

# El Corpus Oral de Madrid (CORMA): materiales para el estudio (socio)lingüístico del español coloquial actual

Renata Enghels (Gent)  
Fien De Latte (Gent)  
Linde Roels (Gent)

**Summary:** CORMA is a corpus of peninsular Spanish including spontaneous conversations recorded in Madrid between 2016 and 2019. The corpus was compiled in order to remedy the scarce documentation of 21<sup>st</sup> century colloquial Spanish. Indeed, a short overview of the corpora of conversational Spanish shows that there is a sharp contrast between the increasing interest in linguistic phenomena representative of this variety and the availability of suitable empirical data. This paper discusses the different steps that were taken to build the dataset, and presents the main characteristics of the final CORMA corpus. It also includes an illustration of how the data can be successfully exploited through a pilot study of the rise and spread of the intensification marker *mazo (de)* in young speak since the beginning of the 21<sup>st</sup> century.

**Keywords:** spoken corpus, Peninsular Spanish, CORMA, intensification, *mazo (de)* ■

Received: 17-09-2019 · Accepted: 04-08-2020

## ■ 1 Introducción

En las últimas décadas ha crecido exponencialmente el interés lingüístico por los rasgos inherentes al lenguaje hablado espontáneo y coloquial.<sup>1</sup> Basta introducir palabras clave como ‘lenguaje oral’ o ‘spoken language’ en el banco de datos MLA International Bibliography (<<https://search.proquest.com>>) para observar que el punto culminante de esta ‘gira ora-

---

1 Siguiendo a Cano Aguilar (2002: 279) entendemos por ‘habla coloquial’: ‘la lengua hablada en situaciones de diálogo (y, en especial, en la conversación informal, espontánea y no planificada), con inmediatez comunicativa, y con todos los rasgos ‘coloquializadores’ (proximidad vivencial, igualdad, finalidad no solo comunicativa sino también de mantener el vínculo social, etc.).’



lista' (López Serena y Borreguero Zuloaga, 2010: 415) se sitúa al inicio del siglo actual. Está claro que se enmarca en el desarrollo general de las áreas de la sociolingüística, la pragmática y el análisis del discurso. Como consecuencia no dejan de publicarse estudios sobre el español coloquial, que se dedican a todos los niveles lingüísticos: la fonética y rasgos suprasegmentales (p.ej. el análisis prosódico de algunos marcadores discursivos por Cabedo Nebot [2013]), la morfología (p.ej. el estudio de Sanmartín Saéz [1999] sobre los sufijos apreciativos en la conversación coloquial), y la sintaxis (p.ej. la descripción detallada de la sintaxis del español coloquial por Narbona Jiménez [2015]). Se destaca también la cantidad de estudios dedicados a los aspectos pragmáticos e interaccionales de la conversación oral, como los marcadores discursivos y pragmáticos (p.ej. la antología de sus características principales por Loureda Lamas y Acín Villa [2010]) y otras unidades parentéticas (p.ej. Schneider, 2007), las formas de tratamiento y la cortesía (p.ej. Blas Arroyo, 1994), estrategias de intensificación (Enghels y Roels, 2017) y atenuación (p.ej. Albelda et al., 2014), etc.

Ahora bien, este continuo desarrollo del campo de investigación contrasta fuertemente con la escasez de datos primarios, indispensables para describir fielmente la realidad oral. Briz (2012) cuenta no menos de 63 corpus (o muestras) orales del español, tanto macrocorpus como microcorpus, que documentan diferentes géneros discursivos orales, dedicados a lenguajes específicos o no, y que están grabados en regiones diferentes. Sin embargo, afirma claramente que “[...] hay muy pocos corpus de conversaciones coloquiales. [...] si bien se ha avanzado mucho, es preciso seguir elaborando corpus orales por el bien de los análisis. [...] Dentro de estos corpus orales, sin duda, se observa un déficit importante de materiales de conversaciones, tanto formales como coloquiales” (Briz, 2012: 123). Esta penuria de corpus representativos del español conversacional espontáneo, tal y como se produce en el siglo XXI, se observa también en el *Panorama de los corpus y textos del español peninsular contemporáneo* (Enghels et al., 2015).

Según nuestro conocimiento, desde entonces, pocos corpus se han añadido a la lista establecida. Es imprescindible mencionar, sin embargo, el proyecto ESLORA (Barcala et al., 2018), un corpus para el estudio del español oral grabado en Galicia entre los años 2007 y 2015 (<<http://eslora.usc.es/>>), cuya digitalización servirá de ejemplo para la publicación de CORMA. Se grabó tanto en contextos conversacionales coloquiales como en contextos de la entrevista semidirigida. El subcorpus de las entrevistas forma parte del macroproyecto PRESEEA. También conviene insistir en el esfuerzo que ha realizado la Real Academia Espa-

ñola para poner a la disposición del lingüista más datos del español oral. En concreto, desde finales del año 2013 está en línea el *Corpus del Español del Siglo XXI* (CORPES XXI) que en la actualidad contiene aproximadamente cuatro millones de formas constituidas por textos orales. En una parte de los textos es posible consultar el sonido alineado correspondiente a la transcripción. El subcorpus oral se compone de transcripciones de programa de televisión y de radio (incluso tertulias), de *YouTube*, de entrevistas en medios de comunicación y de discursos pronunciados en varias instituciones como el Congreso y el Senado. Aun así, no parece contener datos conversacionales.

Así y todo, en las primeras décadas del siglo actual se han observado una serie de cambios socioculturales importantes en las sociedades europeas que, sin duda, han influenciado en las lenguas que usan. Sin entrar en detalles, uno de los hitos de la época actual es la irrefrenable expansión de los medios masivos de comunicación, y en particular los medios sociales (Jenkins, 2009). Las consecuencias lingüísticas se sitúan en, por lo menos, dos campos. Primero, por el fuerte desarrollo del correo electrónico como medio de comunicación principal, y aplicaciones como *Facebook*, *Twitter* e *Instagram*, los hablantes entran continuamente en interacción de manera más indirecta en vez de entrar en una conversación cara a cara. Por consiguiente, las reglas conversacionales ya no se aplican de la misma manera en estas interacciones ‘mediadas’, lo cual puede, a su vez, tener un impacto en las estrategias aplicadas durante una conversación directa. ¿Qué profesor se atrevería hoy día a afirmar que las interacciones que realiza con sus estudiantes no se han vuelto más informales? Segundo, por el avance de plataformas vía *streaming* como *YouTube* y *Netflix*, los hispanohablantes entran más que nunca en contacto con lenguas y culturas extranjeras, encabezadas por el mundo anglosajón. ¿Qué padre no habrá observado la presencia extendida de anglicismos en el lenguaje de sus hijos?<sup>2</sup> Sin embargo, tales cambios lingüísticos ‘recientes’, es decir originados en el siglo XXI, no pueden estudiarse sistemáticamente a menos que el investigador disponga de datos actualizados.

Para hacer frente a este problema de datos, los estudiosos del español coloquial recurren a fuentes alternativas como el habla observada en programas de televisión, y sobre todo telenovelas (p.ej. Matamala, 2008), o en los medios de comunicación (p.ej. Rueda et al., 2013). Primero, las series

---

2 Claro está que los anglicismos entraron en la lengua ya mucho antes pero en la época actual se nota un incremento considerable de su difusión y uso (García Palacios et al., 2016).

televisivas incluyen interacciones de carácter conversacional. Sin embargo, aunque tratan de ofrecer una imagen realista del habla espontánea, las conversaciones están premeditadas, y por lo tanto se sitúan más bien en una posición intermedia entre los géneros escritos y orales espontáneos (véase las discusiones sobre el uso de este tipo de corpus, p.ej. en Bednarek, 2019). Aun así, la comparación entre el habla conversacional espontánea y el habla televisiva se está convirtiendo en una nueva línea de investigación, que merece la pena ser desarrollada. Segundo, se observa que autores recurren cada vez más a datos recopilados en Twitter para analizar diferentes fenómenos lingüísticos. Entre las ventajas ofrecidas por este tipo de datos se destacan, entre otras cosas, su carácter actualizado y la libertad y espontaneidad con que hablan los usuarios. Sin embargo, aquí también se plantea de manera cada vez más acuciante la pregunta si esta variante puede considerarse como una imitación fiable del lenguaje hablado. Hay los que dicen que difiere bastante de la conversación oral y que manifiesta ante todo características de la lengua escrita: existe entre los hablantes una distancia locativo-temporal y los intercambios se suceden linealmente de forma asincrónica a través de un código gráfico (Padilla Herrada, 2016). En su estudio sobre el marcador pragmático *en plan*, De Smet y Enghels (2020) concluyen que los datos de *Twitter* parecen adecuados para el análisis de fenómenos ligados al alto grado de expresividad del lenguaje oral (como la intensificación o la atenuación), pero menos para el estudio de fenómenos lingüísticos condicionados por la interacción entre los hablantes y la alternancia de turnos de habla (como determinados usos de los marcadores pragmáticos y formas de tratamiento). Por lo tanto, es necesario acercarse a este tipo de datos con cautela.

El proyecto CORMA (*Corpus Oral de Madrid*) surgió para llenar esta laguna empírica y para documentar el español conversacional espontáneo tal y como se presenta actualmente en Madrid. El corpus se grabó entre 2016–2019, por el equipo de lingüística española de la Universidad de Gante en colaboración con la UNED,<sup>3</sup> y quiere ser un banco de datos variado incluyendo grabaciones de interacciones orales espontáneas de la vida cotidiana. Por lo tanto, es un corpus ‘directo’, recogido en su contexto natural, frente a los corpus orales ‘indirectos’ como CREA o CORPES XXI oral (Briz, 2012).

---

3 En la composición del corpus participaron hasta la fecha Renata Enghels, Fien De Latte, Linde Roels, Kris Helincks, Sanne Tanghe, Marie Comer y Elena Azofra.

Se ha elegido la variedad capitalina por su prestigio y subsecuente influencia sobre el lenguaje de otras zonas de la Península. Más en concreto, el prestigio del que suele gozar la capital ('the capital's prestige', Jørgensen y Drange, 2012) explica por qué muchas innovaciones lingüísticas emergen en las metrópolis, de donde tienden a extenderse a zonas situadas en la periferia. Por consiguiente, es de suponer que la variedad madrileña presenta los modos lingüísticos más innovadores y de mayor influencia en la norma lingüística en contraste con otras variantes más periféricas.

El presente artículo persigue dos objetivos principales que se desarrollan en sendas partes. Primero, la mayor parte del estudio se dedica a dar a conocer el propio corpus e identificar sus características principales (Sección 2). En segundo lugar, quiere ilustrar las posibilidades de explotación del corpus a través del estudio de caso concreto del intensificador *mazo* (de) (Sección 3).

## ■ 2 Presentación del corpus

En pocas palabras, el corpus CORMA se define por los rasgos siguientes: (a) representa un corpus de interacciones lingüísticas en marcos y actividades cotidianos; (b) se caracteriza por un alto grado de variación situacional y sociolingüística, con un número representativo de participantes de ambos sexos, diferentes generaciones y diferentes niveles socio-culturales; (c) es un corpus de habla oral y conversacional, es decir los participantes interactúan de manera inmediata, cara a cara, están involucrados en diálogos, ante todo cooperativos, con un alto grado de dinamicidad (con muchas alternancias de turno entre los participantes) (Briz, 1995), y (d) es un corpus de habla espontánea, obtenida principalmente mediante la técnica ordinaria. Este apartado tiene como objetivo principal aclarar las decisiones metodológicas concerniente la colección de datos, su transcripción y la elaboración del corpus.

### ■ 2.1 Colección de datos

La recopilación de datos se ha realizado durante tres campañas de trabajo de campo llevadas a cabo entre 2016 y 2019. La primera campaña, que tuvo lugar en abril 2016, permitió recopilar 43 horas de grabación, correspondientes a 320.401 palabras transcritas, que representan el habla de 390

hablantes madrileños.<sup>4</sup> La segunda campaña, realizada en enero 2019, tuvo como principal objetivo ampliar los datos de la segunda generación, sub-representada en el muestreo obtenido en la primera campaña. Esta segunda campaña de trabajo de campo proporcionó ocho horas de grabación, correspondientes a más o menos 88.117 palabras transcritas, e incluye el habla de 84 hablantes adolescentes. Finalmente, en octubre 2019 fue realizada la tercera y última campaña con el fin de equilibrar la composición sociolingüística del corpus, relacionada principalmente con el nivel socio-cultural de los hablantes de la segunda generación. Fueron recopiladas aproximadamente seis horas de grabación, correspondientes a aproximadamente 66.000 palabras, y representan el habla de 55 hablantes.

Con el fin de facilitar la búsqueda de participantes juveniles, colaboramos con distintos institutos en diferentes barrios en Madrid. Visitamos los institutos, presentamos el proyecto en las aulas (principalmente a grupos de 3º de ESO, 4º de ESO, 1º de Bachillerato y 2º de Bachillerato), y los alumnos se responsabilizaban de grabar conversaciones con amigos y/o familiares de la misma edad. En líneas generales, nos concentramos en grabar conversaciones mantenidas en distintas situaciones, que se pueden subdividir en cuatro ámbitos interactivos generales: (a) interacciones entre familiares en la esfera privada, (b) conversaciones entre amigos en la esfera pública o privada, (c) interacciones en contextos comerciales y (d) interacciones regulares entre personas que se (des)conocen en la esfera pública o privada. A lo largo de este apartado, se presentarán en más detalle las condiciones de recopilación, al igual que las características de las conversaciones grabadas.

Las conversaciones se grabaron mediante dos tipos de grabadoras. Primero, en la mayoría de los casos las grabaciones fueron hechas mediante el uso de una grabadora dada prestada por parte de las investigadoras a los participantes. Se disponían de diferentes tipos de grabadoras: Tascam DR-40, Philips Voice Recorder DVT 2000 y ZOOM H1 Handy Portable Digital Recorder. Los participantes también tuvieron la oportunidad de grabar por medio de su propio móvil.<sup>5</sup> Todos los dispositivos disponen de un micrófono integrado y permiten obtener audios de buena calidad. En

---

4 Tres participantes tienen origen extranjero (latinoamericano, marroquí y francés) pero viven desde mucho tiempo en la capital española.

5 El uso de un dispositivo que forma parte esencial de la vida cotidiana moderna facilita el proceso de grabación, además de reducir posibles efectos que afecten a la espontaneidad de la interacción por causa de la presencia de la grabadora y la subsecuente consciencia de ser grabado.

contextos con bastante ruido de trasfondo, el responsable de la grabación también llevaba un micrófono externo para garantizar la calidad e inteligibilidad de las grabaciones (por ejemplo los niños en el parque o el camarero en el bar). Por razones de compatibilidad con los programas de elaboración del corpus (cf. Sección 2.4), los audios se han guardado en formato .wav.

En cuanto al procedimiento de grabación, se pueden distinguir dos métodos distintos: por un lado la grabación (*semi*)*secreta*, en que los participantes no son conscientes del hecho de que sean grabados (y la grabadora no está visible). Por otro, la grabación *ordinaria*, en que los participantes sí son conscientes de ser grabados. En este caso, la grabadora está visible (Briz y Val.Es.Co, 2002: 16). Las grabaciones fueron realizadas a cargo de los *reclutas*, los participantes responsables de las grabaciones. Los reclutas habían sido ‘captados’ siempre mediante una persona de contacto en Madrid, quien era un amigo o conocido mutuo de uno de los miembros del equipo académico y el propio recluta. En principio, solo el recluta estaba informado de antemano sobre la grabación y tenía control sobre ella: decide sobre la duración de la grabación, los participantes, la actividad, al igual que sobre el tema de la conversación. De todos modos, por varias razones prácticas, a menudo resultó imposible realizar las grabaciones en secreto. Así, la presencia de un micrófono externo o de la investigadora, la locuacidad de algunos participantes (p.ej. la peluquera) o la presentación del proyecto al grupo entero de amigos/compañeros de clase/colegas/familiares, son todos factores que explican por qué en muchas ocasiones, todos los participantes estaban informados al inicio de las grabaciones.

Sin embargo, la grabación era secreta para la mayoría de los clientes en los contextos comerciales, más precisamente para los del bar, la tienda de ropa, la floristería y la farmacia, puesto que solo el propietario o responsable del negocio estaba informado sobre las grabaciones. En los contextos de familia, amigos y conocidos, en cambio, la mayoría de los participantes solían ser informados de antemano sobre la grabación, salvo los que intervinieron muy brevemente, de manera que estos contextos se caracterizan mayoritariamente por la grabación ordinaria (cf. Sección 2.2).

Claro está que la grabación secreta garantiza la obtención de datos auténticos y evita eventuales ajustes artificiales del lenguaje o posibles efectos negativos causados por el ‘miedo escénico’. No obstante, esta técnica complica el acceso a información sociológica de los participantes. Así, para varios participantes no está disponible información acerca de su edad,

vivienda, profesión o nivel educativo. Por tanto, sobre todo la especificación del nivel sociocultural resultó problemática para varios informantes.

Al mismo tiempo, cabe señalar que la técnica de la grabación ordinaria no anula la naturalidad de las conversaciones, la cual queda reflejada en el hecho de que los participantes ajusten raramente sus actividades cotidianas o el tema de conversación, a menudo íntimo, a pesar de ser conscientes de la grabación. Además, varios hablantes expresan su sorpresa cuando de repente se acuerdan de la presencia de la grabadora, lo que indica que no le prestan atención. Todos estos aspectos contribuyen evidentemente a la naturalidad, espontaneidad y autenticidad de las grabaciones.

La mayoría de las grabaciones han sido efectuadas sin la presencia de la investigadora, especialmente las entre amigos, colegas y conocidos. En estos casos, los reclutas le informaron a la investigadora, después de la grabación, sobre el número de participantes, sus edades, su papel en la conversación, al igual que sobre los aspectos situacionales en que se desarrolló la interacción. La ausencia de la investigadora contribuyó a la naturalidad de las interacciones, evitando posibles efectos relacionados con la ‘paradoja del observador’ (Labov, 1972). Solo en algunos casos específicos, la investigadora estaba presente como observador no participante, más precisamente en la mayoría de los contextos comerciales (en el bar, la tienda de ropa, la floristería, la farmacia, la peluquería, la tienda de muebles), donde estaba sentada en el fondo para tomar notas de campo. Su presencia permitió inquirir información sobre el sexo y la edad de los participantes y observar más detalladamente el curso de la interacción. Por último, en cinco conversaciones entre amigos, en dos entre familiares y en dos entre conocidos, la investigadora estaba presente como participante porque era conocida de uno de los participantes.

Evidentemente, con el fin de cumplir con las leyes de privacidad europeas, los participantes responsables tuvieron que dar su consentimiento informado explícito. El documento de consentimiento especifica varias informaciones: los objetivos de la investigación, formulados en términos generales para no afectar la espontaneidad de las conversaciones (‘investigación sobre la interacción social entre hablantes españoles’), el alcance de los datos, el tratamiento confidencial y anónimo de los mismos, al igual que el derecho de cada participante a retirar su participación en cualquier momento. En el caso de que los participantes eran menores de edad, también se requirió el consentimiento informado de los padres. Cabe señalar que ninguno de los participantes estaba informado sobre los fines específi-



cos de la composición del corpus, con el fin de salvaguardar la autenticidad de los datos.

En suma, las grabaciones fueron recopiladas de manera no guiada y flexible, lo que se refleja en el autocontrol sobre la grabación y las circunstancias de la misma por parte del recluta. No obstante, este procedimiento implica algunas restricciones en cuanto a las características del corpus, a saber, las grabaciones presentan bastante variación en cuanto a su duración, el número de participantes y la calidad del audio (con más o menos ruido de trasfondo). Luego, la distribución sociolingüística de los participantes no es igualitaria en los distintos contextos de grabación.

## ■ 2.2 Descripción de los datos recopilados

Como nuestro objetivo era recopilar un muestreo representativo del lenguaje madrileño coloquial, aspiramos a componer un conjunto de interacciones diarias o regulares, en las cuales los participantes realizan actividades cotidianas, en su entorno natural. Las conversaciones grabadas fueron mantenidas en la esfera privada o pública, por participantes que se conocen personalmente (p.ej. colega de trabajo, conocido), íntimamente (p.ej. amigo, familiar) o por participantes que se desconocen (p.ej. dependiente y cliente). Evidentemente, se evitó grabar conversaciones entabladas en contextos formales, porque estas van sujetas a normas y restricciones institucionales que frenan la espontaneidad y naturalidad de las interacciones. En otros términos, el objetivo primario de la creación del corpus consistió en recopilar *conversaciones coloquiales*. De todos modos, la definición de este género discursivo y su distinción con otros (sub)géneros no está exenta de retos. En primer lugar, en palabras de Briz (1995: 27–30; 2010: 25, 32), una interacción se considera *conversacional* si es oral, inmediata, dialógica, cooperativa, dinámica, no predeterminada, y con mucha retroalimentación. Si uno o más de estos rasgos faltan o están presentes en menor medida, se trata de una conversación periférica.

En cuanto al parámetro de *coloquialidad*, la mayoría de las conversaciones incluidas en el corpus se caracterizan por la igualdad entre los participantes, el ámbito discursivo familiar, la temática no especializada, la ausencia de planificación, y la finalidad interpersonal del acto comunicativo, de manera que se consideran conversaciones coloquiales prototípicas (Briz, 1998: 41). De todas formas, cabe señalar que estos rasgos ‘coloquializadores’ se presentan en mayor o menor medida, de modo que conviene distinguir entre diferentes grados de coloquialidad y por consiguiente, entre con-

versiones coloquiales prototípicas y periféricas. A modo de ilustración, una conversación entre jóvenes amigos en el parque se considera más coloquial que una transacción entre un farmacéutico y un cliente suyo.

Efectivamente, el corpus incluye algunas interacciones menos informales por su carácter fundamentalmente transaccional, las cuales resultan por tanto menos dinámicas, puesto que siguen un sistema de turnos más predeterminado. Conviene considerarlas conversaciones coloquiales periféricas. Se trata por ejemplo de una conversación entre un distribuidor de muebles y los propietarios de un negocio en torno a la instalación de una cocina. El propósito transaccional y la alternancia de turnos relativamente fija caracterizan igualmente a la mayoría de las conversaciones mantenidas en los contextos de ‘atención al cliente’, por lo que estas también se describen como menos coloquiales. Sin embargo, decidimos incluir todas estas interacciones porque cumplen con todos los demás requisitos de recopilación de datos, incluso la variación socio-situacional y el carácter cotidiano, espontáneo y sumamente conversacional.

Además, tres conversaciones se han mantenido por teléfono,<sup>6</sup> las cuales resultan por tanto menos inmediatas en comparación con las interacciones cara a cara. Aunque no se registraron los turnos del interlocutor, se ha decidido mantenerlas porque siguen siendo relevantes: el hablante grabado produjo suficientes datos lingüísticos, de manera que el tema de conversación, el desarrollo de la misma, al igual que la relación entre los interlocutores son deducibles del material disponible. Además, consideramos que la distancia espacial no compromete el carácter conversacional de la grabación.

Es importante subrayar que, por motivos de representatividad, mucha atención fue dedicada a los metadatos de las conversaciones, es decir, a cualquier tipo de información disponible acerca de las condiciones de la grabación y de las características sociológicas de los participantes. En concreto, registramos sistemáticamente información sobre las siguientes variables situacionales y sociológicas:

- 1) información sobre la grabación:
  - a) fecha
  - b) duración
  - c) lugar de la grabación: barrio y entorno espacial (p.ej. ‘Embajadores’, ‘en casa’)
  - d) investigadora responsable

---

6 Se trata de una conversación entre dos colegas, una entre un abogado y un cliente suyo, y otra entre el propietario de un negocio y un albañil.

- 2) información sobre la conversación:
  - a) tema de la conversación (p.ej. viajar, escuela, trabajo, familia)
  - b) propósito conversacional: interpersonal – transaccional
- 3) información sobre la técnica de grabación:
  - a) papel del investigador: ausente – presente observador – presente participante
  - b) tipo de grabación: (semi)secreta – ordinaria
- 4) información sobre los participantes:
  - a) número de participantes
  - b) información sociológica de cada participante (si está disponible):
    - 1) sexo: masculino – femenino – na
    - 2) edad: generación 1 (0–11) – generación 2 (12–25) – generación 3 (26–55) – generación 4 (+55)<sup>7</sup>
    - 3) nivel educativo: primario – secundario – enseñanza superior
    - 4) profesión
    - 5) rol (basado en la relación con el responsable de la grabación)
    - 6) observaciones adicionales (p.ej. origen extranjero del hablante)
- 5) notas de campo (cualquier información suplementaria relevante para el curso y el análisis posterior de la conversación)

La recopilación resultante de la primera campaña de trabajo de campo puso en evidencia algunos desequilibrios en cuanto a la representación de los distintos grupos sociológicos comprendidos en el corpus. Más en concreto, la distribución reveló una llamativa sobrerrepresentación de la tercera (36,2%; n=141) y cuarta (33,1%; n=129) generación, mientras que la segunda generación quedó subrepresentada (11,8%; n=46).<sup>8</sup> Estas deficiencias se eliminaron gradualmente mediante las dos campañas de trabajo de campo posteriores: la segunda campaña se organizó con el fin de remediar el desequilibrio generacional, mientras que la última campaña tuvo

---

7 Estas categorías se basan en la clasificación sociolingüística tradicional de los grupos etarios (Labov, 1972; Moreno Fernández, 1996 y el corpus PRESEEA): 20–34, 35–54, y +55, con un grupo adicional de ‘niños’ (0–11). Como consecuencia, el segundo grupo, el de los jóvenes, se ha ajustado a los 12 años de edad, considerada como el inicio (psicológico y biológico) aproximado de la adolescencia (Eisenstein, 2005) y la edad de entrada en la enseñanza secundaria. El límite de 26 años para el tercer grupo (adultos: 26–55) fue elegido por ser la edad que significa aproximadamente la transición de la vida estudiantil a la vida laboral, la obtención de la independencia económica y la consolidación del crecimiento físico y psicosocial.

8 Estas proporciones desiguales se reflejan también en el número de palabras por generación: generación 2 = 41.844 palabras; generación 3 = 170.313 palabras; generación 4 = 85.499 palabras.

como principal objetivo remediar el predominio de los hablantes de nivel sociocultural (medio-)alto en el grupo de la segunda generación (cf. *supra*).

De todos modos, cabe señalar la dificultad de precisar de manera objetiva determinadas características, especialmente el nivel sociocultural de los participantes. Más en concreto, para determinar el nivel sociocultural de los participantes, hay que recurrir a diferentes parámetros económicos, sociales, (espacio-)físicos y personales, como el nivel de estudios, la profesión y la vivienda (Helincks, 2016). En el corpus CORMA, hemos recurrido básicamente al nivel socioeconómico del barrio vivencial para determinar el nivel sociocultural de los participantes.<sup>9</sup> El nivel sociocultural de los alumnos de instituto se basa en el barrio donde se encuentran los centros educativos colaboradores. De todos modos, como mencionado (cf. *supra* Sección 2.1), el carácter secreto y/o flexible de muchas grabaciones impide el acceso a información sociológica de muchos participantes (especialmente acerca de su barrio de vivienda, profesión, nivel de estudios), la cual resulta por tanto deficiente en varios casos.

Finalmente, siguiendo los ámbitos interaccionales de grabación, las conversaciones se pueden clasificar en una de las categorías siguientes: ‘familia’, ‘amigos’, ‘interacciones regulares’ y ‘atención al cliente’, y se caracterizan por los siguientes tipos de variación (basado en Helincks [2016] para el *Corpus de Conversaciones Espontáneas del Español en Chile – Iquique* [CCEEC-I]):

- 1) familia (11 conversaciones; 4h 20m 28s; aprox. 39.652 palabras)
  - diferentes familias
  - variación sociolingüística (en cuanto a la edad, el sexo y el nivel sociocultural)
  - variación relacional (en cuanto a los roles funcionales y la composición del grupo)
  - diferentes actividades (comiendo, jugando a las cartas, charlando, visitando a los padres)
- 2) amigos (56 conversaciones; 26h 51m 31s; aprox. 245.274 palabras):
  - diferentes grupos de amigos (compañeros de clase o de la universidad)

---

9 Fidel Oliván (2015) analiza las características sociológicas de cada distrito de la capital en correlación con el comportamiento electoral de los votantes. Las características tomadas en consideración en su análisis son las siguientes: renta media, tasa de desempleo, población extranjera, nivel de estudios, tasa de juventud y tasa de vejez. Con base en estos criterios, hemos elaborado nuestra propia clasificación de los distritos madreños según su nivel socioeconómico (bajo; medio-bajo; medio; medio-alto; alto).

- variación sociolingüística (en cuanto a la edad, el sexo y el nivel socio-cultural)
  - diferentes actividades (en el parque, durante el recreo, jugando a los videojuegos o a los juegos de mesa, almorzando en la facultad, haciendo tareas, cenando en un restaurante, caminando a casa, en el coche, durante un curso de punto y ganchillo, etc.)
- 3) interacciones regulares (9 conversaciones; 5h 28m 36s; aprox. 32.524 palabras):
- interacciones entre colegas durante el trabajo (albañiles, abogados) y encuentros entre conocidos (habitante de un inmueble y el portero, padres durante un partido de rugby de sus hijos, padres de compañeros de colegio, dueños de dos negocios colindantes)
  - diferentes relaciones
  - diferentes grados de formalidad
  - diferentes situaciones con diferentes actividades
- 4) atención al cliente (24 conversaciones; 22h 36m 30s; aprox. 157.156 palabras):
- diferentes situaciones comerciales con diferentes servicios (bar, farmacia, floristería, tienda de ropa, tienda de muebles, peluquería, salón de belleza)
  - variación sociolingüística de los participantes (con respecto al dueño o responsable de los comercios colaboradores).

### ■ 2.3 Transcripción y composición del corpus

La transcripción de las grabaciones fue realizada por hablantes nativos del español mediante el programa Praat (<[www.praat.org](http://www.praat.org)>).<sup>10</sup> Todas las transcripciones (en formato Textgrid) fueron controladas por miembros del equipo de lingüística hispánica. El programa Praat permite crear una capa de transcripción (“transcription tier”) por cada hablante participante y dividir su habla transcrita en diferentes ‘segmentos’ o unidades de habla, correspondientes a enunciados.<sup>11</sup> De esta manera, el programa permite

---

10 Colaboramos con estudiantes hispanohablantes que pasaron una estancia Erasmus en la Universidad de Gante, al igual que con estudiantes nativos de español de nuestra facultad. Todos estudiaban Lingüística (Aplicada) y fueron remunerados por su colaboración.

11 El *enunciado* se define como “el segmento de discurso que corresponde a una acción independiente, a una intención determinada del hablante. Es aislable en tanto posee fuerza ilocutiva propia y es identificable por la presencia de ciertas marcas lingüísticas prosódicas y semánticas” (Albelda et al., 2014: 39).

mostrar el habla simultánea de varios interlocutores en una sola ventana. Sin embargo, por causa del reconocimiento ocasionalmente problemático de las voces, la adscripción del texto al participante correcto constituyó un reto para los transcripores, sobre todo en el caso de conversaciones caracterizadas por un gran número de participantes con el mismo perfil sociológico (p.ej. una conversación entre cinco chicas adolescentes). En el caso de que el texto producido resulte ininteligible, por la deficiente calidad del audio o por el habla simultánea, se lo indica mediante el signo (()). En una fase posterior, todas las transcripciones fueron importadas en Partitur Editor, el primero de los tres subprogramas de Exmaralda (<[www.exmaralda.org](http://www.exmaralda.org)>), que permite la alineación automática de texto y audio por conversación grabada, al igual que la especificación de los distintos tipos de *tiers* por participante (de transcripción o de anotación).

Conviene señalar que las transcripciones son de carácter ortográfico, siguiendo las reglas de la ortografía normativa oficial (RAE, 2010), con la excepción de dos rasgos fonéticos típicos de la oralidad, a saber la aspiración de la *-s* implosiva (González Montero, 1993) y la elisión de consonantes o sílabas (Gómez Molina y Gómez Devís, 2010).<sup>12</sup> Por su carácter uniformador, la transcripción ortográfica genera dos importantes ventajas, simplificadoras del proceso de transcripción: por un lado, reduce la toma de decisiones por parte de los transcripores, que comprometería la uniformidad del corpus. Por otro, implica una representación intacta de las unidades léxicas y estructuras sintácticas, por lo que facilita la posterior búsqueda y recuperación de información en el corpus. Está claro que la homogeneidad y uniformidad del corpus constituyen un aspecto crucial para su utilización como fuente de datos, al igual que para el posterior desarrollo de aplicaciones (automáticas) de búsqueda, recuperación y etiquetación del texto.

La mayor desventaja de la transcripción ortográfica es que no permite la recuperación de aspectos fonéticos particulares. De todos modos, el acceso directo al audio, propio de los corpus orales alineados, hace posible futuros análisis prosódicos y acústicos de los datos. Igualmente, cabe señalar que dentro de los márgenes de la norma lingüística, se admiten ciertas alteraciones al nivel morfológico y léxico, con el fin de obtener una representa-

---

12 Evidentemente, se ha de tomar en cuenta esta representación fonética a la hora de realizar búsquedas en el corpus. Así por ejemplo, si se desea obtener todas las ocurrencias de la conjunción *es que*, hay que introducir las dos posibles grafías, de acuerdo con la pronunciación, a saber *es que* y *eh que*.

ción intacta de esas unidades. Así, variantes morfológicas divergentes del estándar se transcriben de manera fiel a su producción fónica, como por ejemplo casos de leísmo no aceptados por la norma.

Se creó igualmente, además de la línea de transcripción, para cada participante una línea de anotación, destinada a la especificación de información prosódica, extralingüística o paralingüística relevante para el análisis del registro oral. La lista siguiente presenta los signos utilizados para registrar determinados rasgos, de índole variada, manifestados en las conversaciones grabadas:

- línea de transcripción:
  - (( )) palabra o fragmento ininteligible
  - <vamos> cita reproducida
  - lo que p- truncamiento o (auto)interrupción de una palabra<sup>13</sup>
  - cansa'o elisión de consonantes o sílabas
  - pa'
  - eh que aspiración de /s/ implosiva
- línea de anotación:
  - (tos)
  - (risa)
  - (alargamiento vocálico)

Aunque se ha aspirado al máximo a la homogeneidad de las transcripciones mediante sesiones informativas y la elaboración de una guía con instrucciones para los transcripores, la revisión de las transcripciones ha puesto en evidencia algunas deficiencias al respecto. Por consiguiente, en una fase siguiente, se dedicará especial atención a la homogeneización de los datos mediante la implementación sistemática de las reglas de transcripción, antes de proceder a la informatización definitiva del corpus. Un aspecto que ha requerido particular atención es la transcripción de risas y de determinados elementos discursivos fáticos e interjectivos, esenciales para el desarrollo de la conversación cara a cara. Basándonos en los criterios de PRESEEA (Moreno Fernández, 2006) y ESLORA (Barcala et al., 2018), distinguimos los siguientes elementos: *ah*, *ay*, *aha*, *mmm*, *eeh*, *pfh*, *bah*, a los que se pueden añadir otros como *oh*, *jo*, *bof*, *bua*, *mbm*, tal y como se han presentado en el CORMA. Estos elementos se destacan entre los que más dudas y problemas de homogeneidad plantean en el proceso de transcripción.

---

13 Se distinguen dos situaciones: (a) repetición/truncamiento de partes de palabras: indicación mediante un guion (p.ej. *Di-di-dime.*); (b) repetición/truncamiento de palabras enteras: no se pone nada (p.ej. *El el el Juan.*).

Por último, cabe señalar que por motivos de protección de la privacidad de los participantes, todos los datos identificadores se han pseudonimizado, tal y como queda especificado en el documento de consentimiento informado, en que se garantiza el pseudonimato de los participantes. Para conseguir esto, se han pseudonimizado todos los nombres propios mencionados en las conversaciones, tanto los de los propios participantes como los de personas aludidas en las conversaciones, al igual que los topónimos que posibiliten la identificación de una persona.

Igualmente, a cada hablante le hemos asignado un código, con base en su rol conversacional, su edad, género, y orden de intervención. Los códigos también han sido utilizados para etiquetar las líneas de transcripción y anotación correspondientes a cada participante en la fase de transcripción. Así por ejemplo, el hablante con el código de identificación AM2F1 participa en el corpus en cuanto amiga (AM) (= rol), pertenece a la segunda generación (2) (= edad), y es una mujer (F) (= género). El 1 final significa que fue la primera en participar en la conversación. Este código, además de garantizar el carácter anónimo de los datos, nos permite particularizar a todos los participantes en el corpus sin comprometer su privacidad y distinguir entre las diferentes relaciones contraídas entre ellos. Igualmente, se ha creado un código de identificación para cada conversación: AM.GEN3.F.01 constituye la primera conversación (01) grabada entre amigas (AM, F) adultas (GEN3).

En una fase posterior, se ha creado un corpus digital en Coma, el segundo subprograma de Exmaralda, que permite vincular cada una de las transcripciones con el audio correspondiente para así componer el corpus, al igual que añadir los metadatos de la grabación y de los hablantes por conversación incluida, tal y como quedan documentados en las fichas técnicas. Al importar los audios y transcripciones en el programa, Coma genera una segmentación automática de cada transcripción, la cual posibilita la creación de una lista de palabras, al igual que la recuperación del total de palabras por contexto de grabación, por grabación, por hablante, y del corpus entero. Una vez construido en Coma, el corpus se puede consultar a través del último subprograma de Exmaralda, Exakt, que permite realizar búsquedas y análisis detallados, gracias a la especificación de los metadatos y la segmentación de las transcripciones realizadas en Coma. Más precisamente, el programa permite realizar búsquedas combinadas de distintas variables sociales con variables léxicas y gramaticales, siempre con acceso inmediato a los fragmentos de audios correspondientes a las concordancias de la consulta. Además, es posible recuperar información acerca de las fre-



cuencias de los resultados de búsqueda, copiar o descargar las concordancias en otro formato textual (.html o .xml), y ampliar el contexto izquierdo y derecho.

Hasta el momento, el corpus (.coma), todas las transcripciones (Textgrid) y todos los audios (.wav) se guardan en un disco de red de la Universidad de Gante, con acceso limitado. La consulta del corpus se realiza a través de Exakt. De todos modos, con el fin de facilitar el acceso al corpus sin restricciones, estamos trabajando en el diseño de una aplicación de consulta en línea, siguiendo el ejemplo de otros corpus orales de lengua española, como ESLORA (Barcala et al., 2018) y COLA (Jørgensen, 2013). Al igual que Exakt, esta aplicación permitirá realizar búsquedas y análisis específicos y combinados, con acceso al audio alineado, y hará posible la visualización del contexto izquierdo y derecho así como la descarga de los resultados en otro formato textual. También habrá acceso al texto completo de las transcripciones (también disponibles en pdf). Evidentemente, queda mucho trabajo pendiente, por realizar en un futuro (próximo), como la lematización y la anotación morfosintáctica del corpus, que favorecerá la extracción de información de lema y categorías morfosintácticas a partir del texto.

En lo que sigue, se ilustrarán la aptitud y relevancia del CORMA para el estudio de fenómenos lingüísticos típicos del español coloquial contemporáneo.

### ■ 3 Estudio de caso: microdiacronía reciente del intensificador *mazo (de)* en el lenguaje juvenil

A través de un estudio de caso concreto, el presente apartado quiere demostrar que el corpus CORMA abre perspectivas para la detección de cambios lingüísticos recientes.<sup>14</sup> Como el método de recopilación del habla de los jóvenes en el CORMA se equipara al diseño de COLAm (*Corpus Oral de Lenguaje Adolescente de Madrid*), se presentan las condiciones necesarias para llevar a cabo estudios de diacronía reciente. Más en concreto, la comparación del subcorpus del habla juvenil del CORMA, grabado entre 2016 y 2019, con el corpus COLAm, grabado entre 2003 y 2007, permite examinar la extensión y la velocidad del cambio lingüístico en el lenguaje

---

14 Seguimos la asunción de que el cambio lingüístico no siempre afecta el sistema a largo plazo, sino que también puede consistir en cambios lingüísticos recientes que generan una variación sutil en la lengua. Estos cambios actuales (*current changes*) se definen como “changes in the language that have taken place over relatively short spans of time, over decades rather than centuries” (Aarts et al., 2013: 1).

juvenil del siglo XXI. El tema de la intensificación constituye un fenómeno muy apto para este propósito dado el supuesto cambio rápido en el paradigma de los intensificadores, siendo elementos versátiles y expresivos (Tagliamonte, 2008; Palacios y Núñez, 2014). Por su uso recurrente en la lengua hablada, los intensificadores pierden rápidamente su fuerza simbólica y expresiva. Por consiguiente, los hablantes sienten la necesidad de renovar sus estrategias de intensificación, razón por la cual Stenström (2005: 206) define la intensificación como la categoría lingüística “with the most rapid turnover”. Encima, es sobre todo la generación de adolescentes que integra y acepta fácilmente nuevas formas en su paradigma de intensificadores, ya que suelen funcionar como indicios lingüísticos de la identidad joven (Lorenz, 2002). De hecho, el lenguaje juvenil se considera como el catalizador del cambio lingüístico, que prepara e indica nuevas tendencias que más tarde pueden difundirse y estandarizarse (Eckert, 1997; Zimmermann, 2002).

Así las cosas, por falta de corpus orales comparativos, hasta la fecha no fue evidente comprobar estos supuestos cambios en el paradigma de los intensificadores. Gracias a que el subcorpus del habla juvenil del CORMA se ha obtenido de manera paralela al diseño de COLAm, se pueden a partir de ahora efectuar estudios comparativos del habla innovadora juvenil. Para comprobar el ritmo con el cual los intensificadores se renuevan en su lenguaje, efectuaremos un estudio de caso, enfocándonos en el intensificador *mazø* (*de*). Este elemento fue identificado como intensificador propio de la conversación juvenil, puesto que casi no ocurre en la lengua estándar (Stenström, 2008; Martínez López, 2009). Su uso se acerca al funcionamiento de palabras de grado más comunes como *muy* (1) o *mucho* (2).

- (1) IR2F16 Qué va tía, estaba *mazø* triste diciendo <Es que se me ha, se me ha enfadado Miguel> (CORMA)
- (2) IR2F8 Sí, en plan, que yo sudo *mazø*, tía (CORMA)

### ■ 3.1 De nombre colectivo a intensificador multifuncional

*Mazø* (*de*) constituye una innovación de intensificación juvenil del inicio del siglo XXI (Stenström, 2005: 219), y es el resultado de un proceso de gramaticalización (Enghels y Roels, 2017). En su uso pleno y más particularmente en sus acepciones de “conjunto abundante de papeles u otras cosas atadas o unidas formando grupo” y “martillo grande”, el lexema *mazø* ya

incluye referencias a los conceptos de ‘aumento’ o ‘gran cantidad’. Por tanto, se puede considerar como un nombre colectivo que, por un proceso de blanqueamiento semántico, ha perdido su significado léxico de ‘configuración colectiva’ para cumplir funciones gramaticales. Es de suponer que ha sido reanalizado sintácticamente como cuantificador gracias a la convencionalización de la relación entre la construcción cuantificadora binominal y el concepto de cantidad (p.ej. *un montón de, una barbaridad de* + N2).<sup>15</sup> Más en concreto, ha pasado de un nombre *mazo* (N1) a un modificador *mazo de* (N2) de un N1 como se ve en (3) (véase Enghels y Roels, 2017 para más detalles).

(3) MAORE2J01: sí, tiene *mazo de* cosas o sea tiene en plan (COLAm)

Desde allí, *mazo (de)* se ha convertido en un intensificador multifuncional que desempeña diferentes funciones sintácticas. Así actúa entre otros como modificador adverbial de un núcleo verbal (*molar* en [4a]),<sup>16</sup> de una forma adjetival (*bonito* en [4b]), de un adverbio (*bien* en [4c]) o como cuantificador de un nombre (*hambre* en [4d]).

(4a) IR2F5 Ya, ¿a que mola *mazo*?

(4b) IR2F8 Llevaba un abrigo tía, en plan *mazo de* bonito de Pedro Del Hierro tía, y volví con el abrigo hecho mierda

(4c) IR2F9 Y bueno, estuvimos ahí tía, en realidad me lo pasé *mazo* bien tía

(4d) IR2F2 Es que tengo *mazo* hambre o sea, me voy a pedir algo eh contundente (CORMA)

Además, aparece como intensificador independiente en el discurso, un uso que hasta la fecha no ha sido descrito. En (4e), *mazo (de)* llega a cumplir una función pragmática específica como elemento afirmador de la aserción anterior. Mediante este uso, el hablante refuerza y afirma las ideas o sentimientos formuladas anteriormente por las (inter)locutoras. Esta observación parece indicar que, además de su paso de elemento léxico a elemento

15 Los cuantificadores binominales sirven ante todo para cuantificar de manera expresiva el N2 (p.ej. *un montón de dinero, una barbaridad de gente*). Verveckken y Delbecque (2015) señalan que, como la asociación entre la construcción binominal y el concepto de cantidad se ha convencionalizado bastante, cualquier nombre que refiere a una masa o a una multiplicidad de entidades tiene acceso a la construcción.

16 Resulta particularmente frecuente la colocación *molar mazo* que los jóvenes utilizan para expresar que algo les apetece (Enghels y Roels, 2017).

gramatical (gramaticalización), también adquiere una función pragmática (pragmaticalización).

- (4e) IR2F18 Y encima, es superpe-, es superpesada es superpesa-  
 IR2F20 Eee - es una hija de puta  
 IR2F19 *Mazo* (CORMA)

Sin embargo, tanto el proceso de gramaticalización como el uso pragmaticalizado merecen estudios más profundizados, que comparan el uso en diferentes décadas. La elaboración del corpus CORMA abre nuevas perspectivas para llevar a cabo este tipo de estudios diacrónicos, proporcionando los datos orales necesarios para poder indagar en procesos de cambios lingüísticos recientes.

En lo que sigue, llevaremos a cabo un estudio contrastivo del intensificador *mazo* (*de*) en dos corpus juveniles. Más en concreto, estudiaremos la frecuencia y el uso (socio)lingüístico de *mazo* (*de*) en el corpus COLAm y en los datos del habla juvenil del corpus CORMA. Mediante este estudio de caso comparativo de diacronía reciente se pretende estudiar la velocidad del cambio (socio)lingüístico dentro del lenguaje innovador de los jóvenes. Los parámetros analizados son los siguientes: cambios en la productividad del intensificador, en términos de su uso absoluto y la gama de variantes formales (3.2), la función gramatical del intensificador y la categoría gramatical que modifica (3.3), cambios en el conjunto de valores semánticos y pragmáticos que expresa (3.4), y la difusión social de su uso (3.5).

### ■ 3.2 Cambios en la productividad de *mazo* (*de*)

Un estudio comparativo de todas las ocurrencias de *mazo* (*de*) en ambos corpus (cf. Tabla 1) revela que la frecuencia de uso permanece aproximadamente igual, con un ligero aumento en el corpus más reciente, en términos de frecuencias normalizadas.

COLAm		CORMA	
#	fn/10.000	#	fn/10.000
664	14,55	145	16,45

Tabla 1. Frecuencia de *mazo* (*de*) en COLAm y CORMA

A pesar del ligero aumento en la frecuencia de uso del intensificador *mazo* (*de*), se reduce la variación formal del intensificador en el corpus más reciente (cf. Tabla 2).

	COLAm		CORMA	
	#	%	#	%
<i>mazo</i>	373	56	124	86
<i>mazo de</i>	285	43	21	14
<i>mazos de</i>	4	0.5	0	0
<i>maza</i>	2	0.5	0	0
<b>Total</b>	<b>664</b>	<b>100</b>	<b>145</b>	<b>100</b>

Tabla 2. Variación formal de *mazo* (*de*) en COLAm y CORMA

Al inicio del siglo XXI, tanto la forma *mazo* como *mazo* (*de*) se usan frecuentemente. La coexistencia de *mazo de*, al lado de *mazo*, apunta a que el elemento ha llegado a un grado más avanzado en la jerarquía de gramaticalización: *mazo* introduce un sintagma preposicional (encabezado por *de* como en la estructura binominal *mazo de tiempo*) y se incorpora aún más en la estructura gramatical. Este uso más gramaticalizado de *mazo* (*de*) parece que está yendo cuesta abajo dado que en el CORMA surge solo en un 14% de los casos observados, contrariamente a su uso recurrente (43%) en COLAm. En el corpus COLAm, el intensificador generalmente invariable aparece incluso en su forma femenina (*maza*) y plural (*mazos de*), por concordancia con el nombre modificado (p.ej. *modelos* en [5]):

- (5) MALCE2G04: están guapos  
 MALCE2J01: por eso  
 MALCE2G04: y hay mazo hay *mazos* de modelos (COLAm)

### ■ 3.3 Cambio (morfo)sintáctico

En segundo lugar, conviene estudiar la función gramatical que desempeña el intensificador. Ya sabemos que *mazo* funcionó inicialmente como un nombre con significado de, o bien “conjunto abundante de papeles u otras cosas atadas o unidas formando grupo” (DRAE), o bien “un nombre que designa una gran cantidad” como en *me gusta un mazo* (Iglesias, 2003) (cf. Sección 3.1). Si se examina la Tabla 3, resulta claro que hoy en día su uso ya no se restringe a este uso nominal.

	COLAm		CORMA	
	#	%	#	%
Adverbio	465	70	92	63
Adjetivo	139	21	40	28
Partícula	37	5.5	12	8
Dudoso	14	2	1	1
Sustantivo	7	1	0	0
Uso metalingüístico <sup>17</sup>	2	0.5	0	0
<b>Total</b>	<b>664</b>	<b>100</b>	<b>145</b>	<b>100</b>

Tabla 3. Función gramatical de *mazo (de)* en COLAm y CORMA

El uso como sustantivo que indica una gran cantidad (*un mazo*), categoría gramatical inicial de *mazo (de)*, se usa marginalmente en el corpus COLAm (6) pero ya no se documenta en el CORMA.

- (6) MALCE2V01: a ver eh eh luego lo reparto si fuese en litros. si esto fuese en cien litros siempre sería volumen  
MALCE2G02: un *mazo* (COLAm)

Para resumir, en el siglo XXI, *mazo (de)* se gramaticalizó hasta funcionar como adverbio (cf. (4a) *mola mazo*) y adjetivo (cf. (4d) *mazo hambre*). Incluso aparece como partícula independiente, con frecuencia ascendente, uso que se comentará a continuación en más detalle (cf. Sección 3.1, 3.4).

También resulta interesante estudiar el alcance de *mazo (de)* que se define por las categorías gramaticales modificadas por el intensificador. Como muestra la Tabla 4, son muchas las categorías gramaticales modificadas por *mazo (de)*, por lo cual se puede concluir que se ha convertido en un intensificador multifuncional en el lenguaje juvenil madrileño.

17 En estos usos metalingüísticos (p.ej. MAMTE2J01: lo de *mazo* y las expresiones estas que decimos), los jóvenes muestran que son bien conscientes de su lenguaje privativo y no estándar, identificando *mazo* como un elemento típico de su grupo generacional.

	COLAm		CORMA	
	#	%	#	%
Verbo	211	32	47	32
Adjetivo	208	31.5	36	25
Sustantivo	143	21.5	40	28
Independiente	49	7	12	8
Adverbio	40	6	9	6
Dudoso	13	2	1	1
<b>Total</b>	<b>664</b>	<b>100</b>	<b>145</b>	<b>100</b>

Tabla 4. Función gramatical (alcance) de *mazō (de)* en COLAm y CORMA

Así, en ambos corpus *mazō (de)* modifica frecuentemente a verbos (cf. (4a) *mola mazō*), adjetivos (cf. (4b) *mazō de bonito*) y sustantivos (cf. (4d) *mazō hambre*), aunque también aparece con menos frecuencia como modificador de adverbios (cf. (4c) *mazō bien*) o de forma independiente (cf. (4e) *mazō*). Si comparamos ambos corpus, no se observan cambios significativos en el alcance del intensificador.

### ■ 3.4 Cambio semántico-pragmático

Otra pregunta que surge es cuáles son los valores semántico-pragmáticos expresados por el intensificador bajo examen. Los seis valores destacados se resumen en la Tabla 5.

	COLAm		CORMA	
	#	%	#	%
Aumento cualitativo	375	56.5	67	46
Aumento cuantitativo	210	31.5	57	39
Entre cual. y cuant.	6	1	7	5
Afirmador aserción anterior	37	5.5	8	5.5
Muletilla	0	0	4	3
Dudoso	36	5.5	2	1.5
<b>Total</b>	<b>664</b>	<b>100</b>	<b>145</b>	<b>100</b>

Tabla 5. Valores semántico-pragmáticos de *mazō (de)* en COLAm y CORMA

Cuando *mazo* (*de*) subraya la cualidad de la base léxica (p.ej. *mazo de bonito*), tiene el valor semántico de aumento cualitativo (para más informaciones, véanse Enghels y Roels, 2017). Es el valor más frecuente en ambos corpus, aunque sufre un descenso en el CORMA (del 56.5% al 46%). El aumento cuantitativo del elemento modificado (p.ej. *mazo de tiempo*) al contrario, ha subido (del 31.5% al 39%). En algunos casos ambas lecturas aumentativas se presentan a la vez; en (7) la interpretación varía entre una lectura cualitativa (el ataque es intenso) y cuantitativa (el gato ataca varias veces).

- (7) MFA2M1 el gato de Rodri, qué asco. Le pones el dedo así como un gusano y te ataca. Te ataca *mazo* eh (CORMA)

Encima de estos valores semánticos básicos, *mazo* (*de*) también tiene un valor pragmático de afirmador de una aserción anterior en ambos corpus (cf. Sección 3.1). Así en (8) la hablante joven aumenta la fuerza del acto de habla de su interlocutora, afirmando y repitiendo su aserción.

- (8) IR2F2 Ya...Sandra Pesá. ¡Sandra Pesada  
IR2F1 Eh... es que hemos hecho tía-¡*Mazo* tía es que me suena a a pesada! Sandra Pesá (CORMA)

De hecho, en las conversaciones del CORMA, incluso se destaca un uso pragmático de *mazo* que nunca ha sido documentado anteriormente, lo que sugiere que se trata de un valor innovador y reciente en el habla juvenil. Considérense el acto de habla siguiente (9), extraído de una conversación entre dos amigas de 16 años de edad.

- (9) IR2F9 Y luego en plan estuvimos ahí un rato tal, y en realidad mazo bien, o sea, yo es que estaba tía *mazo* en plan de que ni me apetecía quedar. Y es que luego cuando estaba con él es como que tía, me estuve riendo mazo tía, y es como que jolín super bien (CORMA)

En este contexto, *mazo* ha perdido gran parte de su significado conceptual hasta actuar como un tipo de partícula discursiva. De hecho, su uso se acerca a la función de una muletilla que permite al hablante estructurar su discurso, incrementando a la vez la fuerza del contenido proposicional siguiente. Es llamativo que en los cuatro casos encontrados de este valor pragmático, *mazo* siempre coocorra con el marcador de aproximación *en*



*plan*. En suma, la aparición de este nuevo uso en el lenguaje juvenil contemporáneo sugiere que *mazo (de)* se ha pragmaticalizado aún más hasta convertirse en un tipo de muletilla.

### ■ 3.5 Cambio sociolingüístico

Por último, es interesante observar brevemente los cambios que se detectan en el perfil social de los usuarios del intensificador *mazo (de)*. Consideramos el parámetro de género y de clase social.<sup>18</sup> Primero, el uso de *mazo (de)* en el corpus COLAm parece no estar expuesto a la variación de género, puesto que se distribuye de manera equilibrada en el habla masculina (n=323; 49%) y femenina (n=335; 51%).<sup>19</sup> Como hasta la fecha no contamos con un equilibrio genérico en el subcorpus de lenguaje juvenil en el CORMA, recurrimos a las frecuencias normalizadas por 10.000 palabras. Notamos que los chicos recurren al intensificador *mazo (de)* 9.4 veces por cada 10.000 palabras frente a las chicas que lo usan con un promedio de 19.3 veces por 10.000 palabras producidas. Se trata entonces del doble. Por tanto, aunque al inicio de este siglo no se detectó variación genérica, el micro-estudio sugiere que *mazo (de)* hoy en día se usa más frecuentemente por las chicas.

En cuanto a la variación de clase social, *mazo (de)* tiene la siguiente distribución en el corpus COLAm.<sup>20</sup>

---

18 Los jóvenes participantes en el corpus COLAm provienen de distintos colegios y ambientes sociales, por lo que asegura un conjunto de datos más o menos equilibrado en cuanto al género y nivel social (Jørgensen, 2007). El corpus CORMA también se acerca a este equilibrio sociolingüístico, pero este micro-estudio todavía no incluye los datos juveniles de la clase social baja, todavía en proceso de transcripción en el momento de análisis. Además, el subcorpus del habla de las chicas (n=60; 71%) es más amplio que el de los chicos (n=24; 29%). El número de palabras aproximativo del habla masculina es de 25.462 frente a 62.655 palabras en el subcorpus de jóvenes femeninas. *Mazo (de)* tiene 24 ocurrencias en el habla masculina y 121 en el habla femenina.

19 En 6 de los casos analizados el corpus COLAm no indica el género del hablante.

20 En 41 de los casos analizados, el corpus no proporciona informaciones detalladas sobre la clase social del hablante, así que el análisis resumido en la Taula 6 (página siguiente) se basa aquí en un total de 623 ejemplos.

	COLAm	
	#	%
Clase alta	89	14
Clase media	293	47
Clase baja	241	39
<b>Total</b>	<b>623</b>	<b>100</b>

Tabla 6. Variación de clase social

El intensificador *mazo (de)* resulta más frecuente en el habla de jóvenes de la clase media (47%) y baja (39%) mientras que los jóvenes de la clase alta no recurren tanto a este intensificador. Estos datos sugieren que se trata de una innovación vernácula, que ha se ha generado en las clases sociales medio-bajas (Blas Arroyo, 2005).

Por el desequilibrio de la clase social de los hablantes jóvenes del CORMA en el momento de análisis, todavía no se ha podido efectuar un análisis comparativo de las tres clases sociales. Sin embargo, sí se ha comparado la frecuencia de uso de *mazo (de)* en el habla de las chicas de la clase alta en el corpus COLAm (69 ocurrencias) y en el CORMA (121 ocurrencias). Si consideramos las frecuencias normalizadas por 10.000 palabras, se destaca un incremento de uso altamente llamativo de 1,49 usos de *mazo (de)* a 13,73 por 10.000 palabras en el habla de las chicas de la clase alta. Estos resultados sugieren que a lo largo del siglo XXI *mazo (de)* ha alcanzado una difusión social muy significativa, y que su uso está generalizado ampliamente hoy en día en la clase social alta. Por tanto, este micro-estudio de diacronía reciente suscribe la hipótesis del ‘cambio desde abajo’ (Eckert, 2003; Blas Arroyo, 2005) según la cual determinadas innovaciones provienen de clases sociales más bajas.

#### ■ 4 Conclusiones

En resumidas cuentas, el corpus CORMA ha sido compuesto en función de una amplia gama de rasgos sociolingüísticos y situacionales. Por su composición particular se aproxima al máximo al vernáculo (madrileño) que, según Labov (2001), mejor se presta a estudios variacionistas. En concreto, la grabación de conversaciones naturales y espontáneas entre hablantes madrileños, de diferentes géneros, edades y niveles socioculturales —quienes contraen distintos tipos de relaciones entre ellos en distintos tipos de situaciones— ha llevado a un muestreo representativo del habla coloquial

madrileña actual. Siendo lo suficientemente amplio, el corpus se presta tanto a análisis de índole cualitativa como cuantitativa. Además, por su diseño multivariado hace posible el análisis contextualizado de cualquier fenómeno lingüístico, en función de una larga serie de variables socio-situacionales. Dadas su composición y características constitutivas, los datos del CORMA resultan especialmente relevantes para el análisis de fenómenos lingüísticos propios de la oralidad. Además, por la atención dedicada a la representatividad del muestreo del habla juvenil, el corpus permite estudiar este sociolecto en suficientes detalles, además de realizar estudios del cambio lingüístico reciente. Este tipo de estudios micro-diacrónicos se realiza más precisamente mediante la comparación con corpus o muestras comparables, como el COLAm, grabado al inicio del siglo XXI (Jørgensen, 2013).

La relevancia del CORMA para el estudio del cambio lingüístico reciente ha sido ilustrada mediante el análisis comparativo del intensificador *mazo (de)* en la variedad juvenil madrileña, tal y como se representa en los corpus COLAm y CORMA. Primero, en términos de frecuencia, parece que *mazo (de)* sigue siendo de moda entre los jóvenes madrileños. Sin embargo, su variabilidad formal parece contraerse. Los datos muestran claramente que, a través de un proceso de gramaticalización, *mazo (de)* se ha convertido en un intensificador multifuncional. Además, bajo su forma de partícula independiente, parece cumplir cada vez más funciones pragmáticas, incluso la de muletilla. Para continuar, el análisis sociolingüístico ha proporcionado informaciones llamativas con respecto a cambios en el perfil social de sus usuarios. Así, los datos sugieren que la difusión social de *mazo (de)* corresponde a la de un ‘cambio desde abajo’. De hecho, el intensificador vernáculo originó en las clases sociales medias-bajas y a lo largo del siglo XXI se ha generalizado y extendido socialmente, con extensión considerable en el habla juvenil femenina de la clase social alta.

En definitiva, este estudio permitió comprobar la aplicabilidad del corpus CORMA para estudios de cambios lingüísticos (recientes). En el caso del intensificador *mazo (de)*, el corpus resultó útil tanto para observar la difusión social de este elemento lingüístico como para describir los cambios lingüísticos, abriendo por ejemplo perspectivas para el seguimiento de procesos de gramaticalización. ■

## ■ Referencias bibliográficas

- Aarts, Bas / Close, Joanne / Leech, Geoffrey N. / Wallis, Sean (eds.) (2013): *The Verb Phrase in English. Investigating recent language change with corpora*, Cambridge: Cambridge University Press.
- Albelda, Marta / Briz, Antonio / Cestero, Ana María / Kotwica, Dorota / Villalba, Cristina (2014): «Ficha metodológica para el análisis pragmático de la atenuación en corpus discursivos del español. (ES.POR. ATENUACIÓN)», *Oralia* 17, 7–62.
- Barcala, Mario / Domínguez, Eva / Fernández, Alba / Rivas, Raquel / Santalla, María Paula / Vázquez, Victoria / Villapol, Rebeca (2018): «El corpus ESLORA de español oral Diseño, desarrollo y explotación. CHIMERA», *Romance Corpora and Linguistic Studies* 5:2, 217–237.
- Bednarek, Monika (2019): «The multifunctionality of swear/taboo words in television series», in: Mackenzie, J. Lachlan / Alba-Juez, Laura (eds.): *Emotion in Discourse*, Amsterdam: John Benjamins, 29–54.
- Blas Arroyo, José Luis (1994): «Los pronombres de tratamiento y la cortesía », *Revista de Estudios Hispánicos* 22, 439–466.
- (2005): *Sociolingüística del español: desarrollos y perspectivas en el estudio de la lengua española en contexto social*, Madrid: Cátedra.
- Briz Gómez, Antonio (1995): *La conversación coloquial. Materiales para su estudio*, Valencia: Universitat de València.
- (1998): *El español coloquial en la conversación: esbozo de pragmatogramática*, Barcelona: Ariel.
- (2012): «Los déficits de los corpus orales del español (y de algunos análisis)». in: Jiménez Juliá, Tomás et al. (eds.): *Cum corde et in nova grammatica. Estudios ofrecidos a Guillermo Rojo*, Santiago de Compostela: Servicio de Publicaciones e Intercambio Científico da Universidade de Santiago de Compostela, 115–137.
- / Grupo Val.Es.Co. (2002): *Corpus de conversaciones coloquiales*, Madrid: Arco/Libros.
- Cabedo Nebot, Adrián (2013): «Sobre prosodia, marcadores del discurso y unidades del discurso en español: Evidencias de un corpus oral espontáneo», *Onomázein: Revista De Lingüística, Filología y Traducción* 2 (28), 201–213.

- Cano Aguilar, Rafael (2002): «El español coloquial: enfoques y perspectivas de análisis», in: Echenique, María Teresa / Sánchez Méndez, Juan P. (eds.): *Actas del V Congreso Internacional de Historia de la Lengua Española*, Madrid: Gredos, 279–284.
- De Smet, Emma / Enghels, Renata (2020): «Los datos en Twitter como fuente de la oralidad. Estudio de caso del MD *en plan*», *Oralia* 23:2, 199–218.
- Eckert, Penelope (1997): «Age as a sociolinguistic variable», in: Coulmas, Florian (ed.): *The Handbook of Sociolinguistics*, Oxford: Blackwell, 151–167.
- (2003): «Language and adolescent peer groups», *Journal of Language and Social Psychology* 22:1, 112–118.
- Eisenstein, Evelyn (2005): «Adolescência: definições, conceitos e critérios», *Adolesc Saude* 2:2, 6–7.
- Enghels, Renata / Vanderschueren, Clara / Bouzouita, Miriam (2015): «Panorama de los corpus y textos del español peninsular contemporáneo», in: Roegiest, Eugene / Iliescu, Maria (eds.): *Manuel des anthologies, corpus et textes romans*, Berlin / New York: Mouton de Gruyter, 147–170.
- / Roels, Linde (2017): «*Mazo (de)* y otros recursos de intensificación en el lenguaje juvenil madrileño: factores lingüísticos y sociales», *Moderna språk* 111:2, 45–61.
- García Palacios, Joaquín / De Sterck, Goedele / Linder, Daniel / Maroto García, Nava / Sánchez Ibáñez, Miguel / Torres del Rey, Jesús (eds.) (2016): *La neología en las lenguas románicas. Recursos, estrategias y nuevas orientaciones*, New York: Peter Lang.
- Gómez Molina, José Ramón / Gómez Devís, María (2010): «Mantenimiento y elisión de la /d/ intervocálica en el español de Valencia», *Verba: Anuario galego de filoloxía* 37, 89–122.
- González Montero, José Antonio (1993): «La aspiración. Fenómeno expansivo en español: Su importancia en andaluz: Nuevos casos», *Cauce: Revista Internacional de Filología, Comunicación y sus Didácticas* 16, 31–66.
- Helincks, Kris (2016): *Variation and discursive shifting of address forms in Chilean Spanish: formal, socio-situational, and pragmatic analysis of spontaneous conversations*, Ghent: Ghent University (tesis doctoral no publicada).
- Iglesias, José María (2003): *Diccionario de argot español*, Madrid: Alianza Editorial.

- Jenkins, Henry (2009): *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*, Cambridge: MIT Press.
- Jørgensen, Annette Myre (2007): «COLA: Un corpus oral de lenguaje adolescente», *Oralia* 3:1, 225–234.
- (2013): «Spanish teenage language and the COLAm corpus», *Bergen Language and Linguistics Studies* 3:1, 151–166.
- / Drange, Eli-Marie (2012): «La lengua juvenil de las metrópolis Madrid y Santiago de Chile», *Arena Romanística* 9, 74–96.
- Labov, William (1972): *Sociolinguistic Patterns*, Philadelphia: University of Pennsylvania Press.
- (2001): «The anatomy of style-shifting», in: Eckert, Penelope / Rickford, John R. (eds): *Style and Sociolinguistic Variation*, Cambridge: Cambridge University Press, 85–108.
- López Serena, Araceli / Borreguero Zuloaga, Margarita (2010): «Los marcadores del discurso y la variación lengua hablada vs. lengua escrita», in Loureda Lamas / Acín-Villa (eds.), 415–493.
- Lorenz, Gunter (2002): «Really worthwhile or not really significant? A corpus-based approach to the delexicalization and grammaticalization of intensifiers in Modern English», in: Wischer, Ilse / Diewald, Gabriele (eds.): *New Reflections on Grammaticalization*, Amsterdam: John Benjamins, 143–161.
- Loureda Lamas, Óscar / Acín-Villa, Esperanza (ed.) (2010): *Los estudios sobre MDes del discurso en español, hoy*, Madrid: Arco/Libros.
- Martínez López, Juan A. (2009): «Lexical innovations in Madrid's teenage talk: some intensifiers», in: Stendström, Anna-Brita / Jørgensen, Annette Myre (eds.): *Youngspeak in a Multilingual Perspective*, Amsterdam: John Benjamins, 81–93.
- Matamala, Anna (2008): «La oralidad en la ficción televisiva: análisis de las interjecciones de un corpus de comedias de situación originales y dobladas», in: Brumme, Jenny (ed.): *La oralidad fingida: descripción y traducción*, Madrid: Iberoamericana Vervuert, 81–94.
- Moreno Fernández, Francisco (2006): «Información básica sobre el Proyecto para el Estudio Sociolingüístico del Español de España y de América – PRESEEA (1996–2010)», *Revista Española de Lingüística* 36, 385–392.

- Narbona Jiménez, Antonio (2015): *Sintaxis del español coloquial*, Sevilla: Editorial Universidad de Sevilla.
- Oliván, Fidel (2015): *Barrio rico, barrio pobre: un análisis del 24M en Madrid por distritos*, <<https://politikon.es/2015/08/06/barrio-rico-barrio-pobre-un-analisis-del-24m-en-madrid-por-distritos/>> [06.11.2019].
- Padilla Herrada, María Soledad (2016): «Marcadores y partículas discursivas interactivas en el entorno político/periodístico de Twitter», *Philologia Hispalensis* 30:1, 193–212.
- Palacios Martínez, Ignacio M. / Núñez Pertejo, Paloma (2014): «Strategies used by English and Spanish teenagers to intensify language: A contrastive corpus-based study», *Spanish in Context* 11:2, 175–201.
- Real Academia Española (2010): *Ortografía de la lengua española*, Madrid: Espasa Calpe.
- Rueda, Ana Mancera / Alamán, Ana Paro (2013): *El español coloquial en las redes sociales*, Madrid: Arco/Libros.
- Sampson, Geoffrey (2000): *CHRISTINE Corpus: Documentation*, <<http://www.grsampson.net/ChrisDoc.html>> [06.11.2019].
- Sanmartín Sáez, Julia (1999): «A propósito de los sufijos apreciativos en la conversación coloquial: Sus valores semánticos y pragmáticos», *Oralia* 2, 185–219.
- Schneider, Stephan (2007): *Reduced Parenthetical Clauses as Mitigators: A Corpus Study of Spoken French, Italian and Spanish*, Amsterdam: John Benjamins.
- Stenström, Anna-Brita (2005): «He's well nice – Es mazo majo: London and Madrid Teenage Girls' Use of Intensifiers», in: Granath, Solveig / Miliander, June / Wenno, Elisabeth (eds.): *The Power of Words: Studies in Honour of Moira Linnarud*, Karlstad: Karlstad University, 207–216.
- (2008): «Algunos rasgos característicos del habla de contacto en el lenguaje de adolescentes en Madrid», *Oralia* 11, 207–226.
- Tagliamonte, Sali A. (2008): «So different and pretty cool! Recycling intensifiers in Toronto, Canada», *English Language & Linguistics* 12:2, 361–394.
- Verveckken, Katrien / Delbecque, Nicole (2015): «On the development of binominal quantifiers in Spanish: the notion of lexical persistence revisited», *Revue de l'Association Française de Linguistique Cognitive* 13, 1–37.

Zimmermann, Klaus (2002): «La variedad juvenil y la interacción verbal entre jóvenes», in: Rodríguez González, Félix (ed.): *El lenguaje de los jóvenes*, Barcelona: Ariel, 137–164.

- Renata Enghels, Universiteit Gent, Faculteit Letteren en Wijsbegeerte, Blandijnberg 2, B-9000 Gent, <Renata.Enghels@UGent.be>
- Fien de Latte, Universiteit Gent, Faculteit Letteren en Wijsbegeerte, Blandijnberg 2, B-9000 Gent, <Fien.DeLatte@UGent.be>
- Linde Roels, Universiteit Gent, Faculteit Letteren en Wijsbegeerte, Blandijnberg 2, B-9000 Gent, <linde.roels@ugent.be>