uno de los precursores ha sido Val.Es.Co., que propone una segmentación detallada de las unidades conversacionales según las circunstancias de la enunciación: “consta de ocho unidades (*discurso, diálogo, intercambio/ alternancia de turnos, intervención/ turno, acto y subacto*); tres órdenes (*social, estructural e informativo*) y cuatro posiciones (*inicial, media, final e independiente*)” (Grupo Val.Es.Co., 2014, p. 13). Además de estas unidades estructurales, la anotación tiene en cuenta elementos fónicos relevantes en la interacción, como la entonación ascendente, descendente, y mantenida o suspendida; la pronunciación marcada o enfática, las risas, o rasgos complementarios del canal verbal, entre otros.

Cabe mencionar también la plataforma OBSERVAHUMOR, que reúne diversos corpus de géneros humorísticos orales y escritos (*vid.* Ruiz Gurillo, 2026, en el mismo número temático que este artículo). Presenta un sistema de etiquetado pragmático adaptado específicamente a las características de las muestras que contiene cada corpus, mediante el cual pueden realizarse búsquedas de distintos fenómenos pragmáticos. Por ejemplo, en el corpus FEMMES.UP, es posible buscar cuestiones relacionadas con la articulación fónica del mensaje (intensidad vocal, risas, alargamientos...), con el empleo de estrategias de cortesía (intensificadores, atenuadores, sufijación, valorativos, etc.), con distintos estilos de humor, entre otras categorías. También destaca el corpus HumCor (Corpus Oral Multimodal de Humor en Español) (Repede, en línea), que recoge distintos géneros humorísticos, tales como chistes, monólogos, *sketches* y espectáculos de ventrílocuos en español procedentes de varios continentes; entre las distintas opciones de búsqueda que proporciona, tiene en cuenta la gestualidad de los humoristas como un componente adicional del discurso.

En cuanto a los corpus que contienen sistemas de anotación relacionados con los actos de habla, destacan el Corpus Oral Didáctico Anotado Lingüísticamente (C-Or-Di-AL) (Nicolás Martínez, 2012), el Integrated Reference Corpora for Spoken Romance Language, más conocido por su acrónimo, C-ORAL-ROM (Cresti y Moneglia, 2005); el Corpus Multimodal de Actos de Habla (COREMAH) (Vacas Matos, 2017) y el Controlled LAnguage English Spanish (CLANES) (Rabadán, Ramón y Sanjurjo-González, 2021). Todos ellos se presentan en distintos formatos (oral, escrito y multimodal) y han sido codificados mediante diferentes herramientas, como UAM CorpusTool (O'Donnell, 2008) o TreeTagger (Schmid, 1995).

La terminología utilizada también varía entre los distintos corpus. Así, en CLANES se emplea la denominación “pragmatic function” (Rabadán, Ramón y Sanjurjo-González, 2021, p. 216), puesto que la anotación se centra en los objetivos comunicativos de cada mensaje. Por su parte, en C-Or-DiAL (Nicolás Martínez, 2012), un corpus concebido para la enseñanza del español como lengua extranjera, se anotan las “funciones comunicativas”, término adoptado del enfoque comunicativo que hace referencia a las tareas o acciones que llevan a cabo los hablantes en un contexto determinado; por ejemplo: saludar a un amigo, pedir la hora a un desconocido en la calle, comprar en una frutería, etc. Por último, en COREMAH (Vacas Matos, 2017), como su propio nombre indica, se tienen en cuenta “actos de habla”, considerados las unidades básicas de la comunicación lingüística (Searle, 1969) que estudia la pragmática. Todos estos corpus tienen en común la anotación de enunciados que responden a unos objetivos comunicativos específicos y que presentan determinados valores ilocutivos: asertivo, directivo, comisivo, declarativo, entre otros.



El Corpus C-Or-Di-AL contiene conversaciones espontáneas realizadas en Madrid y recoge una minuciosa clasificación de funciones comunicativas, según diferentes temáticas, tomada del *Plan curricular del Instituto Cervantes* (Instituto Cervantes, 2006). Además, por tratarse de un corpus didáctico, las muestras que recopila están organizadas y dispuestas en función del nivel de competencia del alumnado meta. Por este motivo, presenta innumerables ventajas para su explotación didáctica (*vid.*, como muestra, Sampedro Mella y Rodríguez Domínguez, 2026), pero también algunas limitaciones en su empleo con fines de investigación. Por ejemplo, el hecho de que no sea un corpus alineado y requiera gestionar los archivos de audio y transcripción de manera independiente, además de algunos problemas puntuales de calidad acústica que afectan a su inteligibilidad.

El corpus C-ORAL-ROM (Cresti y Moneglia, 2005) constituye uno de los recursos más importantes para el estudio del habla espontánea en las lenguas romances, ya que contiene varios corpus comparables en español, italiano, francés y portugués, anotados a partir de criterios prosódicos. Su diseño permite analizar interacciones reales en distintos contextos comunicativos, tanto formales como informales, e incorpora una segmentación basada en unidades entonativas consideradas relevantes para la interpretación pragmática del discurso. No obstante, no es un recurso abierto, sino que el acceso al corpus es a través de un CD-ROM, por lo que su manejo y consulta resultan menos flexibles en comparación con otros corpus.

En cuanto al Corpus Español Multimodal de Actos de Habla (COREMAH), contiene un total de 108 muestras de hablantes nativos y de no nativos de español de L1 inglés. Es un corpus multimodal, por tanto con acceso a vídeo, y está diseñado en torno a tres actos de habla: el rechazo, el cumplido y la disculpa. En su sistema de anotación tiene en cuenta un amplio repertorio de marcas gestuales y prosódicas, además de subactos: ofrecer ayuda, reprochar, pedir disculpas, expresar deseo o arrepentimiento; también incluye mitigadores e intensificadores verbales y gestuales, o cambios en la voz. Está basado en interacciones dirigidas o *role-plays*, lo que favorece la comparabilidad de las muestras, al haber sido recogido en idénticas circunstancias, si bien esta metodología puede comprometer la espontaneidad de los datos recopilados.

Por último, el CLANES (Controlled LAnguage English Spanish) es un corpus comparable español-inglés que recopila textos escritos en el ámbito de la gastronomía, en los que se anotaron las siguientes categorías: <STATE>, <DIRECT>, <SUGGEST>, <RECOMMEND>, <PRAISE>, <EVIDENCE> (Rabadán, Ramón y Sanjurjo-González, 2021; Pérez Blanco e Izquierdo, 2025). A diferencia de los anteriores, no es un corpus de acceso abierto en formato descargable, sino que está restringido al ámbito del grupo de investigación.

En conjunto, todos estos corpus constituyen avances significativos en el campo de la anotación pragmática del español. Ahora bien, es importante tener en cuenta que, en la actualidad, se está asistiendo a un cambio en las técnicas de anotación, gracias al desarrollo de la inteligencia artificial en el ámbito de la lingüística computacional (ej., Huang *et al.*, 2025; Brocca, Nuzzo y Wang-Kathrein, 2026; Bianco Brocca y Garassino, 2026). Estas técnicas ofrecen nuevas posibilidades de automatización, al reducir el proceso de anotación manual y permitir alcanzar niveles de análisis cada vez más complejos, aunque todavía presentan limitaciones en contextos que requieren mayor interpretación (ej., los casos mostrados en §3.4.3.) y no parece que se hayan aplicado aún en los corpus del español. Así, todavía persisten algunas carencias importantes en este ámbito, como la escasez de corpus con estos sistemas de anotación, el reducido número de muestras incluidas o la falta de estandarización en el etiquetado. En el caso de los corpus orales o multimodales, la falta de



alineación texto-voz dificulta, además, la consulta y el manejo de los materiales. En consecuencia, con el objetivo de contribuir a superar algunas de estas limitaciones, se ha propuesto un sistema de anotación pragmática en el corpus ESLORA centrado en una selección de actos de habla (§3.).

3. METODOLOGÍA

3.1. Descripción del corpus ESLORA

El sistema de anotación pragmática se implementó en la versión 2.3 del corpus ESLORA (octubre de 2024). En su versión actual (2.5), el corpus contiene 60 horas de grabaciones de 53 entrevistas semidirigidas realizadas a hablantes de la ciudad de Santiago de Compostela, y 24 horas de conversaciones coloquiales de hablantes de Galicia, registradas entre 2007 y 2020. Las grabaciones se transcribieron ortográficamente con alineación texto-voz mediante las herramientas ELAN y Transcriber, con el texto codificado en formato XML (Vázquez Rozas y Barcala, 2020). El corpus cuenta, además, con anotación morfosintáctica y lematización, e incorpora una amplia selección de marcas propias del habla oral, como alargamientos, énfasis, risas o cambios de lengua (*vid.*, para más información, Barcala *et al.*, 2018).

La anotación pragmática se realizó exclusivamente en el subcorpus de entrevistas semidirigidas, que forma parte del proyecto PRESEEA⁵. Las entrevistas de ESLORA son desestructuradas o libres (cf. Fernández Sanmartín, 2018, pp. 149-150): “[...] en el caso de la entrevista libre, la acción comunicativa de los entrevistadores tiene como objetivo ir creando una situación interaccional [...] más cercana, informal y flexible”. Así, a diferencia de las entrevistas dirigidas, no se utiliza un guion o cuestionario, sino que “se va[n] co-construyendo turno a turno siguiendo el curso de la conversación y los intereses del informante” (Fernández Sanmartín, 2018, p. 150), lo que favorece una dinámica comunicativa similar a la de la conversación coloquial.

Los entrevistadores de ESLORA, en su mayoría estudiantes de distintos ciclos universitarios, adoptan un papel más activo en la interacción y procuran mostrarse cercanos y romper la estructura prototípica de preguntas y respuestas de una entrevista convencional. Para ello, participan activamente con comentarios, repeticiones, muestras de acuerdo y de sorpresa, interjecciones, vocalizaciones, etc., contribuyendo de este modo a generar un entorno comunicativo más próximo a la conversación. Los temas tratados en estas entrevistas son muy variados, pues responden a los intereses de los informantes, y los cambios temáticos se realizan de forma no necesariamente lineal o lógica, tal y como ocurre en la conversación coloquial. Por último, procurando facilitar la comodidad de los informantes, las entrevistas se realizaron en entornos que les resultaran cercanos, como su domicilio o su lugar de trabajo, o bien, cuando esto no era posible, en la Facultad de Filología de la USC.

3.2. Delimitación y clasificación de los actos de habla

Los actos de habla son las unidades mínimas de la comunicación lingüística (Searle, 1969). Se caracterizan por contener un enunciado (con una secuencia fónica y morfosintáctica), por conllevar una intención comunicativa (fuerza ilocutiva) y por provocar una reacción en el

interlocutor. Su realización (o no) y su expresión varían en función de las características contextuales, de la relación entre los participantes en el discurso y de las normas socioculturales que regulan la interacción.

Para llevar a cabo la anotación de los actos de habla del corpus de ESLORA, se seleccionaron los siguientes: saludos-presentaciones, actos directivos, ofrecimientos y otros; en esta última categoría se incorporaron las citas –que ya se habían etiquetado previamente en el proceso de anotación morfosintáctica del corpus– y las preguntas (*vid.* tabla 1). Estos actos se buscaron e identificaron individualmente en cada una de las entrevistas semidirigidas de ESLORA. Se prescindió del marcado de otros actos, como los de confirmación, evaluación y marcadores fáticos. Si bien, todos ellos cumplen importantes funciones de regulación de la interacción, tales como la retroalimentación del interlocutor o el mantenimiento de la conversación, tanto su frecuencia como su brevedad y su dependencia del contexto fónico dificultan la ejecución de una segmentación precisa y de una clasificación fiable en el proceso de anotación. Por todo ello, se optó por priorizar aquellos actos que fueran más fácilmente identificables y delimitables.

Los actos de habla se organizaron en forma de *pares adyacentes*, a partir de las definiciones y caracterizaciones propuestas por Schegloff y Sacks (1973), Sacks, Schegloff y Jefferson (1974) y Schiffrin (1986); y, en el ámbito hispanohablante, por el Grupo Val.Es.Co. (2014). Así, un par adyacente está compuesto por dos turnos de sendos interlocutores, uno iniciativo y otro reactivo: “a sequentially constrained pair of turns in talk in which the occurrence of a first-pair-part creates a slot for the occurrence of a second-pair-part [...]” (Schiffrin, 1986, p. 48). Por ejemplo, un saludo o una pregunta esperan respectivamente, como turnos reactivos, otro saludo o una respuesta. Esta estructuración, o simplificación, de los actos posibilitó la creación del sistema computacional que se adapta, *grosso modo*, a los datos del corpus, aunque con algunas limitaciones, debidas, entre otras causas, a la falta de turno reactivo (§3.4.1.) o a las secuencias ampliadas (“expanded sequences”, Sacks, Schegloff y Jefferson, 1974, p. 711), también llamadas “actos” (Grupo Val.Es.Co., 2014, p. 14) (§3.4.2).

La Tabla 1 recoge la clasificación de los actos de habla incorporados con ejemplos. En el caso de los actos directivos, los ofrecimientos y las preguntas, nótese que se han agrupado bajo estas categorías actos que forman parte de ellos (ej., peticiones, recomendaciones, advertencias, etc., en los directivos). Dado el reducido número de muestras y la imposibilidad de distinguirlos con precisión en la totalidad de los casos documentados (ej. propuestas e invitaciones), no se han establecido subcategorías específicas, salvo en las interrogativas, que se han subdividido en seis tipos. Esto se debe al elevado número de muestras registradas –consecuencia de anotar un corpus de entrevistas– y a que su tipología suele verse reflejada en su formulación.

Acto de habla	Contenido	Muestras ⁶
Saludo / presentación	Saludos y presentaciones que aparecen de forma independiente o combinados.	[Acto] I: hola / sí hola [Respuesta] E: hola / ¿María Emilia? ¿qué tal? // soy Marta (SCOM_M12_020)
Acto directivo	Actos cuya función es <i>dirigir</i> la acción del oyente. Formulación frecuentemente en imperativo.	

⁶ Para facilitar la lectura de los ejemplos, las barras (/) representan las pausas, mientras que las barras dobles (//) señalan las pausas largas. En cuanto a los participantes, E identifica al entrevistador/a e I al informante (o entrevistado/a). Se marcan en cursiva los fragmentos de discurso directo.



Acto directivo		
– Advertencia	El hablante alerta al interlocutor de un peligro o consecuencia negativa, con el fin de evitar un problema.	[Acto] I: [...] <i>y usted // poniendo este tratamiento / y tal como está / no debería de andar // eeh limpiando las comunidades</i> [sin respuesta] (SCOM_M31_038)
– Recomendación/consejo	El hablante propone o sugiere algo que considera conveniente o adecuado para una situación determinada.	[Acto] I: [...] <i>y ¿por qué no vais todos?</i> [sin respuesta] (SCOM_M33_009)
– Solicitud/petición	El hablante requiere del interlocutor una acción, información o servicio. Se suelen formular de manera atenuada o cortés.	[Acto] E: de eso quería que me hablaras un poquito también / de tus aficiones deportivas // [Respuesta] I: bueno / pues eso / ¡ostra! es que eso / es que eso es lo (SCOM_H32_032)
– Petición	El hablante solicita al interlocutor que realice una acción cuyo beneficio recae en el hablante.	[Acto] E: ¿me podría describir un lugar que le guste aquí de Santiago concreto / donde usted se sienta / bien / relajado? / tranquilo / a gusto / [Respuesta] I: mmm en alguno de los parques ehm <silencio/> el mmm / el parque de Bonaval // (SCOM_H23_003)
– Orden	El hablante manda realizar una acción de forma directa.	[Acto] E: dijo / <i>no hagas ni caso Rosa déjala que salga</i> // [sin respuesta] (SCOM_M31_045)
Ofrecimiento		Actos orientados a beneficiar al interlocutor mediante la propuesta de un producto, servicio o acción.
– Ofrecimiento	El hablante pone algo a disposición del interlocutor.	[Acto] I: vamos a ver / ¿qué quieres? ¿un café? / sí [Respuesta] E: pues no / muchas gracias (SCOM_M33_005)
– Propuesta	El hablante plantea una determinada acción o decisión como una opción posible.	[Acto] I: y eh cuando me vio así se enteró dijo ah <i>¿por qué no te vienes allí para la federación?</i> tal dije [Respuesta] I: ¿yo qué voy a <ruido tipo="golpe"/> a hacer a la federación? // (SCOM_H21_053)
– Invitación	El hablante propone una actividad o realizar una acción	[Acto] E: <ruido tipo="teléfono"/> coge si quieres eh [Respuesta] I: no no no quiero cogerle <silencio/> (SCOM_H11_047)

Otros		
Cita	Fragmento de discurso reproducido en estilo directo.	E: [...] y yo / pues dije [cita] / <i>pues / voy a por todas / el no ya lo tengo // y voy a por el sí // [...]</i> ⁷ (SCOM_M32_025)
Pregunta		
– Interrogativa total	Ofrecen dos opciones de respuesta en su formulación, generalmente antagónicas (<i>sí/no</i>).	[Acto] I: ¿tú vienes en el / coche? [Respuesta] E: sí (SCOM_H12_031)
– Interrogativa parcial	No presentan una opción de respuesta cerrada y suelen ir introducidas por un determinante o adverbio interrogativo (<i>quién, cuándo, qué, donde...</i>).	[Acto] E: ¿y qué tipo de cine te gusta por ejemplo? [Respuesta] I: no eeh fff lo de pagar ocho euros para una / peli / bastante / regular que suele ser lo habitual // salvo excepciones [...] (SCOM_H11_052)
– Interrogativa disyuntiva	Plantean una elección entre varias opciones explícitas o implícitas. Se distinguen de las totales por el uso de una conjunción disyuntiva en su formulación.	[Acto] E: ¿de usted o de tú? [Respuesta] I: de tú mmm mis hijas me tratan de tú y tú no vas a ser menos / (SCOM_H33_007) [Acto] E: pero ¿lo estás sacando porque te vas a trabajar o algo o? [Respuesta] I: espero trabajar / y sobre todo no aquí / en otro lado / que esto es muy pequeño <silencio/> (SCOM_H11_047)
– Interrogativa de eco	Repiten una parte o la totalidad del contenido del turno anterior.	I: sí es que me casé jovencita / me casé de diecisiete años [Acto] E: ¿diecisiete? [Respuesta] I: sí (SCOM_M23_004)
– Interrogativa inconclusa	Enunciados interrogativos incompletos. Suelen acabar con un alargamiento vocálico y una cadencia o anticadencia en el tonema.	[Acto] E: ¿y le gustaría tener <alargamiento>un</alargamiento>? // [Respuesta] I: eso ya es cuestión de ellos / esto es / esto es tan también tan privado que / que eso la libertad de cada uno // [...] (SCOM_H32_033)

⁷ El etiquetado de las citas se limita al fragmento reproducido, sin tener en cuenta la expresión introductora (*dije*, en el caso del ejemplo incluido en la tabla 1). Nótese que se han marcado otros actos de habla (ej., ofrecimientos, saludos, etc.) que puedan aparecer en el contenido reproducido en las citas.



Pregunta		
– Enunciativa con confirmación	Secuencias enunciativas, afirmativas o negativas, seguidas de una partícula interrogativa (generalmente ¿no?), que funciona como petición de confirmación.	[Acto] E: bueno ahora tendrá tiempo para / no sé / para dedicarse a la familia a los hobbies ¿no? / [Respuesta] I: a la familia / o t</palabra_cortada> / tienes que meter / o sea ahora no porque ahora en el verano vendrá toda la / todos <vacilación/> mis nietos y mis hijos y tal (SCOM_M33_005)

Tabla 1. Descripción y ejemplificación de los actos de habla anotados (en negrita) y los subtipos que algunos recogen. Fuente: adaptado de la Guía Anotación Pragmática (ESLORA v. 2.3)⁸

El acto *Saludos / Presentaciones* se denominó de esta manera porque los interlocutores, entrevistador y entrevistado, cuando se saludan por primera vez, suelen presentarse o bien responder al saludo con una presentación. En el caso de las citas directas, como se anticipó, ya estaban anotadas en el corpus, por lo que se recuperaron y añadieron a este listado; cabe mencionar que se etiquetaron también los posibles actos de habla que se registraran en el fragmento reproducido (ej., una cita que refiera una orden, como se muestra en la Tabla 1).

Para la anotación de las preguntas, se tuvieron únicamente en cuenta aquellas que reciben o, al menos, esperan una respuesta por parte del interlocutor. Por tanto, se descartaron las preguntas retóricas o los enunciados en los que el hablante utiliza alguna partícula como marcador de alteridad, para involucrar al receptor (¿no?, ¿eh?), pero no como solicitud de confirmación. Las interrogativas se clasificaron siguiendo a Escandell (1999) en totales, parciales, disyuntivas y de eco. A estas categorías se sumaron las secuencias enunciativas con una partícula con valor confirmativo y las *inconclusas*. Si bien estas últimas no constituyen un tipo de pregunta reconocible en las clasificaciones de enunciados interrogativos, se han distinguido porque aparecen con frecuencia en las entrevistas analizadas. Se consideran interrogativas inconclusas aquellas en que el hablante (generalmente el entrevistador) espera que el interlocutor (típicamente el entrevistado) tome el turno para responder; presentan características fónicas específicas, como un alargamiento marcado y una cadencia o anticadencia finales en el contorno tonal. Este tipo se distingue de las interrogativas que aparecen incompletas, debido a solapamientos con el interlocutor o a interrupciones propias de la lengua coloquial (elipsis, incisos, autocorrecciones, vacilaciones, etc.), que se incluyen en la categoría que les correspondería si apareciesen completas.

3.3. Procedimiento de anotación pragmática

El proceso de anotación pragmática se llevó a cabo en el editor XLM. Para ello, se emplearon las transcripciones de cada entrevista, junto con los metadatos respectivos. La empresa tecnológica NLPgo Technologies S.L. elaboró un etiquetario a demanda adaptado a las necesidades computacionales del sistema, que recogía cada uno de los actos de habla distinguidos (§3.2.); además, en el caso de las interrogativas, al seleccionar esta categoría en el listado era preciso indicar el tipo (Figura 1).

El etiquetario dispone de marcas específicas de apertura y de cierre que debían añadirse de forma manual, al comienzo y final, en cada acto de habla, así como en las respectivas respuestas. Las entrevistas del corpus se encuentran transcritas y almacenadas individualmente en formato XML; la anotación se realizó siguiendo el mismo procedimiento en cada una de ellas: (i) revisión de la transcripción de cada entrevista e identificación de un acto de habla en el documento XML, (ii) etiquetado mediante la inserción de las etiquetas de apertura y cierre al inicio y final del acto dentro del turno iniciativo; (iii) anotación de la intervención reactiva.

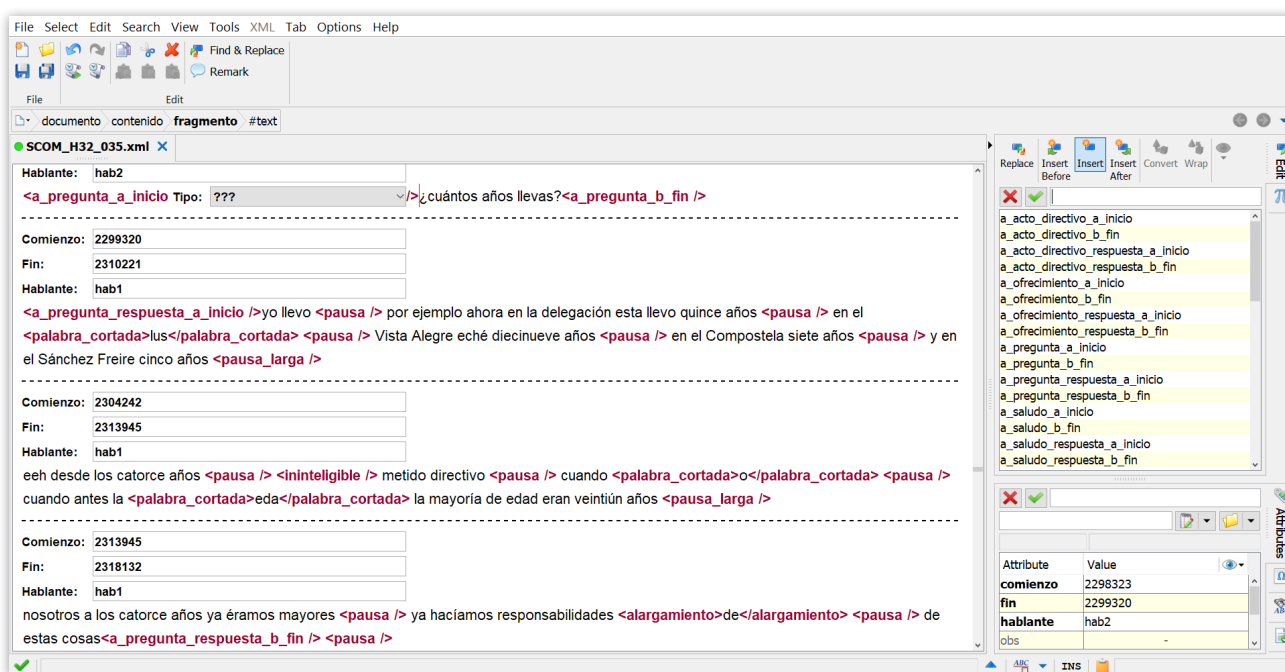


Figura 1. Muestra del entorno de anotación en el editor XML

En la figura 1, a la derecha, puede verse una sección del etiquetario utilizado para la anotación pragmática. El texto reproduce el fragmento siguiente: *–¿cuántos años llevas? (hab2) –yo llevo / por ejemplo ahora en la delegación esta llevo quince años [...] (hab1)*. En la transcripción observamos las marcas de anotación ya existentes, como ‘pausa’, ‘pausa_larga’, ‘alargamiento’, ‘ininteligible’ o ‘palabra_cortada’. A ellas se ha añadido la anotación de la pregunta (<a_pregunta_a_inicio>, <a_pregunta_b_fin>), junto con su ‘tipo’ (interrogativa parcial), que se selecciona en el desplegable. También aparece anotada la respuesta, cuyo inicio (<a_pregunta_respuesta_a_inicio>) comienza en *yo llevo*, y el cierre (<a_pregunta_respuesta_b_fin>) se encuentra en el fragmento *de estas cosas*, en la sección inferior de la imagen.

El diseño de la parte lingüística del sistema de etiquetado se llevó a cabo en el marco del proyecto indicado en la nota 1. Para la anotación y revisión de los datos, se contó además con la contratación de dos investigadoras colaboradoras.



3.4. Retos metodológicos

A lo largo del proceso de etiquetado pragmático del corpus ESLORA surgieron distintos obstáculos metodológicos, que obedecen, fundamentalmente, al procedimiento de anotación manual del corpus, a los requisitos técnicos del sistema de etiquetado y a la variabilidad de la lengua oral. En esta sección repasamos algunos de ellos y las soluciones que se adoptaron.

3.4.1. Ausencia de respuesta (turno reactivo)

Como se explicó en 3.2., el sistema de anotación se organizó en forma de pares adyacentes, por lo que eran siempre necesarios dos turnos –iniciativo y reactivo– de dos interlocutores distintos. La adopción de un modelo más flexible o abarcador, aun cuando permitiría una descripción estructural más precisa (*vid.* §3.4.2.), no fue viable. Como suele ocurrir en tareas de anotación, el aumento del detalle tiende a reducir la precisión y la consistencia del etiquetado, al multiplicarse los casos ambiguos y los problemas de segmentación. Por otro lado, el procesamiento automático de subcategorías también puede resultar problemático para la recuperación de casos, al fragmentar excesivamente los resultados devueltos por el sistema. Por todo ello, se optó por el par adyacente como unidad central de segmentación, a pesar de los problemas expuestos.

Durante el proceso de etiquetado, se comprobó que son numerosos los casos en los que un par se presenta incompleto, es decir, sin el turno reactivo. Los requisitos computacionales del sistema no permiten que haya un acto de habla sin respuesta, como, por ejemplo, un saludo o pregunta sin contestar. Por este motivo, fue necesario incorporar un identificador *ad hoc*, ‘Respuesta: no’ (*vid.* Figura 2), que se situaba en los turnos iniciativos implicados:

Comienzo:	2911306
Fin:	2919174
Hablante:	hab1

<a_pregunta_a_inicio Respuesta: ☒ no /> ¿y qué medio de

Tipo: interrogativa inconclusa

<alargamiento>referencia</alargamiento><a_pregunta_b_fin /> <pausa /> o sea <a_pregunta_a_inicio

Tipo: interrogativa parcial /> ¿qué medio te gusta a ti en ese sentido por su honestidad o

<alargamiento>su</alargamiento>?<a_pregunta_b_fin /> <pausa />

Comienzo:	2919174
Fin:	2920522
Hablante:	hab2

<a_pregunta_respuesta_a_inicio /> la radio <pausa_larga />

Figura 2. Muestra del proceso de anotación de una interrogativa sin respuesta en XLM

En la Figura 2, el contenido del turno anotado es el siguiente: *¿y qué medio de referencia / o sea qué medio te gusta a ti en ese sentido por su honestidad o su /?* (hab2) *la radio* (hab1). Se han distinguido dos preguntas, puesto que hay una pausa entre ellas, una inconclusa y una parcial. Aparecen las etiquetas de apertura y cierre del turno iniciativo en cada una de ellas



(<a_pregunta_a_inicio>, <a_pregunta_b_fin>), junto con la etiqueta de apertura del turno reactivo (<a_pregunta_respuesta_a_inicio>). Puesto que no se responde a la primera pregunta, se ha marcado con el identificador 'Respuesta: no'.

En las muestras de ESLORA, la falta de turno reactivo se documenta frecuentemente en los saludos dirigidos a personas del ámbito del entrevistado que, o bien se encuentran en el domicilio en el momento de la entrevista, o bien interrumpen entrando en la sala en la que tiene lugar, como es el caso de (1): la entrevista se realizó en el despacho del informante, en un colegio en el que trabajaba como administrativo; una persona entra para indicarle que un niño (*un cativo*) va a acudir al día siguiente a recoger un carné escolar y la entrevistada lo saluda, pero él no le responde. La razón podría deberse a la voluntad de no interrumpir –o hacerlo lo menos posible–, o a que no haya oído el saludo:

- (1) I: pero antes hacían un folio / escrito dicen oh jo <ruido tipo="palmada"/> cuánto <risa/> hice / y tal tal pero esto no / esto es / esto es una pantallita
E: ah
A: bueno / perdóname un minuto <risa/> mira / te va a venir mañana un un cativo
E: **hola**
I: un
A: para buscar / un carné escolar // (SCOM_H22_024)⁹

Otra situación habitual en la que el par adyacente queda incompleto se da en las preguntas: el informante está hablando sobre algún tema con visible emoción (enfado, agitación, alegría, exaltación, etc.); el entrevistador le formula alguna pregunta y este la ignora, y prosigue su narración. Muestra de ello son los ejemplos (2) y (3): en (2), la entrevistada relata de manera entusiasta lo mucho que le gustó Florencia pasando por alto la interrogativa de eco de la entrevistadora; en (3), la informante recuerda, con indignación, sus condiciones laborales pasadas, ignorando la pregunta formulada por el entrevistador después de un silencio:

- (2) E: ¿y por ejemplo en / en Italia o o / mmm ¿qué qué visteis por allí?
I: Italia me encantó Florencia
E: ¿**Florencia?**
I: en Italia fuimos a Roma / sobre todo a Roma y Florencia / luego vinimos nos fuimos en coche y vinimos por la Costa Azul toda / estuvimos en Mónaco que me encantó / también / es un mundo aparte claro Mónaco <risa/> / y Flo y mmm / <ruido tipo="chasquido boca"/> esto Florencia me encantó [...] (SCOM_M23_004)
- (3) I: [...] no era / bienvenida a esas cosas / pero a él le interesaba eso y luego me voy a enterar voy al inspector / inspección / y veo que me tiene con una categoría más baja de lo que me comon que lo que me correspondía <silencio/>
E: ¿**y ya decidió irse?**
I: ya / y no tenías nómina o sea te daba el dinero / y no tenías / no figurabas en una nómina ni nada o sea / [...] (SCOM_M33_011)

La falta de respuesta también ocurre cuando los entrevistadores formulan varias preguntas en un mismo turno. Generalmente los informantes solo responden a una, ya sea porque se trata de preguntas contiguas semejantes (4) o porque se reformula la primera (5):

⁹ Las negritas, cursivas y subrayados son nuestros.



- (4) E: y **¿qué / qué problemas viste en el ciclo?** <silencio/> **¿qué es lo que te decepcionó? ¿qué esperabas?**
 I: daba demasiado asignaturas que creo que no eran / muy esenciales / [...] (SCOM_H11_051)
- (5) E: y usted bueno **¿son muchos hermanos? / ¿eran muchos hermanos? /**
 I: nosotros somos cinco / (SCOM_H33_007)

En casos como el de (5), no siempre es posible determinar a qué pregunta corresponde la respuesta dada por el informante. Para solventarlo, se podía crear una nueva etiqueta que agrupara varias preguntas, pero, finalmente, se consideró que se respondía a la última, como criterio homogeneizador en la anotación de todo el corpus, y se empleó el identificador 'Respuesta: no' en las restantes.

3.4.2. Secuencias ampliadas (o encadenamientos) de actos

En el corpus se documentan casos en que los entrevistadores formulan varios actos de habla encadenados de un mismo tipo (ej. (6), preguntas consecutivas) y el informante responde a todos ellos. Para evitar que las respuestas se crucen entre sí o no se asocien correctamente con los turnos iniciativos correspondientes¹⁰, fue necesario insertar identificadores numéricos en cada uno de los actos de habla y de las respuestas obtenidas. De este modo, los identificadores permiten enlazar correctamente cada turno iniciativo con el reactivo.

Esto ocurre, por ejemplo, cuando se formulan dos preguntas seguidas, generalmente una interrogativa total o disyuntiva, seguida de una parcial con una formulación más abierta, como en (6): primero una interrogativa total, que confirma que el marido andaluz de la informante se trasladó a Galicia (*¿vino para Galicia?*); y, a continuación, una parcial más amplia, cuyo objetivo es obtener una narración extensa con información nueva (*¿cómo fue?*).

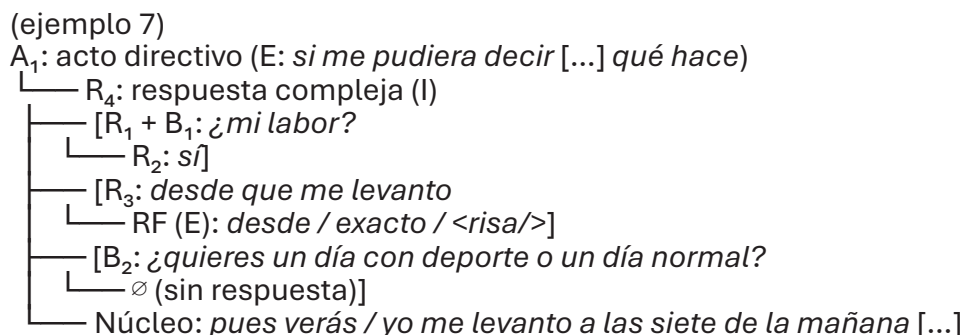
- (6) E: ¿me decías que te casaste con un andaluz?
 [...]
 E: **¿vino para Galicia? ¿cómo fue?**
 I: sí fue simpático yo / eeh / cuando acabé COU con diecisiete años / eeh nosotros nos fuimos de excursión / ¿no? / yo fui con / yo fui de excursión con mi hermana / con el curso de mi hermana / que hacía sexto // [...] (SCOM_M23_001)

Otra situación recurrente, especialmente en un corpus recopilado en Galicia, es que se conteste a un acto directivo o a una pregunta con otra pregunta (7):

- (7) E: bueno / **si me pudiera decir más o menos en un día normal qué qué qué hace** [*acto directivo*]
 I: ¿mi labor? [respuesta al acto directivo + interrogativa total]
 E: sí [respuesta a la interrogativa total]
 I: desde que me levanto
 E: desde / exacto / <risa/>
 I: pues / ¿quieres un día con deporte o un día normal? [sin respuesta] // pues verás / yo me levanto a las siete de la mañana (SCOM_H22_024)

¹⁰ Nótese que no hay riesgo de entrecruzamiento entre actos de habla distintos, ya que cada uno se identifica con un nombre específico (ej., ofrecimiento, pregunta, acto directivo, etc.).

El ejemplo (7) destaca por su riqueza interactiva. Ahora bien, esta misma riqueza atenta contra los límites de un sistema de anotación diseñado y organizado en forma de pares adyacentes, que espera una estructura de estas características: A_1 (acto directivo) $\rightarrow R_1$ (respuesta a A_1); B_1 (acto directivo) $\rightarrow R_2$ (respuesta a B_1). En este tipo de ejemplos encontramos una jerarquía estructurada en varios niveles, como se representa en el siguiente gráfico¹¹:



El hecho de anotar lingüísticamente ejemplos de estas características implica poner vallas al campo, ya que no se puede determinar con certeza dónde comienza la respuesta al acto directivo inicial: ¿en el turno inmediatamente posterior (*¿mi labor?*), o bien en el momento en el que el informante por fin cuenta todo lo que hace en un día normal (*pues verás / yo me levanto a las siete...*)? La solución adoptada en este tipo de casos fue considerar como respuesta el intercambio completo, desde la pregunta aclaratoria inicial, para evitar perder información relevante de la interacción. A su vez, se anotaron individualmente las distintas preguntas y respuestas que se suceden entre ambos interlocutores, por medio de identificadores numéricos, para evitar entrecruzamientos y perder esta información.

Una situación similar a la anterior se produce en el caso de los ofrecimientos, donde comprobamos que es frecuente que el acto de habla no se limite al par “ofrecimiento-aceptación” propiamente. En los casos registrados en ESLORA, algunos entrevistados insisten en ofrecer una bebida al entrevistador cuando lo reciben en su domicilio: el ofrecimiento se reitera sucesivamente como forma de mostrar cortesía, aunque pueda llegar a percibirse como una imposición (Gestido de la Torre, 2006). Así, el ofrecimiento se considera un acto comisivo (Searle, 1969), cuyo beneficio recae sobre el interlocutor. Por este motivo, como señala Ruiz de Zarobe (2001, p. 263), en algunas culturas “[...] no está bien visto aceptar de entrada una invitación, por lo que se establece un “juego” de invitaciones y rechazos hasta que finalmente se acepta la invitación”. Esto es lo precisamente lo ocurre en el ejemplo (8):

- (8) I: [...] (1) **¿qué quieres? ¿un café?** / sí
 E: pues no / muchas gracias
 I: y (2) **¿qué quieres? ¿agua?**
 E: un vag</palabra_cortada> / un vasito de agua sí le agradezco sí
 I: (3) **¿un vasito de agua?** bueno te pongo ahí un vasito de agua / (4) unas galletas / o algo así / ¿quieres?
 E: pues es que a</palabra_cortada> / acabo de comer / muchas gracias
 I: mira si quie</palabra_cortada> / acabas de comer / (5) mira si quieres un nescafé o una infusión <ininteligible/>
 E: no / le agradezco un poquito de agua porque vengo acalorada y todo
 I: sí / vienes / es que hace calor / (SCOM_M33_005)

¹¹ Los gráficos han sido realizados con inteligencia artificial (Chat GPT).



Lingüísticamente, se puede apreciar claramente el *juego* al que hace referencia Ruiz de Zarobe (2001): se ofrece primero preguntando directamente (*¿qué quieres? ¿un café?*) y se rechaza atenuando por medio del agradecimiento precedido de un marcador (*pues no / muchas gracias*); la entrevistada repite entonces la misma estructura proponiendo un cambio de bebida (*y ¿qué quieres? ¿agua?*), a lo que la entrevistadora acepta de manera atenuada con un diminutivo (*un vasito de agua sí le agradezco sí*). Por último, a la entrevistada parece que le resulta insuficiente ofrecerle solo un vaso de agua a su invitada, y entonces propone algo de comer (*unas galletas o algo así ¿quieres?*), pero su interlocutora rechaza y agradece de nuevo, justificando que acaba de comer. La informante aprovecha entonces para ofrecerle una bebida más apropiada para la sobremesa, como una infusión o un Nescafé, pero la entrevistadora insiste en que con el vaso de agua es suficiente (*no / le agradezco un poquito de agua porque vengo acalorada*).

La extensión del ofrecimiento en varios turnos encadenados plantea un reto en lo que respecta a la anotación pragmática: ¿se debería registrar cada uno de los ofrecimientos, de forma individual (8a), o bien el conjunto del intercambio como si se tratara de una extensión del primero (8b)?

(ejemplo 8a)

O₁: ofrecimiento principal (I: *¿qué quieres? ¿un café?*)

└─ R₁: *pues no / muchas gracias*

O₂: reformulación del ofrecimiento (I: *¿qué quieres? ¿agua?*)

└─ R₂: *un vasito de agua sí le agradezco sí*

O₃: ampliación del ofrecimiento (I: *¿un vasito de agua? / ¿unas galletas...?*)

└─ R₃: *acabo de comer / muchas gracias*

O₄: nueva reformulación (I: *si quieres un Nescafé o una infusión...*)

└─ R₄: *le agradezco un poquito de agua porque vengo acalorada [...]*

(ejemplo 8b)

O₁: ofrecimiento inicial (I: *¿qué quieres? ¿un café?*)

└─ R₁: *pues no / muchas gracias*

└─ Expansión de O₁ (secuencia de reformulación de ofrecimiento)

└─ O₂: *¿qué quieres? ¿agua?*

└─ R₂: *un vasito de agua sí le agradezco sí*

└─ O₃: *¿un vasito de agua? / unas galletas...*

└─ R₃: *acabo de comer / muchas gracias*

└─ O₄: *si quieres un Nescafé o una infusión...*

└─ R₄: *le agradezco un poquito de agua porque vengo acalorada [...]*

Al igual que en el ejemplo (7), en que se consideró que la respuesta a la pregunta inicial abarcaba todo el intercambio, aquí se ha optado por establecer como turno reactivo la respuesta al ofrecimiento inicial seguida de la articulación de ofrecimientos-rechazos posterior, que refleja el proceso de negociación entre las hablantes (8b). Así, los ofrecimientos sucesivos no introducen nuevas acciones ilocutivas, sino que funcionan como insistencias o intentos de reparación del ofrecimiento inicial rechazado. Por tanto, desde esta perspectiva, los ofrecimientos sucesivos son *actos* que forman parte de una misma *intervención* (Grupo Val.Es.Co., 2014, p. 14), o bien *secuencias ampliadas* (Sacks, Schegloff y Jefferson, 1974, p. 711), que adoptan la forma de pares adyacentes adicionales expandiendo el par original. Este se mantiene como el foco del intercambio entre interlocutores, de modo que las reformulaciones en forma de nuevos ofrecimientos y los rechazos siguientes no producen un nuevo acto de habla que deba anotarse de manera independiente, sino que se integran en el mismo.

3.4.3. Problemas de identificación de actos de habla

La revisión manual de las transcripciones del corpus, junto con los audios correspondientes, ha revelado algunas dificultades en la identificación de ciertos actos de habla. En el caso de las preguntas, cabe destacar que su reconocimiento, en un corpus ya transcrito, no siempre resulta tan sencillo como cabría esperar¹². En la revisión de las transcripciones, encontramos numerosos enunciados que se interpretan como interrogativos en el contexto de la entrevista, pero carecen de signos de interrogación, o bien de una pronunciación marcadamente interrogativa. Muestra de ello son los ejemplos (9) y (10) de interrogativas de eco:

- (9) I: [...] de vez en cuando pues me junto con eh con algunos amigos / eeh costumbres eeh lo de los dos cafés después de comer en el bar mientras leo los periódicos mmm prácticamente todos los días <silencio/>
E: hm hm / sí sí
I: demasiada televisión <risa/> /
E: **demasiada televisión**
I: sí (SCOM_H23_003)
- (10) I: nació nació en la habitación / tres cientos trece igual / que su hermano <silencio/> pero <silencio/> <ruido tipo="carraspeo"/> /
E: ¡ay! // ¡qué guapa!
I: pero el día trece // pero es que yo nací //
E: **¡el día trece!**
I: sí pero es que yo nací / eeh no mi hijo nació el cuatro de agosto <silencio/> (SCOM_H32_032)

Se ha comprobado que existe cierta variabilidad en la transcripción de las interrogativas de eco (cf. Mas Álvarez y Sampedro Mella, 2023; Sampedro Mella, Sarymsakova y Mas Álvarez, 2026), que afecta a otras clases de interrogativas. Así, estas pueden aparecer transcritas sin signos de interrogación, debido a su articulación poco acentuada, cuando presentan valor confirmativo (9) o, incluso, entre signos de exclamación, cuando se utilizan con una entonación muy marcada para mostrar una actitud ante lo dicho por el interlocutor, generalmente, sorpresa (10). Se considera que se trata de enunciados interrogativos, ya que obtienen una respuesta en forma afirmativa o negativa –también, en ocasiones, una repetición literal o una vocalización con ese mismo sentido–, como si se tratara de preguntas totales. A pesar de la variabilidad prosódica y ortográfica, el contexto discursivo y la dinámica del intercambio indican que estamos ante interrogativas. En consecuencia, la variación en la transcripción de estos enunciados implica una revisión atenta de las transcripciones, acompañada en muchos casos del audio, y no una mera recuperación de casos a través de una búsqueda por signos ortográficos (¿?).

Por otro lado, no siempre encontramos correspondencias unívocas entre los enunciados transcritos entre signos de interrogación y las preguntas. Esto suele ser debido a vacilaciones, falsos comienzos, reformulaciones o muletillas propias de la lengua oral, junto con las variaciones prosódicas en la entonación. Sirvan de muestra los siguientes ejemplos de preguntas formuladas por distintos entrevistadores:

¹² Vid., para más información sobre los problemas de identificación de las interrogativas, Mas Álvarez y Sampedro Mella (2023), a propósito de las interrogativas de eco.



- (11) E: **¿y el máster? ¿de qué era?** / (SCOM_H13_014)
- (12) E: vas / eso / **vas a pescar ¿o vas a solo a mirar o vas a pescar?** (SCOM_H11_047)
- (13) E: y entonces cuando **¿si no estás en el trabajo? // ¿qué haces? // o ¿qué te gusta hacer? o ¿qué te gustaría hacer?** (SCOM_H12_027)
- (14) E: mmm ya / pero ¿qué? / **¿cuáles te gustaría así hay algún perfil / que te guste más?** // (SCOM_H11_051)
- (15) E: pero así **¿no has oído lo q? mmm ¿no? / ¿lo que ha hecho por los medios o? //** (SCOM_H11_051)

Para llevar a cabo la anotación y la clasificación de las interrogativas, se siguieron, por este orden, criterios de carácter pragmatológico y fonopragmático. Primero, se identificó la secuencia que, morfosintácticamente, corresponde al enunciado interrogativo; a continuación, se determinó qué segmento(s) de dicha secuencia presentan un patrón entonativo marcadamente interrogativo. Las negritas de los ejemplos (11-15) reflejan la segmentación, aplicada en cada caso, de acuerdo con estos dos criterios.

En (11), aparecen en un mismo turno dos interrogativas que se completan sintácticamente; en estos casos se etiquetaron juntas, como si se tratara de una sola pregunta, siguiendo el criterio de Escandell (1999). En (12), se ha marcado como interrogativa disyuntiva la totalidad de la secuencia en negrita, incluyendo una parte que no aparece delimitada entre signos de interrogación, pero que forma parte del mismo enunciado interrogativo, desde el punto de vista sintáctico, prosódico y pragmático (*vas a pescar ¿o vas a solo a mirar o vas a pescar?*) En (13), se han distinguido dos interrogativas, atendiendo a la organización sintáctica y a la prosodia de cada una en la intervención: *si no estás en el trabajo // qué haces* y *qué te gusta hacer o qué te gustaría hacer*.

En el caso de (14), el audio revela que, aunque se haya marcado una sola interrogativa, hay dos independientes, integradas en una misma intervención: una primera parcial (*cuáles te gustaría*) y una total (*así hay algún perfil que te guste más*). Por último, en el ejemplo (15), de la misma entrevista, si bien la transcripción identifica dos preguntas, se ha considerado que se trata de una única, clasificada como interrogativa inconclusa a tenor de la entonación suspensa (*¿no has oído lo q? mmm ¿no? / ¿lo que ha hecho por los medios o?*).

Los problemas de identificación nos llevan a cuestionar, finalmente, los límites entre los distintos actos de habla, que no están unívocamente representados por unidades lingüísticas concretas, motivo por el cual la anotación pragmática es más compleja y menos precisa que la morfosintáctica (Weisser, 2015; Milà-García, 2018). Nos detenemos en este caso en las dificultades de clasificación de ciertos enunciados, que podrían ser considerados a la vez actos directivos y ofrecimientos, pese a tratarse de actos contrapuestos (*vid.* Sampedro Mella, 2021b, para más información).

- (16) **vamos a buscar** un sitio donde estemos sin mucho ruido // (SCOM_H11_052)
- (17) **a ver / subimos** porque es la tercera planta mejor por el ascensor (SCOM_H12_031)
- (18) **si quieres cuéntame un poco de ti** y de lo que te gusta (SCOM_H11_052)
- (19) **bueno si te parece bien pasamos a la segunda parte** / que es un cuestionario (SCOM_H12_028)



Los ejemplos (16-19) se han clasificado como actos directivos, a pesar de que su formulación atenuada (18-19), o con un plural inclusivo (16-17), podría hacernos pensar que se trate de invitaciones. Cabe tener en cuenta que, en los actos de habla comisivos, como en los ofrecimientos, el hablante pone a disposición del interlocutor algún beneficio (8). Aquí este no es necesariamente el caso, ya que, en el contexto de la entrevista semidirigida, estos actos, así formulados, permiten guiar o influir en la conducta del receptor, aunque se haga de forma mitigada.

Compárense estos ejemplos con (20-24), clasificados como ofrecimientos:

- (20) mira **¿qué te parece si te sientas ahí?** (SCOM_M33_009)
- (21) **usted siéntese donde más cómodo se sienta** <ruido tipo="sillas"/> (SCOM_H31_043)
- (22) **coge la cómoda / ponte allí ponte allí** (SCOM_H12_028)
- (23) **siéntate** (SCOM_M21_049)
- (24) **pasa** (SCOM_M32_025)

Por su formulación, los ejemplos (20-24) no distan de (16-19), ya que en ambos se emplean formas imperativas, condicionales o expresiones preparatorias (*qué te parece si*). Ahora bien, si atendemos al contenido de los enunciados, en este caso no se pretende organizar el desarrollo de la interacción, sino proponer algo cuyo beneficio redunda en el interlocutor: sentarse, pasar, situarse en un lugar, etc., por lo que, en este contexto, se considera que funcionan como invitaciones: los imperativos no adquieren un valor de orden, sino de ofrecimiento cortés¹³. Por tanto, el criterio establecido para clasificar estos actos es la orientación de la acción, de acuerdo con el papel de los participantes y las características contextuales: por un lado, la necesidad de guiar la interacción; por otro, la invitación al lugar.

4. RESULTADOS: OPCIONES DE BÚSQUEDA¹⁴

El sistema de anotación pragmática del corpus ESLORA permite realizar dos tipos de búsqueda en la aplicación de consulta: 1) búsqueda general de actos de habla y 2) búsqueda de actos de habla en combinación con otros elementos lingüísticos recogidos en el etiquetado morfosintáctico (palabras, lemas, categorías, etc.). Ambas búsquedas se pueden ajustar, a su vez, con los filtros generales del corpus: edad, sexo, nivel de estudios y papel del hablante –entrevistador, informante o audiencia–.

4.1. Búsqueda simple de actos de habla

Para llevar a cabo el primer tipo de búsqueda, se selecciona la categoría *Palab. ortográficas* en la pestaña *Tipo*, que aparece por defecto al abrir la aplicación de consulta. En *Tipo de Resultado* se marca *Intervenciones*, como se muestra en la Figura 3:

¹³ Nótese la diferencia si estos mismos actos se formularan en una comisaría, por ejemplo. Es el contexto, junto con las circunstancias de la interacción y los interlocutores, además de otras variables, como la entonación, los que determinan, en definitiva, cada tipo de acto de habla. Este es uno de los motivos por los que la anotación pragmática requiere interpretación y, por ello, resulta tan compleja.

¹⁴ La información y algunas de las imágenes y ejemplos de esta sección se han tomado de la Guía Anotación Pragmática, disponible en https://eslora.usc.es/guide_pragmatic_annotation. En ella puede encontrarse esta información de manera ampliada.

The screenshot shows the ESLORA search interface. The top navigation bar includes links for Información, Búsquedas, Listados, Guía, Descargas, and Contacto. The search filters are organized into sections: Búsqueda, Resultado, Ordenación, and Tamaño de página. Under Búsqueda, there are dropdowns for Corpus (Cualquiera), Tipo (Palab. ortográficas), and Sensibilidad (Acentos, Mayúsculas). Under Resultado, there is a dropdown for Tipo de resultado (Concordancias, Frecuencia simple, Frecuencia completa, Concordancias, Inventarios, Intervenciones) and a dropdown for Ordenación (Coincidencia). Under Ordenación, there are dropdowns for Desde, Hasta, Nivel de estudios, and Sexo. Under Tamaño de página, there is a dropdown for Tamaño de página (50). The search results section shows a list of results with columns for Tipo de intervención, Desde, Hasta, Nivel de estudios, and Sexo.

Figura 3. Búsqueda de actos de habla

Se puede buscar el par adyacente completo (ej., preguntas + respuestas) o un solo constituyente (ej., respuestas a las preguntas). También se pueden combinar las búsquedas de forma simultánea (ej., saludos, ofrecimientos y respuestas a los actos directivos). Para ello, se selecciona la categoría, o categorías, que se desee obtener en el desplegable *Tipo de intervención*, como se ilustra en la Figura 4:

The screenshot shows the ESLORA search interface with the search filters and results options. The search filters are organized into sections: Búsqueda, Resultado, Ordenación, and Tamaño de página. Under Búsqueda, there are dropdowns for Corpus (Cualquiera), Tipo (Fragmentos), and Sensibilidad (Acentos, Mayúsculas). Under Resultado, there is a dropdown for Tipo de resultado (Intervenciones) and a dropdown for Ordenación (Documento). Under Ordenación, there are dropdowns for Desde, Hasta, Nivel de estudios, and Sexo. Under Tamaño de página, there is a dropdown for Tamaño de página (50). The search results section shows a list of results with columns for Tipo de intervención, Desde, Hasta, Nivel de estudios, and Sexo. The Tipo de intervención dropdown is open, showing a list of options: Saludo / Presentación, Saludo / Presentación inicio, Saludo / Presentación respuesta, Acto directivo, Acto directivo inicio, Acto directivo respuesta, Ofrecimiento, and Ofrecimiento inicio.

Figura 4. Selección del tipo de acto de habla

En el listado aparecen sombreados en gris los pares adyacentes completos, mientras que, en blanco, se puede ver cada uno de los constituyentes del par. El sistema devuelve todos los casos recuperados correspondientes al acto de habla escogido en una nueva pantalla (Figura 5):

Tipo
 Fragmentos

Sensibilidad
☒ Acentos
☒ Mayúsculas

Tipo de intervención
 Saludo / Presentación inicio | Saludo / Presentación respuesta

Filtros

Edad Cualquiera	Desde Cualquiera	Hasta Cualquiera
Papel Cualquiera	Nivel de estudios Cualquiera	Sexo Cualquiera
Hablante Cualquiera		

Resultados 1 a 50 de 124

1 2 3 →

1 Ir a la página

Descargar

1 ► SCOM_H11_047_hab1 y no mola nada ¿sabes? <pausa> mira si era un cantazo tío que me pasaba el portero <pausa> estabas así metiendo rayas <pausa> y al final no te decía nada ¿sabes? pasaba así <pausa> hola hola <pausa> tal <pausa> cortaros <pausa> ¿sabes? <pausa_larga>

2 ► SCOM_E_08 mmm ya bueno <pausa> pues

3 ► SCOM_E_08 hola <pausa_larga>

4 ► SCOM_H11_047_hab1 hola <pausa_larga>

5 ► SCOM_H11_051_hab3 perdón ola <pausa> son Antonio o xefe de estudios

6 ► SCOM_E_08 hola

7 ► SCOM_E_08 sí <pausa> ah hola sí

8 ► SCOM_H12_027_hab1 nos vino a recibir cuando llegamos bueno ¿qué tal? tal <pausa> no sé qué ¿qué tal? y digo ¿de dónde sois? tal y tú eres gallego y yo pss tú también <risa> ¿y de dónde eres? de Santiago y me dice pues <pausa> va a ser que sí ¡joder! y yo ya hablando con él después que que que sí que que que había estudiado allí en la Barcia <pausa_larga>

9 ► SCOM_H12_027_hab1 entonces claro <pausa> nosotros nosotros <pausa> ¿ves? <vacilación> nosotros ya al llegar a los sitios ¿que hai neno? veña pasa tal fun veña para abaixo nosotros ahí siempre estuvimos ahí con con los porteros <pausa> sie siempre entrábamos para donde queríamos por la por la noche <silencio>

10 ► SCOM_E_13 hola Juan <risa> encantada

Figura 5. Listado de casos recuperados de la búsqueda del par adyacente saludo / presentación

4.2. Búsqueda combinada de actos de habla

Para efectuar el segundo tipo de búsqueda, correspondiente a las búsquedas combinadas, en la pestaña *Tipo* se escoge la opción *Elem. gramaticales* y, en el campo *Buscar en* (vid. figura 6), se puede establecer combinaciones entre actos de habla y otras unidades lingüísticas. De nuevo, es necesario seleccionar uno o varios actos de habla, o bien uno o varios constituyentes de dichos actos.

ESLORA | Información | Búsquedas | Listados | Guía | Descargas | Contacto

Búsqueda

Corpus
 Cualquiera

Tipo
 Elem. gramaticales | Palab. ortográficas | Elem. gramaticales | Palab. ortográficas próx. | Elem. gramaticales próx.

Resultado

Tipo de resultado
 Concordancias

Ordenación
 Coincidencia

Tamaño de página
 50

Filtros

Edad Cualquiera	Desde Cualquiera	Hasta Cualquiera
Papel Cualquiera	Nivel de estudios Cualquiera	Sexo Cualquiera
Hablante Cualquiera		

Buscar en
 Cualquiera

Elem. gramaticales

Elemento gramatical	Etiqueta	Lema	Palabra ortográfica
---------------------	----------	------	---------------------

Volver Limpiar Buscar

Figura 6. Búsqueda combinada

Este segundo procedimiento permite formar combinaciones entre actos de habla y unidades simples y complejas con distintas opciones de sensibilidad, al igual que en las búsquedas gramaticales: por ejemplo, un acto de habla junto con una categoría gramatical (ej., actos



directivos + el lema *poder*, con la forma *puedes* o con un verbo en forma imperativa); una combinación de palabras o un sintagma (ej., *es que* + respuestas a ofrecimientos), un prefijo o una terminación, una forma acentuada o no acentuada, entre otras opciones. De este modo, es posible correlacionar el empleo de una determinada unidad o forma lingüística con un acto de habla (o varios), o con algún constituyente del par adyacente, lo que ofrece múltiples posibilidades para la investigación y la enseñanza: el estudio de combinaciones recurrentes entre las distintas opciones de saludo con el lema *bueno*, la parte iniciativa de los ofrecimientos con una forma verbal en distintos tiempos (*quieres*, *te gustaría*, *toma*, etc.), la respuesta a un acto directivo con el adverbio *sí* o con la interjección *vale*, entre otras (vid. Rodríguez Domínguez y Sampedro Mella, 2024, para más ejemplos).

Por último, al igual que ocurre con el procedimiento de búsquedas generales del corpus, y con el mostrado en 4.1., los resultados se presentan en forma de listados (figura 7) que se pueden descargar en un documento en formato Excel o en un fichero CSV. También es posible acceder al audio de cada caso recuperado, clicando en el botón de *play* de la izquierda:

The screenshot displays the ESLORA search interface. At the top, there's a search bar with the text 'Buscar en' and two buttons: 'Saludo / Presentación inicio' and 'Saludo / Presentación respuesta'. Below the search bar, the search term 'qué' is entered in the 'Palab. ortográficas' field. To the right of the search bar are buttons for 'Volver', 'Limpiar', and 'Buscar'. Below the search bar, it says 'Resultados 1 a 30 de 30'. On the right side, there is a 'Descargar' button. The main area shows a list of 15 search results, each with a play button icon, a source identifier (e.g., SCOM_H12_027_hab1), and a snippet of text containing the search term 'qué' with its context and phonetic annotations.

Figura 7. Listado de casos recuperados de la búsqueda del par adyacente saludo / presentación en combinación con la partícula interrogativa *qué*

5. REFLEXIONES FINALES

La pragmática de corpus, como disciplina, ha experimentado un considerable crecimiento en los últimos años, gracias a publicaciones especializadas y a la creación de corpus y herramientas específicas para el etiquetado. Este trabajo, enmarcado en esta disciplina, ha presentado el sistema de anotación implementado en el corpus ESLORA, que permite recuperar de manera automática actos de habla de distintas tipologías, a través de búsquedas generales o combinadas con otros elementos del corpus (lemas, palabras ortográficas, categorías gramaticales, etc.).

Cabe señalar, entre sus limitaciones, que no ha sido posible anotar el subcorpus de conversaciones espontáneas que recoge ESLORA. El etiquetado se llevó a cabo en el marco de un proyecto con una duración limitada y, a diferencia del subcorpus de entrevistas, delimitado



y cerrado, el de conversaciones es abierto y se encuentra en constante actualización. Por tanto, la aplicación de un sistema de etiquetado manual de estas características podría verse interrumpida o quedar desactualizada con la incorporación de nuevas muestras en el corpus. Por todo ello, la anotación se ha centrado en el subcorpus de entrevistas, que ha funcionado como corpus experimental para la validación del sistema propuesto y podría replicarse en el futuro en otros corpus.

Hay que tener también en cuenta que el sistema incorpora un reducido número de actos de habla, debido a las características del género de la entrevista, en la que predominan las preguntas, generalmente orientadas a guiar el discurso de los informantes, así como las citas. Si bien la estructuración en forma de pares adyacentes ha permitido identificar y sistematizar los actos de habla, presenta restricciones en la delimitación de algunos de ellos, como los que contienen secuencias ampliadas. Por otro lado, aunque el etiquetado ha sido validado y revisado independientemente por tres investigadoras diferentes, pueden detectarse fallos. Como en otros proyectos similares en los que intervenga la anotación pragmática, debido al carácter exclusivamente manual del proceso de anotación, puede haber errores que afecten tanto a la recuperación de la información –por problemas técnicos no descubiertos o por la inserción errónea de las etiquetas– como al contenido etiquetado¹⁵.

A pesar de estas limitaciones, el hecho de contar con un sistema de estas características en un corpus oral constituye un avance en la pragmática de corpus, en la lingüística de corpus y en la lingüística computacional, al proporcionar acceso a datos de difícil sistematización. También contribuye a la investigación empírica en el ámbito de la pragmática y del análisis del discurso, ya que no es preciso revisar los corpus completos o lanzar búsquedas basadas en criterios meramente formales para obtener un número significativo de actos de habla. Este recurso puede, asimismo, emplearse con gran aprovechamiento para mejorar el aprendizaje y la enseñanza de las lenguas extranjeras, a través de enfoques como el *data-driven learning*, que favorecen el contacto de los aprendices con datos lingüísticos reales.

Finalmente, como líneas de investigación futuras, sería conveniente solventar las limitaciones descritas mediante la ampliación del tipo de muestras sobre las que se ha aplicado el etiquetado, del número de actos de habla que contiene el corpus y de los niveles de anotación considerados, más allá del par adyacente. También sería importante aprovechar la información obtenida para diseñar sistemas de anotación semiautomática de actos de habla aplicables en otros corpus, por ejemplo mediante herramientas de inteligencia artificial, que faciliten el complejo y costoso proceso de anotación manual.

REFERENCIAS

- Aijmer K. y Rühlemann, C. (2015). *Corpus Pragmatics: A Handbook*. Cambridge University Press.
- Aijmer K. (2026). Corpus Pragmatics. En J. Romero Trillo (Ed.), *The Cambridge Handbook of Language in Context* (pp. 289–313). Cambridge University Press.
- Albelda Marco, M. (2022). Los corpus del español hablado y los estudios pragmáticos. En G. Parodi, L. C. Howe y P. Cantos (Eds.), *Lingüística de corpus en español* (pp. 223–248). Routledge.

¹⁵ Téngase también en cuenta que el proyecto que financiaba la anotación pragmática del corpus no sigue vigente y que el equipo responsable de su elaboración ya no está vinculado a la institución que lo alberga (USC).



- Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., y Weinert, R. (1991). The Hcrc Map Task Corpus. *Language and Speech*, 34(4), 351–366.
- Archer, D., Culpeper, J. y Davies, M. (2008). Pragmatic annotation. En A. Lüdeling y M. Kytö (Eds.), *Corpus Linguistics: An International Handbook* (pp. 613–642). De Gruyter.
- Barcala, M., Domínguez Noya, E., Fernández Sanmartín, A., Rivas Cabanelas, R., Santalla del Río, P., Vázquez Rozas, V. y Villapol Baltar, R. (2018). El corpus ESLORA de español oral: diseño, desarrollo y explotación. *CHIMERA: Romance Corpora and Linguistic Studies*, 5(2), 217–237.
- Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B. y Rauzy, S. (2008). Le CID - Corpus of Interactional Data. Annotation et exploitation multimodale de parole conversationnelle. *Traitement Automatique des Langues*, 49(3), 105–134.
- Bianco, A., Brocca, N. y Garassino, D. (2026). Assessing the Potential of LLM-assisted Annotation for Corpus Pragmatics: The Case of Humor. *Corpus Pragmatics*, 10(33), 1–22.
- Brocca, N., Nuzzo, E. y Wang-Kathrein, J. (2026). AI-driven speech act annotation: accuracy and reproducibility across ChatGPT, LadderWeb and LLaMA. *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses*, 3(1), 1–30.
- Bunt, H., Petukhova, V., Gilmartin, E., Pelachaud, C., Fang, A., Keizer, S. y Prevot, L. (2020). The ISO Standard for Dialogue Act Annotation, Second Edition. *Proceedings of the 12th Language Resources and Evaluation Conference*, 2020, 1–11.
- Escandell, V. (1999). Los enunciados interrogativos. Aspectos semánticos y pragmáticos. En I. Bosque y V. Demonte (Dir.), *Gramática descriptiva de la lengua española* (pp. 3931–3986). Espasa.
- Fernández Sanmartín, A. (2018). La entrevista libre como método para evitar la paradoja del observador. *CHIMERA, Romance Corpora and Linguistic Studies*, 5(2), 141–196.
- Gestido de la Torre, E. (2006). *Outro café? Aproximación ás secuencias de ofrecemento en contexto de sobremesa*. *Madrygal Revista de Estudios Gallegos*, 9, 53–62.
- Grupo Val.Es.Co. (2014). Las unidades del discurso oral. La propuesta Val.Es.Co. de segmentación de la conversación (coloquial). *Estudios de Lingüística del Español*, 35, 11–71.
- Hidalgo Navarro, A. (2019). *Sistema y uso de la entonación en español hablado: Aproximación interactivo-funcional*. Ediciones Universidad Alberto Hurtado.
- Huang, D., Xu, J., Song, Y., Yu, R. (2025). Assessing the Potential of using Large Language Models for Pragmatic Annotation of Historical Texts: A Case of Epistemic Stance in Early Modern English. *Journal of Historical Pragmatics*, 26(3), 406–437.
- Instituto Cervantes (2006). *Plan curricular del Instituto Cervantes. Niveles de referencia*. Biblioteca Nueva.
- Jucker, A. (2018). Corpus pragmatics. En A. Jucker, P. K. Schneider y W. Bublitz (Eds.), *Methods in Pragmatics* (pp. 455–466). De Gruyter.
- Jurafsky, D., Shriberg, L. y Biasca, D. (1997). *Switchboard SWBD-DAMSL Shallow Discourse-Function Annotation: Coders Manual*. Institute of Cognitive Science Technical Report. University of Colorado.
- Kirk, J. M. (2013). Beyond the Structural Levels of Language: An Introduction to the SPICE-Ireland Corpus and its Uses'. En J. Cruickshank y R. McColl Millar (Eds.), *After the Storm: Papers from the Forum for Research on the Languages of Scotland and Ulster triennial meeting*, 207–232.
- Leech, G. (1997). Teaching e Language Corpora: A Convergence. En A. Wichmann et al., *Teaching and Language Corpora* (pp. 1–23). Longman.
- Kohnen, T. (2000). Corpora and speech acts: The study of performatives. En C. Mair y M. Hundt (Eds.), *Corpus Linguistics and Linguistic Theory. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20)*, 177–186.
- Landert D., Dayter, D., Messerli, T. C. y Locher, M. A. (2023). *Corpus Pragmatics*. Cambridge University Press.



- Leech, G. y Weisser, M. (2003). Generic speech act annotation for task-oriented dialogue. En D. Archer, P. Rayson, A. Wilson y T. McEnery (Eds.), *Proceedings of the Corpus Linguistics 2003 Conference*. Lancaster: UCREL Technical Papers, 441–446.
- Leech, G. y Weisser, M. (2013). *The SPAADIA Annotation Scheme*. https://martinweisser.org/publications/SPAADIA_Annotation_Scheme.pdf
- López, C., Renau, I., Nazar, R., Torner, S. (2014). Computer-assisted revision in Spanish academic texts: Peer-assessment. *Procedia-Social and Behavioral Sciences*, 141, 470–483.
- Mas Álvarez, I. y Sampedro Mella, M. (2023). Aproximación a las interrogativas de eco desde la fonopragmática: un estudio de corpus. En R. M. Gómez Pato y A. Ulla (Coords.), *Oralidade(s)* (pp. 115–128). Servizo de Publicacións e Intercambio Científico; Universidade de Santiago de Compostela.
- Maynard, C. y Leicher, S. (2007). Pragmatic annotation of an academic spoken corpus for pedagogical purposes. En E. Fitzpatrick (Ed.), *Corpus Linguistics beyond the Word: Corpus Research from Phrase to Discourse* (pp. 107–115). Brill.
- Milà-García A. (2018). Pragmatic annotation for a multi-layered analysis of speech acts: a methodological proposal. *Corpus Pragmatics*, 2(3), 265–287.
- O'Donnell, M. (2008). The UAM CorpusTool: Software for corpus annotation and exploration. In C. Bretones Callejas et al. (Eds.), *Applied Linguistics Now: Understanding Language and Mind* (pp. 1433–1447). Universidad de Almería.
- O'Keeffe, A. (2018). Corpus-based function-to-form approaches. En A. H. Jucker, K. P. Schneider y W. Bublitz (Eds.), *Methods in Pragmatics (Handbooks of Pragmatics 10)* (pp. 587–618). De Gruyter.
- Pérez Blanco, M., e Izquierdo, M. (2025). Contrasting pragmatic functions in CLANES: <Recommend> and <Suggest>. En R. Rabadán y N. Ramón (Eds.), *Cross-linguistic mediated communication: Hybrid text production English- Spanish* (pp. 51–84). Peter Lang.
- Rabadán, R., Ramón, N. y Sanjurjo-González, H. (2021). Pragmatic annotation of a domain-restricted English-Spanish comparable corpus. *Bergen Language and Linguistics Studies*, 11(1), 209–223.
- Rau, G. y Shih, Y. (2021). Evaluation of Cohen's Kappa and Other Measures of Inter-Rater Agreement for Genre Analysis and Other Nominal Data. *Journal of English for Academic Purposes*, 53, Article 101026.
- Rodríguez Domínguez, A. y Sampedro Mella, M. (2024). Por que crear material didáctico a partir dun corpus oral? Unha proposta para o ensino das funcións comunicativas. En A. Ulla Lorenzo (Ed.), *Dixitalidade(s) e tecnoloxía(s) nos ámbitos literarios, culturais e lingüísticos* (pp. 167–180). Servizo de Publicacións e Intercambio Científico; Universidade de Santiago de Compostela.
- Rojo, G. (2021). *Introducción a la lingüística de corpus en español*. Routledge.
- Rühlemann, C. (2019). *Corpus Linguistics for Pragmatics: A Guide for Research*. Routledge.
- Ruiz de Zarobe, L. (2001). Estrategias de invitación en español e imagen social de los hablantes: Un estudio empírico. *Pragmalingüística*, 9, 261–278.
- Ruiz Gurillo, L. (2026). Digitalizar, compilar y etiquetar corpus de humor. La experiencia con la plataforma OBSERVAHUMOR. *Variación, Revista de Variación y Cambio Lingüístico*, 3, 13–38. <https://doi.org/10.30827/3020.9854rvcl.3.2026.36424>.
- Sacks, H., Schegloff, E. A. y Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4), 696–735.
- Sampedro Mella, M. (2021a). DCT vs. corpus orales: reflexiones metodológicas sobre el estudio de los actos de habla. *Pragmalingüística*, 29, 337–395.
- Sampedro Mella, M. (2021b). Entre el ofrecimiento y la orden: la reproducción de peticiones en secuencias de estilo directo en entrevistas semidirigidas. *Sociocultural Pragmatics. An International Journal of Spanish Linguistics*, 9(1), 26–50.
- Sampedro Mella, M. y Rodríguez Domínguez, A. (2026). Las funciones comunicativas en ELE: de los libros de texto a los corpus con anotación pragmática. En I. Fernández López (Ed.), *Competencia plurilingüe y pluricultural: enfoques desde la Didáctica de la Lengua y la Literatura* (pp. 385–403). Peter Lang.



- Schegloff, E. A. y Sacks, A. (1973). Opening up Closings. *Semiótica*, 8(4), 289–327.
- Schiffrin, D. (1986). Functions of and in discourse. *Journal of Pragmatics*, 10(1), 41–66.
- Schmid, H. (1995). Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*, 1–9.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.
- Vázquez Rozas, V. y Barcala, M. (2020). Computational tools and spoken corpora design: an ongoing dialogue. *Caplletra*, 69, 221–240.
- Vázquez Rozas, V., Barcala, M., Domínguez Noya, E., Fernández Sanmartín, A., Rojo, G. y Santalla, P. (2020). Codificación y anotación del habla en un contexto bilingüe: el corpus ESLORA de español de Galicia. *Verba*, 80, 191–226.
- Weisser, M. (2003). SPAACy – A semi-automated tool for annotating dialogue acts. *International Journal of Corpus Linguistics*, 8(1), 63–74.
- Weisser, M. (2015). Speech act annotation. In Aijmer, K. y Rühlemann, C. (Eds.), *Corpus Pragmatics: A Handbook* (pp. 84–114). Cambridge University Press.
- Weisser, M. (2018). *How to Do Corpus Pragmatics on Pragmatically Annotated Data*. John Benjamins.

CORPUS DEL ESPAÑOL CITADOS

- CHCS – Fernández Juncal, C. (2005). *Corpus de habla culta de Salamanca*. Instituto Castellano y Leonés de la Lengua.
- CLANES – Controlled LANGUAGE English Spanish. Universidad de León.
- COLA – Jørgensen, A. (no disponible). Corpus Oral de Lenguaje Coloquial.
- C-ORAL-ROM – Emanuela Cresti and Massimo Moneglia. 2005. *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. John Benjamins.
- C-Or-DiAL – Carlota Nicolás Martínez. 2012. *C-Or-DiAL (Corpus Oral Didáctico Anotado Lingüísticamente)*. Liceus.
- COREMAH – Vacas Matos, M. (2017). Corpus Español Multimodal de Actos de Habla. <https://coremah-1a37f.firebaseio.com/>
- CSCP – Blas Arroyo, J. L. (coord.), Navarro Morales, B. y Casañ Núñez, J. C. (cols.) (2009). *Corpus sociolingüístico de Castellón de la Plana y su área metropolitana*. Universitat Jaume I.
- ESLORA – Corpus para el estudio del español oral, ESLORA, v. 2.5. <https://eslora.usc.es/>
- FEMMES.UP. – Corpus FEMMES.UP (en línea). Plataforma OBSERVAHUMOR. <https://www.observahumor.com/es/buscador-femmesup>
- HUMCOR – Repede, D. (en línea). HumCor. Corpus oral multimodal de humor en español. <https://humcor.snlt.es/>
- OBSERVAHUMOR – Ruiz Gurillo, L. (Dir.^a). (en línea). OBSERVAHUMOR [Plataforma web]. Universidad de Alicante. <https://griale.dfelg.ua.es/observahumor/>
- PRESEEA Alcalá de Henares. Moreno Fernández, F., Cestero Mancera, A. M., Molina Marcos, I. y Paredes García, F. (2002). *La lengua hablada en Alcalá de Henares*. Universidad de Alcalá de Henares.
- Val.Es.Co. – v. 3 (2022). Valencia Español Coloquial. <https://www.uv.es/corpusvalesco/corpus.html>





PERFIL ACADÉMICO-PROFESIONAL

María Sampedro Mella es investigadora posdoctoral Marie Skłodowska-Curie (MSCA) en el Institut Langage et Communication de la UCLouvain (Bélgica), donde lleva a cabo el proyecto DE-PERSONALisation: *Personal and non-personal forms of address in three contiguous Romance languages: sociolinguistic approach and gender implications* (Ref. 101149488). Ha trabajado como profesora e investigadora en las universidades de Salamanca, Lisboa, Sorbonne Nouvelle y Santiago de Compostela; en esta última fue IP del proyecto *Explotación de corpus generales y de aprendices para la investigación y la enseñanza del discurso en el español como lengua extranjera (ELE)*. Sus intereses de investigación se centran en la pragmática, la lingüística de corpus, la lingüística aplicada y la enseñanza-aprendizaje de ELE.