

¿Texto, audio, emoji o sticker? Elecciones de modo en interacciones por WhatsApp¹

Text, voice messages, emoji, sticker? Mode choice in WhatsApp interactions

Lucía Cantamutto  ¹

Cristina Vela Delfa  ²

¹ CIEDIS - Universidad Nacional de Río Negro/CONICET, Argentina

lcantamutto@unrn.edu.ar

² Universidad de Valladolid, España

cristina.vela@uva.es

ACCESO ABIERTO / OPEN ACCESS

Cita: Cantamutto, Lucía y Vela Delfa, Cristina (2024). ¿Texto, audio, emoji o sticker? Elecciones de modo en interacciones por WhatsApp. *Textos en Proceso*, 10(2), pp. 18-40.

http://doi.org/10.17710/tep.2024.10.2.2cantamutto_vela

Editoras: Esperanza Alcaide Lara (Universidad de Sevilla) y Ana Pano Alamán (Università di Bologna)

Recibido: 12-11-2024

Aceptado: 10-12-2024

Conflicto de intereses: Las autoras han declarado que no poseen conflictos de intereses.

Copyright: © Lucía Cantamutto y Cristina Vela Delfa. Esta obra está bajo licencia internacional [Creative Commons Reconocimiento 4.0](https://creativecommons.org/licenses/by/4.0/)

Resumen

Desde la vertiente sociocultural de la pragmática, este trabajo estudia un aspecto todavía poco desarrollado en el ámbito del Análisis del Discurso Digital: la manera en que los hablantes perciben el cambio de modo en los mensajes intercambiados en las conversaciones por WhatsApp, más específicamente, el paso del texto al audio o a la imagen. Para abordar este objetivo se optó por una aproximación metodológica que combina tres fases. En una primera fase se analizaron los comentarios que circulan en los memes de internet sobre el cambio de modo. En una segunda fase, se identificaron situaciones de cambio de modo en los datos recogidos en el corpus CoDICE. A partir de estas dos fases exploratorias, se diseñó un test de hábitos sociales que buscó identificar las premisas socioculturales subyacentes a estas elecciones. El instrumento fue aplicado a un universo conformado por 74 usuarios de WhatsApp de entre 19 y 70 años, hablantes de español de Argentina y de España. Los resultados indican que los hablantes combinan las tres modalidades (texto escrito, mensajes de audio o emojis u otras imágenes) con fines estratégicos. Asimismo, la selección de una u otra modalidad y, más específicamente, el paso de una a otra en el transcurso de un intercambio

¹ Esta investigación fue financiada por el PI-40-C-928, de la Universidad Nacional de Río Negro (Argentina).

comporta significados sociales cuyos efectos se proyectan en la gestión de los vínculos comunicativos de los participantes.

Palabras clave: discurso digital, emojis, comunicación mediada por computadora, interacción transmodal, mensajes de audio.

Abstract

From a pragmatic sociocultural perspective, this work studies an aspect that is still underdeveloped in the field of Digital Discourse Analysis: how speakers perceive the change of mode in the messages exchanged in WhatsApp conversations, more specifically, the transition from text to audio or image. To address this objective, we combine three phases in our methodological approach. In a first place, we analyse the comments circulating in internet as memes about the change of mode. In a second place, we detect situations of mode change in CoDICE corpus. From these two exploratory steps, we designed a social habits test to identify the sociocultural premises underlying these choices. The instrument was applied to a universe made up of 74 WhatsApp users between 19 and 70 years old, Spanish speakers from Argentina and Spain. The results indicate that speakers combine the forms of communication (written text, audio messages, emojis or other images) for strategic purposes. Likewise, the selection of one or the other modality and, more specifically, the passage from one to the other during an exchange entails social meanings whose effects are projected on the management of the communicative links of the participants.

Keywords: digital discourse, emojis, computer-mediated discourse, transmodal interaction, voice messages

1. Introducción

Entre los memes que circulan por internet, hay una captura de pantalla que recoge una supuesta conversación de WhatsApp. En ella, la primera intervención es un audio extenso y la segunda, la siguiente respuesta: “Si es algo bueno te felicito y si es algo malo lo lamento mucho”. En esta misma línea, lleva tiempo circulando un sticker que, en letras blancas y negras, tiene la siguiente frase: “Perate que estoy buscando el sticker”. Las dos imágenes aluden, desde una perspectiva humorística, a factores que inciden directamente en la alternancia semiótica en los intercambios digitales: la gestión del tiempo y la organización de los turnos conversacionales. Estos ejemplos no son casos aislados. Tal y como ya han abordado algunos trabajos previos (por ejemplo, Sampietro y König, 2023), los comentarios metacomunicativos sobre el cambio de modo son muchos y variados, hasta el punto de conformar una parte fundamental de la lingüística popular, asociada al análisis del discurso digital (Thurlow, 2017).

Así lo hemos comprobado en algunas de nuestras investigaciones previas (Cantamutto, 2018, 2023; Cantamutto y Vela Delfa, 2019, Vela y Cantamutto, 2021), en las que, de forma indirecta, a través de varios de los instrumentos de recogida de datos implementados con otros fines y objetivos, ha emergido información sobre estas percepciones del cambio de modo. Tal circunstancia nos llevó a identificar la necesidad de diseñar un estudio que colaborara en la

sistematización de los significados sociales construidos por los usuarios de aplicaciones de mensajería instantánea, más específicamente de WhatsApp, en torno al cambio de modo y la convivencia polimedia en la actual comunicación digital (Androutsopoulos, 2021).

Estudios previos sobre la comunicación por WhatsApp han dado cuenta de la estructura de sus conversaciones (Alcantará-Plá, 2014) y de los factores que inciden en ella (Benítez-Eyzaguirre, de-Marcos y Acosta-Calderón, 2023). Más específicamente, en lo que atañen a los aspectos multimodales, los graficones o recursos iconográficos han sido analizados ampliamente en lo que concierne tanto en sus aspectos culturales como a sus funciones (Sampietro, 2019, Konrad, Herring y Choi, 2020, Herring y Ge-Stadnyk, 2023). Sin embargo, los audios de WhatsApp constituyen todavía un tema novedoso y poco explorado. Los trabajos de Sampietro y König (2023) y König (2024) son unas de las primeras aproximaciones lingüísticas en profundidad sobre el tema. Llamamos *audio*, siguiendo la denominación extendida en nuestras comunidades de estudio, a aquellas intervenciones que se corresponden con mensajes grabados de duración diversa, que se intercambian en las conversaciones de WhatsApp. En la bibliografía anglosajona se llaman *voice messages*, denominación que, en ocasiones ha sido adaptada al español como *mensajes de voz*.

Sampietro y König (2023) indagan en las percepciones que, sobre el uso de los mensajes de voz de WhatsApp, muestran usuarios españoles y alemanes. Evidencian que muchos de estos mensajes vienen acompañados de intervenciones textuales que funcionan como anclajes metapragmáticos y que varían según se sitúen antes o después de este. El análisis propuesto por estas autoras resulta un antecedente fundamental para nuestro estudio, ya que profundiza en las funciones sociales de estos comentarios que sirven para justificar o resaltar el mensaje de voz y trata de averiguar hasta qué punto los usuarios tienen conciencia sobre la idoneidad de cada modo en relaciones con los distintos contextos comunicativos. Asimismo, el concepto de *interacción transmodal*, propuesto por König (2024), da cuenta de las diferentes combinaciones semióticas en las que se desarrollan las conversaciones digitales para centrarse en el estudio de los mensajes de audio (*voice messages*). Sus resultados se focalizan en la idea de que estos mensajes requieren, en no pocas ocasiones, de apoyos para hacer explícita su relevancia.

Precisamente, este último punto constituye el foco de nuestro trabajo, cuyo objetivo es ahondar en la manera en que se produce la selección de uno u otro modo, dentro de esa *interacción transmodal* y, más específicamente, en las formas en que se negocia el paso de uno a otro en el transcurso de un intercambio. Buscamos analizar si estas alternancias comportan significados sociales que se proyectan en la gestión de los vínculos comunicativos. Además, nuestro estudio no se limita al contraste entre la voz y el audio, sino que incorpora un tercer grupo de elementos: los graficones. Como veremos en el análisis, son los propios informantes quienes consideran que estos juegan también un papel fundamental a la hora de inclinar la elección. En este sentido, nos interesa recuperar los motivos que provocan la alternancia entre el texto y el audio en una conversación por WhatsApp, o la incorporación de recursos gráficos, tanto desde la perspectiva del emisor como del destinatario.

Así, hacemos foco en las dimensiones que inciden en la elección y negociación del modo en la interacción por WhatsApp, entendida como parte de la trama de conversaciones online y offline que las personas establecen entre sí. Estas conversaciones, *porosas* (en el sentido de Blommaert, 2021, p. 41), tienen al sujeto como epicentro de intercambios que se dan de manera simultánea y superpuesta en diferentes espacios interaccionales (Androutsopoulos, 2014), presenciales o en línea (en diferentes aplicaciones o en la misma aplicación, pero en diferentes ventanas o chats). A la complejidad inherente a todas las prácticas comunicativas se le suma una capa más, la de la convivencia de distintos modos semióticos dentro de una misma conversación.

Desde el punto de vista teórico, nuestro abordaje se circunscribe en el enfoque de la Sociolingüística Interaccional (Gumperz, 1982; Tannen, 1996) y toma aportaciones de la Etnografía de la Comunicación (Gumperz y Hymes, 1972, la Microsociología del lenguaje (Goffman, 1981) y la sociolingüística de las interacciones en línea (Holmes y Hazen, 2014; Blommaert, 2021), que estudia la relación entre los repertorios sociolingüísticos y la producción de identidades, en contextos digitales. Además, para describir el uso situado de los recursos comunicativos provistos por un entorno comunicativo concreto seguimos postulados clásicos de la pragmática (Verschueren, 2002) a los que sumamos los modelos ofrecidos por la Pragmática Sociocultural, que agregan la dimensión cultural y la percepción de los hablantes del uso de la lengua en contexto (Bravo y Briz, 2004; Bravo, 2005, 2009, 2020).

Partimos del presupuesto de que, cada vez que se comunican, las personas realizan elecciones (lingüísticas) que tienen en cuenta tanto factores microcontextuales como factores macrosociales. Al mismo tiempo, en este proceso de elecciones en múltiples niveles (Verschueren, 2002), se seleccionan modos semióticos –complementarios– que, en los intercambios por WhatsApp, incluyen mensajes escritos, mensajes de audio o mensajes de video, elementos verbales o multimodales, estrategias de retroalimentación automática para estructurar la conversación o (video)llamadas. Así, son las personas usuarias quienes, en un proceso de negociación que incluye factores del diseño –o propiedades de las interfaces– y necesidades comunicativas, movilizan estratégicamente recursos de sus repertorios sociolingüísticos (Blommaert y Backus, 2013).

Los hablantes realizan continuas elecciones, atravesadas por propiedades del lenguaje (y de las mediaciones inherentes a la comunicación digital), que se constituyen como medulares y jerárquicas, modeladas por la *variabilidad*, la *negociabilidad* y la *adaptabilidad*. En palabras de Verschueren (2002, p. 110), “el uso del lenguaje debe consistir en la *continua elección lingüística*, consciente o inconsciente, por razones internas (p. ej. estructurales) y/o externas” (cursivas en el original). La *variabilidad* refiere a “la gama de posibilidades dentro de las cuales se puede seleccionar” (2002, p. 115), siempre mutable, que se negocia dentro de la misma interacción. Consecuentemente, la *negociabilidad* es la causa de que “las elecciones no se hagan mecánicamente o según unas reglas estrictas o una relación de forma-función fijas, sino basadas en principios y estrategias altamente flexibles” (2002, p. 116). Debe ser, por tanto, permanentemente renegociada. Por último, la *adaptabilidad* es la capacidad de “hacer elecciones lingüísticas dentro de una gama de posibilidades variables de modo que se acerquen a la satisfacción de sus

necesidades comunicativas” (2002, p. 119). Esta propiedad tampoco es unidireccional, es decir, como las otras dos propiedades es dinámica y coyuntural, objeto de reorganizaciones constantes, entre los interlocutores. A las elecciones lingüísticas, en nuestra perspectiva, sumamos los diferentes modos que se despliegan en las interacciones digitales.

Esta matriz teórica nos ofrece un conjunto de conceptos significativos para comprender la construcción y negociación de la identidad interaccional y para explicar los elementos definidores del estilo comunicativo de los grupos sociales. A ellos se prestará especial atención dado el interés que presentan en relación con la significación de la cortesía para la comunidad objeto de estudio y la complejidad que conlleva la articulación de los elementos codificadores de las estrategias pragmáticas y del estilo digital.

Tras esta introducción, en el siguiente apartado presentamos la metodología de la investigación. A continuación, desarrollamos un análisis de la percepción sobre las motivaciones de elección en el repertorio de recursos comunicativos de WhatsApp y algunos aspectos generales que inciden en su elección, a partir de tres conjuntos de datos. Por último, presentamos las conclusiones.

2. Metodología

En el clásico trabajo de Hernández Flores (2002), se propone un esquema de alternativas para acceder a la percepción que tienen los hablantes sobre los valores y creencias que otorgan a determinadas prácticas comunicativas. En su tarea por recuperar esta dimensión, los lingüistas han empleado vías, no excluyentes entre sí, entre las que se encuentran, por ejemplo, la consulta de bibliografía o el análisis de los comentarios que los propios informantes realizan ante determinadas muestras de lengua y que nos permiten acceder a información natural (no dirigida por el investigador) sobre los hábitos y creencias de cada grupo social. Otra vía de acercamiento consiste en el rastreo y la recuperación de información metapragmática que de manera espontánea emerge dentro de los corpus de estudio, aunque no siempre es fácil dar con reflexiones de este tipo y su aparición depende, muchas veces, de la cantidad de datos de los que se disponga. Como cuarta alternativa, Hernández Flores plantea la implementación de cuestionarios, encuestas o test de hábitos sociales (THS) con “preguntas relativas a asuntos del comportamiento social” (2002, p.187).

A partir de esta propuesta, hemos diseñado una aproximación metodológica que incluye tres etapas, en línea con lo planteado por Emery (2018). La primera etapa consistió en el análisis de un corpus de piezas digitales (memes, placas de Facebook, stickers, entre otras), recogido en los últimos diez años, relativo al cambio de modo. Con él buscamos poder acceder a las reflexiones que sobre nuestro objeto de estudio fluyen espontáneamente entre los usuarios de WhatsApp. En segundo lugar, a partir de una revisión del corpus CoDiCE², hemos analizado algunos casos de cambios de modo a partir de dos variables: 1) el tipo de chat (entre dos o grupal), 2) el dominio de uso (laboral, amistad, familiar, de servicio) y 3) el país (Argentina y España). Por último, en la tercera etapa hemos implementado un

² <https://codice.aplicacionesonline.com.ar/>

test de hábitos sociales (THS). En este artículo vamos a presentar, fundamentalmente los resultados obtenidos en la tercera fase y solo de manera muy tangencial y subsidiaria al objetivo principal aludiremos a algunos resultados de las dos fases previas. Sin embargo, conviene recordar que estas dos fases exploratorias resultaron imprescindibles para el diseño del THS cuyos resultados analizamos en este artículo.

Los THS proveen datos sobre cómo los hablantes de una determinada comunidad describen y perciben los diferentes recursos y estrategias de cortesía (Bravo, 2009). Es decir, son un instrumento que permite explicar la (des)cortesía de acuerdo con el contexto sociocultural de ocurrencia y las premisas socioculturales que subyacen a las elecciones de los hablantes.

A partir de este instrumento se puede recabar información sobre los modelos apropiados para la realización de determinados actos de habla, la modificación de los comportamientos en función de variables sociolingüísticas y las premisas socioculturales (Hernández Flores, 2002). Asimismo, los THS aportan datos para apoyar o complejizar las interpretaciones de las muestras de lengua. Con esta técnica es posible evidenciar el uso, a través de situaciones ficticias con las que se busca recuperar las estrategias que los interactuantes utilizan para alcanzar sus metas comunicativas. A través de preguntas orientadas, los THS permiten recabar información sobre la percepción de los hablantes en relación con su conciencia sociolingüística, así como sobre las valoraciones que hacen en torno a la lengua en uso y los fenómenos de (des)cortesía verbal que derivan en valoraciones sobre los usos lingüísticos.

En esta investigación, el estudio se ha realizado a través de un formulario online en el que se interrogaba sobre el cambio de modo en WhatsApp, que se ha hecho llegar, a través de la técnica de bola de nieve, a informantes provenientes de Argentina y de España.

El test contenía un apartado sociolingüístico en el que se consultó tanto sobre la población de nacimiento como sobre la de residencia, atendiendo a los contextos de superdiversidad contemporáneos (Blommaert y Backus, 2013). Las mujeres fueron quienes más respondieron la encuesta (el 75% de los casos). Dado el carácter cualitativo de la investigación, se tendrán en cuenta las respuestas predominantes para los diferentes grupos.

Tabla 1. Descripción de la muestra de Test de Hábitos Sociales

Género	Edad			Total
	A [entre 19 y 35 años]	B [entre 36 y 55 años]	C [entre 56 y 70 años]	
Mujer	21	28	9	58
Hombre	7	11	1	19
Total	28	39	10	77

3. Análisis de los resultados

3.1. El cambio de modo en WhatsApp a través de los memes de Internet

Una de las maneras de acceder a las valoraciones que los usuarios de WhatsApp tienen sobre el repertorio de opciones semióticas disponibles en la aplicación y sobre la adecuación y pertinencia de su uso puede ser la observación de los memes que circulan en internet en relación con ese tema. Por ello, la primera fase de nuestra investigación consistió en el análisis exploratorio de un corpus memes que abordaban la cuestión de la convivencia entre el texto, el audio y la imagen en las interacciones de WhatsApp. Algunos de ellos ya han sido mencionados y comentados en el apartado de la introducción.

Los memes de la cultura digital son signos culturales de naturaleza audiovisual diversa que se comparten a través de las redes sociales. Desde distintas estrategias comunicativas, entre las que destaca el humor, estos signos de la cultura popular abordan y reelaboran diferentes temas de actualidad. No debemos restarles valor exploratorio a estas unidades que de forma tan eficaz recogen y difunden esquemas cognitivos (Goffman, 1979 [1971]), compartidos por los miembros de la comunidad que los hace circular, y que funcionan como marcos interpretativos fundamentales para organizar la experiencia. Por esta razón, la proliferación de estos memes constituía en sí misma una pista clave sobre las representaciones mentales compartidas por los usuarios de mensajería instantánea (MI) en torno a la alternancia texto/audio/imagen y sobre la propia evolución multimodal de la aplicación.

En la actualidad, WhatsApp es la aplicación de MI más empleada, muy por delante de otras como WeChat o Telegram (Statista, 2024). Tanto en su versión móvil como en las alternativas de escritorio o web, WhatsApp alberga una gran cantidad de los intercambios personales privados, semipúblicos e incluso públicos (a través de canales abiertos). En sus comienzos ofrecía únicamente la posibilidad de intercambiar mensajes escritos, pero, con el paso del tiempo, ha incorporado otras alternativas comunicativas, como el audio y la imagen. A pesar de que WhatsApp nunca ha limitado el número de caracteres, sus mensajes siempre han tendido a ser breves (con excepciones justificadas). De hecho, entre los más jóvenes circula una expresión “mucho texto” que se utiliza para comunicar que un texto ha sido ignorado debido a su longitud.

Ilustración 1. Meme "Mucho texto".



En el meme, popularizado desde el año 2020 (Ilustración 1), aparece el personaje de la cultura popular conocido como “el yoda hipster acompañado de la frase ‘mucho texto’” para aludir a la preferencia por los textos breves y a la inadecuación e irrelevancia de las intervenciones demasiado largas en los entornos de MI.

Esta preferencia por la brevedad en WhatsApp o, más bien, inadecuación de los textos extensos puede deberse a su vinculación con otra práctica comunicativa previa, la del envío de SMS, pero también con una tendencia más general del discurso digital, la de la brevedad de los textos. Esta brevedad ha traído aparejado otro interesante fenómeno: la generación de recursos para garantizar la expresividad sin la necesidad de “mucho texto”.

Así, desde la aparición de los primeros sistemas de mensajería instantánea, su evolución, como la de otros entornos de comunicación digital, ha ido en buena medida orientada a la diversificación de las alternativas semióticas. De la etapa inicial, profundamente textual (Kress, 2003; Jewitt, 2013), se ha pasado a la actual, marcada por la multimodalidad. En esta evolución el equilibrio entre texto, audio e imagen ha ido reconfigurándose. Más particularmente, en el caso de WhatsApp, la brevedad lingüística se ha apoyado en distintas estrategias multimodales que han dado lugar a un medio híbrido, que combina múltiples opciones comunicativas.

El 7 de agosto de 2013³ WhatsApp introdujo las versiones primitivas de los mensajes de audio. Esta actualización ofrecía como gancho a los usuarios la posibilidad de acceder a la voz de personas queridas. Sin embargo, más allá de las razones emotivas, el éxito de los mensajes de voz se sustentó en gran medida en la comodidad que supuso no tener que teclear letra a letra los mensajes escritos y poder dar rienda suelta a la expresividad y a la extensión.

Con el tiempo, además, las prestaciones (*affordances*) se ampliaron. Con las mejoras en el sistema de grabación y escucha de los mensajes, así como con la incorporación de opciones para eliminar los mensajes o limitar su reproducción a una única vez, los audios se han popularizado y desplazado, al menos en parte, a los mensajes escritos y a las llamadas telefónicas tradicionales. Tal y como retrata un meme que dice: *Primero el SMS, después vino el WhatsApp, ahora grabas un mensaje de voz, y tu amigo te la graba la respuesta. Si siguen así van a inventar el teléfono*. Este meme da cuenta de una cuarta vía alternativa en las elecciones de los hablantes. Más allá del mensaje escrito, de la imagen o del mensaje de voz también está la llamada telefónica.

La inclusión de los audios en WhatsApp no estuvo exenta de polémica y trajo aparejada la división de los usuarios entre los que adoran los audios y los que los detestan. Los audios no son percibidos igual por todos los hablantes y, aunque puedan resultar la opción preferida en ciertos contextos, en otros ni siquiera son consideradas una alternativa aceptable, sobre todo, cuando, como en el caso retratado en el primero de los memes comentado en la introducción, son demasiado largos.

Pero la idoneidad del audio como elección comunicativa depende mucho del punto de vista: no lo percibe igual el emisor que el receptor. En los datos proporcionados en su último informe por Statista (2023), en relación con las

³ [Introducing Voice Messages - Blog de WhatsApp](#)

preferencias de los usuarios en lo concerniente a la selección de audio o texto en los mensajes, comprobamos que los mensajes escritos siguen siendo el medio preferido por seis de cada diez encuestados (66% cuando envían y 61% cuando reciben mensajes). De hecho, tan solo el 7% de los participantes indicó una inclinación hacia los mensajes de audio. Sorprendentemente, no parece haber diferencias significativas entre el punto de vista de quien los recibe y de quien los envía: al 21% le es indistinto enviar audio o texto y al 27% recibir audio o texto. Sin embargo, los memes de internet no dicen lo mismo y parecen corroborar que las elecciones en la comunicación digital dependen de las diferencias relativas al esfuerzo (cognitivo y físico), implícitas en la interacción humano-interfaz (Scolari, 2004). En este sentido, cabe comentar la cantidad de memes que circulan recogiendo la sensación de desagrado por el esfuerzo suplementario que supone la recepción de un audio largo. Los memes también focalizan en las diferentes percepciones del hablante y el oyente. Hay un meme en el que se contrastan dichas percepciones (Ilustración 2).

Ilustración 2. Meme sobre recepción de audios largos.



Algunos de los cambios de la aplicación han tratado de paliar esta situación al ofrecer, por ejemplo, la posibilidad de escuchar los audios a mayor velocidad, doblando la velocidad (2x) o aumentándola un cincuenta por ciento (1.5x). Muchos memes abordan desde la perspectiva humorística la distorsión que la voz del remitente sufre en tales escuchas.

Las herramientas de Inteligencia Artificial (IA) también permiten transcribir los audios en texto, algo que puede ayudar cuando la elección modal de emisor no coincide con la preferencia del receptor. Tal sería la situación que queda retratada en una serie de memes que ridiculiza la insistencia con el modo audio de uno de los interlocutores a pesar del comentario explícito del otro sobre su inadecuación a la situación comunicativa concreta. Estos suelen tener la siguiente estructura: una captura de WhatsApp en el que se ven varias intervenciones en formato audio que son respondidas por una intervención que dice “no puedo escuchar audios” a la que se contesta con un audio.

Además de la incorporación de la funcionalidad del audio, el uso de imágenes también ha ido ganando protagonismo. Con el paso del tiempo, el reducido número de emojis disponible en las versiones iniciales se ha transformado en un amplio abanico de unidades al que, incluso, se le suma la posibilidad de crear stickers con

IA⁴. Cada vez son más numerosas las alternativas para comenzar, continuar o cerrar las conversaciones mediante recursos gráficos disponibles en los teclados y pantallas. Son tantas que a veces podemos perdernos en ellas, tal y como refiere el segundo de los memes comentados en la introducción.

Además, a este repertorio, se le suman las retroalimentaciones automáticas (*recibido, leído, responder, reenviado, grabando*) y las reacciones, que, de una manera u otra, inciden en la gestión de los vínculos comunicativos. El uso de los graficones resulta menos polémico entre los usuarios de WhatsApp por lo que ha generado menos memes. Aunque algunos van en la línea de explicitar que estos signos, creados aparentemente para precisar la interpretación de ciertas intervenciones, a veces no son entendidos de forma evidente por el interlocutor.

En este breve análisis exploratorio, hemos ido viendo como los memes no solo retratan, desde una perspectiva humorística, los cambios que ha ido implementando WhatsApp, sino que también ilustran los marcos desde los que se aborda socialmente la conciencia que los usuarios tienen de los significados sociales asociados a estas modificaciones. De su análisis se deriva la relevancia que las selecciones semióticas adquieren en estos entornos de comunicación.

3.2. Conversar por WhatsApp: análisis de los cambios de modo y su incidencia en el desarrollo de la conversación en el corpus CoDICE.

En WhatsApp, lo que *la gente hace* es conversar: conversaciones a través de texto, audio, imagen y otros contenidos embebidos (de otras plataformas, enlaces, geolocalización, entre otras). En buena medida, se trata de una conversación siempre abierta, con límites muy difusos, que excede la aplicación de WhatsApp.

Esta complejidad nos lleva a la necesidad de distinguir distintos niveles de *conversación*. Por un lado, tenemos la historia interaccional (Vela Delfa, 2007) o log data, es decir, el chat definido por la interfaz y al que se accede en el inicio de la aplicación. Esta conversación, a la que denominaremos conversación₂, forma parte -entre cosas- del contexto lingüístico compartido entre interlocutores y, de hecho, en el discurrir de un intercambio se traen intervenciones de interacciones previas (como se verá en el ejemplo 3). Según la configuración de temporalidad de los mensajes, puede acercarse más a lo efímero o lo permanente. En el interior del chat, tenemos una sucesión de intercambios –tanto mensajes como reacciones (intervenciones iniciativas y reactivas)–, a las que denominamos conversación₁, a partir de criterios temporales y/o temáticos. Ante un quiebre temporal o ante el inicio de un tópico nuevo se determina el inicio de una nueva conversación₁ entre dos o más personas, según el tipo de chat⁵. De manera sucesiva, cada conversación₁ se inserta en la conversación₂ que queda “oculta” en la parte superior de la pantalla, pero accesible a través del buscador o del deslizamiento de los mensajes. Ambos niveles, además, se proyectan en el repertorio polimedia que podría funcionar como

⁴ https://blog.whatsapp.com/express-yourself-with-more-stickers/?page_source=search&q=sticker [fecha de consulta: agosto de 2024]

⁵ La alternancia dinámica de turnos nos lleva a percibir estos intercambios como conversacionales que se distinguen de los intercambios de difusión, en los que la toma de turno por algunos participantes solo sea recibir los mensajes y, en algún caso, reaccionar (como en los grupos de difusión).

un tercer nivel en el que una conversación₃ se desarrolla a través de distintas aplicaciones y dispositivos (Tagg, 2013; Androutsopoulos, 2014; König, 2024).

En cada conversación₁ está latente todo el repertorio de recursos comunicativos disponibles en la aplicación que, al mismo tiempo, exceden los límites de su arquitectura: la interoperabilidad favorece la convergencia entre aplicaciones. Con frecuencia, tal como veremos en los ejemplos siguientes, los usuarios despliegan y combinan estratégicamente los recursos comunicativos según factores discursivos, contextuales y técnicos, además de otros vinculados a la gestión de los vínculos. Sin embargo, estos cambios de modo se basan en premisas socioculturales en las que el texto escrito es la opción no marcada (o preferente, en línea con la propuesta de Verschueren, 2002) y, a partir de ahí, se adosan y tensionan los otros modos comunicativos.

A los fines de ejemplificación, presentaremos conversaciones₁ en las que se ilustran los cambios de modo extraídas del corpus CoDiCE. Los ejemplos son parte de una muestra mayor de análisis, relativa a la fase 2 de nuestro diseño metodológico, y han sido seleccionados con el objetivo de mostrar parte de la información metapragmática espontánea desplegada en el interior de las interacciones.

Como puede observarse en el ejemplo 1, las dos primeras intervenciones eligen el texto como modo. La primera se corresponde con un vocativo en función alertadora y la segunda con una intervención iniciativa del par pregunta-respuesta. En ese momento se produce un cambio de modo al audio, pero con un paso previo por el texto. La intervención reactiva se inicia por escrito, pero debido al carácter narrativo de la respuesta se cambia al formato audio que permite crear con comodidad una intervención larga. En este sentido, este cambio de modo resulta preferido para completar el par de adyacencia (ejemplo 1). A pesar de ello, como en muchos comienzos de audio, sobre todo cuando este implica un cambio de modo, se realizan comentarios metapragmáticos del cambio de modo.

1) Ejemplo 1

<p>Ámbito: amistad. Contexto: una joven (Vi) escribe a su amiga (Ani) para preguntarle cómo le había ido en un concurso docente.</p> <p>23:07 - Vi: Aniii</p> <p>23:07 - Vi: Cómo fue el concurso?</p> <p>23:08 - Ani: Viiiiiiiiiiii!!!!!!!</p> <p>23:12 - Vi: Audio-3 minutos 12 segundos</p> <p>23:15 - Vi: Audio-2 minutos 17 segundos</p> <p>Transcripción del comienzo del audio: Eh: qué hacés, Vi! Te iba a escribir, pero me di cuenta de que, de que es más rápido un audio (...) no lo quiero hacer muy largo esto. Me fue bien. Yo estoy muy conforme con lo que hice (...)</p>
--

Si analizamos con más detalle el fragmento, observamos que en la respuesta de Ani⁶ hay una primera intervención escrita que funciona como confirmación de recepción y mantiene abierta la conversación₁ hasta que se produce el envío el mensaje de audio, cuatro minutos después, debido a su extensión. La intervención

⁶ Los datos están anonimizados siguiendo el protocolo de CoDiCE, disponible en: [Guía interna para la elaboración de los datos \(PICT-2019-02093-PRÉSTAMO BID\)](#)

escrita podría resultar innecesaria si tenemos en cuenta que WhatsApp tiene marcas de retroalimentación, como por ejemplo el mensaje en verde con la leyenda *grabando audio*, que ofrecen información sobre el tiempo en que el interlocutor está produciendo la respuesta antes de enviarla. Por ello, la aparición del vocativo “Viiii” como primera respuesta podría ser marca de una primera selección de modo que, rápidamente, se desecha por no resultar adecuada para contar lo sucedido. Tras esa nueva elección, se opta por el envío seguido de dos audios extensos. De hecho, la primera parte del audio así lo muestra, ya que afirma “Te iba a escribir, pero me di cuenta de que, de que es más rápido un audio”. La interlocutora afirma que enviar un audio es más “rápido”; pero, sin embargo, consciente de la necesidad de justificar ante su destinataria la elección, sostiene que “no lo quiero hacer muy largo”. Si comparamos el punto de vista de emisor y receptor, contrasta la aparición de la palabra “rápido” con la extensión final de la intervención, casi siete minutos entre los dos mensajes.

El ejemplo anterior ilustra un uso frecuente de los audios: el desarrollo de secuencias narrativas extensas cuya realización en formato texto resulta más dificultosa tanto desde el punto de vista físico/material?, por su tecleado, como por su planificación. En estos audios una misma intervención -audio- despliega diferentes tópicos.

En el siguiente ejemplo, ejemplo 2, desarrollado en el ámbito laboral, se utiliza el audio para explicar la demora en el envío de un comprobante solicitado.

2) Ejemplo 2

Ámbito: laboral. Contexto: conversación entre dos jóvenes profesionales sobre el pago que va a realizar Juan de un trabajo realizado por Manuel.

3/4/18 10:05 - Manuel: archivoadjunto.pdf (archivo adjunto)
consultacbu 000000000000
3/4/18 10:06 -Juan: Recibido
3/4/18 10:06 - Manuel: Ok
3/4/18 10:06 - Juan: Por favor enviame factura así hago el deposito
3/4/18 10:12 - Manuel: <Multimedia omitido>
3/4/18 10:13 - Juan: Barbaro

Desde la perspectiva de quien envía el audio, es una elección de modo basada en la necesidad de realizar un desarrollo extenso de lo que sucede, es decir, de dar en poco tiempo una explicación detallada. En contra partida, y debido muy probablemente al vínculo establecido entre los hablantes, la respuesta vuelve al modo escrito. Como se verá en la siguiente sección de análisis de los THS, el ámbito de uso es determinante en la elección del modo.

En el ejemplo 2, no solamente ilustra un cambio de modo (de texto a audio) sino que da cuenta de otra posibilidad multimedia de la aplicación: embeber en la conversación archivos de diversa naturaleza. En este caso, un archivo PDF. En estas situaciones, al pulsar en el archivo, se abre otro programa, según la configuración del usuario para visualizar archivos, con la consiguiente ampliación polimedia de la conversación.

Tras el comentario de los dos ejemplos precedentes, pasamos a ocuparnos de otro aspecto vinculado con la asistencia tecnológica que intermedia la comunicación

digital. Una diferencia de estas conversaciones con las presenciales radica en que, en la conversación digital, los participantes acceden de un modo simultáneo a la producción y a la recepción de los intercambios y los papeles de emisor/destinatario son fluidos. En los intercambios digitales, la mediación tecnológica imprime modificaciones de diferente índole: los diccionarios predictivos que cambian las palabras, las acciones involuntarias como llamadas que se “disparan” o “stickers” que se envían.

Esta mediación tecnológica otorga al receptor un mayor control sobre el ritmo del intercambio, ya que puede decidir cuándo y cómo accede al mensaje. Desde el punto de vista de las prestaciones de la aplicación (*affordances*), el audio permite al receptor pausar la reproducción y cambiar la velocidad de escucha. La única manera en que el emisor (y el investigador) puede saber cómo fue escuchado el audio es a partir del rastreo de las huellas que se inscriben en la propia conversación: pedidos de disculpas o comentarios, como se verá en los cinco ejemplos de CoDiCE que presentamos a continuación. En ellos, los hablantes introducen comentarios sobre cómo escuchan (y responden) a los mensajes de audio, bien en diferentes velocidades, bien fragmentando las intervenciones de audio que reciben, al contestar parcialmente mientras escuchan.

En el ejemplo 3, se apela al contexto de la conversación₂ para retomar una conversación₁ que había quedado suspendida por no escuchar el mensaje de audio completo.

3) Ejemplo 3

<p>Ámbito: amistad. Contexto: dos amigas que trabajan juntas en la universidad están comentando, en la primera conversación, sobre la presentación de unos datos. En la conversación siguiente, le comenta que se publicó un artículo en coautoría. Luego de responder remite a la conversación anterior para indicar que no había terminado de escuchar el mensaje de audio</p>
<p>30/11/2022 19:01 Emilse: audio_2 minutos y 31 segundos 1/12/2022 00:58 Emilse: Salió <i>nombre_revista_anonimizado</i> 1/12/2022 00: 58 Emilse: Por fin jajaj 1/12/2022 01:40 Belén: Emiiiiii 1/12/2022 01:41 Belén: respuesta_a_Emilse: audio_2 minutos y 31 segundos: No te pude terminar de escuchar 1/12/2022 01:41 Belén: Por eso no te respondí 1/12/2022 01:41 Belén: Ahora te escucho</p>

Algo parecido sucede en el ejemplo 4, en el que, a pesar de su brevedad, el audio no es escuchado en primera instancia, cuando se envía sin texto introductor, pero consigue dar cuenta de su presunción de relevancia en un segundo envío, cuando la emisora lo manda con una pequeña introducción textual en la que recuerda a su receptora que no lo escuchó’ la primera vez. De esta forma el par adyacente tarda cinco días en completarse.

4) Ejemplo 4

<p>Ámbito: laboral. Contexto: dos colegas que trabajan juntas en la universidad. Tras enviar un primer audio que no es atendido por la receptora, la emisora lo reenvía con una introducción textual para llamar la atención y lograr que sea escuchado.</p>
<p>5/9/2023 10:10 Carla: audio_1 minuto 19 segundos</p>

9/9/2023 10:33 Carla: Hola, Marta, te envié un audio y como veo que no lo has escuchado...espero que estés/estéis bien! Te lo reenvío!!
5/9/2023 10:11 Carla: audio_1 minuto 19 segundos
5/9/2023 10:16 Marta: audio_2 minuto 21 segundos

En el ejemplo 5, dada la extensión del audio, quien envía resume el contenido a través de texto escrito e indica que se escuche a gran velocidad.

5) Ejemplo 5

Ámbito: laboral. Contexto: en el marco del chat de un grupo de investigación, una mujer envía un audio resumiendo las entrevistas que había realizado.
13/10/2021 10:20 Agustina: audio_7 minutos 53 segundos
13/10/2021 10:20 Agustina: resumen de las 3 entrevistas y propuesta de caso. Pueden escucharlo en 2x

Además, tenemos otros ejemplos en los que se señala, después de un audio, “Sin apuro 🙌” o “escucha cuando puedas”, entre otras, como en el ejemplo 6.

6) Ejemplo 6

Ámbito: personal. Contexto: dos amigas conversando.
15/9/2022 15:02 Pilar: audio_45 segundos
13/10/2021 15:02 Pilar: Sin apuro

Por su parte, en el ejemplo 7, se produce una estructuración novedosa de la conversación: al mismo tiempo que Euge ocupa el turno de habla, a través de un audio ya enviado, se van intercalando intervenciones reactivas de Hernán a su mensaje.

7) Ejemplo 7

Ámbito: amistad. Contexto: un hombre pregunta a una amiga cómo está su hijo luego de un cuadro de salud.
18/08/2022 17:19 Hernán: Euge, al final no supe más nada, cómo está Germán?qué pasó finalmente en <i>ciudad_omitida</i>
18/08/2022 17:20: Respuesta a mensaje_Euge, al final no supe más nada, cómo está Germán?qué pasó finalmente en <i>ciudad_omitida</i> : Te mando audio
18/08/2022 17:20 Hernán: claro!
18/08/2022 17:20 Euge: audio_40 segundos
18/08/2022 17:21 Euge: audio_35 segundos
18/08/2022 17:23 Hernán: voy respondiendo mientras te escucho
18/08/2022 17:23 Hernán: uhh, no te puedo creer qué escenario para llegar
18/08/2022 17:23 Hernán: claro, ni hablar, llegás lleno de miedo
18/08/2022 17:24 Hernán: a quién escucho de fondo? Jaja
18/08/2022 17:24 Euge: Respuesta a mensaje_ “a quién escucho de fondo? Jaja” Imagen [Foto de un bebé]

3.3. La percepción de los hablantes: factores que inciden en la elección del modo
 Tras la breve presentación de algunos de los resultados de las dos primeras fases de análisis, la reflexión sobre los comentarios espontáneos que circulan en los memes y el análisis de los metacomentarios derivados de situaciones concretas de cambio de modo en el corpus CoDiCE, pasamos a exponer los resultados del THS implementado en la tercera fase de nuestra investigación.

En el THS distinguimos, en primer lugar, situaciones ficticias en el que el foco estuviera puesto en elecciones de modo en la fase de emisor y otras en las que el foco estuviera puesto en la fase de recepción y pudimos comprobar que esta diferencia fue especialmente significativa. En segundo lugar, tras un primer análisis de los datos, encontramos que, de las variables sociolingüísticas estudiadas, el grupo etario se mostró como la más productiva.

Asimismo, en tercer lugar, atendimos a la variable relativa al dominio de uso que resultó ser fundamental en el cambio de modo. La mayoría de las respuestas de la primera situación ficticia, referida a la gestión de un turno médico, dan cuenta de la preferencia por el texto escrito, en los tres grupos etarios y en ambos géneros la elección de este modo fue superior al 87%. En esta situación, solo tres personas indicaron que usarían texto escrito más emoji o emoji solo y otras tres personas respondieron que usarían audio. En esta línea, resaltan algunos comentarios que dan cuenta de cómo se entiende la comunicación en este dominio: *En ningún caso enviaría un audio*⁷. *Sería un texto formal; No le respondería con un audio, lo haría en forma escrita consultándole si puede ser al mediodía*. La distribución de las respuestas que seleccionan un modo diferente al escrito no están relacionadas con la edad ni con la región de origen.

Para estudiar el dominio de uso familiar, se diseñó una situación en la que se produce un mensaje mientras se está cocinando. Ese mensaje trata sobre la organización de la compra de un regalo (situación 2). En la respuesta se propone enviar información acerca de algunas posibles opciones de regalo. En este caso, se observa una mayor preferencia por el audio y por la conjugación/combinación de modos: el 55% de las respuestas indicaban el empleo de audio (y, en un caso, el audio junto con emojis) sin haber diferencias en cuanto al grupo etario, aunque sí en relación con el género: las mujeres optaban por el audio en un 60% mientras que los hombres en un 40%. El 19% de los informantes prefirió responder con texto escrito, aunque otro 24% optó por texto escrito combinado con graficones (emojis, stickers y emoticones). Fueron los hombres quienes, con una diferencia de mínima, eligieron más estas opciones textuales. Destaca, además, el hecho, de que, en estos casos, se indicaba la pertinencia de acompañar el texto del envío de enlaces, a los productos: *Le enviaría el enlace de lo que vi, o se lo describiría. No mandaría un audio porque mis hermanas son mayores que yo y nunca utilizan audios*. La posibilidad de embeber otro tipo de contenidos en la aplicación favorece la inclusión de enlaces a otros sitios web o aplicaciones (como Instagram), como elemento externo que forma parte de la conversación¹. En esta situación también encontramos respuestas que aluden al compromiso (en audio o texto) sobre llamadas posteriores.

Las otras dos situaciones ficticias (situación 3 y situación 4, respectivamente) consideraban la gestión de los vínculos como un factor a la hora de elegir el modo. En ambos casos, un/a amigo/a enviaba un audio extenso de disculpas por haber olvidado el cumpleaños. Sin embargo, en la primera situación el mensaje producía enojo mientras que, en la segunda, el olvido no era percibido como tan importante y no producía conflicto comunicativo.

⁷ Las citas extraídas de las respuestas a los THS se han dejado en su grafía original y se presentan en cursivas.

En el primer caso, la forma preferida para responder por todos los grupos etarios fue el texto escrito, con o sin emojis, tal vez porque permite enmascarar la emoción y mostrar de forma más controlada el enfado. Los emojis funcionaron como mitigadores en esta situación: *No pasa nada amigo 🤔*, *Tranqui + emojis de corazones*; *Mejor tarde que nunca 😊! Gracias 🙏*, *No te preocupes jajaja Muchas gracias 💖*. En estos casos, la función es reparadora. Sin embargo, también pueden ser utilizados en otro sentido, atenuando la intervención: *Odio los audios largos, por qué no me llamas capullo... Y varios emojis, tipo cara enfurecida y similares*.

El grado de confianza entre los hablantes habilitaba la combinación de modos, incluso, en intervenciones diferentes, al combinarse emojis o stickers con mensajes de audio o mensajes con un emoji como único elemento.

En las respuestas brindadas, emergieron valoraciones sobre los audios largos, que coinciden con las preguntas abiertas que realizamos en la segunda parte del THS. Tanto en la situación 3 como en la situación 4, los comentarios iban en líneas semejantes. a, b, c y d son respuestas a la situación 3 y e, f y g a la situación 4:

- a. (si el audio es muy largo, no lo escucharía hasta tener un tiempo libre) culiado cómo te vas a olvidar de venir al cumple? Es de lo único de lo que hablan en el grupo de WhatsApp hace un mes. Te habíamos preparado el humus vegano para vos
- b. NOTA: Si dura más de 2 min y estoy genuinamente molesto, usaría una IA para hacer una síntesis del mensaje. NOTA: En general cuando me enoja tiendo a mandar menos emojis y me parece que uso más signos de puntuación.
- c. Odio los audios largos, por qué no me llamas capullo... Y varios emojis, tipo cara enfurecida y similares
- d. Le pondría un emoji de corazón y listo. Si estoy enfadada con alguien no voy a ponerme un audio largo... le dejo el emoji y listo.
- e. (aunque sea mí madre con la casa prendiéndose fuego, si el audio es extenso lo voy a escuchar cuando tenga tiempo) hola Juan, no pasa nada querido. Te entiendo y te extraño brodie, arreglamos para el Finde que viene;
- f. Esperaría a salir del trabajo para enviar un audio agradeciéndole su mensaje e incitando algún encuentro para ponernos al día;
- g. Igualmente le pondría un emoji con un corazón para salir del paso, ya que me parece molesto que alguien envíe un audio para felicitar algo, pudiendo escribir o llamar.

En la última situación ficticia, el audio también era largo, pero se aceptaban las disculpas. Resulta llamativo que los graficones fueron predominantes para aceptar las disculpas y reparar la situación: muchas de las respuestas combinan texto escrito con emojis (30%), siendo el grupo etario de jóvenes el que más los emplea (36%). Este grupo también combina elementos gráficos con audios. A pesar de la confianza cercana (dado que eran amigo/as) y el empleo del audio como estrategia reparadora ante un olvido, el audio no es el modo elegido en la respuesta en ninguno de los grupos estudiados.

Destaca particularmente el comentario (b) en el que se incluye una nueva capa de asistencia tecnológica en la escucha. En CoDiCE, tenemos ejemplos de

personas que utilizan el asistente de inteligencia LuZIA para transcribir los mensajes de audio, en situaciones en las que no pueden escucharlo (por ejemplo, durante la jornada laboral).

Las respuestas proporcionadas en las cuatro situaciones ficticias apuntan a la baja frecuencia del empleo del audio. La mayoría de las respuestas dan claros indicios de que el audio es una opción marcada y no preferida tanto en chats grupales como individuales, tanto en ámbitos de mayor confianza (amistad) como de menor cercanía (laboral).

Para cerrar el THS, se hicieron preguntas abiertas para que los informantes reflexionaran sobre las alternancias de modo. Se preguntó sobre cuándo y por qué se opta por uno y otro modo de comunicación. Sintetizamos más adelante algunos de los comentarios más destacables con relación a sobre cuándo resulta pertinente enviar un audio.

En primer lugar, un aspecto que destacaron todos los grupos de edad fue que el envío de audios se justificaba por la necesidad de “explayarse” (palabra que salió en varias respuestas). En determinados contextos de confianza, la respuesta más recurrente ha sido la que vincula al audio con un modo más extenso que el mensaje escrito y menos sincrónico que la llamada. Así, por ejemplo, lo apuntan las siguientes respuestas: *conversaciones más sociales que no puedo sostener en un llamado por no coincidir en horario con la persona que quiero hablar. o cuando necesito explicar una situación y q se transmita la intensidad o sensibilidad de lo dicho, que por escrito no siempre comunica bien.; Cuando el chat responde a una conversación con divague o de temas muy personales; Envío audios cuando algo se me complica para explicar en texto o estoy en la calle y tengo que resolver algo.*

Este aspecto se vincula con otro factor señalado como relevante: el menor grado de ambigüedad del audio porque permite ser más preciso con el tono para transmitir las emociones: *Prefiero aclarar con mi propio tono de voz alguna circunstancia; Por priorizar algo en la melodía de la voz, ya sea a propósito o espontáneamente; Cuando tengo que contar algo extenso o que se pueda malinterpretar al solo escribirlo (cada uno lee el texto con la entonación que le parece); Cuando me resulta más agil a mí o cuando hay una situación potencialmente tensionante y prefiero usar la oralidad para matizarla; En la mayoría de los casos trato de enviar audios mas que nada para que sea claro el tono o los acentos del mje y no dejar librado ese tema a la interpretación del receptor cuando lo que tengo que contar algo en el que necesite expresar de manera más natural cómo me siento, porque el audio logra reflejar mejor el estado de ánimo y la entonación unida a ese estado de ánimo*

En relación con el mensaje escrito, los encuestados resaltaron la dificultad que implica interactuar con los teclados, tanto por el tiempo que requiere la acción como por la dificultad de coordinarla con otra actividad. En estos casos, se apela al contexto y a la situación de comunicación como factor que determina la elección del modo: *Cuando tengo mucho para decir y no tengo ganas de tipearlo. Cuando voy caminando y necesito mandar mensajes; Cuando estoy complicada para tipear (estoy con mis hijas o haciendo algo); Porque no puedo detenerme a escribir. Estoy apurada o haciendo algo generalmente; sólo si tengo las manos ocupadas y pido perdón al comienzo del audio, en general. no me gusta que me envíen audios, así*

intento no mandarlos. Cuando voy manejando, cuando estoy haciendo cosas de la casa o cuando es difícil de explicar algo. Porque me resulta más rápido hacerlo.

En estas respuestas, el audio resalta por práctico y, en un mundo *multitasking*, adecuado para simultanearlo con las múltiples tareas que se realizan de manera paralela. Esta explicación ha sido manifestada mayoritariamente por el grupo etario A y B, pero no por las personas adultas.

Otro caso especial es el de las intervenciones reactivas, en las que el modo elegido en el turno inicial desempeña un papel determinante: algunas respuestas indican que, ante una primera intervención en audio, continúan con ese modo: *Cuando me enviaron un audio lo uso como respuesta.; cuando me envían audio por cuestiones puntuales respondo con audios; Si me mandan un audio, a veces, también contesto con uno; Cuando la otra persona lo hace y no tengo la confianza para decirle que no lo haga, le respondo con su moneda.*

Asimismo, otro motivo que justifica la elección del audio puede ser el hecho de que la conversación se esté extendiendo más de lo esperado: *Cuando ya llevamos varios whatsapp intercambiados y lo que tengo que decir es largo o cuándo creo que no me están entendiendo.*

Otra dimensión que emerge en la elección modal es la confianza y cercanía entre los hablantes. Por un lado, porque en las respuestas indican que en determinados ámbitos no usarían el audio. Por otro lado, porque afirman que, si lo llegan a emplear, lo harían con los resguardos que brinda la plataforma: *En algunos casos cuando me solicitaban alguna explicación en el trabajo. (En ese caso permitía el uso de una sola escucha, presionaba el simbolito de 1 en WhatsApp).*

En muchas respuestas, se indica que la confianza propia del ámbito familiar o de la relación de amistad con amigos es la que habilita la utilización del modo: *Cuando tengo mucha confianza con la persona y no es un entorno laboral; Siempre con personas de mi confianza. Sino, sólo mensajes escritos; cuando tengo mucha confianza con el receptor; Sobre todo con gente con quien me siento cómoda.; Cuando estoy en una interacción familiar o de amistad. Enviar un audio implica cierto grado de confianza. A diferencia del texto escrito, en un mensaje oral puedes cometer errores lingüísticos o discursivos (a saber, cohesivos o coherentes) y la comprensión del mensaje puede verse afectada; cuando es la primera vez que me comunico con esa persona, por ejemplo. me parece descortés recibir solo un audio, sin ninguna introducción escrita de quién es el remitente; Pero no me agrada en situaciones en las que no conozco a las personas.*

Por último, tanto en la primera parte del THS como en esta segunda con preguntas abiertas, se observa que el grado de penetración de las IA es tal que encontramos respuestas en las que la asistencia tecnológica atraviesa la instancia de producción: *Si no puedo escribir, le pido al asistente que mande el mensaje o llamo directamente.*

Así como se elige enviar un audio o responder a través de él, los encuestados respondieron sobre las dimensiones en las que no conviene hacerlo. La interrupción que supone escuchar un audio es considerada en muchas de las respuestas, en las que enviar audio supone *molestar* al otro: *En ningún momento corresponde mandar un audio. Estás obligando al receptor del mensaje a tener que dejar lo que está haciendo, tomarse una pausa, buscar los auriculares y procesar el mensaje a la velocidad del audio, que suele ser bastante más lento que leer lo mismo en formato*

escrito; Si necesito enviar un audio y creo que puede no corresponder, aclaro por escrito previamente lo que considere oportuno; Cuando pedis turnos, o cuando el destinatario por alguna circunstancia no puede escuchar audios por qué está trabajando por ejemplo.

Como contracara, también se evidencia que el audio tiene un grado de exposición mayor para quien lo produce: Cuando no quieres que otra persona que esta presente se entere de lo que hablas 😊; Cuando se trata de un tema confidencial. Para evitar riesgos de escape.

La cantidad de personas que participan en los chats resalta como variable que ha de considerarse en la selección modal. Chats muy numerosos, porque recarga a todos y cuando hay información precisa que es importante que que quede registrada (a veces mando audio y reitero por escrito la información precisa); En grupos (muy impersonal). en chats formales, impersonales, cuando sé que la otra persona (aunque sea de mucha confianza) no está en situación de escuchar audios, en chats grupales formales (como de familias de escuelas).

De hecho, en muchas respuestas al THS vinculadas a consideraciones sobre los audios en términos de adecuación, se señala que un audio de más de un minuto es un audio largo. Ante la pregunta de si hizo o le hicieron algún comentario sobre los audios, una mujer joven responde: Siempre, si es un audio de más de 1:30 minuto y tengo cercanía suficiente con la persona para decirle que esto no se hace. En la misma línea, otra joven dice: pero si [sí] me molesta que no haya capacidad de síntesis en los audios. Más de 3 minutos amerita una llamada.

4. Conclusiones

Hasta aquí hemos presentado un análisis sobre las motivaciones subyacentes a los cambios de modo y la manera en que es percibido el audio por hablantes de Argentina y España. Esta mirada atiende a la complejidad inherente a todos los intercambios comunicativos en los que las sucesivas elecciones, en todos los niveles posibles (Verschueren, 2002, p. 110). Dentro del repertorio de estrategias pragmáticas disponibles, en una determinada situación comunicativa, los hablantes elegirán entre aquellas que se ajusten más a su propósito comunicativo: tensionando incluso, en la instancia de recepción, el modo elegido. Así, tenemos personas que leen los mensajes de audio transcritos por un sistema de IA, por ejemplo.

Los cambios de modo se dan en el marco de una serie de elecciones (simultáneas y consecutivas) que los interlocutores realizan en el desarrollo de una conversación. Desde la selección de la aplicación en la que el primer enunciado es enviado a los aspectos más micro de las elecciones lingüísticas, todas las instancias son significativas en el proceso de generación de significado.

Nuestros resultados evidencian que la elección del audio se inscribe en una compleja trama con otros modos (el escrito y el visual), así como en relación con otros medios, extendiendo este concepto no solo a los dispositivos sino a otras aplicaciones o funcionalidades del dispositivo móvil, como la llamada. En la línea de lo que se observó en el análisis de los memes, el audio es ambivalente en tanto que en la fase de producción permite ser más precisos con el tono del mensaje y, a la vez, desarrollar mejor la intervención por su extensión; mientras que, en la fase de recepción, la extensión es una dimensión de relevancia. Los comentarios sobre

la extensión llevan al audio a otros formatos (como el podcast, tal como aparece en nuestro corpus y en los THS) o las llamadas.

La extensión del audio se transforma en una variable relevante: no es solo la elección del modo sino su duración la que, en la instancia de recepción, puede transformar una intervención en despreferida. Esto habilita la justificación (contextual, generalmente) que suele encontrarse al inicio de los audios (como se vio en el ejemplo 1 y 2). En THS implementados en investigaciones previas (Cantamutto, 2018), las razones para elegir el SMS se planteaban en relación con la llamada telefónica y el WhatsApp. Los hablantes identificaban que hay situaciones que requieren llamar como, por ejemplo, saludos de cumpleaños o felicitaciones, disculpas y agradecimientos entre familiares o amigos, mientras que, para realizar pedidos, que no revisten el carácter de urgente, los SMS parecían oportunos en la mayoría de las diadas. Así, los SMS eran una alternativa menos impositiva sobre el interlocutor. A esto se suman las variables del tipo de vínculo entre los interlocutores, dominio de uso y grado de formalidad del mensaje.

La aparición de nuevas aplicaciones en el repertorio polimedia desestabiliza las prácticas comunicativas en las que estas elecciones se insertan. Entre los SMS y los WhatsApp, el primero resultaba más formal (y ahora ha quedado a una función totalmente secundaria y casi de *SPAM*, en el que se reciben publicidades) y el segundo quedaba reducido al ámbito de las relaciones sociales. En la actualidad, las interacciones sincrónicas se han reducido a los intercambios por WhatsApp, y, por consiguiente, las llamadas se han reestructurado en intercambios de audios -en el mejor de los casos- en los que se *conversa* a partir de la yuxtaposición de turnos completos, sin las interrupciones típicas de las conversaciones en presencia, en las que la toma de turno puede ser compartida.

A partir de esta primera aproximación al fenómeno del cambio de modo en la interacción digital escrita en español, se abren nuevas líneas de investigación. Por un lado, el estudio de la recepción del cambio de modo sigue siendo un terreno para explorar, dado el carácter marcado del empleo de audios en determinados ámbitos comunicativos o el consenso en la extensión mínima deseable. Por otro, la exploración desde una perspectiva etnográfica sobre las instancias de producción y recepción de los mensajes permitirá evaluar las estrategias de recepción (en un sentido particular, si se escucha en altavoz, apoyado en el oído, con auricular, en velocidad normal, etc.) y los comentarios orales que puedan emerger ante estas situaciones. Este aspecto requiere de mayor atención: si bien el audio no se puede editar, si es posible “eliminarlo” antes de enviarlo y que el interlocutor no lo sepa.

Considerado este aspecto, que provoca el vacío de orden metodológico señalado anteriormente, el estudio también presenta otras limitaciones. Sería deseable, en futuras implementaciones del test, su realización a través de la misma aplicación de WhatsApp. En estudios previos, hemos comprobado la utilidad de la aplicación de entrevistas y encuestas en el estudio de las representaciones sobre diferentes aspectos de la comunicación digital. Sin embargo, desde una perspectiva pragmática y del análisis del discurso digital, consideramos que los test de hábitos sociales son una herramienta potente para que el investigador recupere usos y percepciones de manera simultánea cuando se los aplica por la misma aplicación sobre la que se está indagando. Si bien en este caso se realizó a través de un

formulario, quienes respondieron mediante sus teléfonos móviles pudieron incluir emojis en sus respuestas.

Referencias

- Alcántara Plá, M. (2014). Las unidades discursivas en los mensajes instantáneos de wasap. *Estudios de Lingüística del Español*, 35, pp. 223–242. <https://doi.org/10.36950/elies.2014.35.8716>
- Androutsopoulos, J. (2014). Computer-mediated Communication and Linguistic Landscapes. En Holmes, J. y Hazen, K. (Eds.) *Research Methods in Sociolinguistics: A Practical Guide* (pp. 75–90). New York, NY: John Wiley & Sons. <https://doi.org/10.1002/9781394260867.ch5>
- Androutsopoulos, J. (2021). Polymedia in interaction. *Pragmatics and Society*, 12(5), pp. 707–724. <https://doi.org/10.1075/ps.21069.int>
- Benítez-Eyzaguirre, L., de-Marcos, C., y Acosta-Calderón, L. (2023). La hiperconversación, el diálogo aumentado del mundo móvil a través de WhatsApp. *Revista Mediterránea de Comunicación: Mediterranean Journal of Communication*, 14(1), pp. 279–294. <https://doi.org/10.14198/MEDCOM.23272>
- Blommaert, J. (2021). Comprender la sociedad a través del lenguaje: una nueva mirada sobre los grupos sociales y la integración (Trad. L. Eisner, trabajo original publicado en 2017). *Enunciación*, 26, pp. 37–54. Doi: <https://doi.org/10.14483/22486798.16908>
- Blommaert, J., y Backus, A. (2013). Superdiverse repertoires and the individual. En de Saint-Georges, I. y Weber, J. (Eds.), *Multilingualism and multimodality: current challenges for educational studies* (pp. 11–32). Rotterdam: Sense. https://doi.org/10.1007/978-94-6209-266-2_2
- Bravo, D. (2009). Pragmática, Sociopragmática y Pragmática sociocultural del discurso de la cortesía. Una introducción. En Bravo, D., Hernández Flores, N. y Cordisco, A. (Eds.), *Aportes pragmáticos, sociopragmáticos y socioculturales a los estudios de la cortesía en español* (pp. 31–68). Buenos Aires: Dunken.
- Bravo, D. (2020). Pragmática sociocultural para el análisis de los aspectos sociales del discurso. En Escandell-Vidal, M.ª V., Amenós Pons, J. y Ahern, A. K. (Eds.), *Pragmática* (pp. 481–497). Madrid: Akal.
- Bravo, D. y Briz, A. (2004). *Pragmática sociocultural: estudios sobre el discurso de cortesía en español*. Barcelona: Ariel.
- Cantamutto, L. (2018). Estrategias pragmáticas de la comunicación por SMS en español bonaerense. Tesis doctoral inédita. Universidad Nacional del Sur, Bahía Blanca.
- Cantamutto, L. y Vela Delfa, C. (2019). Emojis frecuentes en las interacciones por whatsapp: Estudio comparativo entre dos variedades de español (Argentina y España). *Círculo de lingüística aplicada a la comunicación*, 77, pp. 171–186. <https://doi.org/10.5209/CLAC.63282>
- Cantamutto, L. M. (2023). Estilo digital, lenguaje juvenil y gestión de vínculos: Del lenguaje SMS al modo sticker. *Revista Argentina de Investigación Educativa*, 3(5), pp. 199–225.
- Emery, R. (2018). WhatsApp voice messaging as an emergent digital practice: a multi-method analysis. *Cahiers du Centre de Linguistique et des Sciences du Langage*, 55, pp. 135–157. <https://doi.org/10.26034/la.cdclsl.2018.287>
- Goffman, E. (1979 [1971]). *Relaciones en público: microestudios del orden público*. Madrid: Alianza.
- Goffman, E. (1981). *Forms of talk*. Philadelphia, PA: University of Pennsylvania Press.
- Gumperz, J. J. y Hymes, D. (Eds.) (1972), *Directions in sociolinguistics: The ethnography of communication*. Nueva York: Rinehart and Winston
- Gumperz, J. J. (1982). *Discourse Strategies*. Cambridge: Cambridge University Press.

<https://doi.org/10.1017/CBO9780511611834>

Hernández-Flores, N. (2002). Los tests de hábitos sociales y su uso en el estudio de la cortesía: una introducción. En Bravo, D. (Ed.), *Actas del Primer coloquio del programa EDICE "La perspectiva no etnocentrista de la cortesía: identidad sociocultural de las comunidades hispanohablantes"* (pp. 186–197). Escocolmo: Universidad de Estocolmo/Programa EDICE.

Herring, S. C. y Ge-Stadnyk, J. (2024). Emoji and illocutionarity. Structures in Discourse. En Gill, M., Malmvirta, A. y Warvik, B. (Eds.) *Interaction, adaptability, and pragmatic functions* (pp. 124-155). Amsterdam: John Benjamins.

<https://doi.org/10.1075/pbns.345.07her>

Holmes, H. y K. Hazen (eds.) *Research Methods in Sociolinguistics: A Practical Guide* (pp. 75–90). New York, NY: John Wiley & Sons.

Jewitt, C. (2013). Multimodal methods for researching digital technologies. En Price, S., Jewitt, C. y Brown, B. (Eds.), *The Sage handbook of digital technology research* (pp. 250–265). Los Angeles: Sage. <https://doi.org/10.4135/9781446282229.n18>

König, K. (2024). Transmodal messenger interaction—Analysing the sequentiality of text and audio postings in WhatsApp chats. *Discourse, Context & Media*, 62. Doi:

<https://doi.org/10.1016/j.dcm.2024.100818>

Konrad, A., Herring, S. C., y Choi, D. (2020). Sticker and emoji use in Facebook Messenger: Implications for graphicon change. *Journal of Computer-Mediated Communication*, 25(3), pp. 217–235. <https://doi.org/10.1093/jcmc/zmaa003>

Kress, G. (2003). *Literacy in the new media age*. London: Routledge.

<https://doi.org/10.4324/9780203299234-8>

Sampietro, A. (2019). Cómo hacer palabras con emojis: sustitución y enfatización visual de vocablos en WhatsApp. *REDD: Revista de estudios del discurso digital*, 2, pp. 1–33.

<https://doi.org/10.24197/redd.2.2019.1-33>

Sampietro, A., y König, K. (2024). The medium is accountable: Metacommunication and media ideologies about voice messages in WhatsApp chats. *Discourse & Communication*, 18(1), pp. 51–71. <https://doi.org/10.1177/17504813231187109>

Scolari, C. A. (2004). *Hacer clic. Hacia una sociosemiótica de las interacciones digitales*. Barcelona: Gedisa.

Statista (2024). *Most popular global mobile messenger apps as of April 2024, based on number of monthly active users*. Recuperado de:

<https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/>.

Statista (2023). *Share of mobile users worldwide who prefer text or audio messages*.

Recuperado de: <https://www.statista.com/statistics/1459203/mobile-users-preference-text-or-audio-messages/#:~:text=According%20to%20a%20survey%20of%20global%20consumers%20conducted,in%2010%20respondents%20worldwide%20reported%20preferring%20audio%20messages.>

Tannen, D. (1996). *Género y discurso*. Barcelona: Paidós.

Thurlow, C. (2017). Enmarcando el lenguaje Enmarcando el lenguaje de los nuevos medios En Giammatteo, M, Gubitosi, P. y Parini, A. (Eds), *El español en la Red: usos y géneros de la comunicación mediada por computadora* (pp. 27–37). Madrid / Frankfurt am Main: Iberoamericana / Vervuert. <https://doi.org/10.31819/9783954877386-002>

Van Leeuwen, T. (2015). Multimodality. En: Tanner, D., Hamilton, H. E., y Schiffrin, D. (Ed.), *The Handbook of Discourse Analysis* (pp. 447-465). Chichester, West Sussex: Wiley-Blackwell. <https://doi.org/10.1002/9781118584194.ch21>

Vela Delfa, C. (2007). *El correo electrónico: el nacimiento de un nuevo género*. Tesis Doctoral inédita. Universidad Complutense de Madrid.

Vela Delfa, C. (2020). Una aproximación semiótica al estudio de los emojis. *Círculo de*

lingüística aplicada a la comunicación, 84, pp. 153–165.

<https://doi.org/10.5209/clac.69723>

Vela, C. y Cantamutto, L. (2021). *Los emojis en la interacción digital escrita*. Madrid: Arco/Libros.

Verschueren, J. (2002). *Para entender la pragmática*. Madrid: Gredos.