

# ARCHIVAL DYNAMICS : the *Langues De France* PROJECT AND THE BUILDING OF THE JUDEO-SPANISH ORAL ARCHIVE

Pandelis Mavrogiannis (Ph.D. student, CERMOM, Inalco) - Supervisors: M. C. Bornes Varol (Inalco) & F. Descamps (EPHE)

The Judeo-Spanish Oral Archive is accessible at: [https://cocoon.huma-num.fr/exist/crdo/meta/crdo-COLLECTION\\_JSFA](https://cocoon.huma-num.fr/exist/crdo/meta/crdo-COLLECTION_JSFA)

CLARIN Annual meeting, 8-10 october 2018, Pisa, Italy

Contact : [pandelis2000@hotmail.com](mailto:pandelis2000@hotmail.com)

## 1 The story of the JSOA archive

Until 2005, publicly accessible recordings in Judeo-Spanish were inexistent. At that moment, the CNRS ran the *Langues de France*, a CNRS program supported by the French Ministry of Culture's DGLFLF section. The program (2005-2009) was aimed at building oral corpora for endangered languages spoken in France, and had a Judeo-Spanish component, led by Pr. Varol (CERMOM, Inalco).

The value of the data recorded for other social sciences was obvious: what informants were saying to the microphone was more important than the way they were expressing themselves. The initial, linguist-oriented documentation scheme proved narrow, as it was not permitting reuse: the data gathered were concerning the history of migration of Ottoman and post-Ottoman Jews, the building and decay of Sephardic communities, and of course the Holocaust.

The Judeo-Spanish Oral Archive is born from the will to account for these aspects and permit reuse of data.

## JSOA objectives

- Judeo-Spanish: the language of the Jews expelled from Spain in 1492 and established in the Ottoman Empire
- Document Judeo-Spanish, an endangered language and culture, by :
    - localising existing corpora
    - digitising / preserving old recordings
    - linguistic and cross-field documentation
    - providing time-aligned, searchable transcriptions and linguistic annotations
    - providing efficient, cross-field abstracts
    - building a reference corpus for linguists, social scientists, and historians
    - open-access archival (CNRS/CoCoON)

## Inventory of oral resources in Judeo-Spanish (as at 01/10/2018)

Country	Corpus	Corpus	Corpus	Corpus	Total JSOA	Total
Turkey	Varol [60]	Gershon [81]	Spiegel [73*]		141	214*
Bulgaria	Collin [23]	Voyevoda [4]	Kantchev [5*]	Kantchev [130]	32*	162*
Greece	Weich-Shahak [15*]	Gimenez [12*]	Bossong [70**]		27*	97**
Balkans (other)	Weich-Shahak [40**]	Papo [3]			43**	43**
Israel	Varol [38]	Weich-Shahak [160*]	Quintana [200*]	Shaul ; Peretz [400**]	198*	798**
France	Varol, Pfister, Mavrogiannis, Collin G., Collin Gw. ( <i>Langues de France</i> ) [80]	Sephiha ; Algazi: Radio broadcasts (Paris, Lyon, Marseille) [400*]	Sephiha ; Sadak: Language workshops (Paris, Marseille) [420*]		900*	900*
Switzerland	Voyevoda [3*]				3*	3*
Argentina	Mavrogiannis [12]	Lebuitski [18]	Weich-Shahak [50**]		80**	80**
Brazil	Cohen [3*]				3*	3*
Latin Amer. (other)	Weich-Shahak [160**]				160**	160**
USA	Weich-Shahak [30**]	Benmayor [200*]	Kirshner [100**]	Other [200**]	30**	530**
Total					1525**	2998**

## Ethical and legal issues

- New and recent corpora : informants' authorisations
- Old corpora : strategies taking into account the national legal framework
- Conflicts : publish-but-retrieve policy
- Efforts towards compliance to GRDP
- Anonymisation where necessary
- Filter when necessary
- Refrain from on-line publication (some cases)

## Cross-field interest of the JSOA (examples)

- Linguistics : possible to follow various idiolects over a long period (30 years) ; language attitudes ; syntax
- Anthropology : magical-religious practices
- Religion studies : religious syncretism
- Musicology : dozens of undocumented song versions
- Littérature : unveiling unknown theatrical plays
- Sociology : representation of the Other ; community expectations ; social change
- Labor studies : labor structure of Judeo-Spanish communities

## Linguistic and cross-field documentation

### FEATURES OF THE ARCHIVE

- Interoperability : OAI-PMH protocole
- Dublic-Core compliance
- Free, open access (HumaNum - CoCoON)
- Attention paid to ethical issues
- Cartography of resources

### DESCRIPTIVE METADATA

- OLAC metadata
- Linguistic biographies
- EAD description (BNF)
- Adapted, cross-field, medium-sized, multi-lingual (EN/ FR/ ES) abstracts

## Explanation of annotations

- Inventory of corpora completed. Contributes to the JSOA
- Inventory of corpora completed. May contribute to the JSOA
- Inventory completed. The researcher cannot localise his recordings.
- Inventory to be realised.
- [ ] Number of resources known
- [ \* ] Number of resources estimated on good evidence
- [ \*\* ] Number of resources very roughly estimated (only an indication)

## Archival Dynamics : from a small corpus to a comprehensive, cross-field oral archive

JSOA	Phase 1 (2005 - 2008)	Phase 2 (2009 - 2010)	Phase 3 (2017 - 2020)
<a href="https://cocoon.huma-num.fr/exist/crdo/meta/crdo-COLLECTION_JSFA">https://cocoon.huma-num.fr/exist/crdo/meta/crdo-COLLECTION_JSFA</a>			
Institutions	CNRS - DGLFLF	CNRS - DGLFLF	CNRS - BNF - INALCO
Financement rate	25% (DGLFLF)	25% (DGLFLF)	75% (Ph.D. grant)
N° of recordings	70	250	1500
Main development	Interoperability	Geographical scope	Oral history/ Social sciences
N° of corpora	1	6	15
N° of countries	1 country : FR	3 countries : FR - TR - IL	12 countries (Europe, Americas)
Repository	CNRS/CRDO (Ling. )	CNRS/COCOON	CNRS/COCOON
Scientific field	Linguistics	Linguistics	Social Sciences/Oral History/Ling.
Achievements	Interoperable metadata OLAC vocabularies Long-term preservation CRDO support Well-defined objectives	Preservation of old corpora Resources from > 1 country	Better documentation: EAD (BNF) Multi-field docum.: new abstracts Multilingual abstracts (EN-ES-FR) Reuse of corpora encouraged Oral history-oriented PhD research
Limits	No "old" corpora Data limited to France Poor abstracts Abstracts only in FR Reuse not encouraged	Poor abstracts Abstracts only in FR Reuse not encouraged Unresolved ethical issues	Difficulties to resolve ethical issues Misdocumentation risks Geographical bias New version of JSOA not published (work in progress) Suggestions ?

## 4 Developing a methodology for abstracts

Reuse of JSOA oral resources implies providing efficient summaries. The general guideline of 10-15 lines summary proved poor.

- It was necessary to establish formal criteria to arrive at what I call an « optimal abstract » for an oral archive such as the JSOA, by :
- a minimalist but not minimum description
  - regrouping recurrent information
  - accounting of as many aspects interesting linguistics, social sciences and oral history as possible
  - identifying and expliciting in the PhD thesis the information 'lost' as a result of the documentation trade-offs between completeness and efficiency of the abstracts

## Oral History reuse of JSOA: 2 examples

### France: The Sephiha corpus (1980's)

- Emigration from Turkey
- History of French Sephardic communities
- Racial persecution in the Balkans (WWII)
- Racial discrimination in Turkey (WWII)
- Latin America, Balkan and Morocco Sephardic Judaism

### TURKEY : The Varol corpus (1980's)

- « *Vatandaş Türkçe konuş!* » campaigns (1930's)
- Trakya Olayları* (1934)
- The *Vente Klasas* (circa 1942)
- Los Ornos de Balat* (circa 1942)
- Jews in Turkish society

## Bibliography

- ALBERTI-KLEINFORT, E., « La tradición oral sefardí en Sudamérica. Un Archivo Documental de Argentina », *Neue Romania, Judenspanisch* V, n°24, 2001, p. 7-24.
- BAUDE, O. (dir.), *Corpus oraux. Guide des bonnes pratiques*, Paris, CNRS Éditions & Presses Universitaires d'Orléans, 2006.
- BENVENISTE, A., *Le Bosphore à la Roquette: la communauté judéo-espagnole à Paris (1914-1940)*, Paris, L'Harmattan, 1989
- BORNES VAROL, M.C., *Le judéo-espagnol vernaculaire d'Istanbul*, Peter Lang, Berne, 2008
- CNRS/COCOON, *COLLECTIONS de CORPUS ORAUX NUMÉRIQUES (COCOON)*, [En ligne]. <https://cocoon.huma-num.fr/exist/crdo/>. (Page consultée le 05 octobre 2018)
- DESCAMPS, F. *L'historien, l'archiviste et le magnétophone*, Paris, Comité pour l'histoire économique et financière de la France, 2001
- DESCAMPS, F., « Mettre à disposition ou réutiliser des corpus sonores et audiovisuels en SHS ». Introduction à la journée du 19 septembre 2008, Aix-en-Provence, « Collecter, organiser les archives de la recherche en sciences humaines et sociales ».
- JACOBSON, M., *Corpus Oraux en linguistique de terrain, Traitement Automatique des Langues*, 2004, 45/2 : 63-88
- LIEBL, C., *Early recordings of Judeo-Spanish in the Phonogrammarchiv of the Austrian Academy of Sciences, Neue Romania* 37 (2007) (=Judenspanisch XI), pp. 7-26
- MAVROGIANNIS, P., "The Judeo-Spanish oral Archive (JSA): data collection, metadata description, results, and perspectives for development", in STUEMUND-HALEVY, M., LIEBL, M. et VUCINA, I. (dir.), *Sefarad an der Donau : lengua y literatura de los sefardíes en tierras de los Habsburgo*, Barcelona: Tirocinio, 2013, p. 385-40
- STUEMUND-HALEVY, M. et FISCHER S., "Audiovisual Documentation on Bulgarian Judezmo" in Michael Stuemund-Halevy, Christian Liebl et Ivana Vucina Simovic, (dir.), *Sefarad an der Donau. La lengua y literatura de los sefardíes en tierras de los Habsburgo*, Barcelona, Tirocinio, 2013, pp. 407-424