# The LESCO Corpus. Data for the Description of Costa Rican Sign Language

**Alejandro Oviedo[*], Christian Ramírez Valerio[**]**
[*]Department of Deaf Education, University of Cologne
[**]Doctoral Program in Social and Cultural Studies, University of Costa Rica
[*]alejandro.oviedo@uni-koeln.de, [**]xianlesco@yahoo.com

**Abstract**

The LESCO Corpus comprises transcriptions in Spanish glosses and translations into Spanish of videos of the Costa Rican Sign Language (known as LESCO). Transcriptions were made by a team of five Deaf LESCO users. The corpus was produced between 2011 and 2013 and has served as the basis for the development of a basic grammar and a dictionary of LESCO which are available online (www.cenarec-lesco.org. See Costa Rica, 2013a, b). The primary data includes 44 dialogues between two informants. Each film is composed of two or three video files. The corpus lasts approximately 2 hours and have a file volume of 12.6 GB. The sample includes 27 adults (average age: 32 years. 13 women and 14 men). Metadata correspond to age, sex, age of acquisition of LESCO, place of residence and hearing status. The material is protected under a Creative Commons BY-NC-SA license and its use may be requested to the Ministry of Education of Costa Rica (www.cenarec-lesco.org). The authors of this article were in charge of the entire project.

**Keywords:** sign language corpora, Costa Rican Sign Language, LESCO.

## 1. Introduction

In 2008, the government of Costa Rica decided to make a linguistic description of LESCO, the most widely used sign language in the country's urban areas, with the purpose of having a base for the subsequent creation of LESCO teaching materials for both L2 and L1 (for the schools for the Deaf).

The research team consisted of six people, five of them Deaf users of LESCO. The project lasted 27 months (between February 2011 and July 2013) and involved an investment of close to 250 thousand euros. A corpus, a grammar and a dictionary were the final products. All these products are freely distributed under a CC BY-NC-SA license. The grammar and the dictionary has been online since 2014 (Costa Rica, 2013a, b)

## 2. Sociolinguistics of the Costa Rican Deaf

### 2.1. Deaf Community

The single national association, ANASCOR (founded in the 1970s) brings together people from all over the country. The Deaf community is very active politically, and is present in many programs related to Deaf people, especially in the areas of education and interpreter training as well as in the legal field. LESCO was officially recognized in 2012.

There are only two schools for the Deaf in the country, which means that many of them attend regular schools. Schools for the Deaf, until a few years ago, excluded sign language from the classroom, but traditionally, students are allowed to sign outside of class hours. Since the beginning of the decade of 2010, there are some preschool bilingual programs (Spanish-LESCO) and more recently, a Spanish-LESCO bilingual classroom pilot program.

### 2.2. Old LESCO. New LESCO

In Costa Rican cities there are at least two different sign languages in use[1]: the old LESCO (apparently emerged throughout the first half of the twentieth century, with much influence from the Spanish Sign Language and used today by people over 60 years) and the new LESCO (emerged in the second half of the 1970s, with strong influence from the American Sign Language and used by people under 60 years of age (Woodward 1991).

Most people over 60, users of Old LESCO, are fluent users of New LESCO, unlike the younger people, for whom the Old LESCO is unintelligible. Given that the later objective of the project was the Deaf school population, the decision was made to limit the study to the New LESCO, so that informants older than 60 years were not included in the sample. In informal conversations between government officials and members of the Deaf community, it was also proposed to carry out a study to document and describe the Old LESCO. This study has not started yet, but some members of the Deaf community have been filming Deaf community elders[2].

Among the Deaf community of the country, the nominal phrases "Old LESCO" and "New LESCO" are used very rarely and exclusively in contexts where both languages are a topic of conversation. The sign in use that names the sign language in Costa Rica (see Figure 1) designates by antonomasia the language used by the new generations in the urban centers. That is, New LESCO. We follow that

[1] Following Woodward (1991) there would be two other sign languages in use among native American communities: The Bribri and the Brunca Sign Languages. No further study has corroborated the existence of these sign languages.

[2] Personal communication of Alejandro Oviedo with Costa Rican Deaf researchers Christian Ramírez Valerio and Alexander Hernández, in San José, May 2012.

use here and we will understand by LESCO the language that Woodward (1991) designated as New LESCO.

Figure 1: The sign LESCO

### 2.3. How many Users of (New) LESCO Are There?

To estimate the number of LESCO Deaf users, we resorted to a study similar to this one, carried out previously in Venezuela (Oviedo, 2004), whose social conditions were then comparable to those of Costa Rica. Oviedo (2004) estimated the local Deaf population crossing multiple data, such as the numbers of people with hearing disabilities, the number of members of Deaf associations, the school population, and the international percentage of babies born deaf. The estimate was 15,000 Deaf users of the Venezuelan Sign Language, 0.05% of the country's population at that time (26 MM). The present study includes Deaf users of LESCO living in the Greater Metropolitan Area (henceforth, GAM), where it is estimated that more than 60% of the population of Costa Rica lives (about 2.6 MM people (Costa Rica, 2011)). Based on the aforementioned percentage, we assumed that some 1,300 Deaf users of LESCO lived in the GAM. This would be the population of the present study.

## 3. Data Collection

### 3.1. Preparation and Application of the Surveys

To select the sample, 141[3] people between 18 and 60 years old were interviewed between the months of February and March 2011. Interviews were conducted using a written survey designed by a Deaf researcher. Questions were formulated and answered in LESCO, translated into Spanish and recorded in the interview form by the interviewer.
The people surveyed were selected in such a way that each of the main cities of the GAM was represented. The surveys were digitized and archived for further research.

### 3.2. Preparation for Video-Recording

The 141 surveys were subjected to a selection process, carried out by our Deaf researchers. Each potential informant was assigned a score of 1-10 (1 = minimum, 10 = maximum), according to four criteria: age of acquisition, attendance at a school for the deaf, frequency of contacts with other Deaf people and fluency in LESCO. This process allowed the pre-selection of 102 people.
These persons received invitations to come to the Ministry of Education for the filming sessions. They agreed to sign a document in which the Ministry of Education was allowed to use and eventually publish the data. The use of the videos was defined according to a Creative Commons BY-NC-SA license.

### 3.3. Video-Recording, Selection of the Films

#### 3.3.1. Elicitation

Between March and May 2011, 196 filming sessions were made (2000 minutes, 34 hours, 482 video files, format 4:3). Texts were elicited and filmed according to the following scheme:

- An elicited narration (from a cartoon short film).
- A free narration (personal anecdotes).
- Unstructured dialogue between the informant and a Deaf researcher.
- Unstructured dialogue between two informants.
- Structured interview (based on a questionnaire)

All the filming took place in a dialogical situation, mostly with two cameras (Figure 2a), each of which took individual front shots of the participants. Some filming included a third camera directed at the informant's face (the second person present was always, in these cases, a Deaf researcher), with the intention of capturing facial activity in more detail (Figure 2b):

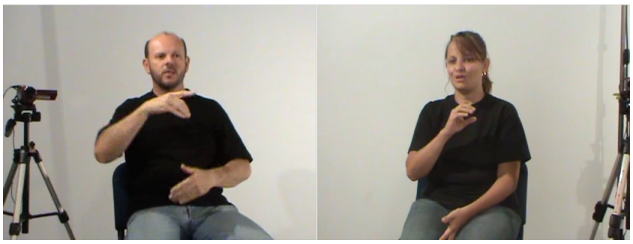

Figure 2a: A filming with two cameras

Figure 2b: A filming with three cameras

#### 3.3.2. Video-Recording and Selection

Videos obtained were subjected to a selection process in which points were assigned (again 1 to 10) according to the following criteria:

- Fluency and intelligibility of discourse,

[3] I.e. 13% of the estimated local Deaf population.

- Technical quality of the video,
- Absence of information that could compromise third parties, and
- Closeness with the LESCO used by our Deaf researchers (since some informants produced a discourse that in the opinion of the researchers was not LESCO, but another signed system like signed Spanish or ASL).

Films receiving more than 7 points were selected. The final result was 44 films, in which a total of 27 informants appear (some of them appear in two, three or four videos). Figure 3 illustrates the distribution of these 27 informants - regarding age of acquisition and year of birth- among the interviewed 141 people:

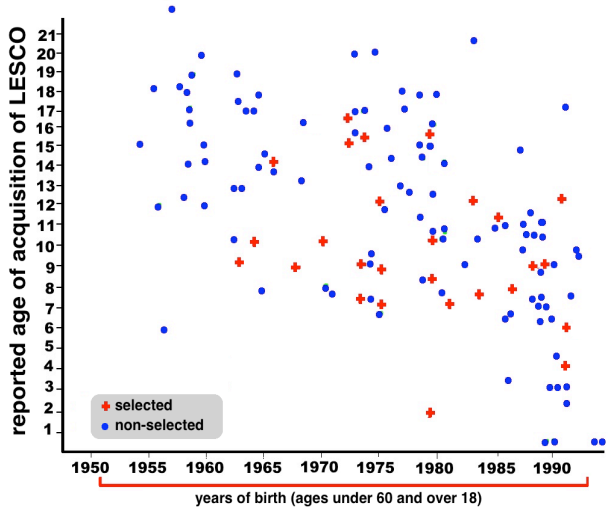


Figure 3. Data of informants

The red crosses in Figure 3 represent the people who were selected as informants for the LESCO Corpus. The blue dots represent the other 114 people who were interviewed in the selection process, but who were not selected as study informants. As can be seen, the relationship between age and age of LESCO acquisition of the selected people corresponds to the sample trend. The younger the age, the younger the age of acquisition. In the sample (141 people interviewed) the average age of acquisition was 11.27 years. Among the 27 selected it is 10.36 years.

Determining the age of acquisition was a complex procedure, since it depended on the assertions of the informant himself, based on childhood memories. In each case, the school and family history that each person reported was also taken into account. Among the 27 informants selected there were several who report relatively late ages of acquisition (between 12 and 17 years), but at the same time they report having Deaf relatives (siblings, uncles, cousins). In such cases, it is legitimate to assume that the actual acquisition ages would be lower than those reported. This would eventually allow us to further reduce the acquisition age of the selected group.

Finally, it should be noted that four Deaf CODAs were included in the initial sample. They correspond to the four blue points near the right end of the horizontal line in Figure 3 above. These people are assigned an acquisition age of 0 years. However, we excluded these four people as informants, since two of them were underage and the other two produced discourse considered by our team as signed Spanish.

# 4. Transcription

## 4.1. Conventions, general description

Corpus files contain a Spanish translation and an ID-Glossing line for each signer that appears in the video. Given glosses generally assumed the written form of the Spanish word(s), which the Deaf community relates to the basic meaning of that sign. Transitions between signs were not marked, but were included as part of the end of a sign and/or the beginning of the next sign. Therefore, only when pauses occur, the annotation line is interrupted. When the pauses were long (400 ms or more), the word PAUSED (pause) was included to fill the space in the ID-Glossing line. Finally, there was a series of annotation lines for manual parameters (i.e. handshape, orientation, location and movement) as well as for the non-manual articulators (head, shoulder, eyebrows, nose, eyes, lips, etc.). Figure 4 illustrates the transcription of a corpus file in ELAN:


Figure 4. Screenshot of a corpus file

## 4.2. Lemmatization

Before the realization of our study, at least two LESCO vocabularies were published (Bravo, 1979; López Gracioso, 1992). They were lists of Spanish glosses, arranged alphabetically and illustrated by means of a drawing or a photo. These works were our first references in the process of lemmatization (Johnston, 2010). Many signs in the corpus, however, did not appear in any of the works mentioned. In those cases, members of the research team assigned temporary glosses. Once the signs were transcribed in the corpus, we exported in ELAN the transcriptions as lists of words and compared the occurrences, taking into account basic form, changes of form observed in the corpus, meaning and use. In this process, errors were corrected or new lemmas were added. The result was a list of 1,541 lemmas and almost 14,000 tokens.

## 5. Some notes about the grammar and the dictionary of LESCO

### 5.1. The Grammar

The grammar of LESCO (Costa Rica, 2013b) consists of a text written in Spanish, divided into four major parts: phonetic-phonological, morphological, syntactic and discursive levels. In each of them, a series of subtopics offers explanations on more concrete aspects. For example, in the case of morphology, it provides information about the meanings and forms that the repetition of the lexical root can carry. In each case, the explanation is illustrated with numerous examples of the corpus, which can be observed in both photos and videos.

### 5.2. The dictionary

The dictionary of LESCO (Costa Rica, 2013a) was elaborated from the lemmas defined in the process of elaboration of the corpus. Out of the 1,541 lemmas defined, about 960 were selected for the dictionary. In some cases, certain semantic fields were not saturated with the signs found in the corpus. It was the case, for example, of the signs corresponding to colors. To complete them, we reviewed the videos that had not been selected for transcription. In this review, many of the signs sought were found and included in the dictionary. If after this process some semantic field still remained unsaturated (as with the signs corresponding to Costa Rican provinces and cities) we consulted members of the Deaf community of San José about the missing signs. Once a consensus was reached, a neutral form of the missing sign was filmed and a new lemma was defined. With this procedure it was possible to complete a list of 1,041 signs for the dictionary. Figure 5 illustrates an example of this (the entry for LESCO in the dictionary):


Figure 5. Screenshot of a dictionary´s entry

Each of those 1,041 lemmas has an entry in the dictionary. The signs can be found through three search criteria:

- The Spanish gloss (in alphabetical order),
- The manual configuration of the active hand (ordered according to the number of active fingers)
- Through a thematic index.

Each entry comprises a video of the neutral form ("forma neutra") as well as one or two videos taken from the corpus, in which examples of use of the sign appear. These examples are glossed and translated for ease of use. Additionally, the entry contains some information about the grammar and the meaning of the sign.

As far as we know, our work on LESCO was the second corpus-based description of a signed language in Hispanic Latin America. A previous experienced was carried out in Colombia between 2000 and 2005 (Oviedo, 2001; Colombia, 2006).

## 6. References


Bravo, E. (1979). *Hacia una nueva forma de comunicación con el sordo.* San José: MEP.

Colombia. Ministerio de Educación Nacional. Instituto Nacional para Sordos (INSOR) / Instituto Caro y Cuervo (ICyC) (2006). *Diccionario Básico de la Lengua de Señas Colombiana*. Bogotá, D.C.: Ministerio de Educación Nacional, Instituto Nacional para Sordos, INSOR.

Costa Rica. (2004). *La discapacidad en Costa Rica: situación actual y perspectivas.* San José: Organización Panamericana de Salud y Ministerio de Salud.

Costa Rica. Instituto Nacional de Estadística y Censos (2011). *Resultados Generales Censo 2011*. San José: INEC.

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. In *International Journal of Corpus Linguistics*, 15(1), pp. 106-131.

López Gracioso, D.M. (1992). *Comuniquémonos mejor. Diccionario Ilustrado de Lengua de Señas Costarricense*. San José: Deisa Internacional.

Oviedo, A. (2004). *A Study on Classifiers in Venezuelan Sign Language*. Seedorf: Signum, 2004.

Oviedo, A. (2001). *Apuntes para una gramática de la Lengua de Señas Colombiana*. Cali: Universidad del Valle/INSOR.

Woodward, J. (1991). Sign language varieties in Costa Rica. In *Sign Language Studies* (20), pp. 329-334.


## 7. Language Resource References


Costa Rica. CENAREC (2013a). Diccionario Básico de la LESCO. Accessed at: http://cenarec-lesco.org/DiccionarioLESCO.php [21/02/2018]

Costa Rica. CENAREC (2013b). Gramática Básica de la LESCO. Accessed at: http://www.cenarec-lesco.org/index.php/ [21/02/2018]