TEN YEARS OF THE ONLINE ALPI
(ATLAS LINGÜÍSTICO DE LA PENÍNSULA IBÉRICA)¹

David HEAP
University of Western Ontario
djheap@uwo.ca

Abstract

After decades with access to only one volume of the Atlas Lingüístico de la Península Ibérica (ALPI), the scholarly community took a few years to get used to the possibility of accessing much more of data from the most complete survey of Iberian Romance dialects. The legacy of Navarro Tomás and his team of fieldworkers, initially made available through a limited web interface, nonetheless allowed for the use original fieldwork transcriptions to produce research in areas of dialect morphosyntax, lexical variables and real-time historical linguistics, be it on the local, regional or national level or on a peninsular scale. After a few years of increasing use of the data, and renewed interest in the atlas, the CSIC fund a new collaborative project to develop and publish the ALPI materials.

Keywords
ALPI, Iberian Peninsula, Navarro Tomás, dialectology, language variation, internet

¹ This work was supported in part by Standard Research Grant 410-2008-1901 (Confronting linguistic theory with non-standard Hispanic pronoun systems) and International Opportunities Fund Grant 861-2009-1072 (International collaborative publication of the Atlas Lingüístico de la Península Ibérica), both from the Social Sciences and Humanities Research Council of Canada, as well as the internal project Elaboración y edición de los materiales del Atlas Lingüístico de la Península Ibérica (ALPI) 200410E604 from the Consejo Superior de Investigaciones Científicas (CSIC, Madrid, Spain). I am grateful to Enrique Pato (Université de Montréal) for his help in gathering a large part of the bibliography of research which uses the ALPI data, and to Stephanie Kelly and Jorge Emilio Rosés Labrada (University of Western Ontario) for their help in reviewing the text. Remaining errors and omissions are all my fault.
DIEZ AÑOS DEL ALPI (ATLAS LINGÜÍSTICO DE LA PENÍNSULA IBÉRICA) ONLINE

Resumen
Tras décadas con acceso a sólo un tomo del Atlas Lingüístico de la Península Ibérica (ALPI), la comunidad científica tardó algunos años en acostumbrarse a la posibilidad de acceder a un mayor número de datos de la encuesta más completa que existe sobre los dialectos ibero-romances. El legado de Navarro Tomás y su equipo de encuestadores, inicialmente accesible mediante una interfaz limitada, posibilitaba sin embargo el uso de las transcripciones originales de las encuestas para producir investigación sobre la morfosintaxis dialectal, las variables léxicas y la lingüística histórica en tiempo real, tanto a nivel local y regional como a escala peninsular. Después de algunos años de un uso cada vez mayor de los datos, y de un interés renovado por el atlas, el CSIC decidió financiar un nuevo proyecto en colaboración para elaborar y editar el ALPI.

Palabras clave
ALPI, Península Ibérica, Navarro Tomás, dialectología, variación lingüística, internet

1. Introduction

When the original field notebooks from the Atlas Lingüístico de la Península Ibérica (ALPI) became available once more to the international scholarly community after decades of silence, the reaction was not immediate. The “rediscovery” announced at the International Congress of Romance Linguistics and Philology in Salamanca (Heap 2001) and in the Revista de Filología Española (Heap 2002) was met initially with more of a trickle of interest than a flood. It took a while to build interest in the scientific community for data which had, in principle, been expected since the appearance of the first volume in 1962.

After the long and troubled history of the project (Navarro Tomás 1975; García Mouton 2006, 2007, 2011b; Heap 2002, 2008ab; much of the inner history is documented by the collected correspondence in Cortés & García Perales 2009), it seemed like a historical duty to that earlier generation of dialectologists to make the ALPI data available to the scholarly community in the fastest, most direct way possible. This meant, in the first instance, abandoning any pretense of linguistic cartography in favour of sharing the raw fieldwork data: scanned images prepared from photocopies of

---

2 One of the very few exceptions being the doctoral thesis by García Perales (2001), which used ALPI data for the País Valencià which had been accessed earlier from the Fons Sanchis Guarner.
field notebook pages were uploaded to a website (alpi.ca which now resolves to http://westernlinguistics.ca/alpi) with a simple search apparatus allowing users to select survey points by province and notebook images by page. This rather basic interface requires consulting the printed questionnaire (http://westernlinguistics.ca/alpi/questionaire.php) in order to determine which questions are to be found on which pages, and then allows users to access and read JPG or PDF images of the original notebook data, exactly as the fieldworkers transcribed them all those decades ago. While limited in many ways (it does not allow for searching a specific form at multiple points, for example), this initial system provided access to most of the data from *Cuaderno I (Fonética y gramática)*, many thousands of pages of original transcriptions, for hundreds of researchers.

Apart from the pragmatic necessity of a rapid and affordable way to make long-awaited ALPI data available to researchers, the website with scanned pages was also inspired partly by Catalán’s suggestion (1964: 233, note 1) that mapping linguistic atlas data can be viewed as an unnecessary luxury, when in fact simply listing data forms by survey point could also in some ways be seen as preferable:

> Como un ilustre colega ha sugerido alguna vez, sería incluso conveniente abandonar la representación cartográfica siempre que la contemplación de los datos sobre un mapa, lejos de facilitar la consulta, la entorpece: Unas listas de resultados ahorrarían, a menudo, tiempo y dinero a los editores y espera y vista a los lingüistas interesados.

Of course, the original ALPI website does not offer the ease of consultation of published lists of forms, requiring instead each page from each point of interest to be accessed individually: users must then compile their own lists of forms of interest. And the hand-transcriptions are much harder to read than printed lists: not only was the highly detailed *Revista de Filología Española* (1915) phonetic alphabet, augmented with additional detailed symbols and diacritics by Navarro Tomás and his team of

---

3 Because the two different formats used for *Cuaderno II: Léxico* (G: *General* and E: *Extendido*) correspond to two different lengths with differing pagination and numbering for the questions, it proved too challenging to make these data accessible using the simple interface developed for the original online ALPI.

4 At its peak, the ALPI website had well over a thousand registered users (though statistics were not kept to monitor how many were regular users or simply casual visitors) and thousands of visits per month.
fieldworkers (http://westernlinguistics.ca/alpi/phonetics.php) in order to capture even more precise articulatory nuances, individual variations between different fieldworkers abound, as do marginal notes, with occasional crossed-out forms and corrections. While immensely rich in details of different kinds, the original field notes are really too detailed to be used as a coherent data set, as can be demonstrated by comparing the notebook data with the small number of maps published in one printed volume (ALPI 1962): there is considerable simplification and uniformization of the transcriptions behind each of the published maps (reflecting in part the “normas simplificadoras” which Sanchis Guarner and Rodríguez Castellano received from Navarro Tomás in New York during their 1950-1951 work trip, Heap 2002: 17). So while making the raw data available directly to researchers was the right decision at the time, this initial online database was clearly never the desired end-point for the ALPI project, but rather a step towards the eventual goal of linguistic cartography (Heap 2002: 18).

Despite these very real limitations, the initial access gave impetus to a certain amount of dialect research using the newly accessible ALPI data. One exemplary early user was Pato, whose doctoral thesis (2003) under the supervision of Fernández-Ordóñez, used the unpublished (and hitherto unaccessible) ALPI data to analyse variation between imperfect subjunctive -se/-ra and conditional / imperfect -ría/-ba verb forms in Northern Peninsular Spanish (ALPI 386. Si tuviera dinero lo compraría, 387. Si estudiase aprendería, 388. Si pudiera la mataría y 390. Ojalá lloviese, all in Notebook I). Since then, much of the research based on ALPI data has a similar focus on dialect morphosyntax, for example on certain pronominal and adverbial constructions. Pato & Heap (2005) and Pato (2010) examine the recategorization of medio in the expression Estaba medio dormida (ALPI, 406), a phenomenon associated with Galician but which also appears in Seville and Cádiz provinces. Pato (2012) examines the conjunction manque in Aunque haga frío saldré (ALPI 393), which appears in just two isolated ALPI survey points but is best viewed as a generalized vernacular form and not characteristic of a particular region. Heap (2006) analyses the ‘inversion’ of the order of clitics in sequences like Se me cayó del bolsillo and ¿Se te calmó el dolor? (ALPI 348 and 349), showing that the vernacular orders me se / te se are nearly categorical in the Spanish-speaking areas of the ALPI. Enrique-Arias (2006, 2011) looks at forms of the second person plural clitic os (sos~sus, us, bos~bus, los, tos

5 ALPI point 504 (Adamuz, Córdoba) and point 408 (Bustantegua, Cantabria).
y se) in expressions like Os vais a caer, Os han engañado, and ¿Os queréis callar? (ALPI 341, 342 and 343).

Verb tense usage is also the object of a number of studies, such as Carter (2003) on the present perfect and Pato & Heap (2009)’s study of the dynamic boundary between the simple preterit and the perfect past being spread from the central zone of the Pensinula. Pato (2010) examines analogical strong preterites (like dijon, quison) using data from Hicieron una caja de madera and Trajeron la harina al horno (ALPI 257 and 320), while Pato & O’Neill (2012) trace the geographic distribution of analogical gerunds in Todo se alcanza teniendo paciencia and Se puede vivir queriendo trabajar (ALPI 326, 327), showing that analogical forms like tuviendo and quisiendo are found especially in the Eastern Peninsula. Heap & Pato (2012) analyse Spanish imperatives and infinitives with plural morphology following the enclitic se (Siéntense (ustedes) and No quieren abrigarse, ALPI 364 and 365), showing that they are characteristic of the Eastern Peninsula, with ‘metathesized’ forms (siéntesen) much more widely used than the reduplicated form (siéntensen).

The ALPI notebook data also allowed for a number of lexical studies, such as Heap & Pato (2007c), which surveys the ‘Asturian domain’ (55 ALPI points, mainly in Asturias but also in Cantabria, León, Zamora, Salamanca and Bragança) with maps for lines 444 (Margarita), 448 (Escaramujo, planta), 468 (Guisantes) and 481 (Bellota), all from Notebook II, while Pato (2011) examines the distribution and lexical areas of terms for ‘weasel’ (mustela, Notebook II, 516), with different variants concentrated in different parts of the Peninsula. These forms, also mapped in Fernández-Ordóñez (2011), reflect the classic ‘Castilian wedge’ theory as proposed by Menéndez Pidal, who first showed the geographic significance of this lexical variable in his Orígenes del español (1926). A thesis dealing with lexical items relating to the human eye in dialects and varieties across the Iberian Peninsula (Julià Luna 2007, 2009) uses ALPI data among other linguistic atlas sources.

At the Universidad Autónoma de Madrid, Fernández-Ordóñez has supervised in recent years a number of research projects based, in whole or in part, on ALPI data, including an MA thesis on appreciative suffixation (diminutives, augmentatives and

pejoratives), 7 Martín Izquierdo (2008), an MA thesis on subject-agreeing clitics (reflexive se, passive-impersonal se, middle voice se, and aspectual se), 8 de Benito Moreno (2009); an MA thesis on the use of ustedes for vosotros in Western Andalusia, 9 Lara Bermejo (2010); and an MA thesis examining intervocalic and final /d/ in ALPI data, 10 Estrada Arráez (2012 and this volume). A number of these studies have been published since as articles (de Benito 2010 and 2011, as well as Lara Bermejo, Estrada Arráez and others in this volume).

From a historical point of view, the ALPI data help us achieve a better understanding of the respective contributions of Castilian and other peninsular dialects to the development of Spanish. The inaugural address delivered by Fernández-Ordóñez on her induction to the Real Academia Española (2011) was historic from a number of perspectives, not least because of the large number of dialect maps presented: 32 maps in all, most of them based on ALPI data, whether lexical 11 or grammatical. 12 Fernández-

---

7 ALPI lines 262-274 in Notebook I: gaitú, gatoz, gatucho, casita, casona, casucha, sombrerito, sombrerucho, arbolito, puertecita, Pedrito, Juanito, Mariquita.
11 ALPI salvado (824), ubre (564), levadura (825), hogar (682), grano suelto (795), nogal (482), cubo de la rueda (777), anamola (443), peonzia (739), morueco (524), zorra (285), regaliz (453), mecer (la cuna) (705), alboroque (728), cría de la oveja (537), cría de la cabra (536), mazorca (278 & 464), and milano (581), all from Notebook II.
12 ALPI 262, 265, 268, 270 y 271, aún (302), -one vs. -aceu (263 y 266), taxu (515), No quieren abrigarse (365), second person plural pronouns (411), and article + possessive (260 y 261).
Ordóñez (2012) also showed the role of the North of the Iberian Peninsula in the history of the language, thanks to maps based on imperative and infinitive forms, among others.\(^{13}\)

Some researchers use ALPI data to address issues of a localized nature regarding the linguistic varieties of specific regions such as Arnal & Castañer (2011), Saura Rami (2006), Viruete Erdozain (2009) on Aragonese, García Perales (2001, 2002), García Perales & Casanova (2012) on Valencian, or the boundaries between traditional linguistic regions, such as Penny (2010) on the Asturian-Galician transitional areas, and González Ferrero (2007) on Leonese. Fernández-Ordóñez (2012) has also been influential in showing the relevance of dialect data like those from the ALPI for linguistic theory, as have others (Heap 2008c, Enrique-Arias 2011).

2. Towards a 21st Century ALPI: Moving forward and back to language history

Of course, as noted above, the field notebook transcriptions, while providing access to raw data which gives impetus to new research using the original (now historical) ALPI data, were never intended to be the end-point of the project, but rather a step on the path towards cartographic analyses which allow for geolinguistic interpretations (Pato & Heap 2007ab). Even those researchers who favour publishing linguistic atlas data as lists recognize the interest of interpretive maps that show the geographic distribution of linguistic variables.\(^{14}\)

Following a series of activities commemorating Navarro Tomás in his native Castilla-La Mancha and in Madrid (Salaberría 2007, see also Torrens Álvarez 2008) which helped bring the ALPI project to national attention, the Consejo Superior de Investigaciones Científicas (CSIC) proposed a new “intramural” (internal) project to publish the ALPI data in electronic format.

Just as Navarro Tomás and his collaborators were at the cutting edge of bringing disciplines like instrumental phonetics to the study of Iberian languages in the early 20\(^{th}\) Century, the ALPI is now poised to become a cutting-edge tool for geolinguistic research in the 21\(^{st}\) Century. While there may well be paper maps produced from the


\[^{14}\] "Junto al archivo, lo que interesaría hallar en el atlas son mapas ya interpretados (con símbolos discernibles en una mirada de conjunto, con límites, etc.)." (Catalán 1964: 233, note 1).
ALPI data, the potential of the electronic GIS database now greatly exceeds that of the printed atlas.

While the methods used to publish the ALPI data reflects the latest innovations in electronic dialectology, the data themselves are of course entirely historical, which of course was already the case when Navarro Tomás (1975) remarked on their historical nature as representing the linguistic situation before the Civil War (for the majority of survey points). One of the exciting possibilities which the ALPI data affords us it the ability to return towards the roots of dialectology in historical linguistics, but now with a privileged position whereby we can study real-time change in comparative corpora which contrast linguistic atlas surveys from successive periods (Kehrin 2012): the ALPI data (mostly from the 1930s) can be used as a baseline to study features which may have changed (or not) in later surveys which employ similar, if not identical, methods (cf. García Mouton & Molina Martos, Rodríguez Lorenzo, Rosés Labrada, Sousa Fernández, in this volume).

The ALPI’s long and often difficult history is far from over, but the last ten years offer us concrete evidence suggesting a much more fruitful future than the final decades of the last century suggested — and its best years may be yet to come.

References


GARCÍA MOUTON, P. & I. MOLINA MARTOS (in press) Atlas Dialectal de Madrid (ADiM), Madrid: CSIC.


métodos informáticos en el tratamiento de las lenguas ibéricas. Tokyo: Universidad Sofía, 77-84.


PATO, E. & P. O’NEILL (in press) “Los gerundios analógicos en la historia del español (e iberorromance)”, Nueva Revista de Filología Hispánica, 60/1.


REVISTA DE FILOLOGÍA ESPAÑOLA (1915) “Alfabeto fonético”. Revista de Filología Española, 2, 374-376.


etnològics, filològics i lingüístics oferts a Artur Quintana i Font, Calaceite: ICM-IEBC-IET, 549-563.