

ÍNDICE

- 5 **Sección temática: Dialectología digital: innovaciones técnicas y metodológicas**
Digital Dialectology: Technical and Methodological Innovations
VERENA WEILAND
- 7 Dialectología digital: innovaciones técnicas y metodológicas
Digital Dialectology: Technical and Methodological Innovations
VERENA WEILAND
- 13 Geolingüística digital y bases de datos: una aproximación al estudio de la variación y del
cambio léxico en español
Digital Geolinguistics and Databases: An Approach to the Study of Lexical Variation
and Linguistic Change in Spanish
CAROLINA JULIÀ LUNA
- 33 Percepción de las realizaciones de <ll> e <y> en las variedades del español: un experimento
en línea
Perception of the Realizations of <ll> and <y> in Spanish Varieties: An Online Experiment
VERENA WEILAND/LINDA BÄUMLER/CARMEN QUIJADA VAN DEN BERGHE
- 59 *Dialectos del español*: presentación de la *app* y primeros resultados
Dialectos del español: Presentation of the App and First Results
MIRIAM BOUZOUITA/MÓNICA CASTILLO LLUCH/ENRIQUE PATO
- 77 La construcción del *Corpus Oral y Sonoro del Español Rural - Anotado y Parseado (CO-*
SER-AP): avances en el etiquetado de partes del discurso
The Construction of the *Annotated and Parsed Audible Corpus of Spoken Rural Spanish*
(*COSE-AP*): Advances in the Annotation of the Parts of Speech
JOHNATAN E. BONILLA/MIRIAM BOUZOUITA/ROSA LILIA SEGUNDO DÍAZ
- 97 **Sección general**
- 99 Fenómenos de contacto entre el aymara y el español: la replicación gramatical de ‘ir a
+ infinitivo’
Contact Phenomena Between Aymara and Spanish: The Grammatical Replication of
‘*ir a + infinitive*’
GERALDINE QUARTARARO

- 125 Paisaje lingüístico en nueve pueblos de origen indígena en el centro-oriente de México
Linguistic Landscape in Nine Villages of Indigenous Origin in Central-Eastern Mexico
LILIANA RUIZ VELASCO D./BARDOMIANO SÁNCHEZ MARTÍNEZ/ANA LILIA TREJO
VÁZQUEZ/DANNA PAOLA ROMANO BECERRA/OBED ANTONIO HERNÁNDEZ CORTÉS/
CECILIA ANTONIO PÉREZ/MARÍA LUISA GARCÍA ESPINOZA/EFRÉN ANTONIO PALOMO
HUAXIN/JOSÉ JUAN EVANGELISTA DÍAZ/MANFRET CRISTIAN HERNÁNDEZ MÁRQUEZ/
ENEDINO DE LA CRUZ DÍAZ
- 161 Variación microparamétrica en el Caribe: una perspectiva sintáctico-pragmática del foco
en el español de Puerto Rico
Microparametric Variation in the Caribbean: Syntactic-Pragmatic Analysis of Focus in
the Spanish of Puerto Rico.
LUIS A. ORTIZ LÓPEZ/MARÍA FERNANDA ESCALANTE
- 193 Reseñas, noticias y estados de la cuestión
- 195 Luciano Canepari/Maurizio Pugliese. *Galician Pronunciation & Accents: Geo-social Applications of the Natural Phonetics & Tonetics Method*. ANDREA GARCÍA-COVELO
- 199 Juan Carlos Conde (ed.): *Una laguna sumergida. Epistolario de Américo Castro & María Rosa Lida de Malkiel*. JOSÉ IGNACIO PÉREZ PASCUAL
- 210 Patricia Gubitosi/Michelle F. Ramos Pellicia (eds.): *Linguistic Landscape in the Spanish-speaking World*. CLARA LAGO CAAMAÑO
- 216 Cristina Lastres López: *From Subordination to Insubordination. A Functional-Pragmatic Approach to if/si-Construccions in English, French and Spanish Spoken Discourse*. VÍCTOR ROYO VIÑUALES
- 220 Isabel Molina Martos/Pilar García Mouton (eds.): *Geolingüística en la Península Ibérica*. MIGUEL RODRÍGUEZ MONTEAVARO
- 224 Alexandre Veiga Rodríguez (2020): *El "pretérito imperfecto" en el sistema verbal español*. BEATRIZ RUIZ GRANADOS
- 228 XXX Congreso Internacional de Lingüística y Filología Románicas: "Romanística y Atlánticidad". DANIEL EDUARDO BEJARANO BEJARANO/ANTOINE PRIMERANO
- 245 Normas editoriales

Dialectos del español: presentación de la *app* y primeros resultados

Dialectos del español: Presentation of the App and First Results

MIRIAM BOUZOUITA
Humboldt-Universität zu Berlin
miriam.bouzouita@hu-berlin.de

MÓNICA CASTILLO LLUCH
Université de Lausanne
monica.castillolluch@unil.ch

ENRIQUE PATO
Université de Montréal
enrique.pato-maldonado@umontreal.ca

| **Abstract:** This article presents *Dialectos del español*, a smartphone app for the study of morphosyntactic variation in the Spanish language, launched in May 2019 and available at www.dialectosdelespanol.org and Google Play. It presents the scientific and technical aspects and challenges of the project (how it came about, its objectives, the 26 questions of the app, the coding of the answers and the functioning of the prediction), reports on its dissemination in the media and public participation and presents some results.

| **Keywords:** Smartphone app, morphosyntactic variation, big data, digital dialectology

| **Resumen:** Este artículo presenta *Dialectos del español*, una *app* para *smartphones* destinada al estudio de la variación morfosintáctica de la lengua española, lanzada en mayo de 2019 y disponible en www.dialectosdelespanol.org y Google Play. Se exponen los aspectos científicos y técnicos del proyecto (cómo surgió, sus objetivos, las 26 preguntas de la aplicación, la codificación de las respuestas y el funcionamiento de la predicción), se informa sobre su difusión en los medios y la participación del público, y se presentan algunos resultados y los retos actuales del proyecto.

| **Palabras clave:** *app* para *smartphone*, variación morfosintáctica, *big data*, dialectología digital

1. INTRODUCCIÓN

Con las herramientas digitales se han abierto en el siglo XXI nuevas vías de acceso a los datos lingüísticos objeto de los estudios variacionistas. Numerosas investigaciones tienen hoy apoyo empírico en las redes sociales (especialmente Twitter y Facebook / Meta), blogs, foros y otras fuentes escritas y orales disponibles en Internet. Además, se están diseñando cada vez más aplicaciones para *smartphones* o encuestas digitales que permiten recopilar *big data* y constituir corpus mediante producción participativa (*crowdsourcing*). En dialectología, el recurso a la ciencia ciudadana (*citizen science*), que promueve la participación del público general en la investigación, es continuador de prácticas tradicionales de la disciplina desde sus inicios. Pero si esta siempre ha necesitado de la colaboración de los locutores que poseen el conocimiento y los usos lingüísticos, lo que ahora cambia es que la participación requerida es masiva y proporciona un ingente caudal de información.

Entre las aplicaciones dialectológicas para *smartphones* destacan las pioneras *Dialäkt Äpp* y *Voice Äpp* (Kolly/Leemann 2015, Leemann 2021) desarrolladas para el estudio de los dialectos del alemán de Suiza, así como *English Dialects App* (Leemann/Kolly/Britain 2018), que investiga el inglés británico; en la estela de estas han surgido otras más durante estos últimos años, de las que se da cuenta en el reciente número especial de *Linguistics Vanguard* editado por Hilton y Leemann (2021).

Este artículo presenta *Dialectos del español*, una aplicación para *smartphones* destinada al estudio dialectológico de la lengua española, lanzada en mayo de 2019 y disponible en www.dialectosdelespanol.org y Google Play. En primer lugar (§ 2), se expone cómo surgió y cuáles son sus objetivos, y se describen las 26 preguntas que contiene la aplicación y los aspectos técnicos del proyecto (la codificación de las respuestas y el funcionamiento de la predicción); en segundo lugar (§ 3), se informa sobre la importancia de la comunicación para lograr la participación del público y también sobre el perfil de los participantes; por último (§ 4), se ofrece un avance de primeros resultados, así como los retos actuales del proyecto. Las conclusiones resumen el contenido de estas páginas.

2. ORIGEN, OBJETIVOS Y ASPECTOS TÉCNICOS DE LA APP

En el 21.º Simposio de Sociolingüística de la *Linguistic Society of America*, celebrado en la Universidad de Murcia en 2016, tuvimos la oportunidad de descubrir dos de las *apps* que acaban de citarse: *Dialäkt Äpp*, diseñada para el estudio de las variedades del alemán de Suiza, y *English Dialects App*, para el estudio de las variedades del inglés británico, presentadas respectivamente por Adrian Leemann y David Britain (véanse Leemann *et al.* 2016 y Britain 2016). Estas aplicaciones se basan en una serie de cuestiones fonéticas y léxicas, los componentes de la lengua más estudiados tradicionalmente, y su principal objetivo es recoger *big data* que facilite el conocimiento

de la evolución de determinadas variables de estas lenguas previamente estudiadas en otros proyectos de dialectología.

Los autores de estas páginas compartieron inmediatamente el interés por desarrollar una aplicación similar para el español, pero para estudiar la variación gramatical, el aspecto menos explorado por la dialectología hispánica, aspirando, además, a llegar a todo el mundo hispanohablante.

Para empezar, diseñamos un estudio piloto que nos permitió poner a prueba un conjunto de preguntas a través de un formulario web (Google Forms), completado por 547 participantes de varios países de habla hispana. En Bouzouita, Castillo Lluch y Pato (2018), expusimos los objetivos y la metodología de nuestro proyecto, que en ese momento estaba en fase de concepción, y presentamos los resultados obtenidos con la serie de preguntas del formulario inicial. Gracias a aquel estudio preliminar pudimos decidir qué variables lingüísticas seleccionar finalmente para las 26 preguntas de la *app*, identificamos algunos problemas que debíamos intentar evitar (por ejemplo, en los enunciados de las preguntas debíamos descartar el léxico no común a España y América para no producir extrañeza en una parte de los participantes) y pudimos tener en cuenta otras peticiones de los usuarios (por ejemplo, incluir “otro” como tercera opción de género, junto a “hombre” y “mujer”).

Con esta primera experiencia y la intervención técnica de Daniel Wanitsch —el mismo informático que diseñó las aplicaciones del equipo de Adrian Leemann y David Britain— desarrollamos la aplicación *Dialectos del español*, disponible en www.dialectosdelespanol.org y en Google Play¹. La aplicación está alojada en la plataforma Wix y los datos recogidos se almacenan en un servidor de la Universidad de Lausana, que garantiza la seguridad y durabilidad de su conservación.

Nuestro principal objetivo es obtener datos que nos permitan estudiar ciertos fenómenos de variación morfosintáctica del español en todo el mundo hispanohablante². Algunos de estos fenómenos ya han sido descritos y analizados anteriormente, en cuyo caso se trata de comprobar si esas descripciones se corresponden con los usos actuales. Otros fenómenos solo se han descrito parcialmente, y en este caso el reto será completar su descripción y análisis. Por último, algunas de las variables gramaticales de nuestra lista no se han investigado hasta ahora, por lo que nos proponemos explorarlas por primera vez.

Las preguntas de la *app* se presentan pidiéndole al participante que seleccione una opción, como se aprecia en el ejemplo siguiente:

¹ Deseamos agradecer a Adrian Leemann su disponibilidad y su actitud abierta para compartir con nosotros información valiosa para desarrollar nuestro proyecto.

² Las preguntas formuladas a los participantes y las respuestas proporcionadas para la predicción de su geolocalización tienen en cuenta las variantes encontradas en Europa y América, pero también aspiramos a contar con la participación de hispanohablantes de África y Asia.



Dialectos del español

Por favor, elige la respuesta que te resulte más natural. Al final de las preguntas, intentaremos predecir de dónde eres.

«Rosa vive lejos. No sé si...»

vendrá esta tarde.

venga esta tarde.

de 26

Las variables gramaticales de las 23 preguntas se resumen a continuación clasificadas en las tres categorías a las que acabamos de aludir:

1. Preguntas clásicas (ya abordadas por la dialectología tradicional):

Perruco / -ino / -ito / -illo / -iño / -ín / -iquio / -ico / -ete

El / la sartén

A Pedro le / lo vi

A los jugadores les / los ves / Ø ves

El libro lo / le tengo

Había(n) muchos estudiantes

Habíamos / estábamos / éramos

¿Qué dices / qué tú dices / qué decís?

A ti / vos te gusta el invierno

Bailas / Bailás / Bailái(s)

Fue en Bogotá donde / que se conocieron...

2. Preguntas menos clásicas (se dispone de alguna información, pero no de resultados para todo el mundo hispanohablante):

Comprásemos / compráramos

Tuviese / tuviera / tendría

Manuel está enfrente de mí / mío / mía

No sé si vendrá / venga

Isa me dijo ayer que viniéramos / vengamos / de venir hoy

Nada más / más nada

Yo ya / Ya yo había salido

3. Preguntas inéditas:

Esta / este agua

Van a venir / vendrán hoy

No la vamos a reconocer / no vamos a reconocerla

*Cuanto / contra / mientras más...
Habla mal de mí / mío / mía*

El principio de este tipo de aplicaciones consiste en solicitar la colaboración del público a cambio de una recompensa lúdica: a partir de las respuestas, la aplicación intenta adivinar el origen dialectal de los participantes. Por lo tanto, un segundo objetivo de nuestro proyecto es identificar correctamente los dialectos de los participantes a partir de sus respuestas. El reto no es trivial en la medida en que muchas de las variables de nuestro estudio no permiten en absoluto la geolocalización de los hablantes (la mayoría de las preguntas “inéditas” y “menos clásicas”), y, cuando permiten la geolocalización, esta es demasiado vaga (por ejemplo, los diminutivos, *el sartén, leísmo, haber* concordado con un objeto plural, *voseo, más nada, ya yo...*). Las respuestas gramaticales que apuntan a un país americano o a una zona de la península ibérica son finalmente excepcionales: (*vos*) *bailái(s) conmigo* (Chile, Venezuela), *Se conocieron en Bogotá fue* (República Dominicana), *enfrente mía* (Galicia, Andalucía, Madrid).

Este contrato con el público, que aparece en la página de inicio de la aplicación (“Dime cómo hablas y te diré de dónde eres”), nos obligó a incluir algunas preguntas léxicas que orientan de forma mucho más fina hacia un país concreto e incluso a veces hacia una región específica. Estas preguntas léxicas son tres: la fórmula utilizada para responder al teléfono y las palabras equivalentes a *guisante* y *lagartija* (basadas en una imagen de la semilla y del animal, respectivamente).

Detrás de la capa visible de 26 preguntas, la aplicación contiene una codificación geográfica de las respuestas, utilizada para predecir el dialecto del participante, y que se pudo establecer a partir de información contenida en la *Nueva gramática de la lengua española* (RAE/ASALE 2009), diversos cuestionarios de los atlas lingüísticos, y monografías y artículos especializados sobre cada cuestión. A continuación, se muestran tres ejemplos concretos de esta codificación:

- (1) ... el agua es potable.
 - a. Esta...
 - b. Este...
 - c. Esta / Este...

- (2) ¿La sartén o el sartén?
 - a. La sartén [ESPAÑA]
 - b. El sartén [AMÉRICA]
 - c. La / el sartén [AMÉRICA]

- (3) Manuel está enfrente...
 - a. de mí. [AMÉRICA: México; AMÉRICA: Guatemala; AMÉRICA: Honduras; AMÉRICA: Nicaragua; AMÉRICA: El Salvador; AMÉRICA: Costa Rica; AMÉRICA: Panamá; AMÉRICA: República Dominicana; AMÉRICA: Puerto Rico; AMÉRICA: Venezuela; AMÉRICA: Colombia; AMÉRICA: Estados Unidos; ESPAÑA: Aragón; ESPAÑA: Islas Baleares; ESPAÑA: Asturias; ESPAÑA: Cantabria; ESPAÑA: Castilla y

- León; ESPAÑA: La Rioja; ESPAÑA: Murcia; ESPAÑA: Navarra; ESPAÑA: País Vasco; ESPAÑA: Castilla La Mancha].
- b. *mío*. [ESPAÑA: Galicia; ESPAÑA: Andalucía; ESPAÑA: Madrid; AMÉRICA: Argentina; AMÉRICA: Uruguay; AMÉRICA: Chile; AMÉRICA: Ecuador; AMÉRICA: Perú; AMÉRICA: Bolivia; AMÉRICA: Paraguay; ESPAÑA: Cataluña; ESPAÑA: Comunidad Valenciana; AMÉRICA: Panamá].
- c. *mía*. [ESPAÑA: Galicia; ESPAÑA: Andalucía; ESPAÑA: Madrid].
- d. *de mí / mío*. [ESPAÑA: Galicia; ESPAÑA: Andalucía; ESPAÑA: Madrid; AMÉRICA: Argentina; AMÉRICA: Uruguay; AMÉRICA: Chile; AMÉRICA: Ecuador; AMÉRICA: Perú; AMÉRICA: Bolivia; AMÉRICA: Paraguay; ESPAÑA: Cataluña; ESPAÑA: Comunidad Valenciana; AMÉRICA: Panamá; AMÉRICA: Costa Rica].
- e. *de mí / mía*. [ESPAÑA: Galicia; ESPAÑA: Andalucía; ESPAÑA: Madrid].
- f. *mío / mía*. [ESPAÑA: Galicia; ESPAÑA: Andalucía; ESPAÑA: Madrid].
- g. *de mí / mío / mía*. [ESPAÑA: Galicia; ESPAÑA: Andalucía; ESPAÑA: Madrid].

Como vemos, la codificación geográfica es a veces imposible (ejemplo 1), ya que ningún estudio previo nos habla de la difusión de las formas femeninas o masculinas del demostrativo ante sustantivos que empiezan por /á/; en otros casos, es muy vaga (como en 2)³, y solo cuando tenemos información muy precisa sobre un fenómeno (como en 3, cf. Salgado/Bouzouita 2017; Marttinen Larsson/Bouzouita 2018, 2022) puede contribuir realmente a la predicción del dialecto de los participantes.

El principio de funcionamiento interno de la *app* para esta predicción es sencillo: la suma de puntos por país americano o provincia/comunidad española, de manera que, por ejemplo, si un participante responde con las respuestas codificadas “[AMÉRICA]” y las que le siguen, será identificado como colombiano:

- *Enfrente de mí*
- *Juan habla mal de mí*
- *Sí, lo vi ayer*
- *Los veo [a los jugadores]*
- *El libro que tengo en mi casa*
- *Arvejas*
- *A ver*
- *No, se conocieron fue en Bogotá*
- *Yo ya / ya yo había salido*
- *A ti / vos te gusta el invierno*
- *Qué dices / Qué decí(s) (tú / vos)*

³ Hemos simplificado deliberadamente en este caso y hemos omitido considerar que en algunos dialectos del sur de España (andaluz, canario) *sartén* también puede ser de género masculino, ya que esta forma no suele ser utilizada por la mayoría de los usuarios de la *app*, que son universitarios (véase *infra*). Sin embargo, la combinación de este rasgo con otros del cuestionario puede predecir que el usuario de la *app* es efectivamente andaluz o canario.

Poco después del lanzamiento de la *app*, reparamos en que su algoritmo aditivo era a veces insuficiente y podía dar lugar a graves errores de geolocalización, lo que nos llevó a añadir algunas condiciones eliminatorias a la hora de calcular la geolocalización. Estas condiciones, implementadas por primera vez en este tipo de aplicación, son tres: 1) si se utiliza el *voseo* pronominal o verbal, España queda excluida de la predicción; 2) la no inversión del pronombre sujeto en las interrogativas (¿*Qué tú dices?*) excluye a España; 3) *ser* focalizador excluye a España.

Tras el cuestionario y nuestra predicción del origen dialectal del participante (tres geolocalizaciones ofrecidas por orden de probabilidad), le solicitamos algunos datos extralingüísticos: su género (hombre / mujer / otro), su edad (-18 / 18-35 / 36-55 / +55), su nivel de estudios (primaria, secundaria, universitaria), si es hablante nativo o no de español (y en este caso, cuál es su lengua materna) y si es nativo de otro idioma además del español. Por último, solicitamos la ayuda del participante para mejorar la aplicación pidiéndole que localice en un mapa el lugar donde nació, donde creció y donde vive actualmente. También le preguntamos cuántas veces ha cambiado de ciudad (nunca / 1 vez / 2-3 veces / más de 3 veces) y de país (*ídem*) a lo largo de su vida. Además, puede calificar con cinco estrellas (en 10 segmentos) la precisión de la predicción de la aplicación, puede dejar comentarios y compartir la aplicación en Facebook / Meta y Twitter. Para nuestros dos objetivos (obtener datos dialectales y que la *app* dé la mejor predicción posible), lógicamente solo son útiles los cuestionarios con los datos cartográficos y extralingüísticos de los participantes. Estos datos nos permiten vincular las respuestas a los lugares y, por tanto, reproducir el funcionamiento de los cuestionarios en la dialectología clásica.

3. IMPORTANCIA DE LA COMUNICACIÓN Y PARTICIPACIÓN DEL PÚBLICO

El éxito de cualquier proyecto de *crowdsourcing* depende estrechamente de la publicidad que se le dé. *Dialectos del español* tuvo la suerte de recibir una cobertura mediática excepcional desde su lanzamiento en mayo de 2019, lo que se tradujo en un número récord de participantes en muy poco tiempo. De hecho, nada más ser anunciada por Lola Pons (2019) en un artículo (“Todos hablamos dialecto y no una lengua”) publicado en *Verne - El País* (28.5.2019), la información en la prensa y la radio españolas se multiplicó y la participación del público se disparó, alcanzando cotas muy superiores a la que habían logrado otras *apps* similares existentes. En un mes, *Dialectos del español* había obtenido más de 300 000 cuestionarios completados, mientras que *Dialäkt Äpp* y *English Dialects App* consiguieron 98 000 y 100 000 cuestionarios respectivamente en el primer año⁴.

⁴ Esta información nos fue proporcionada personalmente por D. Wanitsch en julio de 2019. Actualmente, la *app* ha recogido datos de más de 735.000 participantes (20.10.2022).

Hay que decir que si los medios de comunicación españoles amplificaron muy rápidamente el anuncio publicado en *El País* (entre otros *Heraldo de Aragón* el 29.5.2019, *El Confidencial* el 30.5.2019 y *ABC* el 31.5.2019 en formato impreso y digital; *Rioja 2* el 30.5.2019, *Aragón Radio* el 1.6.2019, *Cadena Ser* el 3.6.2019 y *La Cope* el 20.6.2019 en antena), a pesar de repetidos intentos por nuestra parte, no hemos tenido el mismo éxito con los medios americanos (salvo *Radio Canadá Internacional* el 19.07.2019 y *BBC Mundo* el 2.11.2020, Hernández Velasco 2020). Por ello, todavía el número de participantes americanos representa una parte limitada de los datos válidos recogidos hasta la fecha⁵. Esto nos conduce a poner en práctica estrategias de comunicación específicas para llegar a más participantes hispanoamericanos en la próxima etapa del proyecto.

Otro aspecto que es necesario cuidar es la comunicación de la propia *app*. En nuestro caso, elegimos para el logo la emblemática “ñ” del español y las mariposas naranjas y amarillas que remiten a la difusión y participación multitudinaria, además de evocar el realismo mágico de *Cien años de soledad*. En lo que se refiere a la página de inicio de la *app*, utilizamos el eslogan “Dime cómo hablas y te diré de dónde eres” —que resume bien el funcionamiento de la *app* y el intercambio entre el público y los investigadores— y un texto muy breve dirigido directamente al lector, con frases sugerentes para incitarle a participar, darle las instrucciones necesarias para limitar la paradoja del observador (“te pedimos que reflejes tu modo real de hablar”) y precisar que, cuando se vea ante varias respuestas posibles, deberá elegir la que más usa. En esta misma página de inicio, solicitamos al usuario que comparta la *app* en las redes sociales, lo que sin duda ha tenido un impacto nada desdeñable en su difusión. Por nuestra parte, creamos también una cuenta en Facebook / Meta y otra en Twitter para interactuar con el público: en ellas, por un lado, publicamos la revista de prensa de la *app* y otras informaciones sobre estudios dialectológicos y, por otro, respondemos a preguntas y expresamos nuestro agradecimiento a los participantes que comparten sus resultados de la *app* a través de los medios convencionales de esta red, enviándoles un *like*.

El hecho de que *Dialectos del español* haya superado las 735.000 participaciones (20.10.2022) prueba el gran interés que la sociedad muestra por la variación geográfica del español y la excelente acogida que hace a las herramientas de divulgación acerca de este tema. Es preciso señalar que, en este tipo de aplicaciones, una parte considerable de los datos recogidos no resulta utilizable para la investigación, sobre todo debido a que los participantes no completan toda la información que se les pide o introducen datos geográficos inválidos. En nuestro caso, por ejemplo, de los 302 497 cuestionarios del primer mes, el 51,3 % (155 191) resultaron inutilizables por faltar la información final sobre el lugar en el que nació, se crió y vive el participante. Estos cuestionarios no utilizables han tenido que ser eliminados, lo que ha representado un número impor-

⁵ Fueron menos del 15 % del total de 302 497 cuestionarios en el primer mes, pero la *app* se volvió a viralizar en noviembre de 2020 (tras la aparición de la noticia de *BBC Mundo*) y es muy probable que la participación americana contribuyera a ello (pronto dispondremos de esos datos).

tante de horas de trabajo para el equipo (un total de cuatro personas han intervenido en esta revisión en varias etapas).

En cuanto al perfil de los participantes de *Dialectos del español*, podemos decir que se trata en proporciones bastante equilibradas de hombres y mujeres⁶, pero que predominan los participantes jóvenes (la edad más representada es, de lejos, la de 19 a 35 años, seguida de 36 a 55⁷) y con estudios superiores⁸, como sucede también en proyectos similares (cf. Leemann/Kolly/Britain 2018).

En cuanto a la distribución geográfica de los participantes, en España es generalizada y en todas las provincias la muestra es cuantitativamente representativa del conjunto de la población, ya que supera ampliamente la cifra del 0,025 % —es decir, 25 hablantes por cada cien mil habitantes— establecida por Labov (1966: 170-171). En el continente americano, el número de participantes no alcanza todavía este umbral de representatividad, salvo en el caso de Costa Rica, lo que plantea el reto de realizar en el futuro una campaña más específica en ese continente para obtener un volumen de datos suficiente que permita una descripción y un análisis sobre una base cuantitativa adecuada.

4. AVANCE DE RESULTADOS Y RETOS DEL PROYECTO

En estos momentos aún estamos procesando y analizando los datos recogidos con la aplicación *Dialectos del español*, pero ya tenemos idea de algunos resultados, de los que daremos aquí tres ejemplos: uno referido al mundo hispanohablante (la respuesta al teléfono) y dos con respecto a usos en España (los posesivos en sintagmas adverbiales y verbales, y diversas formas verbales en las prótasis de oraciones condicionales irreales)⁹. A continua-

⁶ De los 125 715 cuestionarios con datos geográficos válidos del primer mes en España se desprende que la proporción de hombres es ligeramente superior a la de mujeres (67 153, 53,4 % y 58 103, 46,2 %, respectivamente) y que también aparece la opción “otro” (459, 0,4 %).

⁷ Por ejemplo, en España —de nuevo refiriéndonos a los 125 715 cuestionarios válidos del primer mes— 77 938 participantes (62 %) tienen entre 19 y 35 años y 34 540 (27,5 %) entre 36 y 55 años, cuando estas franjas etarias representan respectivamente el 18,9 % y el 32,3 % de la población española (INE, datos de 2018). Los mayores de 55 años, al contrario, están subrepresentados en los datos de la *app*, ya que de los 125 715 cuestionarios solo 8 190 (6,5 %) fueron completados por participantes de este grupo etario, aunque el 30,8 % de la población española pertenece a este estrato (INE, datos de 2018).

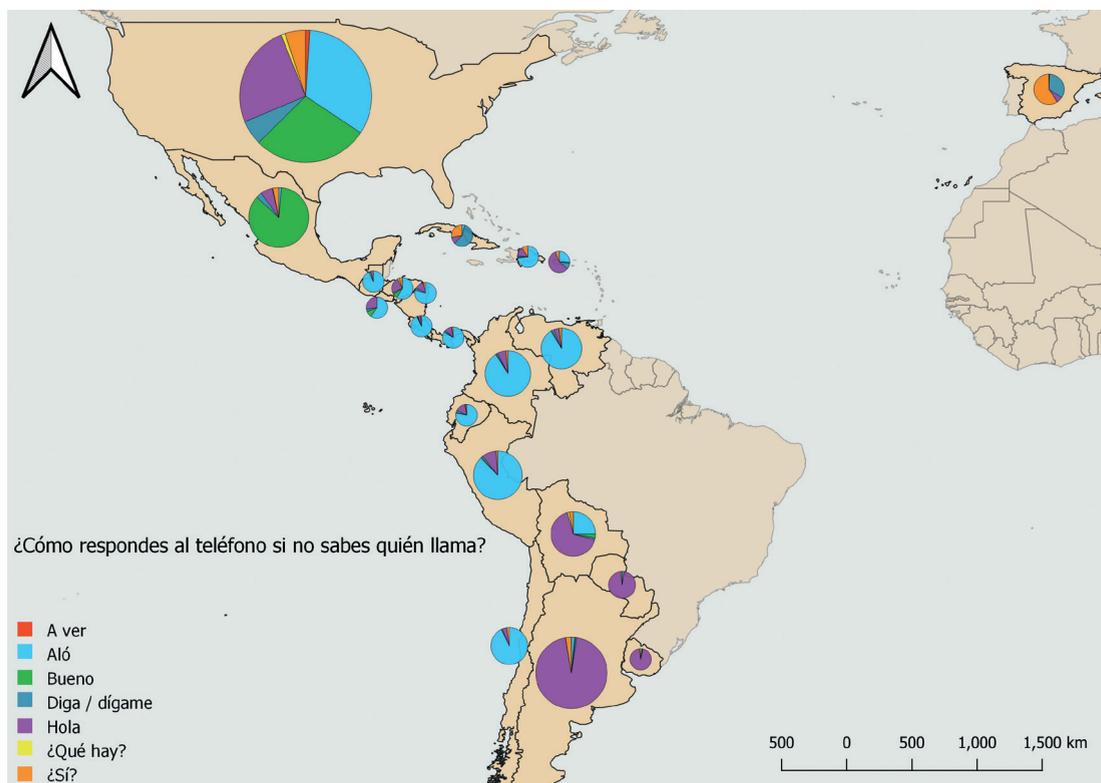
⁸ El 79,3 % de los cuestionarios válidos recogidos en el primer mes para todos los países (113 842/143 563) y el 78,2 % (98 325/125 715) de los cuestionarios de España. En comparación, según los datos del INE, solo el 34,6 % de los hombres y el 39,8 % de las mujeres poseían en 2019 educación superior. Por su parte, los participantes de nivel primario son muy excepcionales (1,3 %, 1 908/143 563) y están subrepresentados en los datos de la *app*.

⁹ El trabajo de extracción de datos fue posible gracias al apoyo financiero de las universidades de Gante, Montreal y Lausana. Los tres estudios de caso se basan en los datos válidos que fueron recogidos en el primer mes desde el lanzamiento de la *app* (hasta el 13.06.2019). Aparte de los cuestionarios sin datos geográficos utilizables, fueron también descartados los siguientes participantes: (i) los menores de edad, (ii) los no nativos de español y (iii) los migrantes internacionales. Para los resultados geográficos, nos centraremos de momento en los proporcionados para el lugar en que vive el participante.

ción, expondremos los retos del proyecto y las publicaciones previstas para divulgar los avances en nuestro conocimiento de la variación gramatical del español gracias a la *app*.

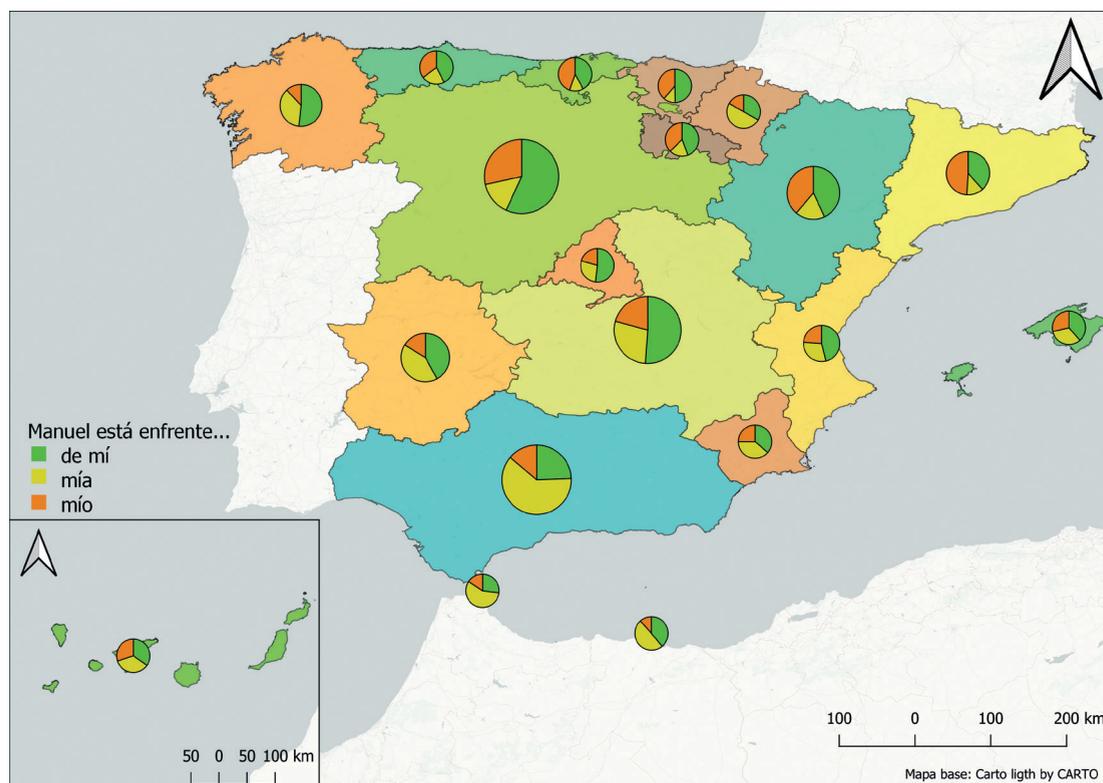
4.1. Algunos fenómenos investigados

Una pregunta como “¿Cómo respondes al teléfono si no sabes quién llama?” permite vislumbrar la variación interna de los distintos países, invisible, por ejemplo, en el DVD *Las voces del español: tiempo y espacio* (RAE/ASALE 2011), que incluye una grabación de audio con una única fórmula por país hispanohablante. El siguiente mapa muestra cómo entre las siete respuestas ¿*Qué hay?*, ¿*Sí?*, *A ver*, *Aló*, *Bueno*, *Diga / dígame*, *Hola*, varias se emplean y alternan en el mismo país. Esta variación ha sido descrita en la *Nueva gramática de la lengua española* (RAE/ASALE 2009: 2508), pero sin indicación de frecuencias y con alguna información no actualizada. Por ejemplo, para España se describe *diga / dígame* como la expresión por defecto, mientras que *sí*, como vemos en el mapa, parece ser ya hoy en día mayoritaria. Por otro lado, *a ver* se indica como fórmula propia de Colombia, pero no aparece por el momento en los datos de la *app* para este país (Bouzouita/Castillo Lluch/Pato 2021).



Mapa 1. La respuesta al teléfono en América y España

Un estudio realizado por Salgado y Bouzouita (2017) que abordó el uso de posesivos con adverbios en locuciones locativas en el español peninsular constató la falta de datos en corpus de referencia para casi todo el norte de España. Gracias a la pregunta “Manuel está *enfrente...* de mí / mío / mía...” de la aplicación, disponemos de un gran volumen de datos sobre la construcción con este adverbio que nos hace descubrir, en primer lugar, que el uso exclusivo de la variante estándar “Manuel está enfrente de mí” es minoritario en España (43,8 %, 37 284/85 084 vs. 56,2 %, 47 800/85 084, para el uso de una o más de las variantes no estándares). Un segundo hallazgo relativo a las variantes no estándares es que hay muchas diferencias diatópicas, como se observa en el mapa 2: mientras que en ciertas zonas predomina el complemento posesivo femenino (*enfrente mía*), en algunas otras se impone el equivalente masculino (*enfrente mío*) y en otras la construcción estándar (*enfrente de mí*). Las cifras más altas de *enfrente mía* se encuentran en Andalucía (45,6 %, 5 766/12 651), Navarra (34 %, 463/1 322), Extremadura (28,3 %, 671/2 370) y Galicia (25,9 %, 1 015/3 923), mientras que *enfrente mío* alcanza el 49,1 % (2 878/5 864) en Cataluña, el 44,8 % (664/1 483) en Cantabria, el 38,9 % (1 529/3 932) en el País Vasco, el 38,8 % (2 876/7 410) en Aragón, el 37,4 % (360/964) en La Rioja y el 35,2 % (1 082/3 078) en Asturias.



Mapa 2. El complemento del adverbio locativo *enfrente* en España

Si damos crédito a los usos declarados por los participantes, existiría un continuo del empleo de la variante posesiva en el norte de España: el morfema {a} parece, pues, un rasgo propio del noroeste y, curiosamente, de Navarra, al contrario del posesivo masculino, que predomina en las regiones nororientales y centrales. Esta observación es nueva, ya que estudios anteriores no proporcionan suficientes datos cuantitativos para llegar a conclusiones robustas sobre el uso del tipo morfológico de sufijo posesivo en las zonas septentrionales (Salgado/Bouzouita 2017).

Junto a estas estructuras adverbiales posesivas, también se ha documentado recientemente el uso de un complemento posesivo en el sintagma verbal: *él habla suyo, ella depende mío, tú gustas mío* (Bertolotti 2017, Bouzouita/Marttinen Larsson 2020, Bouzouita/Pato 2019, Casanova 2020). Este reciente fenómeno, específico del lenguaje hablado espontáneo, tampoco aparece en los corpus lingüísticos convencionales y obliga a los investigadores a recurrir a fuentes de información alternativas que contengan usos no normativos (blogs, foros y redes sociales como Twitter). En este caso, los datos aportados por *Dialectos del español* revelan que el complemento posesivo en lugar del régimen preposicional habitual del verbo *hablar* es extremadamente minoritario en España (solo el 2,7 % de los casos, 2 328/85 084) frente a la variante estándar (“Juan habla mal *de mí*” 97,3 %, 82 756/85 084), y que el uso de la variante con posesivo masculino (“Juan habla mal *mío*”) domina sobre la de posesivo femenino (1,2 %, 1 031/85 084 frente a 0,9 %, 751/85 084), mientras que el 0,6 % (546/85 084) admite ambos complementos posesivos. Galicia presenta el porcentaje más elevado de estas estructuras verbales con complemento posesivo —un 4,3 % de media (169/3 923)—, que llega al 6 % en algunas provincias (La Coruña, 98/1 625) y siempre con preferencia por el posesivo femenino (3,7 %, 60/1 625).

Estos últimos ejemplos muestran cómo una aplicación como *Dialectos del español* puede ser una herramienta muy poderosa para sondear de forma específica variantes gramaticales no estándares que normalmente no aparecen o apenas están documentadas en los corpus lingüísticos de referencia para la lengua española. La profundidad de esta encuesta no tiene parangón con los métodos tradicionales de recogida de datos dialectológicos y puede tener un alcance mayor que la estructura en cuestión: desde luego, con la pregunta *enfrente... de mí / mío / mía* obtenemos datos con una granularidad muy fina para el uso del posesivo en esta locución con este adverbio concreto y no con otros (*detrás, delante, encima...*), pero estos resultados ya nos informan de forma muy detallada sobre un caso del paradigma que puede servir para arrojar luz nueva sobre aspectos interesantes que se podrán explorar más adelante en el conjunto de este con una investigación más amplia. En este sentido, la *app* puede resultar muy valiosa como herramienta de preparación de trabajo de campo, particularmente para estudiar fenómenos no investigados hasta el momento. Su uso previo a una encuesta puede permitirnos conocer la posible extensión de los fenómenos estudiados, así como el perfil del hablante que los emplea (sexo, edad, nivel de instrucción), es decir, puede servir de información, revisión y evaluación preliminares a una investigación realizada con metodología tradicional.

La principal ventaja de la herramienta es, sin duda, su poder para proporcionar en poco tiempo datos masivos de los que antes no se disponía para una forma o estructura lingüística concreta, pero otra de sus cualidades radica en que puede adaptarse a la investigación de posibles correlaciones de fenómenos. Un caso específico en el contexto de nuestra aplicación es el de los complementos posesivos en los dominios adverbial y verbal, que acabamos de describir. El cruce de las variables externas (además de las coordenadas espaciales, la edad, el género, el nivel de educación, la movilidad a lo largo de la vida, la migración, el contacto lingüístico y el espacio urbano / rural) con las distintas respuestas a las variables lingüísticas también proporcionará información esencial para analizar los usos lingüísticos y los cambios en curso (Bouzouita/Castillo Lluch/Pato en preparación).

Otro fenómeno de interés dialectal es el estudiado a través de la pregunta “Si *tuviera / tuviese / tendría* dinero, viajaría”. Las respuestas a esta pregunta tienen dos puntos de interés. En primer lugar, conocer el empleo actual de la forma condicional (*-ría*) en lugar de las formas subjuntivas (*-ra* y *-se*). Y, en segundo lugar, conocer cuál de las dos formas del pretérito imperfecto de subjuntivo es la más empleada en las variedades del español de España. Con respecto a lo segundo, los datos indican que la forma en *-ra* es la preferida por los informantes (57,6 %, 49 023/85 084), seguida de *-se* (40,9 %, 34 812/85 084). Este hecho ya había sido mencionado por varios autores: “los datos cuantitativos dan, siempre en general [...] un claro predominio de las formas en *-ra*” (Rojo/Veiga 1999: 2910, n. 43). Conozcamos, a continuación, los datos en relación al uso de *-ría*, que fue la variante preferida del 1,5 % (1 249/85 084) de los participantes.

Por edad, son las personas mayores de 55 años (157/1 249), seguidos de los hablantes de 36-55 años (380/1 249) y luego los de 19-35 (712/1 249)¹⁰, los que más usan la forma en *-ría*. Para ilustrar por género, hombres y mujeres se comportan del mismo modo en el empleo de la forma del condicional. Por nivel educativo, son los informantes con educación primaria, seguidos de las personas con educación secundaria y los universitarios los que más usan *-ría*.

En cuanto a la movilidad, aquellos que nunca se han movido de su pueblo o ciudad emplean con más frecuencia la forma *-ría*, seguidos de los que han cambiado una sola vez. Las personas que se han movido dos o tres veces, o más de tres veces, emplean con mayor frecuencia la forma en *-ra*. A su vez, por lengua materna, los informantes que solo hablan español emplean más el condicional y la forma en *-ra*, frente a aquellos que tienen otra lengua materna además del español, que emplean con mayor frecuencia la forma en *-se*.

Desde el punto de vista geográfico, los datos de *Dialectos del español* confirman que el uso de *-ría* se mantiene a lo largo del tiempo, con porcentajes de empleo relevantes en La Rioja (22,9 %, 221/964), Navarra (14,9 %, 197/1 322), Vizcaya (8,9 %, 10

¹⁰ Téngase en cuenta que existe un desequilibrio en la muestra de la *app* con respecto a los distintos grupos etarios (véase nota 7). Las cifras 157, 380 y 712 deben ponderarse en relación con el volumen de cada uno de esos grupos.

209/2 351), Cantabria (8,2 %, 121/1 483), Palencia (7,3 %, 58/794), Álava (6,2 %, 43/690), Burgos (5,9 %, 76/1 271) y Guipúzcoa (5,4 %, 48/891); y menos relevantes en Soria (1,4 %, 6/444), Valladolid (0,5 %, 10/2 043), León (0,4 %, 8/1 934), Ávila (0,2 %, 1/418), Zamora (0,2 %, 1/586) y Segovia (0,2 %, 1/592). Estos resultados confirman que es un rasgo típico de los hablantes del español septentrional peninsular (Pato 2003, RAE/ASALE 2009: 1779). Además, lo interesante es que reflejan, 25-30 años después de los datos del COSER, la isoglosa presentada por primera vez (Pato 2003: 104), donde se muestran las tres áreas en la distribución de este uso del condicional en lugar de las formas del pretérito imperfecto de subjuntivo, entre el mar Cantábrico y el río Duero: (i) un trazado oriental que discurre entre la frontera navarro-aragonesa, (ii) un segundo trazado horizontal que atraviesa el sur de La Rioja y el norte de Soria hasta el área más septentrional de la provincia de Valladolid, siguiendo más o menos el cauce del río Duero, y (iii) un tercero y último trazo occidental que se desplaza por las fronteras naturales entre León y Palencia hasta llegar al mar Cantábrico.

4.2. Retos y próximos trabajos

Una aplicación como esta constituye un método de recogida de datos que debería utilizarse de forma complementaria a otros métodos más tradicionales de observación directa del uso espontáneo, ya que no hay que perder de vista que los datos obtenidos con *Dialectos del español*, o cualquier otra aplicación de este tipo, corresponden a un uso declarado (creencia de uso por parte del participante) y no real. Sin embargo, si no se excluye que los participantes no respondan a algunas preguntas reflejando más bien su conocimiento de los usos normativos que su forma real de hablar, el contexto de producir respuestas anónimas y estimulado por la curiosidad de comprobar hasta qué punto pueden obtener de la *app* una predicción precisa de su dialecto, les predispondría a responder de forma más bien realista, de acuerdo con la instrucción inicial “te pedimos que reflejes tu modo real de hablar”. Además, no podemos dejar de mencionar que, a diferencia de la recogida de datos mediante trabajo de campo tradicional, cara a cara con el informante, la obtención de datos con teléfonos móviles u ordenadores conlleva riesgos de respuestas desviadas de los usos reales por la rapidez de lectura de las preguntas, por la manipulación errónea de los dispositivos o por empleo inadecuado de la *app* (contestando aleatoriamente o la misma persona varias veces, por ejemplo). El gran volumen de datos puede compensar, sin embargo, las desviaciones producidas en estos casos.

Por otra parte, el hecho de que las respuestas abiertas sean muy contadas puede ser a veces una verdadera limitación de la aplicación y afectar a la calidad de los datos (el hablante que no encuentre su forma en la lista de respuestas responderá de la manera más cercana, pero la información no corresponderá a un uso “real”). No obstante, decidimos permitir muy pocas respuestas abiertas, ya que estas requieren un procesamiento de datos más sofisticado y laborioso que queríamos evitar en esta primera versión de la aplicación.

Nuestro equipo se enfrenta actualmente a varios retos. Como se ha mencionado anteriormente, todavía esperamos poder obtener con esta versión de la aplicación suficientes datos en América (una muestra que ascienda al menos al 0,025 % de la población de cada país) para permitir la comparación con los datos españoles. Esto requerirá mayores esfuerzos para difundir una información más eficaz sobre la aplicación en los países americanos. Al mismo tiempo, tenemos previsto ajustar algunos detalles, como la codificación de las respuestas, para mejorar el rendimiento de la predicción. Estas mejoras afectarán al contenido —gracias al *feedback* (comentarios explícitos) obtenido de algunos usuarios y a las respuestas ya almacenadas en nuestra base de datos podremos fijar mejor la geolocalización de algunas respuestas— y también a la forma de nuestro código —añadiendo algunas condiciones eliminatorias evitaremos algunas predicciones erróneas—.

Por otra parte, en este momento estamos preparando una monografía sobre la variación gramatical en el español europeo basada en datos recogidos en España (Bouzouita/Castillo Lluch/Pato en preparación), que se completará en una segunda fase con una publicación que presentará los resultados en América. Nuestra idea es difundir noticias divulgativas sobre todos estos resultados en la misma plataforma de la *app* y por sus cuentas en las redes sociales (Facebook / Meta y Twitter), de modo que el gran público pueda acceder fácilmente al conocimiento que nos ha brindado su participación, cumpliendo con los principios del *deber contraído* (Labov 1982: 172) y de *gratitud lingüística* (Wolfram 1993: 227). En el futuro, además, deseáramos desarrollar una versión específica para Hispanoamérica que nos permita explorar su variación gramatical con mayor detalle y que sea más eficaz en términos de predicción (actualmente la *app* solo es capaz de predecir regionalmente para España y nacionalmente para América). En ese momento será necesario diseñar un lenguaje informático más sofisticado que el de *Dialectos del español*, que de momento es básicamente plano. Idealmente, se tratará de producir una aplicación con un diseño arborescente, que permita a los participantes ser geolocalizados progresivamente mientras utilizan la aplicación y continuar con preguntas más específicas sobre la variedad detectada a cada paso. De hecho, este tipo de *app* puede hacerse todo lo sofisticada y especializada que el investigador desee y adaptarse al estudio de zonas con extensión geográfica muy diversa. Nada impide, además, incorporar archivos audio, como ya se ha practicado en otros proyectos de este tipo (cf. Kolly/Leemann 2015; Leemann/Kolly/Britain 2018).

5. CONCLUSIONES

En las páginas anteriores se ha presentado la primera aplicación para *smartphones* desarrollada para estudiar la variación dialectal de la lengua española y algunos ejemplos concretos de fenómenos investigados. *Dialectos del español*, que sigue el modelo de las aplicaciones *Dialäkt Äpp* y *App: English Dialects App* (utilizadas para el estudio del alemán de Suiza y del inglés británico, respectivamente), se centra en la variación gramatical del español europeo y americano. Mediante 26 preguntas, esta aplicación

recoge datos dialectales sobre variables gramaticales de las que tenemos más o menos conocimiento hasta la fecha, y a cambio ofrece a los participantes una predicción de su geolocalización lingüística y un momento lúdico.

Este artículo ha explicado el origen, los objetivos y el contenido de la aplicación, su funcionamiento técnico, la forma en que se ha difundido, la acogida que ha tenido entre el público y ha dado cuenta de sus ventajas a partir de la exposición de algunos ejemplos de los datos obtenidos. También se han discutido las limitaciones de la herramienta, así como las modificaciones previstas para mejorar esta *app* y para producir una futura versión más orientada a América. Por último, nos referimos a las publicaciones previstas con los *big data* obtenidos con *Dialectos del español* gracias a una amplísima participación del público, la cual demuestra el interés y la curiosidad que los hispanohablantes tienen hacia la variación dialectal de su lengua.

REFERENCIAS BIBLIOGRÁFICAS

- BERTOLOTI, Virginia (2017): “Pronombres posesivos y cambios gramaticales en español. Análisis en la variedad rioplatense”. En: Company Company, Concepción/Huerta Flores, Norohella (eds.): *La posesión en la lengua española*. Madrid: Consejo Superior de Investigaciones Científicas, pp. 325-349.
- BOUZOUITA, Miriam/CASTILLO LLUCH, Mónica/PATO, Enrique (2018): “*Dialectos del español*: una nueva aplicación para conocer la variación actual y el cambio en las variedades del español”. En: *Dialectología*, 20, pp. 63-85.
- (2021): “*Dialectos del español*: une application pour l’étude de la variation linguistique dans le monde hispanophone”. En: Thibault, André/Avanzi, Mathieu/LoVecchio, Nicholas/Millour, Alice (eds.): *Nouveaux regards sur la variation dialectale/New Ways of Analyzing Dialectal Variation*. Strasbourg: Éditions de linguistique et de philologie, pp. 291-303.
- (en preparación): *La variación gramatical en el español europeo, hoy. Datos de “Dialectos del español”*.
- BOUZOUITA, Miriam/MARTTINEN LARSSON, Matti (2020): “Variation and Change in the Romance Possessive Constructions: An Overview of Nominal, Adverbial and Verbal Uses”. En: Bouzouita, Miriam/Marttinen Larsson, Matti (eds.): *Possessive Constructions in Romance, Moderna språk*, 114, 3, pp. 1-44, <<https://ojs.ub.gu.se/index.php/modernasprak/article/view/5244>> (13/05/2022).
- BOUZOUITA, Miriam/PATO, Enrique (2019): “¿Por qué no gustas de mí como yo gusto de ti? El verbo *gustar* y el complemento preposicional en las variedades del español actual”. En: *Círculo de Lingüística Aplicada a la Comunicación*, 79, pp. 63-91. <<https://doi.org/10.5209/clac.65654>> (13/05/2022).
- BRITAIN, David (2016): “*Up, App and Away?*: Social Dialectology and the Use of Smartphone Technology as a Data Collection Strategy”, conferencia en el marco del *21 Sociolinguistics Symposium* (Universidad de Murcia, 16/07/2016).
- CASANOVA, Vanessa (2020): “El uso del complemento posesivo verbal por el complemento de régimen preposicional en español actual”. En: Bouzouita, Miriam/Marttinen Larsson, Matti (eds.): *Possessive Constructions in Romance, Moderna språk*, 114, 3, pp. 265-302. <<https://ojs.ub.gu.se/index.php/modernasprak/article/view/5252>> (13/05/2022).

- HILTON, Nanna Haug/LEEMANN, Adrian (eds.) (2021): *Using Smartphones to Collect Data for Linguistic Research, Linguistics Vanguard*, 7.sl. <<https://www.degruyter.com/journal/key/LINGVAN/7/s1/html>> (13/05/2022).
- HERNÁNDEZ VELASCO, Irene (2020): “Lengua, dialecto, geolecto y sociodialecto: ¿hay alguien que hable realmente español?”. En: *BBC News / Mundo*, 02/11/2020. <<https://www.bbc.com/mundo/noticias-53864492>> (13/05/2022).
- KOLLY, Marie-José/LEEMANN, Adrian (2015): “*Dialäkt Äpp*: Communicating Dialectology to the Public – Crowdsourcing Dialects from the Public”. En: Leemann, Adrian *et al.* (eds.): *Trends in Phonetics and Phonology. Studies from German-speaking Europe*. Bern: Peter Lang, pp. 271-285.
- LABOV, William (1966): *The Social Stratification of English in New York City*. Washington D.C.: Centre for Applied Linguistics.
- (1982): “Objectivity and Commitment in Linguistic Science: The Case of the English Trial in Ann Arbor”. En: *Language in Society*, 11, pp. 165-201.
- LEEMANN, Adrian (2021): “Apps for Capturing Language Variation and Change in German-speaking Europe: Opportunities, Challenges, Findings, and Future Directions”. En: *Linguistics Vanguard*, 7.sl, <<https://doi.org/10.1515/lingvan-2019-0022>> (13/05/2022).
- LEEMANN, Adrian/KOLLY, Marie-José/BITAIN, David (2018): “The English Dialects App: The Creation of a Crowdsourced Dialect Corpus”. En: *Ampersand*, 5, pp. 1-17.
- LEEMANN, Adrian *et al.* (2016): “Crowdsourcing Big Data in Dialectology – The Case of Swiss German”, conferencia presentada en el marco del *21 Sociolinguistics Symposium* (Universidad de Murcia, 17/07/2016).
- MARTTINEN LARSSON, Matti/BOUZOUTA, Miriam (2018): “*Encima de mí vs. Encima mío*: un análisis variacionista de las construcciones adverbiales locativas con complementos preposicionales y posesivos en Twitter”. En: *Moderna språk*, 112, 1, pp. 1-39.
- (2022): “Feminine Morphology in Possessive Complements of Adverbial Constructions in Andalusian Varieties”. En: *Círculo de Lingüística Aplicada a la Comunicación*, 91, pp. 157-171.
- PATO, Enrique (2003): *La sustitución del imperfecto de subjuntivo por el condicional simple y el imperfecto de indicativo en el castellano septentrional peninsular. (Estudio de variación dialectal)*. Madrid: Universidad Autónoma de Madrid.
- PONS, Lola (2019): “Todos hablamos dialecto y no una lengua”. En: *Verne – El País*, 28/5/2019. <https://verne.elpais.com/verne/2019/05/21/articulo/1558424530_527443.html> (13/05/2022).
- REAL ACADEMIA ESPAÑOLA/ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2009): *Nueva gramática de la lengua española: Morfología y sintaxis*. Madrid: Espasa.
- REAL ACADEMIA ESPAÑOLA/ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2011): *Nueva gramática de la lengua española: Fonética y fonología. DVD Las voces del español: tiempo y espacio*. Barcelona: Espasa.
- ROJO, Guillermo/VEIGA, Alexandre (1999): “El tiempo verbal. Los tiempos simples”. En: Bosque, Ignacio/Demonte, Violeta (dirs.): *Gramática descriptiva de la lengua española*. Madrid: Espasa, vol. 2, pp. 2867-2934.
- SALGADO, Hugo/BOUZOUTA, Miriam (2017): “El uso de las construcciones de adverbio locativo con pronombre posesivo en el español peninsular: un primer acercamiento diatópico”. En: *Zeitschrift für romanische Philologie*, 133, 3, pp. 766-794.
- WOLFRAM, Walt (1993): “Ethical Considerations in Language Awareness Programs”. En: *Issues in Applied Linguistics*, 4, 2, pp. 225-255.

| **Miriam Bouzouita** es catedrática de lenguas románicas (español) en la Humboldt-Universität zu Berlin (Alemania). Con anterioridad fue profesora en la Universiteit Gent (Bélgica) y Queen Mary University of London (Reino Unido). Ha publicado sobre temas históricos, diatópicos, sociolingüísticos y el contacto de lenguas. Ha coordinado varios proyectos internacionales, entre los cuales destaca el proyecto de creación de infraestructura de investigación en que se anota y parsea morfosintácticamente el corpus COSER (*Corpus oral y sonoro del español rural*).

| **Mónica Castillo Lluch** es catedrática de lingüística hispánica en la Université de Lausanne (Suiza) y antes ha sido profesora titular en Francia (Paris 8, Estrasburgo) y Alemania (Tubinga). Su investigación se centra en la edición y estudio de fueros y documentos antiguos, en la variación gramatical del español del pasado y actual y en la lingüística de la migración. Sus publicaciones están disponibles en <<https://www.unil.ch/esp/monicacastillolluch>>. Se ha formado como dialectóloga en el marco del COSER, participando en numerosas campañas de este proyecto, en el que ha coincidido y colaborado con Miriam Bouzouita y Enrique Pato.

| **Enrique Pato** es *professeur titulaire* (catedrático) en la Université de Montréal (Quebec, Canadá). Doctor en Filología española por la Universidad Autónoma de Madrid, su campo de investigación se centra en el estudio de la gramática de las variedades y dialectos del español. Investigador principal en cinco proyectos internacionales, ha coeditado varias monografías y tiene publicados más de un centenar de trabajos (capítulos de libros, artículos en revistas nacionales e internacionales y actas).