

An interactive linguistic atlas of Andalusian accents (ALIAA): methodology

Alfredo Herrero de Haro

Universidad de Granada, España / University of Wollongong, Australia
ahh@ugr.es ORCID ID: <https://orcid.org/0000-0002-7660-3532>

Submitted: 12/03/2024; Accepted: 24/05/2024; Published online: 21/11/2024

Citation: Herrero de Haro, A. (2024). An interactive linguistic atlas of Andalusian accents (ALIAA): methodology. *Loquens*, 11(1-2), e108. <https://doi.org/10.3989/loquens.2024.e108>

ABSTRACT: The sixth volume of ALEA (*Linguistic and Ethnographic Atlas of Andalusia*) was published in 1973 but the materials were collected in the 1950s; recent studies contradict ALEA's results, which suggests that its content needs to be updated. The present article outlines ALIAA's (*Atlas Lingüístico Interactivo de los Acentos de Andalucía*) methodology. ALIAA aims to update the phonetic and phonological description from ALEA by describing the accent of 500 towns. The data are being collected primarily through online surveys. The audio will be analysed acoustically and statistically, and the results will be connected to a specific point on an interactive map, allowing isoglosses to be created to delimit certain phenomena. ALIAA can help us understand the extent of phonetic phenomena across Andalusia and how their geographical distribution has changed since ALEA. This article also presents some preliminary results to show that ALIAA's methodology can help us develop linguistic atlases in a more effective way than traditional dialectology methods.

Keywords: linguistic atlas; accent variation; interactive linguistic atlas; online data collection.

RESUMEN: *Atlas Lingüístico Interactivo de los Acentos de Andalucía (ALIAA): metodología.* El sexto volumen del ALEA (*Atlas Lingüístico y Etnográfico de Andalucía*) se publicó en 1973 pero los materiales se recogieron en los 1950; algunos estudios recientes contradicen las descripciones del ALEA, lo que parece demostrar que es necesario actualizarlo. Este artículo describe la metodología del ALIAA (*Atlas Lingüístico Interactivo de los Acentos de Andalucía*), un proyecto que pretende actualizar los datos del ALEA describiendo el acento de 500 municipios; los datos se están recogiendo principalmente en línea. El audio se analizará acústica y estadísticamente y los resultados se conectarán a un punto en un mapa interactivo; esto creará isoglosas que delimitarán fenómenos lingüísticos en un mapa interactivo. El ALIAA puede explicar la extensión de distintos fenómenos lingüísticos y cómo ha variado su extensión desde que se creara el ALEA. Este artículo presenta algunos resultados preliminares para demostrar que una metodología como la del ALIAA puede ayudarnos a desarrollar atlas lingüísticos de una manera más efectiva que métodos tradicionales.

Palabras clave: atlas lingüístico; variación de acentos; atlas lingüístico interactivo; recogida de datos en línea.

1. INTRODUCTION

Linguistic atlases contain a great amount of information about language variation but they are very costly, lengthy, and demanding ventures. It took approximately 20 years for a team of three linguists to complete the *Linguistic and Ethnographic Atlas of Andalusia* (ALEA, as per its Spanish acronym). The atlas covered 230 points across Andalusia, it was published in 6 volumes and, although the data were collected between 1953 and 1958, the last volume, which is the one which deals with phonetics and phonology, was not published until 1973. Those data are already 70 years old and recent studies (e.g. Herrero de Haro and Hajek 2022; Regan 2017) have shown that the current distribution of certain linguistic phenomena does not correspond with the isoglosses presented by Alvar et al. (1973) in ALEA, which shows the need to update the atlas. Apart from the fact that the data from Alvar et al. (1973) are 70 years old, it is worth mentioning that the speakers used for the atlas were, on average, 50 years of age, which means that they were born around 1900 - 1910. With this in mind, we can say that Alvar et al. (1973) captured the speech of 20-year-olds from the 1920s.

The discrepancies between the data presented in Alvar et al. (1973) and recent studies have motivated a project called *Atlas Lingüístico Interactivo de los Acentos de Andalucía* (ALIAA) ('Interactive Linguistic Atlas of Andalusian Accents'). ALIAA aims to analyse the accent of 500 towns across Andalusia and to represent that variation in a series of interactive maps. This project also seeks to study how Andalusian accents have changed since the data for Alvar et al. (1973) were collected in the 1950s.

This article describes the methodology developed for ALIAA so that a similar process can be carried out to develop linguistic atlases for other accents or languages in other parts of the world. It also presents the results from some preliminary analyses.

This project is being developed primarily with data gathered through online surveys and it combines traditional dialectology with current methods in phonetics to analyse language variation. Apart from the linguistic value of new descriptions of Andalusian accents, it is expected that a methodology like the one developed for ALIAA can help produce linguistic atlases much cheaper, quicker, and in more effective ways than traditional methods.

1.1. Background

Alvar et al. (1973) contains a total of 1900 linguistic maps and 210 of those focus on phonetic and phonological variation across 230 towns in Andalusia. However, although Alvar et al. (1973)

was published in 1973, the elaboration of the atlas commenced in 1961, and the data collection was completed between 1953 and 1958 by three researchers who visited 230 towns. The researchers interviewed a speaker per town, and it was almost exclusively a male. Their ideal participant was a male speaker who had never left the town, who was between 40 and 60 years of age, and whose family was also from the same town; this is what Chambers and Trudgill (1988: 29) describe as *nonmobile, older, rural males*, or "NORM" for short. It was important to have a set of teeth as complete as possible and it was preferred if the participant was illiterate, to avoid any influence of spelling over his pronunciation. Participants were shown a series of pictures and were given some prompts to utter certain words; the interviewers transcribed the words down based on an impressionistic analysis. Alvar et al. (1973) was published as a collection of loose maps in a format slightly bigger than A3. Nowadays, the data can be collected remotely using an online survey, the analysis of phonetic features can be carried out by software, the maps can be stored in digital format on a website which can be accessed from anywhere in the world, and users can listen to recording samples of the participants to appreciate the nuances of regional accents. Technology has allowed the creation of various electronic atlases in recent years (e.g. Boula de Mareuil et al. 2018; Boula de Mareuil et al. 2021); however, to the author's knowledge, no atlas has been created to date with the level of phonetic analysis which technology allows.

Andalusia, an autonomous region in southern Spain, is divided linguistically into Eastern Andalusian Spanish (EAS) and Western Andalusian Spanish (WAS); these two varieties are differentiated by vowel lowering in the former. However, there are other sources of phonetic variation in Andalusian Spanish. For example, *distinción* (maintaining the contrast /s/ - /θ/), *seseo* (pronouncing /s/ and /θ/ as [s]), and *ceceo* (pronouncing /s/ and /θ/ as [θ]) (Alvar et al. 1973: Map 1705), the pronunciation of /x/ as [h] or [x] (Alvar et al. 1973: Map 1716), and the pronunciation of /tʃ/ as [tʃ] or [ʃ] (Alvar et al. 1973: Map 1709). Furthermore, some recent studies have identified other interesting phenomena in Andalusian Spanish, such as some covariants of gemination (Herrero de Haro and Hajek 2023).

There are some studies which have reviewed the literature regarding Andalusian Spanish (e.g. Herrero de Haro 2017; Mondéjar Cumpián 2006). A review of the literature shows that most studies on Andalusian Spanish have focused on similar issues. In EAS, the focus has been on analysing vowel lowering processes (e.g. Henriksen 2017; Martínez Melgar 1994), gemination (e.g. Herrero de Haro and

Hajek 2023), and vowel harmony processes (e.g. Jiménez and Lloret 2020). In WAS, the affrication of the /st/ sequence has been widely studied (e.g. Ruch and Harrington 2014). Of course, other features have also been studied, such as the perception of certain features; however, those have been studied to a lesser degree. Taking this into account, it seems justified to say that a more thorough analysis of the phonetic qualities and geographical extension of different features of Andalusian accents is needed.

Other studies have also compared atlases to identify phonetic variation across time. For example, Loporcaro et al. (2021) analysed the AIS atlas of the 1920s and updated it with data from the 2010s. Amongst other changes, they saw that [ʃ] had changed into [tʃ]. Something similar has happened in Andalusia, as research shows that, in some parts of Andalusia, [ʃ] has been replaced by [tʃ] (Melguizo Moreno 2007). Amongst other factors, Loporcaro et al. (2021) point towards normalised spelling or pressure from the spelling system as a possible explanation of pronunciation changes; this could also be causing changes in Andalusian Spanish.

The discrepancy between the descriptions from Alvar et al. (1973) and recent studies also shows the need to update the linguistic atlas of Andalusia. For example, Regan (2017) studies *ceceo* in Huelva, and Herrero de Haro and Hajek (2022) describe the accent of Eastern Andalusia, and both works show how the extension of the spread of *ceceo* differs from what Alvar et al. (1973) describe in their atlas. The literature on Andalusian Spanish is extensive and new studies show how the descriptions contained in Alvar et al. (1973) need to be updated. Furthermore, technology and new advances in linguistics, such as software to analyse audio acoustically, allow for a much deeper description of phonetic features. All this justifies the creation of a new linguistic atlas of Andalusian accents. Some recent atlases have taken advantage of recent technological advances but it is believed that linguistic atlases can push this further and, as such, ALIAA aspires to be an example of what can be achieved with current technology.

1.2. Aims and objectives

ALIAA aims to analyse the accent of 500 towns across Andalusia and represent the results from language variation analyses through interactive maps which can be interpreted easily by non-specialists. With this in mind, the present article has two main objectives:

- 1) To introduce the ALIAA project and its methodology so that similar projects can be carried out for other regions, countries, and languages.

- 2) To present some preliminary analyses as a way of proving the effectiveness of the methodology.

2. METHODOLOGY

2.1. Initial design

ALIAA was designed to update the data from Alvar et al. (1973) and to extend the analysis of phonetic and phonological variation in Andalusia.

The first step was to analyse Alvar et al. (1973) both, in regards to what type of information it contained, and how the information had been presented. A list of the words and phenomena analysed in Alvar et al. (1973) was collated. The list of phenomena was extended to incorporate linguistic phenomena which have been reported after Alvar et al. (1973) (e.g. covariants of gemination in EAS, as per Herrero de Haro and Hajek 2023).

After careful consideration, it was decided to analyse the following features:

- All phonemes word-initially, word-medially and word-finally
- Vowels before /-s/, /-r/, /-θ/, and /-n/
- Vowel harmony processes
- Gemination (word-initially and word-medially)
- Pronunciation of /x/
- The contrast /k/ - /j/ word-initially and word-medially
- Maintenance of the /s/ - /θ/ contrast or the merging into /θ/ (*ceceo*) or into /s/ (*seseo*)
- Types of /s/
- Types of /θ/
- The contrast /r/ - /l/ in coda
- Vowel lowering in Andalusia
- Unstressed /-as/
- Pronunciation of /a/ as [e] before deleted /-s/, /-l/, /-r/, and /-θ/
- Types of /j/
- Types of /tʃ/
- Aspiration of /p/, /t/, and /k/
- Word-initial Latin /f/
- /sb/, /sd/, /sg/, and /sj/

Once a list of phenomena was prepared, a few words which could be used to investigate each phenomenon were selected. The initial list contained over 400 words; however, as the survey was going to be run online, the list was considered too long. To shorten the word list, it was decided to use words which could investigate more than one phenomenon at the same time. For example, the word *osos* ‘bears’ was chosen over *pocos* ‘few’ as it could be used to investigate /-os/, vowel harmony, maintenance of /s/

or *ceceo*, and types of /s/. Words were also chosen based on phonetic principles. For example, it was preferred to have vowels next to voiceless stops to aid segmentation; nasals and laterals were only used when strictly necessary, as it is not as easy to set boundaries between these consonants and vowels. The final list contained 123 words and a 119-word version of *The North Wind and the Sun*. Alvar et al. (1973) contains a total of 378 maps covering different aspects of language variation. It covers the pronunciation of vowels and consonants, phonetic and phonological phenomena, nominal morphology, verbal morphology, personal pronouns, syntax, and phraseology. ALIAA only covers the pronunciation of vowels and consonants and phonetic, and phonological phenomena; however, other researchers can adapt the methodology and tools created for this project to analyse other aspects of language (e.g. lexical variation across Andalusia).

Alvar et al. (1973) gathered data from 230 towns in face-to-face interviews; ALIAA aims to gather data from 500 out of the 785 local councils in Andalusia. Audio samples will be gathered using online surveys in the first instance and in-person recordings at a later stage to complement the data. Table 1 includes the number of data points studied per province in Navarro Tomás et al. (1962) (ALPI), Alvar et al. (1973) (ALEA), and the expected data points for Herrero de Haro (in preparation) (ALIAA).

Table 1: Data points studied in ALPI and ALEA compared with the points expected to be studied in ALIAA.

Province	Publication		
	ALPI	ALEA	ALIAA
Almería	8	30	67
Cádiz	4	17	30
Córdoba	7	25	51
Granada	10	46	112
Huelva	6	24	52
Jaén	9	31	64
Málaga	8	26	67
Sevilla	9	31	69

ALEA studied one speaker for every 24334 inhabitants of Andalusia, one data point for every 379 km², and 230 towns (29.29% of Andalusian towns). In contrast, ALIAA plans to study one speaker for every 4200 inhabitants, one point for every 175 km², and 500 towns (63.69% of Andalusian towns and

88.18% of the 567 Andalusian towns with more than 1000 inhabitants).

It is worth mentioning that, while Alvar et al. (1973) and other atlases usually organise data based on towns, ALIAA has been designed to organise the data by postcodes. Some studies have shown the complex linguistic reality of cities such as Málaga, Granada, and Huelva (e.g. Alvar et al. 1973: Map 1705; García Mouton 1992; Melguizo Moreno 2007; Regan 2017), where the maintenance of the /s/ - /θ/ contrast, or the realisation of /tʃ/ as either [tʃ] or [ʃ] are governed by complex sociolinguistic characteristics and are also neighbourhood-dependent. Therefore, a breakdown by postcode allows zooming into bigger cities and provides an analysis of different neighbourhoods, when needed. Spain has a five-digit-postcode system, and it is also easier to organise towns by digit than by name, as that way every point in the map can be linked to a five-digit postcode. The purpose-built tool created to visualise the interactive linguistic atlas has been built from scratch, and the coordinates for the area covered by each postcode were added onto a map.

2.2. Data collection

The data collection process has been designed to be carried out mainly online, although a series of fieldtrips will be needed to collect data for those areas of Andalusian for which insufficient data have been collected through online surveys. A variety of options was considered for the online survey. The first decision was whether or not to design an app to host it. It was decided not to use an app for various reasons. First of all, programming apps is expensive and time-consuming. Apps can stop working as soon as there is a software update by different phone companies, and it costs money and time to keep them updated and working. It also costs money to distribute apps via Google Play or Apple Store. Furthermore, research has shown that people are more likely to do a survey online if they only have to click on a button than if they have to install an app on their phone. Installing an app might be considered quite intrusive by some users and it could deter people from completing the survey. Furthermore, some users might not be technologically competent enough to install an app, so it was decided to host the survey on a site to make the online data-gathering process as simple as possible.

Different online options were explored and, after careful analysis, it was decided to use a website called *Phonic* (<https://www.phonic.ai/>). The process of comparing different options was carried out between February and April 2023. An exhaustive list of the online applications or sites compared for

the data-gathering process will not be provided as some of the issues that were found with some sites might have been corrected since then. To avoid any possible litigation, only the issues found will be mentioned but not the name of any other website or app. Some free options were available to collect audio data online, but programming knowledge was needed to set it up and to maintain them and/or a separate server was needed to store the audio data. Other online survey platforms allowed audio responses, but these had to be recorded first and then uploaded; this would have made the online survey more complex, and it could have deterred possible participants. Other online survey platforms had the option to record audio answers but there was either a limit on the size of the audio, or it was a beta functionality which was not finalised. The website <https://www.phonic.ai/> was chosen; this site is being used for the online data collection process and all the feedback provided in this article is objective and not sponsored. It should be noted that the site has had no involvement with the writing of this article and I am not receiving any compensation from <https://www.phonic.ai/> for the content of this article.

Phonic allows the researcher to host a survey online and potential participants can complete it without having to register or provide any personal data. The survey shows a series of prompts, the participants can record their answers online for each question, and the site stores the audio securely. The site specialises in capturing audio and video responses and there are different options for audio formats and quality. The researcher can download the audio from Phonic's server together with a spreadsheet which has anonymous information for each user (e.g. a randomly assigned 20-digit alphanumeric code and a breakdown of answers for the sociolinguistic part of the interview (e.g. age and gender)). Additionally, the site also generates an automated transcription for each audio response. It has been found that the automated transcription is, in general, not very accurate with Andalusian accents. This could be due to the great variety of phonetic outputs included in the responses and, although the analysis of these transcripts will not be part of the atlas project, an analysis of them can highlight common issues with automated transcription tools when it comes to dealing with innovative varieties of Spanish. Phonic assigns a code to each speaker and to each recording from each speaker. The survey has a total of 128 questions for which oral answers are given and each speaker is divided into a total of 128 audio tracks. Phonic gives each track another random code and it can be very confusing and time-consuming to navigate this naming system. For this, the author of this paper designed a script and the Information

Technology research assistant programmed it so that each audio and folder for each speaker would be renamed using a code devised by the researcher. Each participant receives a code which starts with a five-digit number (the speaker's postcode), then a letter for the gender (*m* *mujer* 'woman'; *h* *hombre* 'man'; *o* *otro* 'other'), then the age of the participant and, finally, the level of instruction (p – primary education, b – secondary education or baccalaureate, or u – university). For example, a speaker code could be 11600H28U; this would be a 28-year-old male from the postcode 11600 (Ubrique, Cádiz), and university-educated. If two participants had the same code, an additional number was given. For example, 11600H28U1 would be given to another speaker with the same characteristics as 11600H28U. Initially, a “.” was inserted before the final number (e.g. 11600H28U.1) but it was decided not to use “.” as some scripts changed it to “_” and this created issues with coding later on.

The audios are coded with the following format, for example, Q28-Paz-11600H28U; this means that this audio track is for Question 28, which contains the word *paz* 'peace' uttered by speaker 11600H28U. This system allows us to organise the speakers and the audios in a very simple way. It should be noted that, as the aim of ALIAA is to analyse the speech of 500 towns, four speakers per town (two males and two females), and as each speaker's survey is divided into 128 audio tracks, it is expected to gather and to have to organise a total of 256,000 audio tracks.

The survey has four sections. In the first section, the participants have to answer sociolinguistic questions such as their age, gender, postcode and postcode, time outside that town, and information about other aspects such as level of instruction. This information is useful to know the sociolinguistic characteristics of the speaker, to create the speaker code, and to geotag the subsequent phonetic analysis to the coordinates of that postcode. Some researchers might have reservations regarding the veracity of the data/answers by potential jokers who want to boycott the project. However, it takes approximately 25-30 minutes to complete the survey and it is highly unlikely that people who want to play a trick will spend this amount of time completing the survey. The second part of the survey focuses on spontaneous speech; Question 11 asks participants to say the name of their town and the demonym for their town. Question 12 asks participants to describe their town, customs, and tourist or general information which might be of interest to potential visitors. Question 13 asks speakers to summarise the last film they have watched or the last book they have read. Questions 16 to 139 ask the participants to say one word, generally by showing an object and

asking them to say the word (e.g. *casa* ‘house’), as in Figure 1; when it is not possible to show a picture, a short definition is used (e.g. the opposite of clean is ...). Question 140 asks participants to read a 119-word version of *The North Wind and the Sun* adapted to Andalusian accents by, for example, changing the word *abrigo* ‘coat’ to *chaqueta sucia* ‘dirty jacket’ to investigate how speakers say /tʃ/, /s/, and /θ/.

Figure 1: Screenshot of Question 20 prompting participants to say the word *casa* ‘house’. The screenshot reads *Esto es una ...* ‘this is a ...’.



The survey has a total of 140 questions. Due to space constraints, it is not possible to show all the questions. However, Table 2 contains the structure of the survey together with the words which the participants were expected to say.

After testing the online survey with a few close contacts, the link to the survey was distributed on social media in early June 2023. The project received wide attention from members of the public outside linguistics and the posts have been shared and viewed over 69,000 times. The project has also attracted media attention and news about the project has featured on regional and national radio, newspapers, and TV. News about the project and results from preliminary analyses have been shared on the project’s site (<https://acentosandaluces.com/>) and via social media to promote the survey, and the public has engaged with some of the maps published on X Twitter. After spending five months distributing the survey through organic posts on social media, a marketing company was hired to advertise the survey link on X Twitter and Facebook. The adverts reached, on average, over 700,000 impressions, over 7000 clicks, and 350 survey completions per month; the money spent on advertising was worth it.

From 1st June 2023 until 1st February 2024, 1781 participants have completed the survey. The mean age is 32.68 years. A sociolinguistic breakdown of participation is included in Table 3, and a breakdown by each of the eight Andalusian provinces is included in Table 4.

Table 2: Breakdown of the online survey. Questions 11 – 140 require spoken answers.

1. Permission to record	19. Sucio ‘dirty’	37. Pie ‘foot’
2. Age	20. Casa ‘house’	38. Pies ‘feet’
3. Gender	21. Caza ‘hunt’	39. Grafiti ‘graffiti’
4. Gender: details	22. Zeta ‘zed’	40. Grafitis ‘graffiti’ (plural)
5. Level of instruction	23. Oso ‘bear’	41. Tele ‘TV’
6. Postcode	24. Osos ‘bears’	42. Teles ‘TVs’
7. Time spent outside hometown	25. Papá ‘dad’	43. Gato ‘cat’
8. Explaining time outside hometown	26. Papás ‘dads’	44. Gatos ‘cats’
9. Is Spanish your first language?	27. Tapar ‘to cover’	45. Gata ‘female cat’
10. Other languages	28. Paz ‘peace’	46. Gatas ‘female cats’
11. Town and demonym	29. Bati ‘I beat’ (preterite)	47. Raya ‘line’
12. Describe your town	30. Batís ‘you beat’	48. Racha ‘period’
13. Summarise a book or film	31. Ve ‘(s)he sees’	49. Ralla ‘(s)he grates’
14. Example of how to answer	32. Ves ‘you see’	50. Yegua ‘mare’
15. Example of how to answer 2	33. Lo ‘it’ (direct object)	51. Llega ‘(s)he arrives’
16. 56	34. Los ‘them’ (direct object)	52. Caja ‘box’
17. 16	35. Champú ‘shampoo’	53. Jabalí ‘wild boar’
18. 65	36. Champús ‘shampoos’	54. Los ojos ‘the eyes’

55. Humo ‘smoke’	85. Está ‘(s)he is’	114. Cataratas ‘cataracts’
56. Hambre ‘hunger’	86. Hecha ‘made’ (feminine)	115. Nuca ‘nape’
57. Coser ‘to sew’	87. Este ‘East’	116. Nucas ‘napes’
58. Municipal ‘municipal’	88. Vaca ‘cow’	117. Azúcar ‘sugar’
59. Cocer ‘to boil’	89. Vacas ‘cows’	118. Zetas ‘zeds’
60. Dos horas ‘two hours’	90. Esté ‘(s)he is’ (subjunctive)	119. Capi ‘captain’ (informal)
61. Albañil ‘bricklayer’	91. Vasca ‘Basque woman’	120. Lápiz ‘pencil’
62. Calcetín ‘sock’	92. Vascas ‘Basque women’	121. Fernando ‘proper name’
63. Arcoiris ‘rainbow’	93. Capa ‘cape’	122. Fernández ‘surname derived from Fernando’
64. Árbol ‘tree’	94. Resbalar ‘to slip’	123. Batir ‘to beat’
65. Pito ‘whistle’	95. Caspa ‘dandruff’	124. Matiz ‘shade’
66. Pitos ‘whistles’	96. Dos batas ‘two robes’	125. Calor ‘heat’
67. Tela ‘clothing material’	97. Dos dedos ‘two fingers’	126. Veloz ‘fast’
68. Telas ‘clothing materials’	98. Dos gotas ‘two drops’	127. Singapur ‘Singapore’
69. Osa ‘female bear’	99. Dos llaves ‘two keys’	128. Chapuz ‘shoddy work’
70. Osas ‘female bears’	100. Dos yeguas ‘two mares’	129. Piel ‘skin’
71. Lupa ‘magnifying glass’	101. Dos hielos ‘two ice cubes’	130. Azul ‘blue’
72. Lupas ‘magnifying glasses’	102. Desde ‘since’	131. Atún ‘tuna’
73. Ver ‘to see’	103. Musgo ‘moss’	132. Pez ‘fish’
74. Ven ‘they see’	104. Vosotros, ustedes ‘you plural informal, formal’	133. Ñoño ‘dull’
75. Va ‘(s)he goes’	105. Peine ‘comb’	134. Naranja ‘orange’
76. Vez ‘once’	106. Hielo ‘ice’	135. Dedo ‘finger’
77. Van ‘they go’	107. Huevo ‘egg’	136. Coche ‘car’
78. Vas ‘you go’	108. i,e,a,u,o ‘i,e,a,u,o’	137. Cayó ‘(s)he fell’
79. Gasta ‘(s)he spends’	109. Pera ‘pear’	138. Tizne ‘soot’
80. Bar ‘bar’	110. Vara ‘stick’	139. Lobezno ‘wolf cub’
81. Gastas ‘you spend’	111. Perra ‘bitch’ (female dog)	140. Reading of <i>The North Wind and the Sun</i>
82. Esta ‘this’	112. Barra ‘bar’	
83. Cesta ‘basket’	113. Catarata ‘cataract’	
84. Cestas ‘baskets’		

Table 3: Participation in the ALIAA online survey until 1st February 2024.

Gender		Level of instruction	
Males	39.75%	Compulsory education	4.16%
Females	59.48%	Baccalaureate	30.82%
Other	0.77%	University degree	65.02%

These figures are in line with other crowdsourced studies. For example, in Avanzi (2023), out of 37000 participants, 67% were university graduates and the average age was 27; the age in Avanzi’s (2023) study is lower than in ALIAA as Avanzi (2023) required participants to download an app and this would not appeal to older audiences, while the data for ALIAA was first gathered through an online survey. Furthermore, in Scherrer et al. (2015), 72% of participants were women. As we can see from the data in Table 3, women’s participation was 20% higher than men’s. This is an interesting pattern, as Alvar et al. (1973), and other European atlases, had

been carried out studying men almost exclusively. For example, in Gilliéron (1902-1910), 8.4% of the informants were women and in Jaberg and Jud (1928-1940) female informants made up 8.2% of the sample. Table 4 shows that speakers from rural areas are underrepresented in the data sample; this is to be expected as other crowdsourced linguistic projects have had similar issues (Boula de Mareuil et al. 2021). The fieldwork carried out as part of ALIAA is designed to overcome this issue.

The online adverts were modified as data were received to target certain profiles (e.g. men over 50). Data gathered through online surveys results in self-selected participants and it was expected for most participants to be between 20 and 60 years of age and competent Information Technology users, as in other studies for which they have used crowdsourced data (e.g. Avanzi 2023). The fieldworks are designed as a way of mitigating this limited sociolinguistic profile but, given the extensive area to be surveyed, this approach was considered the best one and it was decided to accept limited sociolinguistic profiles

Table 4: Breakdown of participation by each of the eight provinces of Andalusia until 1 February 2024.

Province	Percentage of participation	1 participant for every ... inhabitants	Inhabitants per province	Percentage of participants who are from the capital city of the province
Almería	9.17%	10144	740,534	36.98%
Granada	20.55%	5621	921,987	17.39%
Jaén	9.02%	8663	623,761	20.25%
Córdoba	9.89%	9778	772,464	39.02%
Málaga	16.04%	13418	1,717,504	47.27%
Sevilla	14.03%	17396	1,948,393	6.94%
Cádiz	14.41%	10841	1,246,781	39.06%
Huelva	6.89%	9613	528,763	53.57%
Total			8,500,187	33.58%

in favour of data from a more varied and extensive area. It should be noted, however, that an approach like this might not work for languages with greater sociolinguistic diversity. For example, recent studies have analysed sociolinguistic variation in Seville (León-Castro Gómez and Jiménez Fernández 2023) and Córdoba (Perea Siller 2023) and they have found non-significant differences within their samples based on age, gender, or level of instruction. Furthermore, Herrero de Haro and Hajek (2022) describe the accent of Eastern Andalusia and identify features which are considered stigmatised, such as *heheo*, even in highly educated people. The level of instruction and career progression, when people have not moved away from their town, seems not to modify the accent of the speaker and this allows for regional accent features to be studied even in the speech of educated people; these are the ones who tend to do the interview. The data gathered via online surveys might not be enough to develop a sociolinguistic profile of the 500 towns to be analysed, but the data allow us to describe the local accent in terms of which features are present and what the acoustic characteristics of those features are. While, for example, there is a clearly defined sociolect of English linked to speakers of higher sociolinguistic strata (Received Pronunciation), and speakers of English in England tend to modify their speech the higher they climb the social ladder, this is not the case in Spain. As Moreno Fernández (2005: 55) explains, unlike what happens in the English-speaking world, climbing up the social ladder does not necessarily need to result, although it can, in a change in the pronunciation of a speaker; regardless of how high a speaker's sociocultural status is, it is fairly easy to identify if the speaker is from northern Spain, from the Canary Islands, or from the

Caribbean, for example. Some steps, however, have been taken to improve the reliability of the sample. When there were enough surveys collected from a certain town, priority was given to older speakers, speakers with a lower level of instruction, and speakers who had spent less time outside their town.

The situation is very different in English. Research shows that, as people climb up the social ladder in the United Kingdom, their accent loses characteristics of their regional accent and tends to merge with the accents of the higher strata of speakers (Trudgill 1974: 41). This means that the approach to gather data via online surveys, although useful for the current project, might not work, for example, for British English. This might also not work for other languages for which changes in a speaker's sociocultural status might result in changes in pronunciation patterns.

2.3. Data analysis

It is normal to discard crowdsourced data due to issues such as background noise and, at this stage, it is not possible to know how much data will be discarded. The proportion of completed surveys to clicks is usually around 28%. There is a big drop in participants when the audio response questions start (Question 11, name of your town and demonym). After that, there is a big drop in people continuing beyond Question 12 (describe your town). Interestingly, between Question 13 and Question 140, the drop-out rate is very low, 32%. This all-or-nothing pattern in answering questions could, perhaps, be explained by a few factors. Many of the people who clicked on the survey link initially might not have had the intention of completing the survey, but they simply clicked on it to see what

the survey was like. The percentage of clicking-to-completion went down abruptly once we started advertising campaigns on Facebook and X Twitter. This could have been by people who qualified for the survey (i.e. from Andalusia) or people who did not (i.e. people from other places who wanted to see how the survey or project was structured); with more exposure on social media, the adverts could have reached people who did not qualify to complete the survey. To minimise the number of clicks needed for survey completion, additional information and a video explaining the project were added. The online adverts also said very clearly that the survey took 25-30 minutes to complete. On many occasions, we can see that someone clicked on the survey, entered the initial information (e.g. age, gender, location, etc.) then disappeared, and someone with exactly the same profile emerged in the answers a few hours or a day later. The assumption here is that people clicked on the survey and had the intention of completing it, but they waited until they had more time available to do it, or perhaps they realised that they were in a noisy environment.

Once the data are received, participants are chosen depending on different factors. Not all the audio responses which were received will be analysed for the atlas but no data will be deleted; audio not used for this project has been kept for other possible future projects. People who had been away from their postcodes for up to a year were examined to determine their suitability. The speech of each speaker was also analysed in terms of naturalness to decide whether or not to include it in the corpus for the atlas. For example, research shows that codas are deleted in Andalusian Spanish and, if a speaker retained codas (e.g. *casas* ‘houses’ pronounced as in conservative varieties [‘kasas]), which does not conform to Andalusian Spanish pronunciation, the speaker was discarded. The pronunciation of isolated words was also compared with the spontaneous speech from Questions 12 and 13 (i.e. describe your town and summarise the last film you have seen or book you have read). This allowed the researcher to identify a mismatch between the more natural speech and the laboratory speech which we aimed to capture during the survey.

2.3.1. Audio analysis

The audio data were downloaded from Phonic and coded using the code devised for each speaker. To ensure consistency, the analysis was organised by feature, not by town, so that all tokens relevant to a specific feature (e.g. pronunciation of /s/ in *casa* ‘house’) were analysed first, and then another feature.

Audios were converted to mono using a Praat (Boersma and Weenink 2020) script, and all the audio was automatically segmented using Web MAUS (Kisler et al., 2017). For this, a script designed by the author of this paper and written by the Information Technology technician, Edgar Huaranga, was written so that a Word file with specific content could be created for each of the audio files within a folder. For example, if there were 300 audio files with the word *casa* ‘house’, by having a Word file called *original.doc* in the folder with the word *casa* ‘house’ in that file, the script would create the corresponding 300 Word files with the same name as each audio file; this was needed in order to run the automatic segmenting tool on Web MAUS (Kisler et al., 2017).

After that, a Praat script copied the last tier in the first position. That way, the automatically segmented tier was left untouched in the last position of the Textgrid, and the duplicated tier, in the first position, was manually corrected by the researcher. Every single segmentation has been hand-corrected by the author.

As a general rule, all segmentation has been done following some general principles for all types of tokens, although additional steps will be incorporated for specific words. For example, vowel onset was marked at the closest *F2* (second formant) and *F3* (third formant) abrupt increase in energy, as per Herrero de Haro and Alcoholado Feltstrom (2024) and vowel offset was marked at the closest *F2* and *F3* sudden decrease in energy. Marking vowel onset at the increase of *F1* (first formant) results in vowels being too long and it could produce errors by confusing *F1* with the voicing bar. Marking vowel onset or offset according to *F3* results in vowels being too short. *F2* is usually used as a guide for segmenting vowels and, to increase accuracy, it was decided to use *F2* and *F3* so that it was easier to identify relevant points. Scripts will be used to extract different acoustic measurements such as length, formant information, *f0*, and intensity.

For this paper, some preliminary analyses are carried out as a first exploration and exploitation of the data. For this, 70 tokens of *casa* ‘house’ and 67 of *caza* ‘hunt’ are analysed to study the /s/ - /θ/ contrast; 69 samples of /tʃ/ are analysed in the word *champú* ‘shampoo’ and 56 in the word *champús* ‘shampoos’. The samples are from 70 different postcodes and one speaker has been analysed per postcode; the mean age is 28.8 years and the standard deviation 10.6. This analysis was first performed impressionistically and then using spectrographic analysis. However, no acoustic features have been taken yet (e.g. centre of gravity) since this will be done once all samples have been gathered.

2.4. Data representation

A software has been developed to represent linguistic data on a series of interactive maps. This tool is now available, and a screenshot of the tool has been provided in Figure 2. However, this software is still only locally available on the researchers' computers. Once the project is finished in December 2026, it will be activated on a website, and it will also be shared with the research community as a free open-access source for researchers to use for their own purposes. For this paper, however, the linguistic data have been displayed using QGIS (QGIS Development Team).

The maps are divided into two types: word map and phenomenon map. The user can choose one of these two types on the main page of the atlas. Word maps will display all the towns which data have been gathered for and show the phonetic transcription of the same word across all those points; users can choose a phenomenon of interest and the phenomenon maps will show how a specific phenomenon is distributed across Andalusia (e.g. the pronunciation of /x/ as [x] or [h]). In both types of maps, users can click on a town and this will bring up an information box with additional comments (e.g. this phenomenon is more common in men than in women), and it will also let users play a recording from a speaker from that town pronouncing the transcribed word in the word map, or a word to exemplify that phenomenon in the phenomenon map.

The interactive map tool has already been created; however, it has not been uploaded onto the project's

website yet. The coordinates of all the postcodes from Andalusia have been added to the map. A certain postcode can be activated by entering the postcode in the database of the map. Then, some information can be entered, such as phonetic transcription for the word maps, or the type of phenomenon for a phenomenon map.

2.5. Data curation

A corpus will be created for this project; it will be called *CALIAA* (Corpus del *Atlas Lingüístico Interactivo de los Acentos de Andalucía* 'Corpus of the Interactive Linguistic Atlas of Andalusian Accents'). This audio corpus will be publicly available to download as a large folder but, in order to protect the data, access to the audio corpus will only be given to people who register via an official registration form. The corpus will include the data used to elaborate the atlas. It is expected that *CALIAA* will be a very useful resource for researchers interested in phonetic and phonological variation in Spanish. It will not be necessary to register to use the online map which displays linguistic data and reproduces audio examples for each data point.

The initial aim is to include audio data from 500 towns, four speakers per town, and a maximum of 128 audio files per speaker (a total of 256,000 files). The survey has two open-ended questions (i.e. describe your town and summarise the last book you read or film you watched), a series of picture naming tasks, and a reading of *The North Wind and the Sun*.

Figure 2: Purpose-built program designed to represent the linguistic data on a series of interactive maps.



The importance of this corpus goes beyond research interests. It covers aspects of Andalusian life (e.g. things to do in different towns) and culture, such as customs of different areas.

3. PRELIMINARY RESULTS

To demonstrate the suitability of the method developed for ALIAA, it has been decided to show two preliminary analyses for two different phonetic features. The maps have been created using QGIS, calculating a Voronoi polygon from the data points which have been analysed.

3.1. The contrast /s/ - /θ/ in Andalusia

Spanish has a phonemic contrast /s/ - /θ/, which distinguishes between words such as *casa* /'kasa/ 'house' and /'kaθa/ 'hunt'. However, while a relative minority of the Spanish-speaking world maintains this contrast, only in Spain, in all Latin America, the Canary Islands, and some parts of Andalusia, the phonemes /s/ and /θ/ have merged into /s/; this is called *seseo* and in these accents the words /'kasa/ 'house' and /'kaθa/ 'hunt' are pronounced ['kasa]. In addition to this, in some parts of Andalusia, /s/ and /θ/ have merged into /θ/; this is called *ceceo* and, in these accents, the words /'kasa/ 'house' and /'kaθa/ 'hunt' are both pronounced ['kaθa].

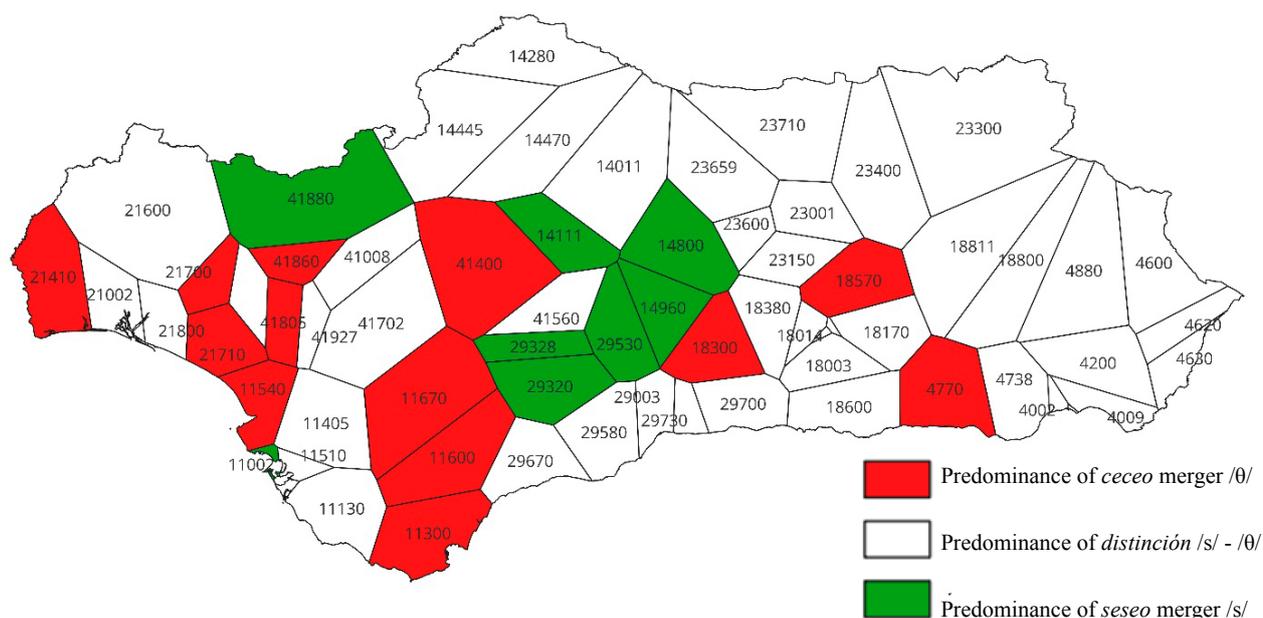
The /s/ - /θ/ contrast is very important in Andalusia, as this is the only Spanish-speaking area where *distinción* (phonemic distinction of /s/ and /θ/), *seseo* (merger of these into /s/) and *ceceo* (merger of

these into /θ/) coexist. Figure 3 represents the spread of *distinción*, *seseo*, and *ceceo* across Andalusia; to simplify the map, only the categories of *distinción*, *seseo*, or *ceceo* are used and no information is given regarding whether there is a mixture of more than one phenomenon. Some recent studies have shown that the spread of *distinción*, *seseo*, and *ceceo* has changed since the data for Alvar et al. (1973) were gathered between 1953 and 1958. For example, while the area coded 4738 on the map is identified as having *ceceo* in Alvar et al. (1973), recent studies in this part of Andalusia have found that *ceceo* is no longer the norm there (e.g. Herrero de Haro and Hajek 2023). Similar changes have been found in other parts of Andalusia, such as Huelva (Regan 2017) Huelva. Figure 3 is a preliminary map created with the data analysed so far for ALIAA. Spanish postcodes have five digits and the postcodes for Almeria province start with 04; however, QGIS drops the initial 0 of these postcodes, and that results in Almería postcodes coming up as four-digit postcodes on these QGIS maps. The map uses the term “predominance” to show the most common realisation, but that does not mean that other realisations do not coexist.

3.2. Types of /tʃ/ across Andalusia

The phoneme /tʃ/ has several realisations in Andalusia. Although, traditionally, two main pronunciations have been posited, [tʃ] and [ʃ], a preliminary analysis of the samples shows a more complex situation. There seems to be a continuum in terms of the level of affrication, and these realisations

Figure 3: Distribution of *distinción*, *seseo*, and *ceceo* across Andalusia. Data from 70 locations.



can be divided into six categories; Figure 4 shows the spread of these pronunciations across Andalusia. These categories are:

1. [tʰ] This consonant displays high intensity in the occlusive moment, which can be identified by a strong aspiration after the explosion bar.
2. [tʃ] This consonant displays the expected level of intensity in the occlusive part of the affricate.
3. [tʃ] This affricate is pronounced further back than it would normally be expected.
4. [tʃ̺] This consonant is visibly affricate in the spectrogram, but the explosion bar is weak and, typically, it only goes down to around 2000 – 3000 Hz.
5. [tʃ̺] This consonant seems to be formed by a [t] with no audible release and a fricative moment. Although there is no explosion bar, the spectrogram shows two articulatory moments in this consonant. The first one seems to be the beginning of [t] and then, without an audible or visible explosion, it moves onto [ʃ̺].
6. [ʃ] This is a fricative consonant.

Apart from the importance of knowing how this feature extends across Andalusia, these data can be used to understand phonemic processes in other languages. For example, the change from [tʃ] to [ʃ] in French could have followed a similar evolution to the one we can see in Andalusian Spanish: [tʃ] → [tʃ̺] → [tʃ̺] → [ʃ].

4. DISCUSSION

4.1. The contrast /s/ - /θ/ in Andalusia

As we can see from Figures 3, the general pattern of *distinción*, *seseo*, and *ceceo* seems to be along the

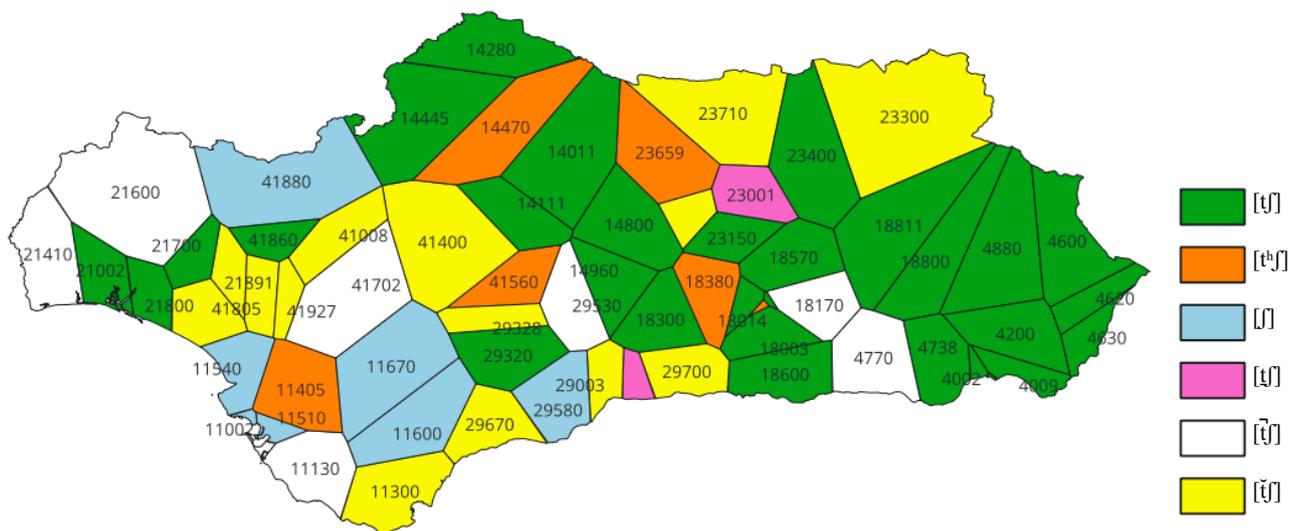
same lines as the one identified in Alvar et al. (1973). The eastern part of Andalusia has *distinción*, and this seems to spread further west in the northern part of Andalusia than the south. This pattern has made some linguists describe *ceceo* as more prevalent in the coastal areas. As in Alvar et al. (1973: Map 1705), there is an area with *distinción* close to the border with Portugal; this area is also visible in Figure 3.

As some researchers have explained, some areas where /s/ and /θ/ had merged into /θ/ have reintroduced the distinction. This goes against Garde’s principle (Labov 1994: 311), which says that mergers are irreversible by linguistic means; however, we seem to have several examples of this across Europe (Regan 2017). It is believed that what is causing this is the exposure to media, travelling, and contact with speakers of more prestigious varieties where the distinction /s/ - /θ/ is maintained. It is worth noting the reduction of the *ceceo* area in the southern part of Andalusia. This reduction is noticeable between the southeastern and south-central parts of the map (between postcodes 18600 and 29670). It is worth noting an area where *ceceo* is still preserved in the borders of Granada and Almería province (4770 on the map). Regarding western Andalusia, this part has experienced more changes than the eastern part. Areas which were identified as having *seseo* around central Andalusia now have *distinción* or *seseo* with some *distinción*.

4.2 Types of /tʃ/ across Andalusia

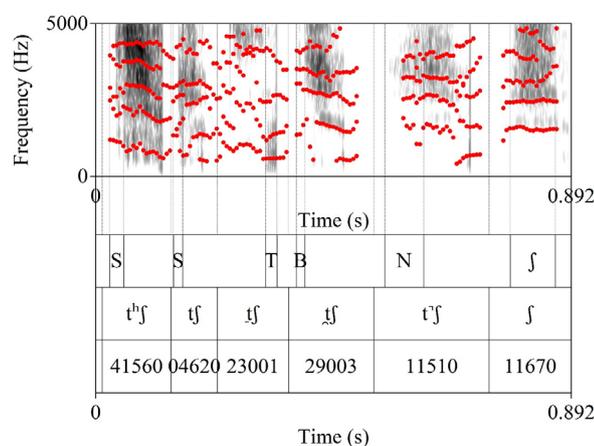
The pronunciation of /tʃ/ across Andalusia has a similar pattern to that of *distinción*, *seseo*, and *ceceo*.

Figure 4: Types of /tʃ/ across Andalusia.



A closer pattern to that of central-northern Spain is found in the east and it reaches further west in the north than in the south. The area furthest west, closer to Portugal, seems to have a similar pronunciation to that of eastern Andalusia. While there is a similar pattern in Alvar et al. (1973) and the current map, we can see some differences. The area where /tʃ/ is pronounced [ʃ] in Alvar et al. (1973: Map 1709) clusters around three-quarters to the west, mainly around the provinces of Cadiz and Seville (postcodes starting with 11 or 41). That area still continues to be where [ʃ] is found the most. However, a big area around central Andalusia, which was identified as areas where [tʃ] and [ʃ] could be found in Alvar et al. (1973), now we find [ʃ]. It is interesting to note some intermediate pronunciations which have been identified. There is a weakened /tʃ/ with a softer burst [t̃ʃ] and another type of realisation with no audible or visible burst in the spectrogram but with evidence of tongue movement; it seems to be a fricative which starts by being pronounced in the same place as [t], and then it moves to where [ʃ] is pronounced. It has been decided to transcribe this allophone as [t̃ʃ]. The types of /tʃ/ can be found in Figure 5.

Figure 5: Allophones of /tʃ/ across Andalusia. The bottom tier includes the speakers' postcodes, the middle tier includes the proposed phonetic transcriptions, and the top tier includes relevant details. "S" stands for "stop", "T" for "transition into vowel", "B" for "burst", and "N" for "no visible burst".



4.3. General discussion

As we can see from Figures 3 and 4, the general pattern of variation across Andalusia seems to be roughly the same in Alvar et al. (1973) and in the current atlas. We continue having two big sections, eastern and western Andalusia, with the area furthest west having some similarities with eastern Andalusia, and with the isoglosses being situated further west in the northern than in the southern part of Andalusia.

Some authors (Villena Ponsoda 2008) consider eastern Andalusia as being closer to those realisations of central-northern Spain than western Andalusia. Those features are moving west and it might not be due to a linguistic influence of eastern Andalusia into the west, but rather as an influence caused by the media and more prestigious realisations from central-northern Spain.

5. CONCLUSION

The data analysed in the present paper highlight important information. The EAS vs WAS division is still obvious today; some features from EAS have spread further west since the data for Alvar et al. (1973) were gathered in the 1950s. However, it is not clear what is causing that shift, whether it is eastern Andalusian features spreading west, or central-northern Spanish features spreading south.

ALIAA is also gathering information which may be relevant to understanding phonetic and phonological historical processes in other languages. For example, the change from /tʃ/ to /ʃ/ in French could have followed a similar process to the one we have seen in this paper. Decisions regarding statistical models will be made for each feature once the in-depth analysis commences for each feature. Some steps will be taken in order to normalise data, either by normalising measurements (e.g. using Z-score) or by coding speakers as a random variable in the statistical model and by including random slopes and random intercepts for each speaker.

This article introduces a linguistic atlas project which is being carried out with a new methodology. The way the data are gathered, analysed, and displayed is different to what has been done before and a preliminary analysis of the data shows that this project is highly viable. The method presented in this article can allow us to make linguistic atlases much cheaper and faster than traditional dialectology methods.

DATA AVAILABILITY

The audio used for this article will be made publicly available once the interactive linguistic atlas is published in December 2026 and it will be called "CALIAA" (*Corpus of the Atlas Lingüístico Interactivo de los Acentos de Andalucía*). A reader can request access to the audio files for research purposes before that date by emailing the author.

ACKNOWLEDGMENTS

I would like to thank all the participants in this study; they have given up their time to contribute to

this atlas. I would also like to thank everyone who has made the online data collection process possible by forwarding the information to their contacts.

DECLARATION OF COMPETING INTEREST

The author of this article declares that they have no financial, professional or personal conflicts of interest that could have inappropriately influenced this work.

FUNDING SOURCES

The project *Atlas Lingüístico Interactivo de los Acentos de Andalucía* has been funded by the *Consejería de Universidad, Investigación e Innovación* of Junta de Andalucía (Spain) awarded to the author with code EMC21_00042.

AUTHORSHIP CONTRIBUTION STATEMENT

Alfredo Herrero de Haro: Funding acquisition; Conceptualisation; Methodology; Validation; Formal Analysis; Investigation; Resources: Data Curation; Writing – Original Draft: Writing – Review & Editing; Visualisation.

REFERENCES

- Alvar, M., Llorente, A., & Salvador, G. (1973). *Atlas lingüístico y etnográfico de Andalucía* (Vol. 6). Granada: Universidad de Granada/Consejo Superior de Investigaciones Científicas.
- Avanzi, M. (2023). *The Français de nos régions app: method, development, and first results*. Paper presented at the Université de Neuchâtel.
- Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer* (Version 6.1.14). Retrieved from www.praat.org.
- Boula de Mareüil, P., Bilinski, E., Vernier, F., de Iacovo, V., & Romano, A. (2021). For a mapping of the languages/dialects of Italy and regional varieties of Italian. In A. Thibault, M. Avanzi, N. Lo Vecchio, & A. Millour (Eds.), *New Ways of Analyzing Dialectal Variation* (267-288). Strasbourg: ELiPhi.
- Boula de Mareüil, P., Vernier, F., & Rilliard, A. (2018). *A Speaking Atlas of the Regional Languages of France*. Paper presented at The Eleventh International Conference on Language Resources and Evaluation, Miyazaki, Japan.
- Chambers, J. K., & Trudgill, P. (1988). *Dialectology* (2nd ed.). Cambridge: Cambridge University Press.
- García Mouton, P. (1992). *El atlas lingüístico y etnográfico de Andalucía. Hombres y mujeres*. Paper presented at the Congreso Nacional de Dialectología IKER.
- Gilliéron, J. (1902-1910). *Atlas linguistique de la France*. Paris: Champion.
- Henriksen, N. (2017). Patterns of vowel laxing and harmony in Iberian Spanish: Data from production and perception. *Journal of Phonetics*, 63, 106-126. <https://doi.org/10.1016/j.wocn.2017.05.001>
- Herrero de Haro, A. (2017). The phonetics and phonology of Eastern Andalusian Spanish: A review of literature from 1881 to 2016. *Ikala, Revista de Lenguaje y Cultura*, 22(2), 313-357. <https://doi.org/10.17533/udea.ikala.v22n02a09>
- Herrero de Haro, A. (In preparation). *Atlas Lingüístico Interactivo de los Acentos de Andalucía*. [Linguistic atlas]
- Herrero de Haro, A., & Alcoholado Feltstrom, A. (2024). Anti-hiatus tendencies in Spanish: rate of occurrence and phonetic identification. *Linguistics*, 62(1), 203-228. <https://doi.org/10.1515/ling-2021-0228>
- Herrero de Haro, A., & Hajek, J. (2022). Illustrations of the IPA: Eastern Andalusian Spanish. *Journal of the International Phonetic Association*, 52(1), 135-156. <https://doi.org/10.1017/S0025100320000146>
- Herrero de Haro, A., & Hajek, J. (2023). Covariants of Gemination in Eastern Andalusian Spanish: /t/ following Underlying /s/, /k/, /p/ and /ks/. *Languages*, 8(2), 1-27. <https://doi.org/10.3390/languages8020099>
- Jaberg, K., & Jud, J. (1928-1940). *Sprach- und Sachatlas Italiens und der Südschweiz, Zofingen, Ringier, 8 vols*. Zofingen: Ringier.
- Jiménez, J., & Lloret, M.-R. (2020). Vowel harmony. In S. Colina & F. Martínez Gil (Eds.), *The Routledge Handbook of Spanish Phonology* (100-128). Oxford/New York: Routledge.
- Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, 45, 326-347. <https://doi.org/10.1016/j.csl.2017.01.005>
- Labov, W. (1994). *Principles of Linguistic Change, Vol. 1: Internal Factors*. Cambridge (MA): Blackwell.
- León-Castro Gómez, M., & Jiménez Fernández, R. (2023). *Pronunciación de la /-s/ implorativa en el habla de Sevilla: análisis cuantitativo en el nivel instruccional bajo del corpus PRESEEA-Sevilla*. Paper presented at the Congreso Internacional sobre el español en Andalucía, Canarias e Hispanoamérica, Málaga.
- Loporcaro, M., Schmid, S., Zanini, C., Pescarini, D., & Donzelli, G. (2021). AIS reloaded: a digital dialect atlas of Italy and Southern Switzerland. In A. Thibault, N. Lo Vecchio, & A. Millour (Eds.), *Nouveaux regards sur la variation dialectale, Editions de Linguistique et de Philologie* (111-136). Strasbourg: ELiPhi.
- Martínez Melgar, A. (1994). El vocalismo del andaluz oriental. *Estudios de fonética experimental*, 6, 11-64.
- Melguizo Moreno, E. (2007). La fricativación de /ç/ en una comunidad de hablantes granadina. *Interlingüística*, 17, 748-757.
- Mondéjar Cumpián, J. (2006). *Bibliografía sistemática y cronológica de las hablas andaluzas*. Málaga: Servicio de Publicaciones e Intercambio Científico de la Universidad de Málaga.
- Moreno Fernández, F. (2005). *Principios de Sociolingüística y Sociología del Lenguaje*. Barcelona: Ariel.
- Navarro Tomás, T., Espinosa, A. M., Lindley Centra, L. F., de Borja Moll, F., Nobre de Gusmão, A., Otero, A.,... & Sanchis Guarner, M. (1962). *Atlas Lingüístico de la Península Ibérica* (Vol. I. Fonética). Madrid: CSIC.
- Perea Siller, F. J. (2023). *Materiales para un corpus de español hablado en la provincia de Córdoba*. Paper presented at the Congreso Internacional sobre el español en Andalucía, Canarias e Hispanoamérica.
- QGIS Development Team. (2021). QGIS geographic information system. QGIS Association. <https://www.qgis.org>
- Regan, B. (2017). A study of ceceo variation in Western Andalusia (Huelva). *Studies in Hispanic and Lusophone linguistics*, 10(1), 119-160.
- Ruch, H., & Harrington, J. (2014). Synchronic and diachronic factors in the change from pre-aspiration to post-aspiration in Andalusian Spanish. *Journal of Phonetics*, 45, 12-25. <https://doi.org/10.1016/j.wocn.2014.02.009>
- Scherrer, Y., Boula de Mareüil, P., & Goldman, J.-P. (2015). *Crowd-sourced mapping of pronunciation variants in European French*. Paper presented at the 18th International Congress of Phonetic Sciences (ICPHS).
- Trudgill, P. (1974). *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Villena Ponsoda, J. A. (2008). Sociolinguistic patterns of Andalusian Spanish. *International Journal of the Sociology of Language*, 193/194, 139-160. <https://doi.org/10.1515/ijsl.2008.052>