



*DIALECTOS DEL ESPAÑOL:*  
**APRESENTAÇÃO DA APLICAÇÃO  
E PRIMEIROS RESULTADOS**

Miriam Bouzouita  
Mónica Castillo Lluch  
Enrique Pato

***DIALECTOS DEL ESPAÑOL:***  
**APRESENTAÇÃO DA APLICAÇÃO E PRIMEIROS RESULTADOS**

*DIALECTOS DEL ESPAÑOL:*  
PRESENTATION OF THE APP AND FIRST RESULTS

Miriam Bouzouita  
Humboldt-Universität zu Berlin

Mónica Castillo Lluch  
Université de Lausanne

Enrique Pato  
Université de Montréal

**Resumo**

Este artigo apresenta *Dialectos del español*, uma aplicação para *smartphones* destinada ao estudo da variação morfossintática da língua espanhola, lançada em maio de 2019 e disponível em [www.dialectosdelespanol.org](http://www.dialectosdelespanol.org) e *Google Play*. Descrever-se-ão os aspetos científicos e técnicos do projeto (como surgiu, os seus objetivos, as 26 perguntas da aplicação, a codificação das respostas e o funcionamento da previsão), relatar-se-á a sua divulgação nos meios de comunicação e a participação do público, e apresentar-se-ão alguns resultados e os desafios atuais do projeto.

**Palavras-chave:** Aplicação para *smartphone*, variação morfossintática, *Big data*, vialetologia digital.

**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](http://www.dialectosdelespanol.org) and Google Play. It presents the scientific and technical aspects 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. It further presents some results and the current challenges of the project.

**Keywords:** Smartphone app, morphosyntactic variation, Big data, digital dialectology.

## 1. INTRODUÇÃO

No século XXI, as ferramentas digitais abriram novas formas de acesso a dados linguísticos que são objeto de estudos variacionistas. Muitas investigações são agora empiricamente apoiadas por redes sociais (especialmente Twitter e Facebook/Meta), blogs, fóruns e outras fontes escritas e orais disponíveis na Internet. Além disso, cada vez mais aplicações de *smartphones* ou inquéritos digitais estão a ser concebidos para recolher *big data* e construir *corpora* através do *crowdsourcing*. Na dialetologia, o uso da *citizen science*, promovendo a participação do público em geral na investigação, é uma continuação das práticas tradicionais na disciplina desde o seu início. Mas se isto sempre exigiu a colaboração dos falantes que possuem os conhecimentos e usos linguísticos, o que está agora a mudar é que a participação necessária é maciça e fornece uma enorme quantidade de informação.

Entre as aplicações dialetológicas para *smartphones* estão as pioneiras *Dialäkt Äpp* e *Voice Äpp* (Kolly & Leemann, 2015; Leemann, 2021) desenvolvidas para o estudo dos dialetos suíços alemães, bem como o *English Dialects App* (Leemann et al., 2018), que investiga o inglês britânico. Na sequência destas, outras surgiram nos últimos anos, relatadas na recente edição especial de *Linguistics Vanguard* editada por Hilton e Leemann (2021).

Este artigo apresenta *Dialectos del español*, uma aplicação para *smartphone* para o estudo dialetológico da língua espanhola, lançada em maio de 2019 e disponível em [www.dialectosdelespanol.org](http://www.dialectosdelespanol.org) e Google Play. Em primeiro lugar (§ 2), descrever-se-á como surgiu, os seus objetivos e as 26 perguntas que a aplicação contém, bem como os aspetos técnicos do projeto (a codificação das respostas e o funcionamento da previsão). Em segundo lugar (§ 3), relatar-se-á a importância da comunicação para alcançar a participação do público e o perfil dos participantes. Por fim (§ 4), apresentar-se-á uma previsão dos primeiros resultados, bem como os desafios atuais do projeto. As conclusões resumem o conteúdo destas páginas.

## 2. ORIGEM, OBJETIVOS E ASPETOS TÉCNICOS DA APLICAÇÃO

No 21.º Simpósio de Sociolinguística da *Linguistic Society of America*, realizado na Universidade de Múrcia em 2016, tivemos a oportunidade de descobrir duas das aplicações acima mencionadas: *Dialäkt Äpp*, concebida para o estudo de variedades suíças alemãs, e *English Dialects App*, para o estudo de variedades do inglês britânico (ver Leemann et al., 2016; Britain, 2016). Estas aplicações baseiam-se numa série de perguntas fonéticas e lexicais, os componentes mais tradicionalmente estudados da língua, e o seu principal objetivo é recolher

*big data* que facilite o conhecimento da evolução de certas variáveis destas línguas previamente estudadas noutros projetos de dialetologia.

Conhecendo estes projetos, pareceu muito interessante desenvolver uma aplicação semelhante para o espanhol, mas com o objetivo de estudar a variação gramatical, o aspeto menos explorado pela dialetologia hispânica, e com a aspiração, além disso, de chegar a todo o mundo hispanófono.

Para começar, concebemos um estudo piloto que nos permitiu testar um conjunto de perguntas através de um formulário web (*Google Forms*), preenchido por 547 participantes de vários países de língua espanhola. Em Bouzouita, Castillo Llach e Pato (2018), expusemos os objetivos e a metodologia do nosso projeto, que nessa altura ainda se encontrava em fase de conceção, e apresentámos os resultados obtidos com o conjunto de perguntas do formulário inicial. Graças a esse estudo preliminar, fomos capazes de decidir quais as variáveis linguísticas a selecionar para as 26 perguntas da *app*, identificámos alguns problemas que deveríamos tentar evitar (por exemplo, na redação das perguntas tivemos de descartar termos lexicais não comuns a Espanha e América para não produzir estranheza numa parte dos participantes) e fomos capazes de ter em conta outros pedidos dos utilizadores (por exemplo, incluir *outro* como terceira opção de género, juntamente com *homem* e *mulher*).

Com esta primeira experiência e a intervenção técnica do mesmo informático que concebeu as aplicações da equipa de Adrian Leemann e David Britain, desenvolvemos a *app Dialectos del español*, disponível em [www.dialectosdelespanol.org](http://www.dialectosdelespanol.org) e *Google Play*<sup>1</sup>. A aplicação está alojada na plataforma *Wix* e os dados recolhidos são armazenados num servidor da Universidade de Lausanne, o que garante a segurança e durabilidade da sua conservação.

O nosso principal objetivo é obter dados que nos permitam estudar certos fenómenos de variação morfossintática do espanhol em todo o mundo de língua espanhola<sup>2</sup>. Alguns destes fenómenos já foram descritos e analisados no passado e, neste caso, o objetivo é verificar se estas descrições correspondem à utilização atual. Outros fenómenos só foram descritos parcialmente e o desafio é completar a sua descrição e análise. Finalmente, algumas das variáveis gramaticais da nossa lista não foram investigadas até agora, pelo que nos propomos explorá-las pela primeira vez.

Estas três categorias de perguntas com as variáveis que estudámos são as seguintes:

---

<sup>1</sup> Gostaríamos de agradecer a Adrian Leemann pela sua disponibilidade e atitude aberta para partilhar connosco informações valiosas para desenvolver o nosso projeto.

<sup>2</sup> As perguntas feitas aos participantes e as respostas fornecidas para a previsão da sua geolocalização têm em conta as variantes encontradas na Europa e na América, mas também aspiramos à participação de falantes de espanhol de África e da Ásia.

(1) Perguntas clássicas (sobre variáveis já abordadas pela dialetologia tradicional):

- Diminutivos
- *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) Perguntas menos clássicas (sobre variáveis para as quais existe alguma informação mas sem resultados para todo o mundo de língua espanhola):

- *Comprásemos / compráramos*
- *Tuviere / 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) Perguntas inéditas (sobre variáveis não estudadas até agora):

- *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*

O princípio deste tipo de aplicações é pedir a colaboração do público em troca de uma recompensa lúdica: com base nas respostas, a *app* tenta adivinhar a origem dialetal dos participantes. Portanto, um segundo objetivo do nosso projeto é identificar corretamente os dialetos dos participantes a partir das suas respostas. O desafio não é trivial na medida em que muitas das variáveis do nosso estudo não permitem, de todo, a geolocalização dos falantes (a maioria das perguntas “inéditas” e “menos clássicas”) e, quando permitem a geolocalização, é demasiado vaga (por exemplo, os diminutivos, *el sartén*, *léismo*, *haber* concordado com um objeto plural, *voseo*, *más nada*, *ya yo...*). As respostas gramaticais que apontam para um país americano ou para uma área da Península Ibérica são finalmente excepcionais: *(vos) bailái(s) conmigo* (Chile, Venezuela), *Se conocieron en Bogotá fue* (República Dominicana), *enfrente mía* (Galiza, Andaluzia, Madrid).

Este contrato com o público, que aparece na página inicial da aplicação (“Dime cómo hablas y te diré de dónde eres”), obrigou-nos a incluir algumas questões lexicais que são muito mais finamente orientadas para um país concreto e, por vezes, até para uma região específica. Existem apenas três perguntas lexicais: a fórmula utilizada para responder ao telefone e as palavras equivalentes a *ervilha* e *lagarto* (com base numa imagem da semente e do animal, respetivamente).

Por trás das 26 perguntas, a aplicação contém uma codificação geográfica das respostas, utilizada para prever o dialeto do participante, que poderia ser estabelecida com base na informação contida na *Nueva gramática de la lengua española* (RAE / ASALE, 2009), vários questionários dos atlas linguísticos, monografias e artigos especializados sobre cada pergunta. Três exemplos específicos desta codificação são mostrados abaixo:

(1) ... el agua es potable.

- (a) Esta...
- (b) Este...
- (c) Esta / Este...

(2) ¿La sartén o el sartén?

- (a) La sartén [ESPANHA]
- (b) El sartén [AMÉRICA]
- (c) La / el sartén [AMÉRICA]

(3) Manuel está enfrente...

- (a) de mí. [AMÉRICA: México; AMÉRICA: Guatemala; AMÉRICA: Honduras; AMÉRICA: Nicarágua; AMÉRICA: El Salvador; AMÉRICA: Costa Rica; AMÉRICA: Panamá; AMÉRICA: República Dominicana; AMÉRICA: Porto Rico; AMÉRICA: Venezuela; AMÉRICA: Colômbia; AMÉRICA: Estados Unidos; ESPANHA: Aragón; ESPANHA: Ilhas Baleares; ESPANHA: Astúrias; ESPANHA: Cantábria; ESPANHA: Castela e Leão; ESPANHA: La Rioja; ESPANHA: Múrcia; ESPANHA: Navarra; ESPANHA: País Basco; ESPANHA: Castela-La Mancha].
- (b) mío. [ESPANHA: Galiza; ESPANHA: Andaluzia; ESPANHA: Madrid; AMÉRICA: Argentina; AMÉRICA: Uruguai; AMÉRICA: Chile; AMÉRICA: Equador; AMÉRICA: Peru; AMÉRICA: Bolívia; AMÉRICA: Paraguai; ESPANHA: Catalunha; ESPANHA: Comunidade Valenciana; AMÉRICA: Panamá].
- (c) mía. [ESPANHA: Galiza; ESPANHA: Andaluzia; ESPANHA: Madrid].
- (d) de mí / mío. [ESPANHA: Galiza; ESPANHA: Andaluzia; ESPANHA: Madrid; AMÉRICA: Argentina; AMÉRICA: Uruguai; AMÉRICA: Chile; AMÉRICA: Equador; AMÉRICA: Peru; AMÉRICA: Bolívia; AMÉRICA: Paraguai; ESPANHA: Catalunha; ESPANHA: Comunidade Valenciana; AMÉRICA: Panamá; AMÉRICA: Costa Rica].
- (e) de mí / mía. [ESPANHA: Galiza; ESPANHA: Andaluzia; ESPANHA: Madrid].

- (f) *mío / mía*. [ESPAÑA: Galiza; ESPAÑA: Andalucía; ESPAÑA: Madrid].
- (g) *de mí / mía / mía*. [ESPAÑA: Galiza; ESPAÑA: Andalucía; ESPAÑA: Madrid].

Como podemos ver, a geocodificação é por vezes impossível (exemplo 1), uma vez que nenhum estudo anterior descreve a difusão de formas femininas ou masculinas do demonstrativo antes dos substantivos que começam com /á/; noutros casos, é muito vaga (como em 2)<sup>3</sup>, e apenas quando temos informações muito precisas sobre um fenómeno (como em 3, cf. Salgado & Bouzouita, 2017; Martinen Larsson & Bouzouita, 2018) pode realmente contribuir para a previsão do dialeto dos participantes.

O princípio de funcionamento interno da *app* para esta previsão é simples: a soma de pontos por país americano ou província/comunidade espanhola, de modo que, por exemplo, se um participante reagir com as respostas codificadas “[AMÉRICA]” e as que se seguem, a aplicação somará pontos para América e Colômbia e, finalmente, identificará o participante 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)*

Pouco depois do lançamento da *app*, percebemos que o seu algoritmo aditivo era por vezes insuficiente e podia levar a graves erros de geolocalização, o que nos levou a acrescentar algumas condições eliminatórias no cálculo da geolocalização. Estas condições, implementadas pela primeira vez neste tipo de aplicação, são três: 1) se for utilizado o *voseo* pronominal ou verbal, 2) se a posição do pronome do sujeito não for invertida em interrogativas (¿Qué *tú* dices?) e 3) se for utilizado o *ser* focalizador, a Espanha é excluída da previsão.

<sup>3</sup> Simplificámos deliberadamente neste caso e omitimos considerar que em alguns dialetos do sul de Espanha (andaluz, canário) *sartén* também pode ser masculino, uma vez que esta forma não é normalmente utilizada pela maioria dos utilizadores da *app*, que são estudantes universitários (ver *abaixo*). No entanto, a combinação desta característica com outras no questionário pode prever que o utilizador da *app* é de facto andaluz ou canário.

Após o questionário e a nossa previsão da origem dialetal do participante (3 geolocalizações oferecidas por ordem de probabilidade), pedimos alguns dados extralinguísticos: o seu género (masculino / feminino / outro), a sua idade (-18 / 18-35 / 36-55 / +55), o seu nível de educação (primário, secundário, universitário), se são ou não falantes nativos de espanhol (e, em caso negativo, qual é a sua língua materna) e se são falantes nativos de outra língua para além do espanhol. Finalmente, pedimos a ajuda do participante para melhorar a *app*, pedindo-lhe para localizar num mapa o local onde nasceu, onde cresceu e onde vive atualmente. Também perguntamos quantas vezes mudou de cidade (nunca / 1 vez / 2-3 vezes / mais de 3 vezes) e de país (*idem*) durante a sua vida. Além disso, pode classificar com 5 estrelas (em 10 segmentos) a precisão da previsão da aplicação, pode deixar comentários e partilhar a aplicação no Facebook/Meta e Twitter. Para os nossos dois objetivos (obter dados dialetais e para que a *app* dê a melhor previsão possível), logicamente apenas os questionários com o mapeamento dos participantes e dados extralinguísticos são úteis. Estes dados permitem-nos relacionar as respostas aos locais e assim reproduzir o funcionamento dos questionários na dialetologia clássica.

### 3. IMPORTÂNCIA DA COMUNICAÇÃO E DA PARTICIPAÇÃO DO PÚBLICO

O sucesso de qualquer projeto de *crowdsourcing* depende estreitamente da publicidade que lhe for dada. *Dialectos del español* teve a sorte de receber uma cobertura mediática excepcional desde o seu lançamento em maio de 2019, o que resultou num número recorde de participantes num período de tempo muito curto. De facto, assim que foi anunciada por Lola Pons (2019) num artigo (“Todos hablamos dialecto y no una lengua”) publicado em *Verne – El País* (28.5.2019), a informação na imprensa e rádio espanholas multiplicou-se e a participação do público aumentou, atingindo níveis muito mais elevados do que tinha sido alcançado por outras aplicações semelhantes existentes. No espaço de um mês, *Dialectos del español* teve mais de 300.000 questionários preenchidos, enquanto *Dialäkt Äpp* e *English Dialects App* tiveram 98.000 e 100.000 questionários, respetivamente, no primeiro ano<sup>4</sup>.

Os meios de comunicação espanhóis ampliaram muito rapidamente o anúncio publicado em *El País* (entre outros *Heraldo de Aragón* em 29.5.2019, *El Confidencial* em 30.5.2019 e *ABC* em 31.5.2019 em formato impresso e digital; *Rioja 2* em 30.5.2019, *Aragón Radio* em 1.6.2019, *Cadena Ser* em 3.6.2019 e *La Cope* em 20.6.2019 no ar), mas apesar das repetidas tentativas, não conseguimos a mesma difusão nos meios de comunicação social

---

<sup>4</sup> Esta informação foi-nos fornecida pessoalmente pelo nosso informático, que também desenvolveu as outras aplicações, em julho de 2019. Atualmente, a aplicação recolheu dados de 638.113 participantes (24.03.2022).

americanos (exceto na *Radio Canada International* em 19.07.2019 e na *BBC Mundo* em 2.11.2020, Hernández Velasco 2020). Por conseguinte, o número de participantes americanos ainda representa uma parte limitada dos dados válidos recolhidos até à data<sup>5</sup>. Isto leva-nos a implementar estratégias de comunicação específicas para alcançar mais participantes hispano-americanos na próxima fase do projeto.

Outro aspeto que precisa de ser tratado é a comunicação da própria *app*. No nosso caso, escolhemos para o logótipo o emblemático “ñ” espanhol e as borboletas laranja e amarelas, que se referem à disseminação e participação em massa, e evocam o realismo mágico de *Cien años de soledad*. Quanto à página inicial da *app*, utilizámos o slogan “Dime cómo hablas y te diré de dónde eres” – que resume bem como funciona a aplicação e o intercâmbio entre o público e os investigadores – e um texto muito breve dirigido diretamente ao leitor, com frases sugestivas para encorajá-lo a participar e dar-lhe as instruções necessárias para limitar o paradoxo do espectador (“te pedimos que reflejes tu modo real de hablar”) e para especificar que, quando confrontados com várias respostas possíveis, devem escolher a que mais utilizam. Na mesma página inicial, pedimos ao utilizador que partilhe a *app* em redes sociais, o que certamente teve um impacto significativo na sua divulgação. Quanto a nós, criámos também uma conta Facebook/Meta e uma conta Twitter para interagir com o público: por um lado, publicamos a revista de imprensa da *app* e outras informações sobre estudos dialetológicos e, por outro lado, respondemos a perguntas e agradecemos aos participantes que partilham os seus resultados da aplicação.

O facto de *Dialectos del español* já ter ultrapassado (24.03.2022) 638.113 participações é prova do grande interesse demonstrado pela sociedade na variação geográfica do espanhol e da excelente receção que dá aos instrumentos de divulgação de informação sobre este tema.

Contudo, é de notar que, neste tipo de aplicações, uma parte considerável dos dados recolhidos não é utilizável para a investigação, principalmente devido ao facto dos participantes não completarem toda a informação solicitada ou introduzirem dados geográficos inválidos. No nosso caso, por exemplo, dos 302.497 questionários no primeiro mês, 51,3% (155.191) provaram não ser utilizáveis, porque a informação final sobre onde o participante nasceu, cresceu e viveu estava em falta. Estes questionários inutilizáveis tiveram de ser eliminados, o que representou um número significativo de horas de trabalho para a equipa (um total de quatro pessoas estiveram envolvidas nesta revisão em várias fases).

---

<sup>5</sup> Foi menos de 15% do total de 302.497 questionários no primeiro mês, mas a *app* voltou a ser viral em novembro de 2020 (após o aparecimento da notícia em BBC Mundo) e é muito provável que a participação americana tenha contribuído (teremos esses dados em breve).

Quanto ao perfil dos participantes de *Dialectos del español*, podemos dizer que estão em proporções bastante equilibradas entre homens e mulheres<sup>6</sup>, mas que os jovens participantes predominam (de longe a faixa etária mais representada é de 19-35 anos, seguida de 36-55<sup>7</sup> anos) e com formação superior<sup>8</sup>, como é também o caso em projetos semelhantes (cf. Leemann et al., 2018).

Quanto à distribuição geográfica dos participantes, em Espanha é generalizada e em todas as províncias a amostra é quantitativamente representativa da população como um todo, pois está bem acima do valor de 0,025% –ou seja, 25 falantes por cem mil habitantes– estabelecido por Labov (1966, pp. 170-171). No continente americano, o número de participantes ainda não atinge este limiar de representatividade, exceto no caso da Costa Rica, o que coloca o desafio de realizar uma campanha mais específica nesse continente no futuro, a fim de obter um volume de dados suficiente para permitir uma descrição e análise numa base quantitativa adequada.

#### 4. PROGRESSO DOS RESULTADOS E DESAFIOS DO PROJETO

Neste momento, ainda estamos a processar e a analisar os dados recolhidos com a aplicação *Dialectos del español*, mas já temos uma ideia de alguns resultados, dos quais daremos aqui três exemplos: um referente ao mundo de língua espanhola (atendimento telefónico) e dois referentes a usos em Espanha (possessivos em sintagmas adverbiais e verbais, e várias formas verbais na prótase de condicionais irreais)<sup>9</sup>. A seguir, apresentaremos os desafios do projeto e as publicações previstas para divulgar os avanços no nosso conhecimento da variação gramatical em espanhol, graças à *app*.

<sup>6</sup> Dos 125.715 questionários com dados geográficos válidos no primeiro mês em Espanha, a proporção de homens é ligeiramente superior à das mulheres (67.153, 53,4% e 58.103, 46,2% respetivamente) e a opção *outro* também aparece (459, 0,4%).

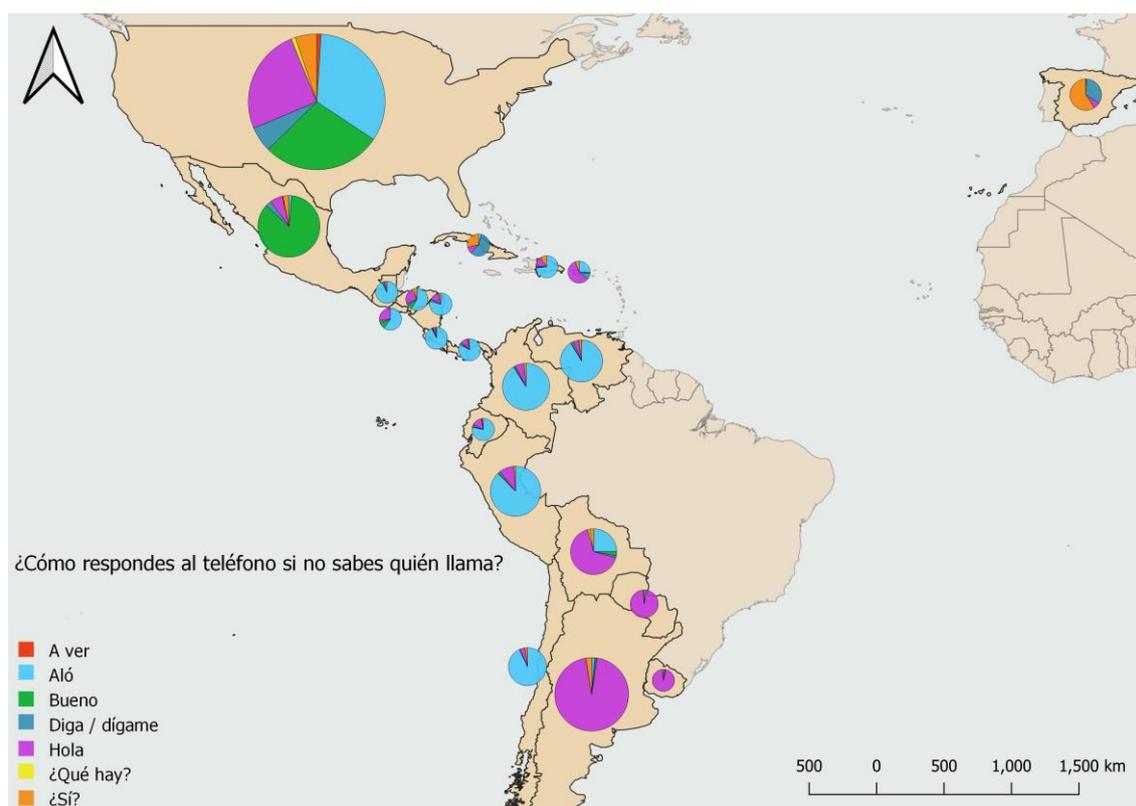
<sup>7</sup> Por exemplo, em Espanha –referindo-se novamente aos 125.715 questionários válidos no primeiro mês– 77.938 participantes (62 %) têm entre 19 e 35 anos e 34.540 (27,5 %) entre 36 e 55 anos, quando estes grupos etários representam respetivamente 18,9 % e 32,3 % da população espanhola (INE, dados de 2018). As pessoas com mais de 55 anos, pelo contrário, estão sub-representadas nos dados da *app*, pois dos 125.715 questionários apenas 8.190 (6,5 %) foram preenchidos por participantes desta faixa etária, embora 30,8 % da população espanhola pertença a este estrato (INE, dados de 2018).

<sup>8</sup> 79,3 % dos questionários válidos recolhidos no primeiro mês para todos os países (113.842/143.563) e 78,2% (98.325/125.715) dos questionários para Espanha. Em comparação, de acordo com dados do INE, apenas 34,6% dos homens e 39,8% das mulheres tinham formação superior em 2019. Por outro lado, os participantes do nível primário são muito excecionais (1,3 %, 1.908/143.563) e estão sub-representados no nosso corpus.

<sup>9</sup> O trabalho de extração de dados foi possível graças ao apoio financeiro das universidades de Ghent e Montreal. Os três estudos de caso baseiam-se em dados válidos que foram recolhidos no primeiro mês desde o lançamento da *app* (até 13.06.2019). Além dos questionários sem dados geográficos utilizáveis, os seguintes participantes foram também descartados: (i) menores, (ii) falantes não nativos de espanhol e (iii) migrantes internacionais. Para os resultados geográficos, concentrar-nos-emos por agora naqueles previstos para o local onde o participante vive.

#### 4.1. Alguns fenómenos investigados

Uma pergunta como “¿Cómo respondes al teléfono si no sabes quién llama?” dá um vislumbre da variação interna entre países, invisível, por exemplo, no DVD *Las voces del español: tiempo y espacio* (RAE/ASALE, 2011), que inclui uma gravação áudio com uma única fórmula por país de língua espanhola. O mapa seguinte mostra como entre as sete respostas *¿Qué hay?*, *¿Sí?*, *A ver*, *Aló*, *Bueno*, *Diga / dígame*, *Hola*, várias são utilizadas e alternadas no mesmo país. Esta variação foi descrita na *Nueva gramática de la lengua española* (RAE/ASALE, 2009, p. 2508), mas sem qualquer indicação de frequência e com alguma informação desatualizada. Por exemplo, para Espanha, *diga / dígame* é descrito como a expressão padrão, enquanto *sí*, como podemos ver no mapa, parece ser a mais comum hoje em dia. Por outro lado, *a ver* é indicado como uma fórmula específica para a Colômbia, mas não aparece de momento nos dados da *app* para este país (Bouzouita et al., 2021), onde *aló* é a resposta mais comum.



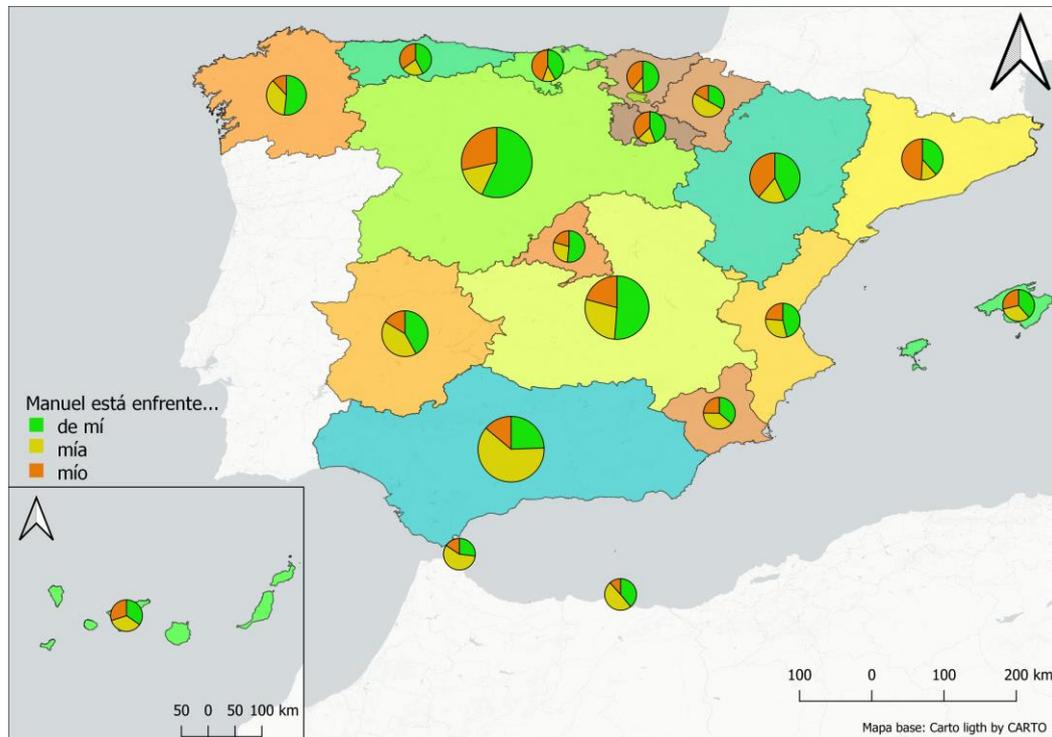
Mapa 1. Atender o telefone na América e Espanha

Relativamente a *a ver*, é impressionante que os poucos casos registados para Espanha (314) estejam acumulados no Norte, especificamente no País Basco (concretamente nas

províncias de Guipúzcoa e Biscaia) e nas Astúrias, o que corresponde a um uso tradicional que talvez esteja a ganhar prestígio escondido entre um setor mais jovem da população. Em relação às diferentes utilizações entre gerações, tendo em conta os dados recolhidos na *app*, podemos afirmar que se, em Espanha, a resposta mais comum é *sí* é porque é promovida pelos mais jovens (19-35 anos: 67,6 %, 37.444 / 55.373), uma vez que a sua utilização diminui à medida que a idade aumenta (36-55 anos: 46,2 %, 11.013 / 23.825 e +55 anos: 35,3 %, 2.079 / 5.886). Pelo contrário, o uso de *diga* / *dígame* é mais comum entre os maiores de 55 anos (60,5 %, 3.560 / 5.886) e progressivamente menos entre os mais jovens (36-55 anos: 48,5 %, 11.548 / 23.825 e 19-35 anos: 24,3 %, 13.444 / 55.373). Outro facto revelador é que *hola* é utilizado em Espanha com mais frequência pelos mais jovens do que pelos mais velhos (19-35 anos: 7,4 %, 4.101 / 55.373, 36-55 anos: 4,6 %, 1.093 / 23.825 e +55 anos: 3 % 181 / 5.886). Há sem dúvida vários fatores que influenciam estas mudanças. Por um lado, não é surpreendente que os mais jovens tendam a evitar uma forma que implica o uso dos imperativos (*diga* / *dígame*) e, por outro lado, que estes mesmos interlocutores usem uma fórmula informal de saudação (*hola*) para estabelecerem contacto com um interlocutor não identificado, adotando a mesma expressão que usariam quando o dispositivo revela o nome do interlocutor. Em relação aos dados obtidos para esta pergunta em Espanha, deveríamos acrescentar que teria valido a pena oferecer a possibilidade de escolher a fórmula combinada *sí dígame*, uma vez que nos teria permitido verificar se a sua frequência de utilização excede a do simples imperativo neste país, o que, por experiência, suspeitamos.

Um estudo de Salgado e Bouzouita (2017), que abordou a utilização de possessivos com advérbios em locuções locativas em espanhol peninsular, notou a falta de dados em *corpora* de referência para quase todo o norte de Espanha. Agora, no entanto, graças à pergunta “Manuel está *enfrente...* de *mí* / *mío* / *mía* / *mía...*” da *app*, temos um grande volume de dados sobre a construção com este advérbio que nos leva a descobrir, em primeiro lugar, que o uso exclusivo da variante *standard* “Manuel está *enfrente de mí*” é minoritário em Espanha (43,8 %, 37.284/85.084 dos participantes utilizam a variante padrão vs. 56,2 %, 47.800/85.084 que utilizam uma ou mais das variantes padrão). Uma segunda descoberta relativa às variantes não-*standard* é que existem muitas diferenças diatópicas, como mostra o Mapa 2: enquanto que em algumas áreas predomina o complemento possessivo feminino (*enfrente mía*), noutras o equivalente masculino (*enfrente mío*) e noutras a construção padrão (*enfrente de mí*) prevalece. Os números mais elevados de *enfrente mía* encontram-se na Andaluzia (45,6 %, 5.766/12.651), Navarra (34 %, 463/1.322), Extremadura (28,3 %, 671/2.370) e Galiza (25,9 %, 1.015/3.923), enquanto que *enfrente mío* atinge 49,1 % (2.878/5.864) na

Catalunha, 44,8 % (664/1.483) na Cantábria, 38,9 % (1.529/3.932) no País Basco, 38,8 % (2.876/7.410) em Aragão, 37,4 % (3/60.964) em La Rioja e 35,2 % (1.082/3.078) nas Astúrias.



**Mapa 2.** O complemento do adverbial locativo *enfrente* em Espanha.

Considerando o acima exposto, parece haver um *continuum* do uso da variante possessiva no norte de Espanha: o possessivo feminino {a} parece, então, ser uma característica do Noroeste e, curiosamente, de Navarra, ao contrário do possessivo masculino que predomina nas regiões centrais e do Nordeste. Esta observação é nova, uma vez que estudos anteriores não fornecem dados quantitativos suficientes para chegar a conclusões sólidas sobre a utilização do tipo morfológico do sufixo possessivo nas zonas do Norte (Salgado & Bouzouita, 2017).

A par destas estruturas adverbiais possessivas, o uso de um complemento possessivo na frase verbal também foi documentado recentemente: *él habla suyo, ella depende mío, tú gustas mío* (Bertolotti, 2017; Bouzouita & Marttinen Larsson, 2020; Bouzouita & Pato, 2019; Casanova, 2020). Esta inovação, específica da linguagem espontânea falada, também não aparece nos *corpora* linguísticos convencionais e obriga os investigadores a recorrer a fontes alternativas de informação contendo usos não-normativos (blogues, fóruns e redes sociais como o Twitter). Neste caso, os dados fornecidos por *Dialectos del español* revelam que o complemento possessivo no lugar do regime preposicional habitual do verbo *hablar* é

extremamente raro em Espanha (apenas 2,7% dos casos, 2.328/85.084) em comparação com a variante padrão (“Juan habla mal *de mí*” 97,3%, 82.756/85.084), e que a utilização da variante com possessivo masculino (“Juan habla mal *mío*”) domina sobre a variante com possessivo feminino (1,2%, 1.031/85.084 em comparação com 0,9 %, 751/85.084), enquanto que 0,6 % (546/85.084) admitem ambos os complementos possessivos. A Galiza tem a maior percentagem destas estruturas verbais com complementos possessivos – uma média de 4,3% (169/3.923) –, atingindo 6% em algumas províncias (A Coruña, 98/1.625) e sempre com preferência pelo possessivo feminino (3,7 %, 60/1.625).

Estes últimos exemplos mostram como uma aplicação como *Dialectos del español* pode ser uma ferramenta muito poderosa para sondar especificamente variantes gramaticais não padronizadas que, normalmente, não estão ou dificilmente estão documentadas nos *corpora* linguísticos de referência para a língua espanhola. A profundidade deste levantamento é inigualável pelos métodos tradicionais de recolha de dados dialetológicos e pode ter um âmbito mais vasto do que a estrutura em questão: evidentemente, com a pergunta *enfrente... de mí / mío / mía* obtemos dados específicos para o uso do possessivo nesta locução com este advérbio concreto e não com outros (*detrás, delante, encima...*). Mas estes resultados já nos informam de forma muito detalhada sobre um caso do paradigma que pode servir para lançar uma nova luz sobre aspetos interessantes que podem ser explorados no futuro em todo este processo com uma investigação mais ampla.

A principal vantagem do instrumento é sem dúvida o seu poder de fornecer, num curto espaço de tempo, dados massivos anteriormente indisponíveis para uma forma ou estrutura linguística específica, mas outra das suas qualidades reside no facto de poder ser adaptado à investigação de possíveis correlações de fenómenos. Um caso específico no contexto da nossa aplicação é o dos complementos possessivos nos domínios adverbial e verbal, que acabámos de descrever. O cruzamento de variáveis externas (para além das coordenadas espaciais, idade, género, nível de educação, mobilidade, migração, contacto linguístico e espaço urbano/rural), com as diferentes respostas às variáveis linguísticas, fornecerá também informação essencial para analisar os usos linguísticos e as mudanças em curso (Bouzouita et al., em preparação).

Outro fenómeno de interesse dialetal é o que se estuda através da pergunta “Si tuviera / tuviese / tendría dinero, viajaría”. As respostas a esta pergunta têm dois pontos de interesse. Em primeiro lugar, conhecer a utilização atual da forma condicional (*-ría*) em vez das formas subjuntivas (*-ra* e *-se*). Em segundo lugar, para descobrir qual das duas formas do pretérito imperfeito do subjuntivo é a mais comumente utilizada nas variedades de espanhol em

Espanha. Em relação a estas últimas, os dados indicam que a forma *-ra* é preferida pelos inquiridos (57,6%, 49.023/85.084), seguida de *-se* (40,9%, 34.812/85.084). Este facto já tinha sido mencionado por vários autores: “los datos cuantitativos dan, siempre en general [...] un claro predominio de las formas en *-ra*” (Rojo & Veiga, 1999, p. 2910, n. 43).

Passemos agora aos dados sobre a utilização da forma *-ría*, que era a variante preferida de 1,5% (1.249/85.084) dos participantes.

Por idade, as pessoas com mais de 55 anos (157/1.249), seguidas pelas de 36-55 anos (380/1.249) e depois as de 19-35 anos (712/1.249)<sup>10</sup>, são as que mais utilizam a forma em *-ría*. Para ilustrar por género, homens e mulheres comportam-se da mesma forma na utilização da forma condicional. Por nível educacional, os informantes com educação primária, seguidos pelos com educação secundária e os com educação universitária são os que mais usam *-ría*.

Em termos de mobilidade, aqueles que nunca se mudaram da sua aldeia ou cidade usam com mais frequência a forma *-ría*, seguidos por aqueles que se mudaram apenas uma vez. Aqueles que se mudaram 2 ou 3 vezes, ou mais de 3 vezes, usam com mais frequência a forma *-ra*. Por sua vez, por língua materna, os inquiridos que só falam espanhol usam mais frequentemente a forma condicional e a forma *-ra*, em comparação com aqueles que têm outra língua materna para além do espanhol, que usam mais frequentemente a forma *-se*.

De um ponto de vista geográfico, os dados de *Dialectos del español* confirmam que a utilização da forma *-ría* se mantém ao longo do tempo, com percentagens relevantes de utilização em La Rioja (22,9 %, 221/964), Navarra (14,9 %, 197/1.322), Biscaia (8,9 %, 209/2.351), Cantábria (8,2%, 121/1.483), Palencia (7,3%, 58/794), Álava (6,2 %, 43/690), Burgos (5,9 %, 76/1.271) e Guipúzcoa (5,4 %, 48/891); e menos relevante em Soria (1,4%, 6/444), Valladolid (0,5 %, 10/2.043), Leão (0,4 %, 8/1.934), Ávila (0,2 %, 1/418), Samora (0,2%, 1/586) e Segovia (0,2 %, 1/592). Estes resultados confirmam que esta é uma característica típica dos falantes de espanhol do Norte da Península (Pato, 2003; RAE/ASALE, 2009, p. 1779). Além disso, o que é interessante é que refletem, 25-30 anos após os dados do projeto COSER, a isoglossa apresentada pela primeira vez (Pato, 2003, p. 104). Esta mostra as três áreas na distribuição da utilização do condicional em vez das formas do pretérito imperfeito do subjuntivo, entre o mar Cantábrico e o rio Douro: (i) um traçado oriental entre a fronteira entre Navarra e Aragão, (ii) um segundo traçado horizontal que atravessa o sul de La Rioja e o norte de Soria até à zona mais a norte da província de

---

<sup>10</sup> Note-se que existe um desequilíbrio na amostra da *app* em relação aos diferentes grupos etários (ver nota 7). Os valores 157, 380 e 712 devem ser ponderados em relação ao volume de cada um destes grupos.

Valladolid, mais ou menos seguindo o curso do rio Douro e (iii) um terceiro e último traçado ocidental ao longo das fronteiras naturais entre Leão e Palencia até chegar ao mar Cantábrico.

#### 4.2 Desafios e futuros trabalhos

Idealmente, se o tempo e os recursos permitirem, uma aplicação como esta deveria ser utilizada para recolher dados de forma complementar aos métodos mais tradicionais de observação direta de uso espontâneo, uma vez que não se deve perder de vista o facto de que os dados recolhidos com *Dialectos del español*, ou qualquer outra *app* deste tipo, correspondem à utilização declarada (crença de utilização por parte do participante) e não à utilização real. Contudo, embora não seja impossível que alguns participantes possam responder a certas perguntas refletindo os seus conhecimentos sobre usos normativos em vez da sua forma de falar real, o contexto de produzir respostas anónimas e estimuladas pela curiosidade de verificar até que ponto podem obter da *web-app* uma previsão exata do seu dialeto, predispô-los-ia a responder de forma bastante realista, de acordo com a instrução inicial “te pedimos que reflejes tu modo real de hablar”. Por outro lado, o facto de as respostas abertas serem em número muito reduzido pode, por vezes, constituir uma verdadeira limitação da aplicação e afetar a qualidade dos dados (o falante que não encontrar a sua forma na lista de respostas escolherá a forma mais semelhante, mas a informação não corresponderá a uma utilização “real”). No entanto, decidimos permitir muito poucas respostas abertas, pois estas requerem um processamento de dados mais sofisticado e moroso que queríamos evitar nesta primeira versão da aplicação.

A nossa equipa está atualmente a enfrentar vários desafios. Como mencionado acima, ainda esperamos poder obter com esta versão da *app* dados suficientes na América (uma amostra de pelo menos 0,025% da população em cada país), para permitir a comparação com os dados espanhóis. Isto exigirá esforços adicionais para divulgar informação mais eficaz sobre a aplicação do outro lado do Atlântico. Ao mesmo tempo, planeamos ajustar alguns detalhes, tal como a codificação das respostas, para melhorar a previsão geográfica. Estas melhorias afetarão o conteúdo – graças ao *feedback* (comentários explícitos) obtido de alguns utilizadores e às respostas já armazenadas na nossa base de dados, poderemos definir melhor a geolocalização de algumas respostas – mas também a forma do nosso código – adicionando algumas condições eliminatórias, evitaremos algumas previsões erradas.

Por outro lado, estamos atualmente a preparar uma monografia sobre a variação gramatical do espanhol europeu baseada em dados recolhidos em Espanha (Bouzouita et al., em preparação), que será completada numa segunda fase com uma publicação que

apresentará os resultados na América. No futuro, gostaríamos também de desenvolver uma versão específica para este continente que nos permita explorar a sua variação gramatical em maior detalhe e ser mais eficaz em termos de previsão (atualmente a *app* só é capaz de fazer previsões a nível regional para Espanha e a nível nacional para a América). Nesse momento, será necessário conceber uma linguagem informática mais sofisticada que a de *Dialectos del español*, que por enquanto é basicamente plana. Idealmente, o objetivo será produzir uma aplicação com um desenho em forma de árvore, permitindo aos participantes serem progressivamente geolocalizados à medida que utilizam a aplicação e continuarem com perguntas mais específicas sobre a variedade detetada em cada etapa.

## 5. CONCLUSÕES

As páginas anteriores apresentaram a primeira aplicação de *smartphone* desenvolvida para estudar a variação dialetal na língua espanhola e alguns exemplos concretos dos fenómenos investigados. *Dialectos del español*, modelada nas aplicações *Dialäkt Äpp* e *English Dialects App* (utilizadas para o estudo do suíço alemão e do inglês britânico, respetivamente), centra-se na variação gramatical do espanhol europeu e americano. Através de 26 perguntas, a *app* recolhe dados dialetais sobre variáveis gramaticais, das quais temos mais ou menos conhecimento até à data e, em troca, oferece aos participantes uma previsão da sua geolocalização linguística e um momento lúdico.

Este artigo explica a origem, objetivos e conteúdo da aplicação, como funciona tecnicamente, como foi divulgada, como foi recebida pelo público e dá conta das suas vantagens, mostrando alguns exemplos dos dados obtidos. As limitações da ferramenta são também discutidas, bem como as modificações planeadas para melhorar esta *app* e para produzir uma versão futura mais orientada para a América. Finalmente, referimo-nos às publicações previstas com os *big data* obtidos com *Dialectos del español* graças a uma grande participação do público, o que põe em evidência o interesse e a curiosidade que os falantes de espanhol demonstram em relação à variação dialetal da sua língua.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Bertolotti, V. (2017). Pronombres posesivos y cambios gramaticales en español. Análisis en la variedad rioplatense. In C. Company Company & N. Huerta Flores (eds.), *La posesión en la lengua española* (pp. 325-349). Consejo Superior de Investigaciones Científicas.

- Bouzouita, M., Castillo Lluch, M. & Pato, E. (2018). *Dialectos del español*: Una nueva aplicación para conocer la variación actual y el cambio en las variedades del español, *Dialectología*, 20, 63-85.
- Bouzouita, M., Castillo Lluch, M. & Pato, E. (2021). *Dialectos del español*: Une application pour l'étude de la variation linguistique dans le monde hispanophone. In A. Thibault, M. Avanzi, N. LoVecchio & A. Millour (eds.), *Nouveaux regards sur la variation dialectale/New Ways of Analyzing Dialectal Variation* (pp. 291-303). Éditions de linguistique et de philologie.
- Bouzouita, M., Castillo Lluch, M. & Pato, E. (em preparação). La variación gramatical en el español europeo, hoy, *Datos de Dialectos del español*.
- Bouzouita, M. & Marttinen Larsson, M. (2020). Variation and change in the Romance possessive constructions: An overview of nominal, adverbial and verbal uses. In M. Bouzouita & M. Marttinen Larsson (eds.), *Possessive Constructions in Romance. Moderna språk*, 114 (3), 1-44.  
<https://ojs.ub.gu.se/index.php/modernasprak/article/view/5244>
- Bouzouita, M. & Pato, E. (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. *Círculo de Lingüística Aplicada a la Comunicación*, 79, 63-91. <https://doi.org/10.5209/clac.65654>
- Britain, D. (2016). *Up, app and away?: Social dialectology and the use of smartphone technology as a data collection strategy* [conferência 16/07/2016]. 21<sup>st</sup> Sociolinguistics Symposium, Universidad de Murcia.
- Casanova, V. (2020). El uso del complemento posesivo verbal por el complemento de régimen preposicional en español actual. In M. Bouzouita & M. Marttinen Larsson (eds.), *Possessive Constructions in Romance. Moderna språk*, 114 (3), 265-302.  
<https://ojs.ub.gu.se/index.php/modernasprak/article/view/5252>
- Hilton, N. H. & Leemann, A. (2021). Using Smartphones to Collect Data for Linguistic Research, *Linguistics Vanguard*, 7, s1.  
<https://www.degruyter.com/journal/key/LINGVAN/7/s1/html>
- Hernández Velasco, I. (2020). *Lengua, dialecto, geolecto y sociodialecto: ¿Hay alguien que hable realmente español?* BBC News / Mundo (02/11/2020).  
<https://www.bbc.com/mundo/noticias-53864492>
- Kolly, M. J. & Leemann, A. (2015). *Dialäkt Äpp*: Communicating dialectology to the public – crowdsourcing dialects from the public. In A. Leemann et al. (eds.), *Trends in Phonetics and Phonology. Studies from German-speaking Europe* (pp. 271-285). Peter Lang.
- Labov, W. (1966). *The Social Stratification of English in New York City*. Centre for Applied Linguistics.
- Leemann, A. (2021). Apps for capturing language variation and change in German-speaking Europe: Opportunities, challenges, findings, and future directions. *Linguistics Vanguard*, 7, s1. <https://doi.org/10.1515/lingvan-2019-0022>
- Leemann, A. et al. (2016). *Crowdsourcing Big Data in dialectology – the case of Swiss German* [conferência 17/07/2016]. 21<sup>st</sup> Sociolinguistics Symposium, Universidad de Murcia.
- Leemann, A., Kolly, M. J. & Britain, D. (2018). The English Dialects App: The creation of a crowdsourced dialect corpus. *Ampersand*, 5, 1-17.
- Marttinen Larsson, M. & Bouzouita, M. (2018). *Encima de mí vs. encima mío*: Un análisis variacionista de las construcciones adverbiales locativas con complementos preposicionales y posesivos en Twitter. *Moderna språk*, 112(1), 1-39.

- Martinen Larsson, M. & Bouzouita, M. (no prelo). Feminine morphology in possessive complements of adverbial constructions in Andalusian varieties. *Círculo de Lingüística Aplicada a la Comunicación*.
- Pato, E. (2003). *La sustitución del imperfecto de subjuntivo por el condicional simple y el imperfecto de indicativo en el castellano septentrional peninsular*. Universidad Autónoma de Madrid.
- Pons, L. (2019). *Todos hablamos dialecto y no una lengua*. Verne – El País (28/05/2019). [https://verne.elpais.com/verne/2019/05/21/articulo/1558424530\\_527443.html](https://verne.elpais.com/verne/2019/05/21/articulo/1558424530_527443.html)
- RAE/ASALE = Real Academia Española/Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española: Morfología y Sintaxis*. Espasa.
- RAE/ASALE = 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*. Espasa.
- Rojo, G. & Veiga, A. (1999). El tiempo verbal. Los tiempos simples. In I. Bosque & V. Demonte (dirs.), *Gramática descriptiva de la lengua española* (vol. 2, pp. 2867-2934). Espasa.
- Salgado, H. & Bouzouita, M. (2017). El uso de las construcciones de adverbio locativo con pronombre posesivo en el español peninsular: un primer acercamiento diatópico. *Zeitschrift für romanische Philologie*, 133(3), 766-794.