

ConRam: metodología de un corpus de conversaciones con rastreo de mirada

ConRam: methodology for a corpus of conversations with gaze tracking

Asela Reig Alamillo

Centro Interdisciplinario de Investigación en Humanidades, Universidad Autónoma del Estado de Morelos
Cuernavaca, México
assela.reig@uaem.mx
ORCID: <https://orcid.org/0000-0001-6010-8177>

Lisandra Barrios Arias

Centro de Investigación en Ciencias Cognitivas, Universidad Autónoma del Estado de Morelos
Cuernavaca, México
lisandra.barrios@uaem.edu.mx
ORCID: <https://orcid.org/0000-0003-4071-2953>

Resumen

Este artículo presenta el Corpus de conversaciones con rastreo de mirada (ConRam) y la metodología empleada para su creación. El Corpus de conversaciones con rastreo de mirada es el primer corpus de conversaciones en español recogidas mediante tecnología de rastreo visual. Se trata de un corpus multimodal de conversaciones diádicas entre hablantes con un alto grado de cercanía y una relación simétrica, recogidas en contexto de laboratorio. El artículo justifica la pertinencia de obtener un corpus multimodal de estas características y, específicamente, el interés por obtener datos de la mirada de los hablantes durante la conversación para el estudio de la lengua en la interacción.

Palabras clave: corpus; multimodalidad; rastreo visual; interacción cara a cara; mirada

Abstract

This article presents the Corpus de conversaciones con rastreo de mirada (Corpus of conversations with eye tracking), ConRam, and the methodology used to create it. The Corpus of Conversations with Eye Tracking is the first corpus of conversations in

Recepción: 24-09-2025 | Aceptado: 4-12-2025
Publicado: 24-12-2025

Acceso abierto

Esta obra está bajo licencia Creative Commons Atribución-NoComercial 4.0 Internacional (CC BY-NC 4.0) <https://creativecommons.org/licenses/by-nc/4.0/deed.es>

Citación:

Reig, Asela, et al. "ConRam: metodología de un corpus de conversaciones con rastreo de mirada". *Estudios del Discurso* 11.2 (2025): 1-28.

DOI: <https://doi.org/10.30973/esdi.2025.11.2.224>

Spanish collected using visual tracking technology. It is a multimodal corpus of dyadic conversations between speakers with a high degree of closeness and a symmetrical relationship, collected in a laboratory setting. The article discusses the relevance of obtaining a multimodal corpus of these characteristics and, specifically, the interest in obtaining data on the gaze of speakers during conversation for the study of language in interaction.

Keywords: corpus; multimodality; eye tracking; face-to-face interaction; gaze

1. Introducción

Este artículo presenta el Corpus de conversaciones con rastreo de mirada (ConRam), un corpus conversacional de diálogos que se registran en audio, video y rastreo visual para capturar los patrones de mirada de los hablantes durante la interacción lingüística.

La interacción lingüística es multimodal: en ella se activan simultáneamente modalidades auditivas y visuales y, junto con el canal verbal, la comunicación emplea otros canales semióticos como los gestos paralingüísticos, la postura o la mirada (Brône y Oben, 2018; Brône et al., 2017). El estudio de los gestos paralingüísticos ha recibido una creciente atención desde el trabajo pionero de Kendon (1967) (Keevallik, 2018; Nevile, 2015). En cambio, el comportamiento de la mirada y la forma en que esta interacciona con la expresión lingüística en la conversación son aspectos menos desarrollados y que requieren de datos que, en la actualidad, no están disponibles en español y apenas lo están en otras lenguas.

Este artículo presenta la metodología de recogida de datos de ConRam, un corpus de conversaciones de parejas de hablantes en el que se incluye el comportamiento de la mirada de los hablantes recogido con rastreo visual. Aunque el corpus no está todavía completo —la revisión y anotación están en proceso—, consideramos adecuado dar a conocer, desde esta etapa, tanto la existencia del corpus como el método de recogida y tratamiento de los datos para fomentar el diálogo y la colaboración entre grupos de investigación interesados en el estudio de la interacción oral cara a cara.

El resto del artículo se organiza de la siguiente manera: el apartado 2 se dedica a la multimodalidad de la interacción lingüística y, específicamente, al papel de la mirada en la conversación y sus diversas funciones (2.1). El apartado 3 hace una revisión breve

de los corpus existentes para el estudio de la interacción oral en conversación y de aquellos con rastreo visual en otras lenguas, para mostrar la carencia que cubre el presente corpus. En el apartado 4, se detallan las características del Corpus de conversaciones con rastreo de mirada. Por último, el apartado 5 ofrece datos respecto a la naturalidad de las conversaciones recogidas, reflexión pertinente dado el contexto de laboratorio en el que estas tienen lugar. El artículo se cierra con las conclusiones.

2. La multimodalidad de la interacción y el papel de la mirada en la conversación

La comunicación cara a cara es multimodal (Brône et al., 2017; Kendon, 1967; Kibrik y Fedorova, 2018; Levinson, 2006, 2010; Levinson y Holler, 2014; Poyatos, 1994). Aunque el principal canal de comunicación humana es verbal, esto es, involucra el diseño del mensaje lingüístico en todos sus niveles, otros canales no verbales participan de ella: empleamos expresiones faciales, gestos, postura, mirada, y también por estos medios nos comunicamos.

Desde hace algunos años, lingüistas, analistas del discurso y de la conversación, y estudiosos de la interacción lingüística desde otras disciplinas han avanzado en describir cómo los diferentes canales funcionan y actúan en la interacción cara a cara (Gironzetti, 2020). En especial, el diálogo entre la lingüística y los estudios de otras modalidades involucradas en la comunicación cara a cara es relevante para comprender fenómenos lingüísticos que se relacionan con el funcionamiento de la interacción: pragmáticos, sociolingüísticos e interaccionales. Por ejemplo, la expresión lingüística, las expresiones faciales y gestuales, la gestión de la mirada de los participantes y la fluidez verbal participan en la gestión de turnos, de información nueva y conocida, en la organización de secuencias, el establecimiento y la actualización del contexto discursivo, el monitoreo de la comprensión y del acuerdo y desacuerdo, la lectura de intenciones, la obtención de inferencias, la gestión de la atención del interlocutor y la atención conjunta o en aspectos sociales de la interacción, entre otros (Levinson, 2006, 2010). Pensemos en la obtención de inferencias, la cual constituye una operación cognitiva básica en la comunicación lingüística y en los estudios pragmáticos: esta operación también puede estar afectada por señales no verbales. Si, por ejemplo, un hablante dice “Todo está bien”, pero su expresión facial y dirección de la mirada no están

alineados con el contenido emitido lingüísticamente, el receptor puede inferir que, para el emisor, todo no está realmente bien.

La investigación sobre la gestualidad paralingüística ha recibido una cierta atención en las últimas décadas (Goldin Meadow y Alibali, 2013; Keevallik, 2018; Nevile, 2015) y aún es un campo de estudio fructífero. El corpus que aquí presentamos permitirá abordar estudios desde este enfoque, aunque, por el momento, la anotación de gestos es una labor pendiente. Nos centraremos, aquí, en justificar la pertinencia de un corpus que, además de tener un video que capture la gestualidad, contenga información sobre la mirada de los participantes, por ser este un aspecto menos desarrollado en el estudio de la investigación en la comunicación lingüística y la mayor novedad que aporta este corpus.

2.1. Patrones y funciones de la mirada en la conversación

Aunque ya hace décadas que se señala el papel relevante de la mirada en la comunicación lingüística cara a cara, las preguntas sobre cómo funciona en la comunicación y cómo interactúa con la lengua en la conversación son, todavía hoy, preguntas abiertas.

Kendon (1967) ya describe el patrón general de la mirada en una conversación diádica: mientras el oyente mantiene mayormente contacto visual con el hablante, este último mira menos al interlocutor y alterna esa conducta con el desvío de la mirada. Este hallazgo ha sido respaldado por varios estudios posteriores (Brône y Oben, 2018; Goodwin, 1980; Hirvenkari et al., 2013; Vertegaal et al., 2001) que apuntan también a que esos patrones pueden depender del acto de habla realizado, entre otros factores (Rossano et al. 2009). Se piensa que el mayor desvío de la mirada por parte del hablante está relacionado con la gestión de la carga cognitiva. Algunos autores defienden que mantener contacto visual prolongado con el interlocutor es altamente costoso para el hablante y, por tanto, disminuye la fluidez del discurso (Beattie, 1981; Kendrick y Holler, 2017; Novick et al., 1996). De esta manera, los hablantes desvían la mirada para reducir la carga cognitiva que involucraría atender a estímulos ambientales, como el rostro del interlocutor (Glenberg et al., 1998), que no son necesarios para la tarea específica que se esté realizando (planeación, selección léxica, etcétera).

La mirada participa en la regulación de la toma de turnos en la conversación, un sistema de organización básico en las interacciones sociales humanas. La mirada permite a los hablantes señalar qué hablante es seleccionado para el siguiente turno y

confirmar si está listo para intervenir; a su vez, el desvío de mirada —dejar de mirar al interlocutor— tiende a ocurrir cuando se está a punto de tomar el turno (Auer, 2018; Brône et al., 2017; Lerner, 2003).

Otro conjunto de actividades esenciales en la conversación en el que participa la mirada es el monitoreo de estados atencionales y emocionales del interlocutor. Las funciones de monitoreo del interlocutor, constantes en la conversación, pueden satisfacerse verbalmente: por ejemplo, lanzamos a nuestro oyente, a lo largo de la conversación, preguntas cortas como *¿sí?*, esperando una confirmación de su atención y comprensión. Sin embargo, la mirada parece ser especialmente útil para observar las reacciones del interlocutor y ajustar el discurso convenientemente (Degutte y Astell, 2021). Así, la mirada al otro facilita la emisión y monitoreo de retroalimentadores, señales verbales o no verbales que se utilizan para mostrar que se está siguiendo el discurso, así como señales de las reacciones emocionales del interlocutor que pueden provocar un ajuste en los contenidos o la forma de expresarlos (Bertrand et al., 2007).

La mirada contribuye, asimismo, a la lectura de intenciones. Para lograr una comunicación exitosa, los participantes monitorean el comportamiento del interlocutor, predicen sus intenciones y ajustan su propio comportamiento en consecuencia (Huang et al., 2015). Por ejemplo, si durante el diálogo una persona mira su reloj, el interlocutor puede inferir que la persona está preocupada por la hora, con base en experiencias previas con esa acción (Romani et al., 2005; Sebanz y Shiffrar, 2009; Shiffrar y Freyd, 1990). La mirada deíctica o referencial está sincronizada, temporalmente, con expresiones verbales que refieren al mismo objeto (Griffin, 2001; Meyer et al., 1998). Es decir, cuando un interlocutor se refiere a un objeto presente, es común que la mirada se pose en el objeto un poco antes de que se produzca la mención verbal. Esta mirada referencial permite al interlocutor recuperar las intenciones del hablante (Staudte y Crocker, 2011). De este modo, la dirección de la mirada se considera un recurso esencial, pues permite deducir el foco de atención del interlocutor y las acciones que se deriven este.

El desvío de la mirada desempeña, además, un papel clave en la regulación emocional durante la interacción. Exline y colaboradores (1965) observaron que los participantes evitan la mirada directa con su interlocutor al recibir preguntas personales sobre sus deseos o miedos. Kendon (1967) observa este patrón conductual cuando la intensidad emocional llega a cierto umbral y reporta que el desvío de la mirada funciona como un reductor de la activación emocional tanto en el oyente como en el hablante. Este autor denominó dicho comportamiento la función expresiva de la dirección de la mirada,

la cual se interpreta como estrategia para mitigar emociones socialmente negativas como la vergüenza (Stanley y Martin, 1968). De hecho, estos autores también señalan que el contacto visual prolongado está relacionado con un aumento de la ansiedad, lo que convierte el desvío de la mirada en una herramienta eficaz para reducirla.

Por otro lado, a la función expresiva también corresponde el empleo de la mirada para establecer afiliación social. Este fenómeno se observa en la mirada mutua —los hablantes se observan uno al otro— y en la atención conjunta —los hablantes fijan su atención visual sobre un mismo elemento— que tiende a incrementar durante interacciones cooperativas (Kendon, 1967; Kendrick y Holler, 2017). Kendrick y Holler (2017) integran esta idea en la estructura del par adyacente pregunta-respuesta y sugieren que la mirada al interlocutor sirve como recurso expresivo que acompaña las respuestas preferidas (cooperativas/afiliativas) y el desvío de mirada se elige más en las no preferidas (potencialmente no cooperativas/no afiliativas).

En definitiva, resulta evidente que el comportamiento visual participa de diversas funciones en la conversación cara a cara, pero todavía quedan muchas preguntas abiertas respecto a qué funciones específicas realiza y, más concretamente, respecto a cómo interactúa con el comportamiento lingüístico en la interacción.

3. Corpus orales multimodales y con rastreo visual

El estudio de la lengua y la interacción lingüística se apoya, en buena medida, en materiales recogidos en corpus. Un corpus es una compilación de muestras de habla o de escritura recogidas en su contexto natural de enunciación (Rojo, 2016). En concreto, el estudio del español oral y de la interacción en español se beneficia, hasta el momento, de corpus de interacciones orales que, desde hace algunas décadas, se han recogido en las diversas zonas del mundo hispanohablante (Briz y Albelda, 2009; Carcelén, 2024; Rojo, 2016; Briz y Samper, 2022; Solís, 2018).

Por el tipo de interacción que se recoge, diferenciamos aquellos de conversaciones espontáneas o semiespontáneas, desarrollados principalmente desde el análisis de la conversación (Briz y Val.Es.Co, 2002; Albelda y Estellés, 2020; Vila, 2001; Vila y Etxebarria, 2005), de aquellos que se componen de entrevistas semidirigidas, principalmente con un enfoque sociolingüístico (Martín Butragueño y Lastra, 2011-2015; PRESEEA, 2014, Rodríguez et al., 2010). Los últimos tienen mayor énfasis en el control del equilibrio

de la muestra respecto a los parámetros sociodemográficos que se seleccionan (género, edad, nivel de instrucción, etcétera), mientras que los primeros suelen incluir información sociodemográfica de los hablantes, pero no necesariamente equilibran la muestra, pues se prioriza la grabación en contextos espontáneos o semiespontáneos.

Los corpus orales mencionados hasta ahora ofrecen al investigador la transcripción del material oral y suelen tener el audio accesible: son corpus unimodales o monomodales (Adolphs y Carter, 2013), enormemente valiosos y empleados por investigadores de diversas subdisciplinas lingüísticas. Frente a los corpus unimodales, uno multimodal se define como “una colección anotada de contenido coordinada de canales de comunicación que incluyen el habla, la mirada, gestos manuales y lenguaje corporal, y que se basa generalmente en comportamiento humano grabado” (Foster y Oberlander 307-308). Hasta la fecha, son escasos en cualquier lengua los corpus que ofrecen, además de grabación de audio, imágenes de video del intercambio lingüístico. En lenguas diferentes al español, podemos mencionar el *Corpus of Interactional Data for French* (Bertrand et al., 2007; Blache et al., 2008); *NMMC Nottingham Multimodal Corpus* (Knight et al., 2008, 2009), el *Corpus D64* (Oertel et al., 2013), el *Swedish Spontaneous Dialogue Corpus* (Edlund et al., 2010) o el *IFA Dialogue Video Corpus* (van Son et al., 2008). En el caso del español, la recogida de materiales lingüísticos con video es, hasta donde sabemos, muy limitada: tenemos noticia del corpus GestINF (Rodríguez Rosique, 2021) y algunos investigadores reportan en sus publicaciones análisis de datos conversacionales con video (Vázquez Carranza, 2017).

Dentro del reducido grupo de corpus con información de video, son aún más escasos aquellos recogidos con tecnología de rastreo visual, la cual permite tener precisión respecto a los patrones de mirada de los participantes durante la interacción (Brône y Oben, 2018). En este reducido grupo se encuentran, hasta donde tenemos noticia, *The InSight Interaction Corpus*, en holandés (Brône y Oben, 2015), el corpus de lengua inglesa recogido en el Max Planck Institute for Psycholinguistics in Nijmegen (Holler y Kendrick, 2015) y el *Russian Pear Chats and Stories* (Kibrik y Fedorova, 2018). El *InSight Interaction Corpus* consiste en quince conversaciones espontáneas de dos hablantes nativos del idioma flamenco. Las conversaciones se grabaron en contexto de laboratorio y con los participantes —parejas que se conocían previamente— sentados uno frente a otro. Las conversaciones tienen una duración aproximada de veinte minutos cada una y se graban con grabadora externa y con dos dispositivos de lentes de rastreo visual, uno en cada participante. Las grabaciones fueron anotadas en ELAN, incluyendo

anotaciones lingüísticas, gestuales y de comportamiento de la mirada. El corpus de Holler y Kendrick (2015) consiste en grabaciones de conversaciones espontáneas de tres personas, en inglés, con una duración aproximada de veinte minutos. Las conversaciones tienen también lugar en contexto de laboratorio, donde los participantes están sentados en una disposición de triángulo, y cada conversación se graba con tres cámaras externas y tres lentes de rastreo visual, uno por cada participante. Las grabaciones fueron anotadas para material verbal y no verbal (mirada y gestos) en el programa ELAN. Por último, el *Russian Pear Chats and Stories* recoge 24 conversaciones entre tres personas hablantes nativas de ruso, en contexto de laboratorio. La duración promedio de las conversaciones es de veinticuatro minutos. En este caso, no se trata de conversaciones espontáneas, sino que los participantes relatan una historia que han visto, previamente, en una breve película. Las conversaciones fueron grabadas con tres videocámaras externas y dos de los participantes llevaban lentes de rastreo visual. Las grabaciones se anotaron por rasgos verbales, gestos manuales y mirada en el programa ELAN.

El Corpus de conversaciones con rastreo de mirada de México se une a estos últimos corpus. Los objetivos, justificación y método de recogida y anotación se detallan en los siguientes apartados.

4. El Corpus de conversaciones con rastreo de mirada (ConRam)

El ConRam consiste en 15 grabaciones de conversaciones diádicas, de aproximadamente 30 minutos de duración, en las que los participantes llevan lentes de rastreo visual mientras mantienen la conversación sentados uno frente al otro. La conversación se desarrolla en un ambiente controlado de laboratorio. El resultado, como se detallará, es un corpus de conversaciones en español de México, con rasgos de espontaneidad, aunque en un ambiente controlado. Este combina el audio y su transcripción, la captura visual de los participantes interactuando y de su espacio circundante, así como la captura del campo de visión de cada uno de ellos con el comportamiento de su mirada durante la conversación.

4.1 Objetivo, justificación y posibles aplicaciones

El objetivo del ConRam es compilar material novedoso y original de conversaciones cara a cara en el que se registre tanto el audio del habla como el video de los

participantes, su contexto situacional, campo de visión y el comportamiento de la mirada de cada uno de ellos. Hasta donde sabemos, el ConRam constituirá el primer corpus multimodal con información de rastreo visual en español y complementará los corpus conversacionales existentes hasta el momento.

El empleo de rastreo visual durante las conversaciones permite capturar con precisión el comportamiento de la mirada de los dos hablantes a lo largo de la conversación. Específicamente, el corpus ConRam permitirá, por un lado, ahondar en el estudio de la relación entre aspectos lingüísticos y no lingüísticos involucrados en la comunicación lingüística y, por otro, al obtener datos sincronizados de los dos participantes en el diálogo, servirá para abordar preguntas específicas respecto a la relación estrecha entre el procesamiento y la producción durante los intercambios cara a cara, y el ajuste o *alineamiento* de los participantes en un intercambio lingüístico (Garrod y Pickering, 2009). Como señalan Brône y Oben (2018), el interés multidisciplinario por atender la mirada en la conversación tiene que ver con su papel esencial en la realización de la acción conjunta y coordinada, base de la interacción en general y, en concreto, de la interacción lingüística. Así, este corpus ofrecerá datos potencialmente interesantes para estudios multi o interdisciplinarios sobre la interacción lingüística.

Desde el punto de vista del lingüista y el estudio de la comunicación verbal, resulta inequívoco que muchas de las descripciones y análisis de fenómenos lingüísticos y, específicamente, pragmáticos, o bien, asumen, o bien, hacen predicciones acerca de aspectos no lingüísticos de la interacción: la accesibilidad de elementos en el contexto, la atención del hablante o el oyente, los estados emocionales, la coocurrencia de expresiones determinadas con comunicación gestual paralingüística, entre otros. Si bien la lista no es exhaustiva, se enumeran aquí algunos de los fenómenos para los que pueden resultar relevantes los datos del corpus con rastreo de mirada.

1. Manejo del turno de palabra. Señalización de los lugares adecuados para la transición y el papel conjunto de las expresiones lingüísticas descritas como marcación de final de turno junto con la mirada de los participantes.
2. Patrones de mirada de los participantes en los diversos tipos de actos de habla, así como en la distinción entre turnos iniciativos y turnos reactivos.
3. Patrones de mirada complementarios y suplementarios a expresiones lingüísticas de inicio de turno que señalan acuerdo y desacuerdo o respuestas preferidas y no preferidas.

4. Pausas llenas y patrones de visión asociados con la alta carga cognitiva involucrada en planear el mensaje lingüístico.
5. Elementos lingüísticos señaladores de prominencia informativa y su relación con la atención visual de los participantes.
6. Preguntas breves con función de mantenimiento de turno y de búsqueda de acuerdo, y patrones de visión asociados con sus diversas funciones.
7. Retroalimentación y patrones de visión de los participantes durante la retroalimentación lingüística y paralingüística.
8. Accesibilidad contextual de los referentes y las expresiones referenciales escogidas en virtud del grado de accesibilidad.
9. Reformulación lingüística y el papel de la comunicación multimodal del interlocutor como disparador de la reformulación.
10. Patrones de mirada y convergencia de mirada como indicadores del alineamiento entre hablantes en fragmentos discursivos de mayor relevancia o en la identificación de información nueva y mantenimiento del *suelo de la interacción*.
11. Posicionamiento de los participantes y patrones de mirada.

4.2. Participantes

Han formado parte del corpus 30 participantes (12 hombres, 17 mujeres, 1 persona no binaria), agrupados en 15 parejas. Todos ellos tienen el español mexicano como lengua materna. La edad promedio de los participantes es de 25.7 años. Todos viven actualmente en Morelos (México). De los 30, 20 son nacidos en Morelos; 3 proceden de Guerrero; 2, de la Ciudad de México; 2, de Puebla; 1, de Guanajuato; 1, de Oaxaca, y 1 nació en Estados Unidos, pero vive en Morelos desde la infancia.

Las parejas de participantes tienen siempre una relación de cercanía (amistad, parentesco o relación sentimental). Este requisito se estableció con la idea de facilitar una conversación espontánea a pesar del contexto poco natural en el que se recogen los datos: en el laboratorio y con los lentes de rastreo visual puestos (Brône y Oben, 2015). La tabla 1 recoge los datos de los participantes y el tipo de relación de cada pareja.

Código	Edad	Género	Ocupación	Tipo de relación
01A	28	M	Estudiante/profesora	Amigos
01B	31	H	Estudiante	
02A	32	H	Empleado	Amigos
02B	27	M	Estudiante	
03A	20	M	Estudiante/trabajadora en la universidad	Amigos
03B	41	M	Hogar	
04A	22	H	Escritor	Amigos
04B	22	M	Estudiante	
05A	24	No binario	Estudiante	Hermanos
05B	21	M	Estudiante	
06A	25	H	Trabajador	Amigos
06B	21	H	Auxiliar de cocina	
07A	22	M	Estudiante	Amigos
07B	21	M	Estudiante	
08A	22	M	Estudiante	Amigos
08B	21	M	Estudiante	
09A	23	M	Estudiante	Amigos
09B	22	M	Estudiante	
10A	22	H	Estudiante	Pareja sentimental
10B	28	H	Estudiante	
11A	28	M	Estudiante	Pareja sentimental
11B	32	H	Estudiante	
12A	29	M	Estudiante	Amigas
12B	25	M	Veterinaria, zootecnista	
13A	47	H	Periodista	Amigos
13B	21	M	Estudiante	
14A	30	No binario	Corrector de estilo	Amigos
14B	25	H	Docente	
15A	19	M	Estudiante	Amigos
15B	19	M	Estudiante	

Tabla 1: Datos de los participantes.

4.3. Consideraciones éticas

El proyecto de investigación que incluye la recogida de este corpus fue aprobado por el Comité de ética en Investigación del Centro de Investigación Transdisciplinar en Psicología de la Universidad Autónoma del Estado de Morelos. De acuerdo con lo establecido en el proyecto, los participantes fueron informados con detalle de las condiciones de su participación en el proyecto, así como de la posibilidad de abandonar la grabación o solicitar que la conversación o parte de ella fuera eliminada en cualquier momento. Ninguno de los participantes hizo esta solicitud. Los participantes dieron consentimiento informado tanto para participar en la grabación como para autorizar el empleo de sus grabaciones y/o de los videos en productos y eventos académicos. Asimismo, autorizaron que la grabación se incorpore al ConRam y que pueda ser compartida con la comunidad académica. Por último, los archivos que componen el corpus no contienen ninguna información personal de los participantes: desde el momento en que aceptan su participación, se asigna a cada participante un código y este es empleado para identificar todos los archivos.

4.4. Recogida de datos

La invitación a los participantes se realizó por medio de un cartel divulgativo que se difundió con la técnica de amigo del amigo y en redes sociales del centro de investigación. La recogida de datos tuvo lugar en el Laboratorio de Lenguaje y Cognición de la Universidad Autónoma del Estado de Morelos (México). La sala de grabación contaba con dos sillas, ubicadas una frente a otra, una grabadora de voz y una videogramadora externa. Se trata de un espacio aislado al ruido exterior y cerrado.

Los participantes que voluntariamente responden al anuncio entran el laboratorio, reciben la información del proyecto y, en caso de estar conformes, firman su consentimiento informado. Desde ese momento, se les colocan los lentes de rastreo visual (Tobii Pro Glasses 3) para que se adapten a estos y comuniquen cualquier incomodidad, si fuera el caso.

Durante la presentación del proyecto, se explica a los participantes que el objetivo de la sesión es que hablen de la forma más natural y espontánea que sea posible y que el tema puede ser libre o, si lo prefieren, pueden usar alguno de los temas que se les ofrecen como detonadores para empezar la conversación, sin tener que mantener ese tema a lo largo de la misma. Los temas entre los que podían elegir, en caso de así solicitarlo, son los siguientes:

1. En México no hay diferencias entre géneros en cuanto a la seguridad.
2. Las redes sociales tienen más desventajas que ventajas.
3. La mayoría de los géneros musicales populares en México promueven valores negativos.

De las 15 parejas, 5 escogieron uno de los tres temas propuesto para comenzar la conversación y 10 hablaron de temas de su elección de forma espontánea.

Una vez firmado el consentimiento y con los lentes colocados, cada uno de los participantes completó, en una computadora, un cuestionario de información socio-demográfica. Posteriormente, si no reportaron incomodidad con los lentes, entraron al cuarto de grabación. Los dos participantes se sentaron en sillas colocadas una frente a la otra. Se les recordó que podían moverse con libertad en la silla, de forma espontánea, durante su conversación. El investigador realizó la calibración de los lentes con el software Tobii Pro Controller y se comenzó la grabación tanto en los dos lentes como en la cámara externa y en la grabadora de audio. El investigador abandonó el cuarto y los participantes llevaron a cabo la conversación. Pasados aproximadamente 30 minutos, se tocó la puerta para indicar que la grabación ha finalizado y los voluntarios pasaron a participar en una prueba dirigida para un proyecto complementario.

Mientras la conversación se lleva a cabo, los investigadores monitorean fuera de la habitación que no se interrumpa la grabación de los lentes en el software Tobii Pro Controller. En una conversación se desconectó uno de los lentes y hubo que interrumpir la conversación y reiniciar el lente. Esta grabación duró 15 minutos antes de la interrupción y se grabaron otros 30 minutos después de la misma.

Desde el momento en que se completó el consentimiento informado, se asignó a cada participante un código. Este se empleó en el cuestionario para preservar los datos personales de los participantes, así como en el almacenamiento de todos los archivos relacionados con la sesión (grabaciones, archivos del rastreador visual, videos, audios, archivos de ELAN). Ninguno de los archivos se etiquetó con el nombre o apellido de los participantes.

4.5. Instrumentos de grabación y programas empleados

Para la grabación de los intercambios lingüísticos se utilizaron los siguientes instrumentos de grabación:

1. Dos lentes de rastreo visual Tobii pro glasses-3 (50Hz): cuentan con una cámara de escena (1080p; 1920x1080, a 30 fotogramas por segundo) y un micrófono.

Con ellos, los participantes pueden cambiar de posición libremente durante sus intercambios.

2. Cámara de videogramación externa: se utilizó en la grabación de la escena completa, es decir, del espacio en que se encontraban los participantes (1080p; 1920x1080, a 30 fotogramas por segundo).
3. Grabadora de audio Tascam Dr-05X: Permite grabar dos pistas de audio de alta resolución, en un formato de onda sin comprimir (.wav) a 24bits/44.1kHz.

Se emplearon los siguientes programas para los distintos procesos involucrados en la creación del corpus.

1. Tobii Pro Controller: se emplea para la calibración de los lentes de rastreo y para la grabación de las conversaciones con los lentes.
2. Tobii ProLab: permite la extracción de videos de los lentes con la información de rastreo visual. Dado que se busca que la consulta del corpus no dependa de programas de pago, el corpus aquí presentado no depende del uso de este programa.
3. Adobe Premiere Pro: se emplea para sincronizar las grabaciones.
4. Praat. Software de análisis fonético (Boersma y Weenink, 2024). Se emplea para realizar la segmentación automática del audio.
5. ELAN: software libre desarrollado en el Max Planck Institute for Psycholinguistics (Nijmegen) (Brugman y Russel, 2004; Lausberg y Sloetjes, 2009). En él se realiza la segmentación y transcripción de la conversación, así como la anotación de la dirección de la mirada.

4.6. Tratamiento de los datos

Una vez recogidos los datos, es necesario sincronizar las grabaciones de las grabadoras y los lentes de los dos participantes. Para ello, se empleó el software de edición Adobe Premier Pro y los archivos se sincronizaron a partir de la frecuencia de onda. Se comprobó, posteriormente, que no hubiera ningún desfase en la sincronización. En todos los casos, la sincronización con este software fue suficiente. Se sincronizaron los archivos y se recortó el comienzo de las grabaciones de forma que todas comenzaran en el momento en que los participantes se quedaron solos en el cuarto de grabación, sincronizadas, para importarlas a ELAN.

El programa ELAN permite observar a la vez las tres grabaciones y escuchar la conversación. La figura 1 muestra una captura de pantalla de la interfaz de ELAN con

uno de los archivos del corpus (cortado a la derecha). Actualmente, en cada archivo hay cinco líneas:

1. dos líneas de transcripción del habla, una para cada participante (H1 transcripción y H2 transcripción);
2. dos líneas de anotación de la mirada, una para cada participante (H1 mirada y H2 mirada);
3. una línea para anotar la coincidencia de mirada.

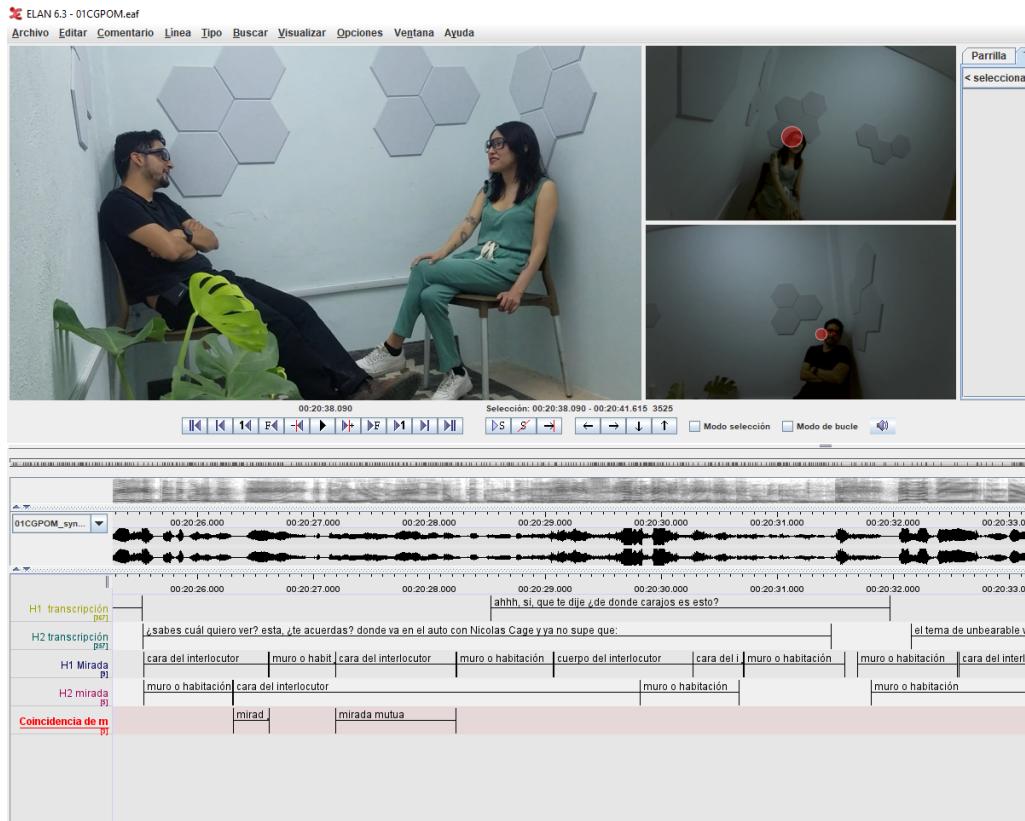


Figura 1. Captura de pantalla (recortada) del software ELAN.

La figura 1 muestra una captura de pantalla, ligeramente recortada para mostrar con mayor claridad la disposición de los tres videos en ELAN: la imagen más grande es el video grabado por la cámara externa. Las dos imágenes de menor tamaño muestran los videos capturados por los lentes de los dos participantes, los cuales registran su campo de visión. En cada uno de ellos, se aprecia un punto rojo que muestra dónde está fijando su mirada el hablante que lleva puestos los lentes. Es decir, en el video, en

la zona superior, se captura el campo visual del hablante vestido de negro y, a lo largo del video, el punto rojo indica dónde fija su mirada. En el momento de la captura, está mirando al rostro de su interlocutora. En la imagen se observan, asimismo, las cinco pistas de transcripción y anotación descritas: H1 transcripción y H2 transcripción incluyen la transcripción ortográfica de cada hablante, y las tres líneas dedicadas a la anotación de la mirada.

Para la transcripción del habla, la segmentación se realizó automáticamente con el programa Praat (umbral de silencio, -35 dB; intervalo mínimo de silencio, 0.25 s). Una vez importada la segmentación a ELAN, se realizó, en este programa, la transcripción de los segmentos de habla de cada uno de los participantes en la pista correspondiente. Esta primera transcripción es ortográfica y no incluye símbolos de transcripción jeffersoniana, que potencialmente podránadirse en versiones posteriores de la transcripción. Si hay fragmentos no comprensibles, se marcan con un doble paréntesis: (()).

Respecto a la anotación de la mirada en las tres pistas restantes, el primer elemento que se busca determinar para anotar el comportamiento de la mirada es qué cuenta como una fijación. Se acepta, de forma estándar, en la bibliografía, que la duración mínima de la fijación es de alrededor de 120 ms (Brône y Oben, 2015; Jacob y Karn, 2003; Vertegaal et al., 2001). Usamos la tasa de fotogramas del video, 30 fotogramas por segundo, para definir la unidad de análisis menor (1/30 o 33.3ms). Cada fotograma del video es de 33.3 ms, por lo que, para anotar una fijación de mirada, el participante debe enfocarse en un punto por cuatro fotogramas o más (133 ms). Las fijaciones que duran menos de cuatro fotogramas no se anotan en ELAN.

Las líneas en las que se anota la mirada de cada participante cuentan con un vocabulario controlado (una serie de valores posibles de los que se escoge uno para cada anotación). Las opciones de anotación son las siguientes: mirada a la cara del interlocutor; mirada al cuerpo del interlocutor; mirada al gesto manual del interlocutor; mirada al gesto manual del propio hablante; mirada al muro o resto de la habitación.

En caso de ser de interés para estudios futuros, el último de estos valores (mirada al muro o resto de la habitación) podría descomponerse en otros valores más específicos. Por ejemplo, Persyn (2023) diferencia, en sus anotaciones, la mirada arriba y abajo, así como derecha e izquierda. No obstante, en este punto de la investigación dicha anotación más detallada no se ha considerado pertinente.

Por último, la línea Coincidencia de mirada tiene dos valores:

1. Mirada mutua: H1 mira a H2 y H2 mira a H1;

2. Mirada conjunta a mismo referente: H1 y H2 miran a la vez el mismo objeto o punto de la habitación, o a las manos de uno de ellos mientras realiza un gesto significativo.

Este último valor sería más interesante en un contexto conversacional donde fuera frecuente que los participantes se refieran a elementos del contexto situacional, algo que en esta conversación es poco frecuente, pero no imposible (en algunos momentos, comentan la decoración de la habitación, se refieren a la grabadora, etcétera).

Los lentes de rastreo visual capturan información muy detallada sobre el comportamiento de la mirada (características de las fijaciones, movimientos o sacadas y datos de pupilometría). Esos datos están disponibles para su análisis con el software de pago de Tobii, pero no se integran en el corpus anotado en ELAN.

5. Conversaciones en laboratorio y naturalidad del habla

Las circunstancias en las que se recogen las conversaciones para el corpus ConRam hacen necesaria una reflexión respecto al tipo de conversación que se obtiene y, en concreto, respecto a su grado de naturalidad o espontaneidad. En este caso, es indudable que hay varios factores situacionales que alejan esta interacción de una conversación en contexto cotidiano, pero, como veremos, también hay varios otros que sí favorecen un alto grado de naturalidad en el diálogo.

Como señala Kendrick (2017), consideramos conveniente pensar en la naturalidad de la conversación grabada no en términos de una dicotomía en la que los datos son naturales o no naturales, sino como un continuo. Kendrick propone que este continuo de naturalidad puede descomponerse en varios elementos: quiénes son los participantes en el estudio, qué hacen los participantes, dónde tiene lugar la interacción y cómo se graba la interacción. En este sentido, el primero de los componentes (*quiénes son los participantes*) propicia, en nuestro corpus, una conversación espontánea: las parejas de participantes mantienen entre sí relaciones de cercanía e igualdad, lo cual se considera un *rasgo coloquializador*, es decir, favorece la aparición del registro coloquial (Gallardo Paúls, 1996). Además, como ya se señaló, los participantes son personas jóvenes. Consideramos, por la conversación mantenida con ellos, que puede existir una diferencia generacional en el efecto que tienen los instrumentos de grabación en los

sujetos: las personas jóvenes, presumiblemente por su intenso contacto con las redes sociales y la tecnología asociada, parecen más acostumbradas y menos inhibidas por el hecho de ser grabadas y saber que esos datos pueden ser compartidos (en este caso, solo con fines académicos), en comparación con personas de otras generaciones.

Respecto a qué hacen los participantes, todas las interacciones son de tipo interaccional y no transaccional (Gallardo Paúls, 1994). Aunque cada conversación es diferente, los participantes tienen libertad para escoger el tema o temas y pasar de uno a otro, lo que también propicia conversaciones naturales. Varias de las parejas continuaron con la conversación que mantenían de camino a la grabación y se recuperó la espontaneidad de su charla coloquial inmediatamente, o bien, se involucraron muy rápido en un diálogo tipo *chisme*, muy basado en la relación cercana y el conocimiento compartido, y que favorece la espontaneidad. Las parejas que escogen temas propuestos en el laboratorio derivan, a lo largo de la conversación, hacia temáticas de su vida o su experiencia privada con narraciones anecdóticas.

Por otra parte, los dos últimos componentes del continuo propuesto por Kendrick (2017), dónde y cómo se graba la conversación, ubican esta interacción en el extremo del continuo correspondiente a menos natural. El diálogo se grabó en un espacio cerrado que no era familiar para los participantes, ubicado dentro de un laboratorio en un centro de investigación. Antes de la conversación, los participantes habían recibido la explicación del proyecto de investigación y firmado el consentimiento informado, lo que aumenta la experiencia de tipo experimental. Además, en el espacio de la grabación, las personas están sentadas en una disposición fija, saben de la presencia de la videogramadora y grabadora de voz que están a la vista y llevan puestos los lentes de rastreo visual que involucran, además de los lentes en la cara, un pequeño aparato de almacenamiento y batería que se engancha en la ropa de cada participante.

Es pertinente, por tanto, admitir abiertamente que el proceso de investigación influye de cierta manera en las interacciones. El estudio detallado de los datos del corpus permitirá analizar y establecer exactamente cómo y a qué aspectos de la conversación afectan estas circunstancias externas. No obstante, una primera observación de los datos para su transcripción y anotación permite ya señalar que, a pesar de lo dicho, se lograron materiales con un alto grado de espontaneidad. Como es frecuente en el caso de interacciones cuyos participantes son conscientes de la grabación, los primeros minutos de la conversación se caracterizan, sobre todo en algunas parejas, por ser poco naturales: los participantes están sentados en posiciones rígidas, hay silencios largos o

relativamente largos, aparecen preguntas explícitas para buscar acuerdo respecto al tema de conversación (*¿de qué hablamos?*) y risas que denotan nerviosismo. El fragmento (1) recoge el comienzo de una conversación en la que esta extrañeza inicial es muy patente. Esta transcripción incluye los símbolos para pausas de menos de 200 ms (/), pausas de entre 200 y 500 ms (//) y de entre 500ms y 999 ms (///). Las pausas de un segundo o más se anotan con el número de segundos entre paréntesis. Además, se emplean los corchetes para marcar traslapes del habla.

(1)

G: aay qué nervios///

O: mmhmm

G: (risas) (1) bueno/ ¿de qué quieres platicar?

O: esa es la peor manera de iniciar una conver[sación]

G: [(risas)]///

O: pues no sé// eehh (6)

G: (risas) ///

O: no sé / no sé//

G: plátícame del nuevo disco de Gorillaz // yo no lo he escuchado/ solo escuché la colaboracióón

O: [¿con quién?]

G: [con Bad] Buny (1) ahhhh // o sea no está mal/ no está mal

O: pues a mí me gusta // ¿qué esperabas? (1) pues [es Bad Bunny]

G: [no sabía] qué esperar/ realmente

O: pues yo esperaba oro y encontré/ orísimo [(risas)]

G: [(risas)]

O: oro/ oro por dos// a mí sí me gustó (1) pero (2.5) pues no sé/ si no la has escuchado
creo queee (1.5) comoo (1) no sé/// no sé/// mejor cuéntame dee (5)

G: (risas)// ¿de qué?

O: aaahhh/// ya vamos a hablar de fútbol (risas)

G: nooo

O: ¿no?/ bueno/ a ver/ estee// ¿has visto algo en mubi? (1)

G: algo en mubi

O: ¿has visto algo en mubi?

G: sí/ claro que sí/ siempre veo cosas en mubi// la última que vii// fuee// ¿cuál fue la última que vi?///

Sin embargo, en pocos minutos, la conversación va adquirió rasgos de espontaneidad: se estableció un tema conversacional que fue cambiando de forma natural, los turnos de habla fueron menos rígidos, aumentó su dinamismo, hubo interrupciones, reformulaciones, bromas, risas reactivas al turno de interlocutor, disminuyeron los silencios prolongados, aparecieron las pausas llenas y los retroalimentadores, aumentó la velocidad del habla, etcétera. Estos patrones de comportamiento propios de la conversación espontánea se prolongaron el resto de la conversación. El ejemplo (2) corresponde a un fragmento bastante posterior, del minuto 20 de la misma conversación, e ilustra esa mayor espontaneidad:

(2)

O: ¿tú viste mandalorian?///

G: (risas) no (risas)

O: está chido, además seguro te gusta Pedro Pascal

G: eeh [eh]

O: [¿no?] ¡pero es chileno!

G: es que: o sea/// el algoritmo lo ha hecho muy bien/// porque [está]

O: [¿algoritmo?]

G: (risas) está en TikTok, por todos lados

O: sí

G: y de alguna forma [terminé]

O: [no es el algoritmo/] es su manager

G: de alguna forma (risas) terminé en el lado de Pedro Pascal

O: ajá

G: entonces/ o sea// me parece/ mmm/ equis [no es como que me super guste]

O: [pero ¿has visto algo de Pedro Pascal?]

G: Eee[eehh]

O: [¿o estás] hablando únicamente de?

G: si/ en realidad he visto los pedacitos/ que me presentan ahí

O: ah es que// bueno/// si [no]

- G: [¿sabes] cuál quiero ver? esta/ ¿te acuerdas de este que te mandé donde va en el auto con Nicolas [Cage? ya no te dije (risas)]
- O: [Ah sí/ sí que te dije ¿de dónde] carajos es esto?
- G: Se llama the unbearable weight of massive talent
- O: la insoportable levedad del no// la insoportable peso [del talento]
- G: [del talento] masivo

A reserva de un análisis del comportamiento de mirada y gestos que se corresponde con este cambio en el grado de naturalidad de la conversación, la postura de los participantes y la cantidad de gestos paralingüísticos también se modifica: en los primeros minutos, la postura de los hablantes es rígida, se sientan rectos en la silla y apenas hacen movimientos con las manos, brazos y cabeza. Esta rigidez se pierde posteriormente, las posturas se relajan y aumenta la gesticulación paralingüística. La figura 2 muestra este contraste entre el minuto 0 de la conversación (A: cinco tomas consecutivas en intervalos de un segundo entre toma y toma) y el minuto 20 (B: cinco tomas consecutivas, con intervalos de un segundo entre cada toma).



Figura 2. Diferencias en la postura de los hablantes en el minuto 0 y en el minuto 20.

Mientras que, en A, primer minuto de la conversación, los participantes muestran prácticamente la misma postura rígida en los cinco segundos consecutivos, las cinco imágenes capturadas en el minuto 20 muestran movimiento de ambos participantes: gestos manuales (la hablante mueve la mano mientras habla) y cambios de postura a lo largo de este breve espacio de tiempo. Esta figura solo ilustra la diferencia en la rigidez de la postura y la mayor aparición de gestualidad y cambios de posición cuando la conversación se hace más natural y espontánea pasados los primeros minutos, pero solo un análisis detallado del corpus ofrecerá datos más precisos sobre estos comportamientos.

Conclusiones

Cada corpus lingüístico se recoge con una finalidad primera e inmediata, aunque, en todos los casos, tiene el potencial de convertirse en objeto de estudio de investigaciones de muy diversa índole. En todo caso, la finalidad con la que fueron recogidos los datos impone tanto las fortalezas y particularidades del corpus como sus inevitables limitaciones. Este es también el caso en el corpus que aquí se presenta.

Este trabajo ha descrito el ConRam, justificando la pertinencia de enriquecer los corpus conversacionales existentes en español con un corpus multimodal que recoja, entre otros aspectos, el comportamiento de la mirada de los participantes. Aunque el corpus está todavía en proceso de revisión y no está, por el momento, disponible en línea, consideramos adecuado darlo a conocer para favorecer el diálogo académico desde esta etapa inicial.

En este artículo, se ha revisado la importancia de la mirada en la comunicación cara a cara y la línea de investigación que reclama que los análisis lingüísticos y de la interacción puedan beneficiarse de observar la multimodalidad inherente a la conversación lingüística y, en concreto, el papel de la mirada en el monitoreo y alineamiento entre los hablantes en una conversación.

El centro del artículo se ha dedicado a detallar la metodología de recogida del corpus: los participantes, procedimientos, instrumentos, características de las conversaciones, entre otros aspectos relevantes.

Las limitaciones del corpus son, al menos, las siguientes. En primer lugar, como ya se comentó, el diseño de recogida de datos impone ciertas condiciones a la conversación que van en contra, al menos en principio, de la obtención de una conversación espontánea y relajada. Hemos argumentado que, a pesar de las restricciones impuestas por la grabación y sus equipos, sí se logra un nivel de naturalidad alto al menos en la mayoría de los datos. Sobre este punto, es pertinente señalar que sería posible y, de hecho, factible obtener un corpus de conversaciones con rastreo visual en contextos menos constreñidos —fuera del laboratorio, por ejemplo, con más personas que puedan incorporarse espontáneamente a la conversación, etcétera—, pero esto quedará pendiente para labores futuras.

En segundo lugar, el corpus consiste, actualmente, en quince conversaciones. No es un corpus muy extenso, aunque es comparable en cantidad de conversaciones a los corpus en inglés, flamenco y ruso con los que dialoga. En todo caso, se espera, en el

futuro, ampliar la muestra de conversaciones tanto en número como en variedad sociodemográfica de los hablantes.

Por último, el corpus en su estado actual solo cuenta con la transcripción del habla y la anotación de la mirada (en proceso). Vemos esta como una limitación temporal: el software ELAN permite ir anotando el corpus en diferentes etapas, según los objetivos de los investigadores. Así, este se presta a ser anotado para cualquier fenómeno lingüístico o gestual, por mencionar dos de las grandes dimensiones de estudio de la multimodalidad en la conversación.

En resumen, el ConRam se crea para contribuir al estudio y comprensión de la conversación lingüística cara a cara, y la interacción entre la lengua y los demás canales de comunicación desplegadas en la interacción lingüística. El corpus, que se concibe como un proyecto de largo alcance que crecerá en sucesivas etapas, pretende ser una aportación relevante y novedosa en la colección de corpus lingüísticos en español, y se recoge con el espíritu de incidir en investigaciones lingüísticas e interdisciplinarias sobre la interacción lingüística. **D**

Referencias

- Adolphs, Svenja y Ronald Carter. *Spoken Corpus Linguistics: From Monomodal to Multimodal*. Routledge, 2013. <https://doi.org/10.4324/9780203526149>
- Albelda, Marta y María Estellés, coordinadores. *Corpus Ameresco*. Universitat de València, 2020. www.corpusameresco.com
- Auer, Peter. "Gaze, addressee selection and turn-taking in three-party interaction". *Eye-tracking in Interaction Studies on the role of eye gaze in dialogue*. De Gruyter, 2018, pp. 197-232. <https://doi.org/10.1075/ais.10.09aue>
- Beattie, Geoffrey W. "A further investigation of the cognitive interference hypothesis of gaze patterns during conversation". *British Journal of Social Psychology*, vol. 20, 1981, pp. 243-248. <https://doi.org/10.1111/j.2044-8309.1981.tb00493.x>
- Bertrand, Roxane, et al. "Backchannels Revisited from a Multimodal Perspective". *Proceedings of Auditory-Visual Speech Processing*, Hilvarenbeek, 2007. <https://hal.science/hal-00244490v1>

- Blache, Philippe, et al. "Creating and Exploiting Multimodal Annotated Corpora". *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC)*. 2008. <http://www.lrec-conf.org/proceedings/lrec2008/>
- Boersma, Paul y Davud Weenink. Praat: making phonetics by computer, 2024. [Programa informático]
- Briz, Antonio y Grupo Val.Es.Co. *Corpus de conversaciones coloquiales*. Arco Libros, 2002.
- Briz, Antonio y Marta Albelda. "Estado actual de los corpus de lengua española hablada y escrita: I+D". *El español en el mundo. Anuario del Instituto Cervantes 2009*. Instituto Cervantes, 2009, pp. 165-226.
- Briz, Antonio y María Samper Hernández. "Estudio de la variación situacional en corpus orales del español". *Lingüística de corpus en español / The Routledge Handbook of Spanish Corpus Linguistics*. Routledge, 2022, pp. 310–324.
- Brône, Geert y Bert Oben. „InSight Interaction: A Multimodal and Multifocal Dialogue Corpus“. *Language Resources and Evaluation*, vol. 49, núm. 1, 2015, pp. 195-214. <https://doi.org/10.1007/s10579-014-9283-2>
- Brône, Geert y Bert Oben. *Eye-Tracking in Interaction*. John Benjamins Publishing Company, 2018. <https://doi.org/10.1075/ais.10>
- Brône, Geert, et al. "Eye gaze and viewpoint in multimodal interaction management". *Cognitive Linguistics*, vol. 28, núm. 3, 2017, pp. 449-483. <https://doi.org/10.1515/cog-2016-0119>
- Brugman, Hennie y Albert Russel. "Annotating Multi-media/Multi-modal Resources with ELAN". *Proceedings of LREC*, 2004, pp. 2065-2068. <http://www.lrec-conf.org/proceedings/lrec2004/pdf/480.pdf>
- Carcelén Guerrero, Andrea. "Novedades en los corpus digitales para el estudio del español oral". *Normas*, vol. 14, núm. 1, 2024, pp. 241-260. <https://doi.org/10.7203/Normas.v14i1.29929>
- Degutyte, Ziedune y Arlene Astell. "The Role of Eye Gaze in Regulating Turn Taking in Conversations: A Systematized Review of Methods and Findings". *Frontiers in Psychology*, vol. 12, 2021, pp. 1-22. <https://doi.org/10.3389/fpsyg.2021.616471>
- Edlund, Jens, et al. "Spontal: A Swedish Spontaneous Dialogue Corpus of Audio, Video and Motion Capture". *Proceedings of LREC*, 2010, pp. 2992-2995.

- Exline, Ralph, et al. "Visual Behavior in a Dyad as Affected by Interview Content and Sex of Respondent". *Journal of Personality and Social Psychology*, vol. 1, 1965, p. 201. <https://doi.org/10.1037/h0021865>
- Foster, Mary Ellen, and Jon Oberlander. "Corpus-based generation of head and eyebrow motion for an embodied conversational agent." *Language Resources and Evaluation* 41.3 (2007): pp. 305-323.
- Gallardo Paúls, Beatriz. "Conversación y conversación cotidiana: sobre una confusión de niveles". *Pragmalingüística*, vol. 2, 1994, pp. 151-194.
- Garrod, Simon y Martin J. Pickering. "Joint Action, Interactive Alignment, and Dialog". *Topics in Cognitive Science*, vol. 1, núm. 2, 2009, pp. 292-304. <https://doi.org/10.1111/j.1756-8765.2009.01020.x>
- Gironzetti, Elisa. "Eye-Tracking Applications for Spanish Pragmatics Research". *The Routledge Handbook of Spanish Pragmatics*. Routledge, 2020, pp. 517-531.
- Glenberg, Arthur M., et al. "Averting the Gaze Disengages the Environment and Facilitates Remembering". *Memory and Cognition*, vol. 26, núm. 4, 1998, pp. 651-658. <https://doi.org/10.3758/BF03211385>
- Goldin-Meadow, Susan y Martha Wagner Alibali. "Gesture's role in speaking, learning, and creating language". *Annual Review of Psychology*, vol. 64, núm. 1, 2013, pp. 257-283. <https://doi.org/10.1146/annurev-psych-113011-143802>
- Goodwin, Charles. "Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning". *Sociological Inquiry*, vol. 50, núm. 3-4, 1980, pp. 272-302. <https://doi.org/10.1111/j.1475-682X.1980.tb00023.x>
- Griffin, Zenzi M. "Gaze Durations during Speech Reflect Word Selection and Phonological Encoding". *Cognition*, vol. 82, núm. 1, 2001, pp. 1-16. [https://doi.org/10.1016/S0010-0277\(01\)00138-X](https://doi.org/10.1016/S0010-0277(01)00138-X)
- Hirvenkari, Lotta, et al. "Influence of Turn-Taking in a Two-Person Conversation on the Gaze of a Viewer". *PLoS ONE*, vol. 8, núm. 8, 2013. <https://doi.org/10.1371/journal.pone.0071569>
- Holler, Judith y Kobin H. Kendrick. "Unaddressed Participants' Gaze in Multi-person Interaction: Optimizing Recipiency". *Frontiers in Psychology*, vol. 6, 2015, pp. 1-14. <https://doi.org/10.3389/fpsyg.2015.00098>
- Huang, C. M., et al. Using gaze patterns to predict task intent in collaboration. *Frontiers in Psychology*, vol. 6, 2015, pp. 1-12. <https://doi.org/10.3389/fpsyg.2015.01049>

- Jacob, Robert JK, and Keith S. Karn. "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises." *The mind's eye*. North-Holland, 2003, pp. 573-605.
- Keevallik, Leelo. "What Does Embodied Interaction Tell Us about Grammar?" *Research on Language and Social Interaction*, vol. 51, núm. 1, 2018, pp. 1-21. <https://doi.org/10.1080/08351813.2018.1413887>
- Kendon, Adam. "Some Functions of Gaze-Direction in Social Interaction". *Acta Psychologica*, vol. 26, 1967, pp. 22-63. [https://doi.org/10.1016/0001-6918\(67\)90005-4](https://doi.org/10.1016/0001-6918(67)90005-4)
- Kendrick, Kobil H. "Using Conversation Analysis in the Lab". *Research on Language and Social Interaction*, vol. 50, núm. 1, 2017, pp. 1-11. <https://doi.org/10.1080/08351813.2017.1267911>
- Kendrick, Kobil H. y Judith Holler. "Gaze Direction Signals Response Preference in Conversation". *Research on Language and Social Interaction*, vol. 50, núm. 1, 2017, pp. 12-32. <https://doi.org/10.1080/08351813.2017.1262120>
- Kibrik, Andrej A. y Olga V. Fedorova. "An Empirical Study of Multichannel Communication: Russian Peer Chats and Stories". *Psychology. Journal of the Higher School of Economics*, vol. 15, núm. 2, 2018, pp. 191-200. DOI: 10.17323/1813-8918-2018-2-191-200
- Knight, Dawn, et al. "The Nottingham Multi-Modal Corpus: A Demonstration". *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC)*, 2008.
- Knight, Dawn, et al. "HeadTalk, HandTalk and the Corpus: Towards a Framework for Multi-modal, Multi-media Corpus Development". *Corpora*, vol. 4, 2009, pp. 1-32. <https://doi.org/10.3366/E1749503209000203>
- Lausberg, Hedda y Han Sloetjes. "Coding Gestural Behavior with the NEUROGES-ELAN System". *Behavior Research Methods*, vol. 41, núm. 3, 2009, pp. 841-849. <https://doi.org/10.3758/BRM.41.3.841>
- Lerner, Gene H. "Selecting next speaker: The context-sensitive operation of a context-free organization". *Language in Society*, vol. 32, núm. 2, 2003, pp. 177-201. <https://doi.org/10.1017/S004740450332202X>
- Levinson, Stephen C. "On the Human 'Interaction Engine'. *Roots of Human Sociality: Culture, Cognition and Interaction*, Routledge, 2006, pp. 39-69. <https://doi.org/10.4324/9781003135517-3>

- Levinson, Stephen C. "Interactional Foundations of Language: The Interaction Engine Hypothesis". *Human Language: From Genes and Brain to Behavior*, The MIT Press, 2010, pp. 189-200.
- Levinson, Stephen C. y Judith Holler. "The Origin of Human Multi-Modal Communication". *Philosophical Transactions of the Royal Society B*, vol. 369, 2014, pp. 1-9. <https://doi.org/10.1038/nature13748>
- Martín Butragueño, Pedro y Yolanda Lastra, coordinadores. *Corpus Sociolingüístico de la Ciudad de México (CSCM)*. El Colegio de México, 2011–2015.
- Meyer, Antje S. et al. "Viewing and Naming Objects: Eye Movements during Noun Phrase Production". *Cognition*, vol. 66, núm. 2, 1998, pp. 25-33. [https://doi.org/10.1016/S0010-0277\(98\)00009-2](https://doi.org/10.1016/S0010-0277(98)00009-2)
- Nevile, Maurice. "The Embodied Turn in Research on Language and Social Interaction". *Research on Language and Social Interaction*, vol. 48, núm. 2, 2015, pp. 121-151. <https://doi.org/10.1080/08351813.2015.1025499>
- Novick, David G. et al. "Coordinating Turn-Taking with Gaze". *International Conference on Spoken Language Processing (ICSLP) Proceedings*, vol. 3, 1996, pp. 1888-1891. <https://doi.org/10.21437/icslp.1996-485>
- Oertel, Catharine, et al. "D64: A Corpus of Richly Recorded Conversational Interaction". *Journal on Multimodal User Interfaces*, vol. 7, núm. 1, 2013, pp. 19-28. <https://doi.org/10.1007/s12193-012-0108-6>
- Persyn, Yne. "Gaze Aversion and Response Preference in Dutch Face-to-Face Conversation: A Corpus Study." 2023. Universidad de Radboud, tesis de maestría.
- Poyatos, Fernando. *La comunicación no verbal*. Ediciones AKAL, 1994.
- PRESEEA. *Corpus del Proyecto para el Estudio Sociolingüístico del Español de España y de América*. Universidad de Alcalá, 2014. <http://preseea.uah.es>
- Rodríguez, Lidia, et al. *Corpus El habla de Monterrey – PRESEEA*. Secretaría de Educación Pública, Monterrey, 2010.
- Rodríguez Rosique, Susana. "Creencia, activación y comunidades culturales: GestInf, o la pertinencia de un corpus basado en el conocimiento compartido". *Oralia*, vol. 24, 2021, pp. 151-186. <https://doi.org/10.25115/oralia.v24i1.6484>
- Rojo, Guillermo. "Los corpus textuales del español". *Enciclopedia lingüística hispánica*, editado por Javier Gutiérrez-Rexach, Routledge, 2016, pp. 285-296.

- Romani, Michela, et al. "Motor Facilitation of the Human Cortico-Spinal System during Observation of Bio-mechanically Impossible Movements". *NeuroImage*, vol. 26, núm. 3, 2005, pp. 755-763. <https://doi.org/10.1016/j.neuroimage.2005.02.027>
- Rossano, Federico, et al. *Gaze, questioning and culture*. Cambridge University Press, 2009.
- Sebanz, Natalie y Maggie Shiffrar. "Detecting Deception in a Bluffing Body: The Role of Expertise". *Psychonomic Bulletin and Review*, vol. 16, núm. 1, 2009, pp. 170-175. <https://doi.org/10.3758/PBR.16.1.170>
- Shiffrar, Maggie y Jennifer J. Freyd. "Apparent Motion of the Human Body". *Psychological Science*, vol. 1, núm. 4, 1990, pp. 257-264. <https://doi.org/10.1111/j.1467-9280.1990.tb00210.x>
- Solís García, Inmaculada. "Corpus dialógicos para el análisis de la conversación". *Chimera*, vol. 5, 2018, pp. 117-129.
- Stanley, Gordon y Donald S. Martin. "Eye-Contact and the Recall of Material Involving Competitive and Noncompetitive Associations". *Psychonomic Science*, vol. 13, núm. 6, 1968, pp. 337-338. <https://doi.org/10.3758/BF03342614>
- Staudte, Maria y Matthew W. Crocker. "Investigating Joint Attention Mechanisms through Spoken Human-Robot Interaction". *Cognition*, vol. 120, núm. 2, 2011, pp. 268-291. <https://doi.org/10.1016/j.cognition.2011.05.005>
- Van Son, R.J.J.H., et al. "The IFADV Corpus: A Free Dialog Video Corpus". *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC)*, 2008.
- Vázquez Carranza, Ariel. "If Vegetables Could Talk...: A Structural and Sequential Analysis of Buying and Selling Interactions in a Mexican Fruit and Vegetable Shop". *Discourse Studies*, vol. 19, núm. 6, 2017, pp. 711-731. <https://doi.org/10.1177/1461445617727185>
- Vertegaal, Roel, et al. "Eye Gaze Patterns in Conversations: There Is More to Conversational Agents than Meets the Eyes". *Proceedings of the Conference on Human Factors in Computing Systems*, 2001, pp. 301-308. <https://doi.org/10.1145/365024.365119>
- Vila Pujol, María Rosa. *Corpus del español conversacional de Barcelona y su área metropolitana*. Grupo GRIESBA, Universitat de Barcelona, 2001.
- Vila Pujol, M. Rosa y Maitena Etxebarria. "Corpus para el estudio de las interferencias lingüísticas: los corpus de Barcelona, Lérida y Bilbao". *Oralia*, vol. 8, 2005, pp. 213-242. <https://doi.org/10.25115/oralia.v8i1.8351>