Intangible cultural heritage of the Portugal-Spain border: a contribution from language documentation

Xosé A. Álvarez-Pérez
Universidade de Alcalá

Online archive:

http://www.frontespo.org/en

This paper describes the audio-visual corpus developed within the research project Spanish-Portuguese border: linguistic and bibliographic documentation (FRONTESPO). This contribution intends to demonstrate its vast potential for the safeguarding and the study of the Intangible Cultural Heritage of the Portugal-Spain border, both as regards the rich linguistic landscape of this territory, as well as ethnographic materials, information on traditional life and memories linked to the border: smuggling, cross-border immigration, and the like. The first phase of the oral corpus has explored nine survey areas distributed along the entire border, with diverse geographic and linguistic characteristics. Each area comprises normally two pairs of Spanish and Portuguese localities, and within each locality, at least three persons are interviewed. FRONTESPO is firmly committed with open access; therefore, all the interviews will be freely available on the net. In this way, the team seeks to maximize the scientific impact, and to facilitate return to border communities.

The human and linguistic landscape of the borderland

The border between Spain and Portugal is one of the oldest boundaries in the world, and is also one of the most stable because it has remained practically unaltered since the Treaty of Alcañices (1297). "Relationships between the neighbouring towns on each side of the border have always been intense over the centuries (Amante, 2010), and are usually "stronger than is the contact with localities placed further inland in the same country (Beswick, 2014). Consequently, over the centuries, many borderland communities, with specific feelings of identity and shared experiences (commerce, smuggling, illegal cross-border migration, etc.) have developed (Medina, 2006).

This intense human contact for centuries and the geographic location of this territory, far away from the two countries capitals and regional centres of prestige,[1] have led to a very rich linguistic landscape with extremely interesting features, such as the following:

– Linguistic borders often do not match the political boundaries (Boller, 1995), especially with regard to lexical circulation.
– Along the Spanish side of the border, there are some Portuguese-speaking localities that originated as a result of diverse historical situations (military conquests, displacement of boundaries, immigration, and the like).

---

[1] Major cities that are linguistic role models because of their administrative and economic importance.

– The continued cross-border relationship and the relative isolation from the rest of the country have sometimes resulted in the emergence of specific linguistic varieties, such as the *barranquenho*, a mixed Portuguese-based variety with a strong influence from southern Spanish dialects (Navas Sánchez-Élez, 2011).
– The dialectal varieties spoken along the borderland preserve archaic linguistic traits or lexical items that have disappeared from the rest of the linguistic domain because most border communities are composed by rural localities that are far from prestige-centres.

However, it must be noted that this cultural and linguistic heritage of the borderland is currently undergoing an intense transformation. Some converging processes are shared with other rural areas: a) depopulation of rural territories, b) loss of the traditional way of life; c) stronger exposure to standard languages (school, mass media, and so on), d) linguistic harmonisation with regional dialects, because of a more intense mobility to the cities for work, shopping, and the like. Nevertheless, there are also specific factors related to the nature of border area. The entry into force of the Schengen Agreement has greatly favoured mobility between Spain and Portugal: the number of cross-border commuters has risen significantly (Falagán *et al.,* 2013), as has cross-border mobility to make everyday purchases. Obviously, this daily human mobility means increasing exposure to other linguistic models and promotes linguistic interference.

A project of linguistic and bibliographical documentation of the Portugal-Spain border
Considering the aforementioned precarious situation of the intangible heritage of border communities, in 2015 the university of Alcalá (Spain) launched the project FRONTESPO (*Frontera hispano-portuguesa: documentación lingüística y bibliográfica* - "Spanish-Portuguese Border: linguistic and bibliographical documentation"), a research enterprise aimed at obtaining a comprehensive view of the linguistic configuration of the border between those states, at the present time and throughout history.

Nevertheless, it should be made clear at the outset that the project does not restrict its focus to the linguistic aspect. In this sense, among other actions, it should be mentioned explicitly that the work plan foresees a specific line heading to the development of a multidisciplinary reference database. It will integrate bibliographical records, websites and profiles in social networks aimed to divulge and to study the Portugal-Spain border, paying special attention to the historical and geographical factors that have determined the linguistic reality of the borderland, such as medieval settlements or local cross-border migrations.

A core element of FRONTESPO is the collection of an oral corpus. It will be a valuable tool for obtaining linguistic and sociolinguistic information. Nevertheless, a good part of the collected materials also documents other elements of the cultural heritage of this territory, as detailed below. Thanks to our open access policy, other scholars (and, obviously, the border communities themselves) may study the information provided by the oral corpus. In any case, regardless of which of the disciplines will exploit the information, this project will be a great opportunity to preserve a huge amount of cultural data that are vanishing at an accelerating pace.

Oral Corpus of the Portugal-Spain border

Contribution to the knowledge of Cultural Intangible Heritage in this territory
With respect to the linguistic heritage of the borderland, I should like to mention the following priorities: a) safeguarding of the traditional language that is disappearing at an increasing rate and is of extraordinary interest for scientific research; b) documentation of the language of younger and middle generations, who are more exposed to standard languages and regional centres of prestige, and an examination of contrast with the linguistic features of older speakers; c) evidences of linguistic change driven by increased exposure to standard languages (labour mobility all over Portugal and Spain, more frequent travels to the cities, mass media, universalisation of education and the like). This circumstance shows itself

in linguistic interferences, loss of dialectal linguistic features or lexical reduction; d) current information about the linguistic and sociolinguistic situation of enclaves at the Spanish side of the border, which shows a steady decline of Portuguese language in these localities; e) speakers' linguistic attitudes and identities, taking into account differences among generations: reasons for the abandonment of dialectal varieties, considerations about local language, neighbour varieties or standard, etc.

Regarding other cultural contents, the interviews about traditional life will provide with extensive ethnographic information: agriculture, livestock, fabrication of bread and wine, local festivities, etc. Besides, a strong component of this corpus are the speakers' own voices and direct experiences of cross-border mobility and relationships with the neighbouring country, in open and clandestine contexts, which are essential to the construction of border identity: smuggling and commerce, exiles and fugitives, mixed marriages, popular pilgrimages, and the like.

Geographical area. Informants and structure of the interviews

For the first phase of the project, nine survey areas have been established (see Fig. 1). They cover a representative sample of different linguistic situations, since this initial network is a pilot run to be expanded in future stages. In general, each of the survey areas consists of four localities (two Spanish and two Portuguese villages). One criterion was that one of the locations chosen in each country should have been previously explored in dialectal atlases or monographs, in order to make diachronic comparisons.

Figure 1. Map showing the initial survey areas.

During the execution of fieldwork, 57 Spanish and Portuguese localities were explored and 178 individual and group interviews were conducted, with 237 main and secondary informants. Altogether, the corpus exceeds 225 hours of raw video recordings, plus 53 hours of only audio recordings (because informants refused to be filmed).[2] Additionally, there are 90 gigabytes of complementary material: photographs of objects, recordings at ethnographic museums, etc.

Within each survey point, at least three informants are interviewed, with representation of both sexes and age stratification in three ranges: a) speakers over 75 years; b) speakers between 50 and 75 years; c) speakers under 50 years.

Our dialectal interviews are usually organised in three blocks:

1. The central core is a semi-structured conversation about various semantic fields related to traditional life: agricultural works, bread making, livestock, wildlife, etc. In order to be able to establish systematic comparisons throughout the network, the interviewer has a small questionnaire, with about a hundred concepts to be asked if the informant do not name them spontaneously during the interview;

2. The second section examines the informants' linguistic behaviour and identity. That is, we explore issues such as the (perceived) degree of divergence with the standard and other dialects, informant's opinion about linguistic change within the community, the degree of understanding of the language spoken at the other side of the border, etc.;

3. The third block aims to obtain a more spontaneous record by talking about topics such as the smuggling, the relationships between Portuguese and Spaniards or the contrast between present life and former days, which are well-known issues for any informant. In some cases, we have opted by group interviews where the informants can dialogue between them. In that way, as both speakers are members of the linguistic community, we can expect that a "more real" language will be produced.

Edition and digital publication

The project's website in Drupal is the central point for the recordings. [3] There is a map where all the survey points are located and a sheet for each locality and interview with information about the villages and the informants; it is possible to perform searches according to several criteria (country, age range, sex and the like).

The files are stored externally in different specialised repositories: YouTube and Vimeo, for videos, and SoundCloud, for audio. External hosting permits savings in our IT infrastructure. Besides, as our corpus is available under open access, other sites may copy these files on their servers, increasing dissemination and preservation. The issue of the long-term preservation of audio-visual materials will be addressed in the following months using two complementary procedures, namely storing them in Wikimedia Commons and language documentation repositories.

The raw materials from each interview are being edited, to eliminate non-pertinent parts of the recordings and to group the diverse informants' testimonies thematically. The original MTS video files are segmented using TsMuxeR, which allows for the cropping, cutting and merging of files without re-encoding, and will be edited using Adobe Premiere. The audio in WAV is edited using Adobe Audition.[4]

The original unedited files are kept in our archives. However, in order to edit and store them in external repositories it is imperative to make some conversions. Video files are uploaded to YouTube and Vimeo in MP4 format, but, because of intellectual property issues, hosting in Wikimedia Commons requires the conversion to WebM, a royalty-free container. Similarly, audio recordings are stored in the original WAV format in Soundcloud, but before being uploaded to Wikimedia Commons, they are transformed to OGG.

---

[2] Videotaped interviews provide some advantages. They capture the multimodality of language: voice, but also facial gestures, movement of eyes and body, etc. Besides, informant's gestures allow a better understanding of the explanation of some concepts or activities. Last, but not least, it is a friendlier format for online dissemination.

[3] http://www.frontespo.org/en/grabaciones

[4] More precisely, there are two sources of audio recording: the videorecorder (sampled at 48 kHz with 16-bit depth) and an audio recorder connected to a high-quality cardioid microphone; as the latter is sampled at 48 kHz with 24-bit depth, it will be more useful for phonetic research.

Figure 2. Screenshot of the section of recordings (*beta* version).

Transcription.

The oral corpus is being fully transcribed. This will be a tremendous asset for research on this corpus and for consultation by the public, since it will be possible to read the contents of the entire interview without needing to listen the audio track; besides, it will allow textual searches.

The team will use ELAN, a widely documented professional tool that allows time-aligned annotation of audio or video streams and that uses a system of overlapping tiers. The team will undertake an orthographic transcription of the speech and will identify the main structural elements of the conversation (identification of speakers, pauses, conversational sequences and the like). However, as the files containing the audio, the video and the transcripts in ELAN will be freely available on the site, other groups may download them and add other tiers according to their research needs and theoretical orientations without having to start from scratch. Detailed metadata (cataloguing, descriptive, structural, technical and administrative) will follow IMDI guidelines, ensuring the correct description in databases and repositories.
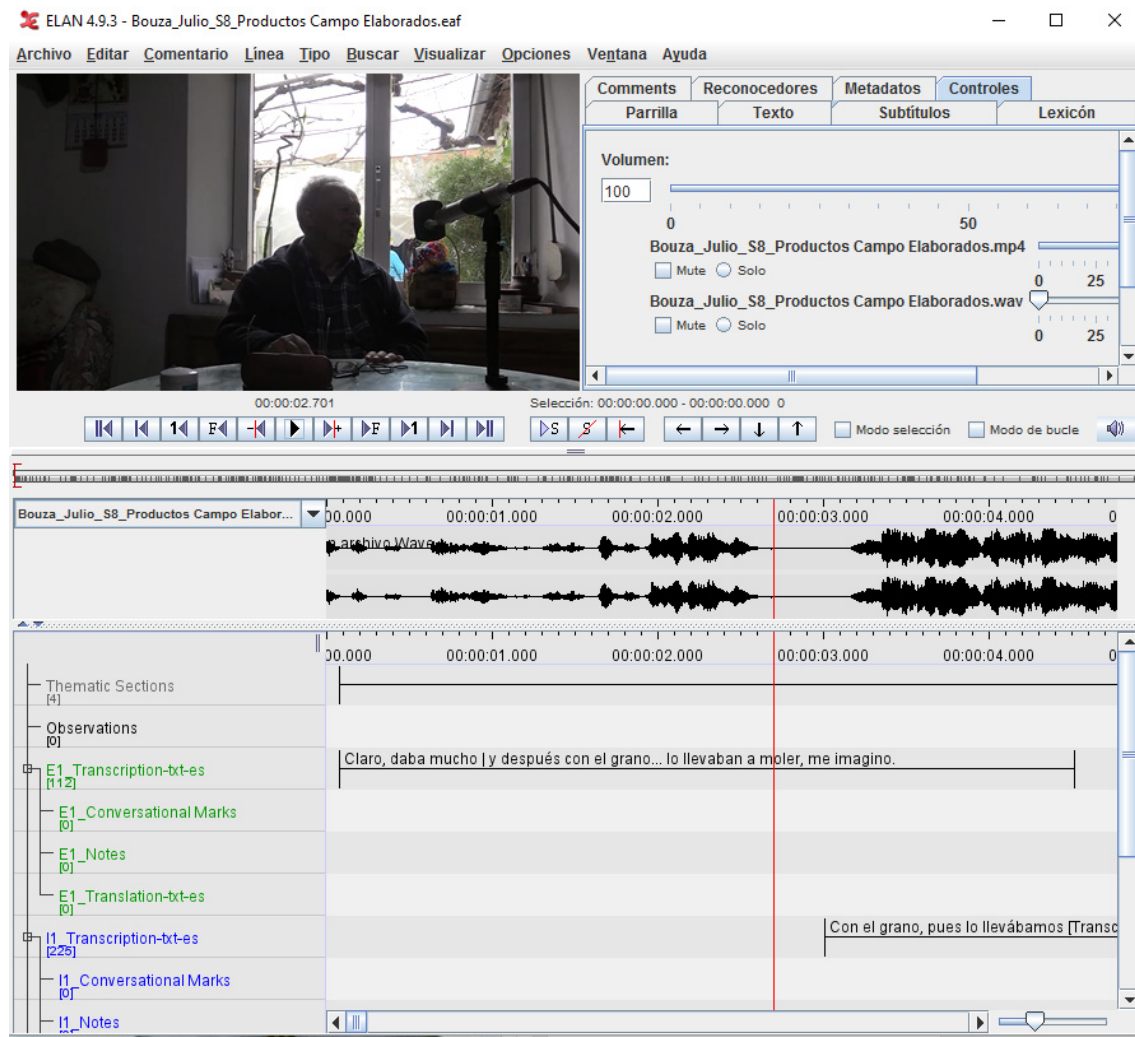
Figure 3. Time-aligned transcription with ELAN.

Publication and dissemination

This project fully embraces the principles of action stated in the *Manifeste des Digital humanities* (2010).[5] Accordingly, our dissemination plan will pay special attention to open access to the primary data collected during our project, to ensuring the long-term preservation of data and metadata, to the establishment of interdisciplinary partnerships and to the involvement with local communities. In order to guarantee free access to the data and the use thereof, the materials gathered or produced by the team are available under a Creative Commons Attribution–ShareAlike 4.0. license.

This project is committed to integrating local communities in the work plan. The stakeholders have been usually relegated to a passive role: they are research subjects, but the results of the research are normally inaccessible to them, since the findings are published as technical papers in specialised journals. Our proposal aims to create dialogue between local communities and scholars through specific actions: informative talks about border culture, creation of adapted materials on local language, the planning of revitalisation measures —including didactic units for schools, the promotion of traditional heritage in bilingual villages and the aim of fostering the generational transmission of language—, and the presentation of primary materials to local speakers in order to obtain verification, refutation or supplementary data. The main risk involved is the need to engage two very different worlds in dialogue, since our materials must be attractive and accessible without sacrificing scientific rigour. The close relationship with cultural associations will allow refining the approach of our materials and activities.

---

[5] http://tcp.hypotheses.org/category/manifeste

References

· AMANTE, Maria Fátima. (2010). "Local discursive strategies for the cultural construction of the border: The case of the Portuguese–Spanish border". In *Journal of Borderlands Studies* 25:1. 99-114.

· BESWICK, Jaine. (2014). "Borders within borders: Contexts of Language Use and Local Identity Configuration in Southern Galicia", Dominic Watt / Carmen Llamas (eds.). *Language, borders and identit*y, Edinburgh: Edinburgh University Press. 105-117.

· BOLLER, Fred. (1995). *Die Isoglossenstaffelung in der galicisch-portugiesisch-spanischen Kontaktzone und im Lombada-Aliste-Grenzgebiet*. 2 vols. Kiel: Westensee-Verl.

· FALAGÁN MOTA, Jorge; DE CARLOS VILLAMARÍN, Pablo; LORENZO ALONSO, Pedro. (2013). "Movilidad transfronteriza de trabajadores entre Galicia y el norte de Portugal". In *Contabilidad y negocios*, vol. 8, no. 15. 77-94.

· MEDINA GARCÍA, Eusebio. (2006). "Orígenes históricos y ambigüedad de la frontera hispano-lusa (La Raya)." In *Revista de Estudios Extremeños* 62 (2).713-724.

· NAVAS SÁNCHEZ-ÉLEZ, María Victoria. (2011). *El barranqueño: un modelo de lenguas en contacto*. Madrid: Universidad Complutense.