Choguita Rarámuri (Tarahumara) language description and documentation: a guide to the deposited collection and associated materials

Gabriela Caballero
University of California, San Diego

Choguita Rarámuri (Tarahumara) is a Uto-Aztecan language spoken in Northern Mexico of great typological, theoretical, and historical significance. This paper presents an overview and background of the Choguita Rarámuri language description and documentation project and provides a guide to the documentary collection emerging from this project. This collection, deposited in the Endangered Languages Archive, is the result of collaboration with community members with the long-term goals of aiding in language preservation efforts and the development of a reference grammar of the language. While the production of linguistic analysis in the form of the reference grammar and other publications motivates a significant amount of the documentary corpus, the collection was also theorized from the perspective of a variety of audiences and provides an example of community-based design of documentary materials. This paper provides details on the development of the project, which allows readers to contextualize the scope and nature of the resulting corpus. This paper also discusses current restrictions on access to the collection, as well as an overview of existing associated materials and work underway that seeks to provide direct links between the deposited collection and products of linguistic analysis.

1. Introduction

With the advent of language documentation as a discipline and growing awareness of language endangerment across the globe, linguistic description and analysis of lesser-studied languages has brought about increased development of language documentation corpora and grammatical descriptions based on documentation-based language data gathering. However, as Mosel (2014) points out, corpus-based linguistic analyses of previously undescribed languages for the most

---

1 I owe my deepest gratitude to the people of Choguita, and in particular to Luz Elena León Ramírez, Bertha Fuentes Loya, Sebastián Fuentes Holguín, Guillermina Fuentes Moreno, Giltro Fuentes Palma, María Dolores Holguín, Morales Fuentes Hernández, Jesusita Loya Guerra, María Guadalupe Díaz and Crispín Bustillos, for their insights, patience, and generosity when teaching me their language and their hospitality in their homes and their community. Wàrú matétala ba. I am also indebted to Lucien Carroll for his comments on this manuscript and his continued support to the CR documentation and description project, as well as Gary Holton and an anonymous reviewer for helpful feedback. All errors and omissions are, however, my sole responsibility. The work reported here was supported by the Endangered Language Documentation Programme and by the Documenting Endangered Languages program of the National Science Foundation (grant #1160672). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Licensed under Creative Commons Attribution-NonCommercial 4.0 International
part do not provide many details of the content and structure of the corpus on which they are based. Furthermore, language documentation corpora on their own constitute a valuable resource for multiple audiences and, as such, must be accessible to all potential users (Woodbury 2014). Thus, there is a need for greater contextualization of the contents and design of particular documentary corpora (as exemplified in Salffner 2015) and for the development of a clearly articulated relationship between these corpora and associated linguistic analyses.

In this article, I provide an overview of the contents of the language documentation collection of the Choguita Rarámuri (Tarahumara) language, and background to the ongoing research project that gave rise to the collection. The materials, currently housed at the Endangered Languages Archive at SOAS, University of London (http://elar.soas.ac.uk/deposit/0056), represent the output of language description and documentation carried out together with community members since 2003. This documentary corpus includes approximately 130 hours of digital audio recordings, 10 hours of video recordings, and a substantial amount of digital transcription and annotation of a wide range of genres of speech, including different kinds of elicitation, conversations, narratives, interviews with elders by native speakers, language teaching sessions, and ritual/ceremonial speech. In addition to this collection, the project has produced linguistic publications focusing on the analysis of different aspects of the language. Current work seeks to integrate the contents of this collection (and materials subsequently obtained) with a comprehensive grammatical description of the language. This paper seeks to situate the documentation collection and associated materials into the larger context of its design, its evolution, and the work yet to be completed to make the materials more accessible to a variety of audiences.

The structure of the paper is as follows: in §2, I provide background to the language and its speakers, existing documentation of Rarámuri varieties, the research project that generated the documentary collection, and the collection conventions. In §3, I describe the content of the collection, including materials that have an exclusively linguistic focus (§3.1), materials created for both linguistic and community-oriented purposes (§3.2), and materials designed and created by community members of rituals and other community events (§3.3). In §4, I discuss the current restrictions on access of the collection materials. In §5, I present an overview of work in progress and future developments. I conclude in §6.

2. Background

2.1 The language and its speakers Rarámuri (Tarahumara) is a Uto-Aztecian language belonging to the Taracahitan subgroup of Southern Uto-Aztecan (Campbell 1997; Mithun 1999; cf. Hill 2011).2 Rarámuri, together with Guarijío, belongs to a Tara-Guarijío subgroup of Taracahitan. The place of Rarámuri within Uto-Aztecan is

2The term “Rarámuri” (or “Ralámuli”) is the one Rarámuri speakers will use to refer to themselves, while “Tarahumara” is the term used by the Spanish speaking population, as well as references such as the ethnologue and the Mexican government census.
shown in Figure 1. The geographic location of the different branches of Uto-Aztecan is shown in Figure 2.

There is currently no consensus as to the genetic status of Southern Uto-Aztecan and Taracahitan (for recent discussion see Hill 2011 and Merrill 2013), and this is indicated with parentheses in these branches.
Rarámuri is spoken by an estimated 85,000 to 100,000 people (INEGI 2010; Embriz Osorio & Zamora Alarcón 2012) in the Sierra Tarahumara, a mountainous range in the northern Mexican state of Chihuahua belonging to the Sierra Madre Occidental. The language is typically classified into five dialect areas, Central, Western, Northern, Southeastern, and Southwestern (Simons & Fennig 2017), though there is no consensus between different dialect surveys about the precise boundaries of these areas (INALI 2010). Choguita Rarámuri (henceforth CR) is part of a dialect continuum within a “Central” dialect area ([tar]; INALI 2008). The location of CR with respect to neighboring Uto-Aztecan varieties is shown in Figure 3.

Figure 3. Location of Choguita Rarámuri and neighboring Uto-Aztecan varieties

The sociolinguistic situation in Rarámuri communities is complex, and the vitality of each Rarámuri variety depends on several factors, including the relative isolation of each community from the Spanish-speaking mestizo (non-indigenous Mexican) population. Some communities display interrupted intergenerational transmission of the language, while others remain completely monolingual. In the mid-1990s, Paciotto (1996) reported most communities had varying levels of bilingualism, though the sociolinguistic situation has been changing drastically in the past decades due to increased political instability in the area.
Choguita is the head of the ejido of Choguita, a federally recognized territory containing Choguita and other small villages or rancherías. Choguita is the largest community of the ejido and can be easily accessed, with roads and public transportation connecting it to other Rarámuri villages and the larger mestizo towns in the sierras. Children in Choguita are for the most part monolingual until they attend elementary school, where they learn Spanish. There are still older monolingual adults, but most community members are proficient in Spanish and are able to interact with government officials and other Spanish-speaking outsiders. While several local government and religious activities are primarily conducted in Rarámuri, community members often switch to Spanish in these contexts. Due to increasingly difficult political and economic circumstances affecting the area, many people have relocated to larger towns in the Sierras or to Chihuahua, the state capital. Children who have grown up in diaspora settlements tend not to be proficient in Rarámuri, communicating mainly in Spanish when visiting the community with family members. Thus, while Rarámuri has a relatively large number of speakers for the area, the domains of usage of the language are rapidly contracting, severely threatening intergenerational transmission of the language.

2.2 Existing documentation and description of Rarámuri

Aside from brief early grammatical descriptions (Guadalajara 1683), most existing documentation of Rarámuri has been produced during the twentieth century. This includes several grammatical descriptions, dictionaries, vocabularies, and texts for several varieties. The most comprehensive grammar of Rarámuri to date is David Brambila’s 1953 description of the Norogachi variety (a Northern-Central dialect), a description structured like grammars produced during colonial times, which also contains many examples from texts.

More recent linguistic work includes a PhD thesis on Urique Tarahumara syntax (Valdez Jara 2013) and a master’s thesis analyzing basic clause structure and other syntactic aspects of a Central variety, Rochéachi Rarámuri (Morales Moreno 2016). The work of Morales Moreno, a native speaker of Rochéachi Rarámuri, is notable in that the analysis provided is based on a carefully annotated corpus of 11 texts (amounting to approximately ten hours of audio recordings) recorded with several native speakers of this variety.

In addition to the publications that have a linguistic focus, much work has been carried out in the state of Chihuahua documenting traditional narratives, poetry, and other forms of verbal art of the Rarámuri nation by Dolores Batista† and Martín Makawi, Rarámuri poets and language activists, and by Enrique Servín Herrera, linguist and poet in charge of a Program for Indigenous and Minority Languages hosted by the state government’s Cultural Development Office. Their efforts have produced

⁴The ejido system is a Mexican land usage system in place since the early 20th century and administered by the federal government, where rural land plots are devoted for collective use by community members (ejidatarios).

⁵A list of published references on the Rarámuri language is available in Caballero 2008 (appendix 1).

⁶Programa Institucional de Atención a las Lenguas Indígenas y Minoritarias.
several publications in different Rarámuri varieties and both bilingual and monolingual books and resources for Rarámuri speakers and second language learners (Batista 1994; Servín Herrera 2002; Makawi 2012; inter alia).

Language documentation and description of Choguita Rarámuri began in 2003, and has been supported by two Endangered Language Documentation Programme research grants (IG50042, IPF0138) and a Documenting Endangered Languages grant from the National Science Foundation (1160672), currently underway. Details of the motivation and impact of the project are provided next.

2.3 The CR research project  The long-term goal of the CR description and documentation research project is two-fold: (i) to carry out deep, detailed language documentation on an endangered language where a full range of speech genres, lexical, and grammatical knowledge are still appreciated in older speakers; and (ii) to produce linguistic analysis of a highly theoretically, typologically, and comparatively/historically relevant language. Below, I provide details of these two aspects of the project.

2.3.1 Language documentation and community involvement  The research project began with an agenda purely defined by my interest in writing a dissertation on the language and producing language documentation, but the description and documentation of CR has had an impact on a growing community initiative to reverse the contraction of Rarámuri. Community members have expressed interest in creating a record of the speech and linguistic practices of elders, as well as the community’s historical past and receding ritual, cultural, and artistic practices. Language documentation thus becomes part of a larger project to document and preserve the history and culture of a community that has undergone drastic changes in the past decades.

To date, thirty-four community members have participated in the project, mainly as language teachers and consultants, but also as contributors of annotations, analysis, and video documentation. The main members of the documentation team and their roles in the project are listed in (1) (in alphabetical order).

(1)  CR language description and documentation project members

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>Gabriela Caballero&lt;br&gt;linguist and principal investigator</td>
</tr>
<tr>
<td>b.</td>
<td>Lucien Carroll&lt;br&gt;linguist and corpus developer</td>
</tr>
<tr>
<td>c.</td>
<td>Rosa Isela Chaparro Gardea&lt;br&gt;language consultant, annotation</td>
</tr>
<tr>
<td>d.</td>
<td>Sebastián Fuentes Holguín&lt;br&gt;language consultant, annotation, video documentation</td>
</tr>
<tr>
<td>e.</td>
<td>Bertha Fuentes Loya&lt;br&gt;language consultant, annotation</td>
</tr>
<tr>
<td>f.</td>
<td>Guillermina Fuentes Moreno&lt;br&gt;language consultant, annotation</td>
</tr>
<tr>
<td>g.</td>
<td>Giltro Fuentes Palma&lt;br&gt;language consultant, annotation, video documentation, local school curriculum</td>
</tr>
<tr>
<td>h.</td>
<td>Luz Elena León Ramírez&lt;br&gt;language consultant, annotation</td>
</tr>
</tbody>
</table>

This research has also been supported by fellowships from UCMEXUS and the Hellman Foundation.
This level of involvement has allowed for documenting aspects of linguistic variation in the community, through the creation of a comprehensive and heterogeneous record of the language in terms of genres of speech and age, gender, social role, and level of literacy and bilingualism of speakers.

Interest from community members on documentation arose gradually, after learning about the project through information I provided in local meetings presided over by local authorities, where I would also request community-wide permission to conduct my research. In every visit, I addressed community members to inform them of the activities carried out with individual consultants for the project and confirm their support. After a few visits, a group of interested community members requested to be trained on how to carry out their own video documentation in order to create a record of the history of Choguita, as well as verbal art and traditional knowledge that the younger generations no longer have command of. The community-developed repository of materials would then be mobilized in the community with the explicit goal of having younger generations become acquainted with an important aspect of their community.

In response to this request, several video documentation workshops with adults and high school students were held with the help of a videographer (Jorge Esteban Moreno Romero) and through funding from ELDP. In these workshops, participants designed and carried out their own video documentation projects, which included interviews with elders, recording of pedagogical materials for children, and recording of different community events (more details about these materials is provided in §3 below). Figures 4 and 5 show still pictures of project members carrying out video documentation.

Materials produced were digitized and copies given to individual creators of materials, as well as local authorities (siríame or governors). Copies of these video recordings, along with recording equipment and a projector, were deposited at the local school, where Mr. Giltro Fuentes Palma (Figure 6) would be in charge of using these materials in the school curricula. Mr. Fuentes Palma received video documentation training and was appointed by local authorities to have custody of these materials and lead the continuation of the video documentation project and mobilization of documentation outcomes in the community.

Some of the participants who received the initial training were no longer able to continue carrying out documentation, as some left the community and others acquired time-consuming obligations, including Mr. Fuentes Palma. A standing challenge of this project has thus been to enable a sustainable infrastructure within Choguita for continued community-based language and culture documentation and safekeeping and use of materials by community members.

2.3.2 Typological and theoretical relevance of CR With its highly agglutinating morphology and complex morphophonological processes and prosodic system, CR provides a unique opportunity to explore critical questions about the nature of the phonology-morphology interface, the typology of word prosodic systems, as well as the interplay between processing and distributional factors in agglutinating morpho-
Figure 4. Sebastián Fuentes Holguín making a video recording of a traditional ceremony (photograph: Gabriela Caballero)

Figure 5. Sebastián Fuentes Holguín and Francisco Fuentes Moreno record ritual pinto dancers during the Easter celebration (documentation4.jpg) (photograph: Jorge Esteban Moreno)
logical systems. CR is also of great comparative/historical importance: while several analytical works of Uto-Aztecan languages of Northern Mexico have been produced in the last years (Guerrero Valenzuela 2006; García Salido 2014; Reyes Taboada 2016; Morales Moreno 2016; inter alia), many varieties still lack comprehensive linguistic description and documentation.

The linguistic products of the project to date include a PhD dissertation completed by the author in 2008 (Choguita Rarámuri (Tarahumara) phonology and morphology) and several publications focusing mainly on the phonology and morphology of the language, given the typological and theoretical relevance of this language in these domains. Phenomena and patterns that have been addressed include complex prosodic patterns (both at the word and phrase level), morphological conditioning on stress and other phonological phenomena, variable affix order patterns, and multiple (extended) exponence.

Description of particular word prosodic systems has traditionally received a great deal of attention in Uto-Aztecan studies, given their complexity across the language family. CR features a highly complex stress system with elaborate lexical and morphological conditioning governing its distribution and an initial three-syllable stress window, a pattern that has been documented in only four other languages of the world outside this language family (Caballero 2011a).

In addition to stress, CR possesses a tone system that contrasts three tones (high, low and falling) in stressed syllables. While the development of tonal contrasts has
been documented for a number of Uto-Aztecan languages, no variety of Rarámuri had previously been described as featuring lexical tone. Languages with independent stress and tone contrasts in their word-prosodic systems (“hybrid”) are still underdocumented cross-linguistically, though they have developed in other Uto-Aztecan languages (e.g., Balsas Nahuatl (Aztecan)) (Guion et al. 2010). Research on these hybrid systems has emphasized the need to investigate the acoustic correlates of stress and tone in addition to their phonological properties in order to confirm they are in fact independent systems (Everett 1998; Remijsen 2002; Remijsen & van Heuven 2005). As part of an ongoing project that investigates the structural and phonetic properties of prosodic phenomena in this language, I have examined the phonological distribution and morphological role of CR tone, as well as its interaction with the stress system of the language. This research has so far shown that stress and tone are not only phonologically independent systems in CR, but they are also acoustically distinct: duration and intensity are correlates of stress, whereas fundamental frequency (Fo) is the primary correlate of tone (Caballero & Carroll 2015).

Finally, in addition to featuring a tonal contrast, CR also exploits Fo intonationally. Preliminary examination of acoustic data reveals that lexical and morphological tones take precedence over phrase-level tones if these tones conflict (Caballero et al. 2014; Aguilar et al. 2015; Garellek et al. 2015). There are no in-depth studies of intonational properties of the languages of the area, so this project seeks to fill this gap.

In terms of its morphology, CR exhibits a case of variable affix order where alternative orders are determined by semantic, phonological and purely morphological factors (Caballero 2010). CR also exhibits complex patterns of multiple (extended) exponence (ME), a one-to-many mapping between morphological categories and their formal expression. The CR case offers an important opportunity to explore the properties of this morphological phenomenon, given that ME in CR: i) involves derivational information (contra suggestions that ME is exclusively displayed by markers of inflectional categories (Matthews 1972; Stump 2001)); and ii) that it involves categories in specific areas of the layered structure of the verb with characteristic morpho-prosodic properties which make them difficult to parse and prone to be reanalyzed as part of the stem (Caballero 2008; 2011). Finally, CR provides a relevant testing ground for investigating the potential functional role that complex morphophonological patterns in morphologically complex languages may have. I have investigated the perceptual functionality of ME in CR through a perception experiment with CR speakers (Caballero & Kapatsinski 2015). Our results show a significant effect of adding a redundant marker: the redundant (ME) pattern helped with recognition of the cued meaning when this meaning is unexpected from context, but it was detrimental when the meaning was expected from context. We interpreted these results as evidence of a mechanism of pragmatic inference at play in morphological processing, whereby listeners expect the speaker to produce as little as possible while successfully transmitting the intended information.
A summary of these and other linguistic publications focusing on the phonology and morphology of CR are listed in (2).

(2) Publications and manuscripts on the phonology and morphology of CR

b. Word prosody: distributional properties and acoustic correlates of stress and tone (Caballero & Carroll 2015); morphological conditions on stress assignment (Caballero 2011a); tonal and non-tonal correlates of lexical tone (Caballero et al. 2014), prosodic loanword adaptation (Caballero & Carroll 2014), and morphological role of tone (Caballero in prep. b)
c. Intonation: interaction between (morpho-)lexical and post-lexical tone (Garellek et al. 2015) and phonetic implementation of lexical tones in different intonation contexts (Aguilar et al. 2015)
d. Multiple exponence: description and analysis (Caballero 2011b), and functionality in morphological processing (Caballero & Kapatsinski 2015)
e. Variable affix order and its conditioning (Caballero 2010)
f. Voicing alternations, lenition, and prosodic conditioning (Caballero in prep. a)

In addition to these phenomena, CR is also typologically relevant given a rich set of argument structure changing morphology, and a complex system of case marking and lexical distinctions to refer to spatial and topographic properties of the landscape (Caballero & Pintado 2012), also documented for closely-related Guarijío (Miller 1996). These and other aspects of the structure of CR will be described in a reference grammar currently underway, described in more detail in §5.

2.4 The collection conventions

This section provides an overview of the conventions adopted in the recording of the data, gathering of metadata, and annotations of audio and video recordings.

2.4.1 Data recording and recording formats

Audio recordings of elicitation sessions, text collection and other speech events were digitally recorded on a solid-state recorder (Marantz PDM 660) at a sampling rate of 44.1 Khz. Individual speakers were recorded with a lavalier condenser cardioid microphone (Audio Technica AT8531b or AT898). Groups of speakers were recorded with a condenser omnidirectional shotgun microphone (Audio Technica AT897). Video recordings were made on mini-DV tapes using a Canon Vixia HV30 camera and the shotgun microphone using a mini-XLR adapter and a 16-bit sampling rate. Both audio and video recordings were transferred into archival CD format to distribute copies with collaborators,
traditional authorities and for deposit within the community. Video recordings were
digitized onto AVI and archival versions onto MPEG-2 format for deposit to ELAR.

A subset of elicitation notes exists in digital format, in PDF format, as well as in
time-aligned, XML-encoded annotations in transcriber (TRS) and ELAN (EAF) files
(Barras et al. 2001; Sloetjes & Wittenburg 2008). Handwritten notes (originally in
acid-free, archiving quality paper) have been converted to digital format (scanned
copies in PDF format), but have yet to be deposited with the rest of the collection.

2.4.2 Metadata Item-level metadata of the collection is currently in a Microsoft Excel XLS spreadsheet. The file name of each document is a unique identifier with con-
tinuous numbering and an abbreviation encoding the type of document involved, e.g.
el45, where “el” = elicitation, “tx” = text, “co” = conversation, “in” = interview, “tr” =
free Spanish translation, “en” = elicitation notes, “te” = language teaching, and “mu” =
music.⁸ This metadata describes the content, participants involved, and recording
circumstances of each deposited item. In addition, the following information is given:

(3) Item-level metadata information for CR documentation collection

a. File format (wav, mpeg, pdf, eaf, etc.)
b. Date recorded
c. Duration (for media files)
d. Languages featured in the document (CR and/or Spanish)
e. Initials of contributor(s) (native speaker participants recorded)
f. Initials of creator of document
g. Title
h. Document type (elicitation, procedural text, historical description,
etc.)
i. Brief description of the content of the document
j. Linguistic topic for elicitation files
k. Description of recording circumstances
l. Access rights
m. Date of last update

In addition to item-level metadata, the deposit includes metadata on cross-referen-
ces between related documents. Figure 7 shows how associations between files are
encoded in the deposited metadata page. As shown in this screenshot, this component
of the metadata identifies different kinds of files by their unique identifier plus a brief
description of the nature of the association.

Relationships between deposited documents are also reflected in the document
bundles available in ELAR, where each bundle corresponds to resources that form a
coherent set, e.g., one or several media files (audio and/or video) that are thematically
related and corresponding transcription, annotation, elicitation notes, etc. Examples

⁸There are a few exceptions to this file naming system, mostly involving video recordings and still pho-
tographs.
Figure 7. Screenshot of cross-reference tab in deposited metadata file, where relationships between documents are explicitly described.

of document bundles are provided below in the description of the contents of the collection.

Finally, the metadata file also contains brief biographical information about each project participant, including role(s) taken (consultant, participant, creator, or analysis), gender, approximate birthdate and birthplace, as well as parents’ birthplace. When available, additional information about age of acquisition of Spanish (if bilingual), literacy, and schooling is provided.

2.4.3 Annotations  In addition to audio and video recordings, the collection includes written annotations carried out by the documentation project members listed in (1) above, containing minimally a broad phonetic transcription and a Spanish free translation. In addition to this, some annotations additionally contain English free translations, morphological glosses, and fine phonetic transcriptions. Most annotations include a commentary specifying recording circumstances, issues in transcription and translation of particular segments, comments by speakers, and relevant grammatical or cultural information associated with any piece of data. All annotations also provide relevant metadata describing the content, associated media file(s), and recording circumstances of the document annotated.

Most transcriptions use the Americanist convention of transcription, though some files contain a transcription using certain symbols from a Spanish-based orthographic convention (e.g., <ch> for the alveopalatal affricate and <rr> for the alveolar trill). Transcriptions also encode pauses as “…” and grammatical or general comments in angled brackets. Stress is marked with an acute accent diacritic. Tones are not marked. This omission is due to the complex interaction between lexical tone, intonation, and morphological structure. Work in progress seeks to elucidate these interactions (Ca-
Annotations also include databases of lexical items and prosodic and morphological properties encoded in XLS files. These databases have now been migrated to a database in SIL’s FieldWorks Language Explorer (FLEx).³

3. The content of the collection  The deposited collection comprises the materials listed in (4).

(4) Choguita Rarámuri documentary collection contents
   a. Over 750 audio and video recordings
   b. ELAN and Transcriber transcription files
   c. Word files with transcriptions, translations, analyses, and annotations
   d. Excel databases of lexical items and prosodic and morphological properties
   e. Still pictures
   f. Metadata including descriptions of documents, recording circumstances, and cross-references between related documents
   g. A doctoral dissertation on the phonology and morphology of the language¹⁰

Recordings of the collection amount to approximately 130 hours of digital audio and 10 hours of video recordings, plus digital transcriptions and annotations of about forty documents, including both texts and elicitation sessions. These materials are organized in 764 document bundles. With participation of over twenty community members at the time of the last deposit made, the documentation work with Rarámuri speakers includes staged communicative events, different kinds of elicitation, and participant observation. Speech genres documented include myth narratives, oratory, historical narratives, interviews, conversations, procedural texts, ritual song and prayer, and language teaching, among others. Video recordings include elicitation sessions on the language of space and landscape terms, recordings of agricultural practices, and descriptions in situ of culturally relevant events, such as races, traditional games, ritual music performances, and community meetings and celebrations. Sets of related documents are organized in the ELAR collection as bundles. There are currently over one hundred document bundles available in the deposited collection.

Documentation of CR for this project was carried out following the approach laid out in §2.3.1 above, but the sections below provide the details of the contents of

³FLEx is available at http://software.sil.org/fieldworks/download/.
¹⁰More recent publications have not yet been deposited, but they will be included in the next update to the collection.
the collection in terms of how the materials were theorized at the moment of recording and whether their creation was motivated by a linguistically-defined agenda, a community-defined agenda, or both. Materials with both a linguistic focus and a community focus include documents that provide a glimpse of linguistic variation within the community through the recording of historically and culturally relevant events carried out by multiple participants. Details and relevant examples of each set of documents are provided below.

3.1 Materials with an exclusively linguistic focus

3.1.1 Elicitation sessions A large portion of the elicited materials contained in the collection is made up of sessions focusing on the verbal morphology, phonology, and phonology-morphology interactions. In addition to eliciting morphological paradigms, these materials include elicitation of valence-changing morphology (causatives, applicatives, passives), general affix order patterns, phonological conditions on multiple exponent or allomorph selection, as well as tonal and other prosodic phenomena in morphologically complex words. As the linguistic analysis goal of the project broadened to include more general aspects of grammatical description, a wider range of topics were increasingly covered in elicitation sessions.

The most common elicitation methodologies undertaken were translation elicitation with Spanish-Rarámuri bilinguals, as well as contextualized and text-based elicitation. In contextualized and text-based elicitation, particular constructions encountered during participant observation or in the process of annotating individual texts would serve as starting points for exploring grammatical or lexical aspects of the constructions in question. Text-based elicitation is currently listed as a genre in the deposited collection. An example of this type of elicitation is found in the bundle “Romára elicitation,” which was recorded after a video documentation session of the traditional game *romara* (called *quince* in Spanish). In this session, elicitation focused on vocabulary and grammatical constructions that arose during transcription and annotation of the original video recording.

Other kinds of contextualized elicitation involved sessions where speakers undertook the role of language teachers, with full control of the content of the material covered, which allowed for greater contextualization of the data elicited (these sessions are described in more detail in §3.1.3 below). Finally, elicitation sessions also involved recording of metalinguistic judgments and the use of culturally specific visual props, such as pictures of rituals or agricultural practices in other Rarámuri communities. An example of this kind of elicitation is found in the bundle “lexical verbs 9,” in which the session was carried out using pictures of several landscapes of the Rarámuri area.

Elicitation sessions were mostly recorded using only audio, but some sessions were video recorded. These video elicitation sessions focused on in-situ descriptions of agricultural terms and the language of space, including deictic terms, topographic terms, as well as landscape-based standardized place names. An example is *deictics1* and *deictics2*, two sessions recorded with project member Bertha Fuentes Loya, where
she describes Choguita place names from different viewpoints within the community. Figure 8 provides a screenshot of one of these documents.

![Figure 8](image)

**Figure 8.** Bertha Fuentes Loya describes deictic terms and place names (“deictics2”)

As mentioned in §2.3.3 above, all recordings have corresponding hand-written notes. Some deposited elicitation notes are in digital format and include minimally Spanish translations and grammatical notes, with some documents also including morpheme glosses.

3.1.2 Conversations  Approximately fifty documents in the collection involve audio- and/or video-recorded conversations. In these sessions, speaker participants engaged in unstructured or semi-structured conversation, either with other native speakers or with me (as a second language user of Rarámuri). Unstructured conversations focused on linguistic constructions, lexical items (such as kinship terms), sociolinguistic aspects of the community, biographical information, or culturally relevant aspects of linguistic constructions that had emerged during elicitation sessions. One such example is file *co131*, which was recorded after a session of text annotation carried out by Mrs. Luz Elena León Ramírez and myself. In this recording, Mrs. León Ramírez explains to me in both Rarámuri and Spanish the metaphorical use of a lexical term to explain a particular human behavior. A different kind of session is exemplified in *co832*, where Mrs. Guillermina Fuentes Moreno and myself have a conversation about the kind of work women carry out in the community in both farming and in the household and in comparison to men’s work. In both of these examples, the recordings feature both Rarámuri and Spanish.
In some cases, the recorded conversations were semi-structured. These sessions were designed to elicit particular lexical items or morphological constructions with monolingual speakers. This is the case with files co753-7, a series of audio recordings made with Mr. Morales Fuentes Hernández, elder and ritual singer from the community. These recordings focused on target tonal minimal pairs in different morphological constructions from a list that was collected with bilingual speakers using different elicitation methodologies.

Finally, in a third set of conversation documents, the exchanges were carried out between several native speaker participants talking mainly in Rarámuri. This is the case of co290 and co291, two recordings of Mrs. María Dolores Holguín, Mr. Morales Fuentes Hernández, and Sebastián Fuentes Holguín discussing the meaning of lexical items and constructions elicited using culturally relevant still pictures. Transcriptions of these recordings are available in transcriber files (TRS), but the majority of the recorded conversations only have corresponding hand-written annotations.

Since all these interactions took place in the context of elicitation sessions or typically involved topics and constructions that were recorded for the purpose of informing linguistic analysis, they are part of the set of materials from the deposited collection that have an exclusively linguistic focus.

3.1.3 Language teaching sessions Another set of documents that served a linguistic agenda were recordings of language teaching sessions, where native speaker participants explained lexical items, constructions and expressions, and their cultural contextualization to me. In some cases, I prompted the topics of discussion, which included overheard expressions and terms recorded during elicitation sessions. In most of these sessions, however, speakers were just asked to teach me anything they decided would be appropriate for me to learn as a second language learner. In an example of this latter type of session (te464), Sebastián Fuentes Holguín corrects an erroneous use I had made of a Rarámuri expression and clarifies its use with examples of appropriate contexts for its utterance. Other topics covered in these sessions include politeness formulas, vocabulary associated with rituals and traditional agricultural practices, culturally relevant contextualization for the use of some terms and colloquial expressions, and discussions about dialect differences between CR and neighboring varieties.

3.2 Materials with both community and linguistic-based focus

3.2.1 Texts For the purpose of the CR collection, the term “texts” encompasses a series of documents that mainly involve monologic speech, of a kind closer to the planned speech event end of a spontaneity continuum. These texts include genres such as general descriptions of events, historically contextualized descriptions, myth narratives, procedural texts, and narratives from visual prompts, recorded (in both audio and video) with single speakers.

Audio recordings of texts were typically made in indoor, private spaces, while video-recorded texts generally involve procedural texts and historical descriptions.
made in situ. In some cases, speakers had their families as an audience when recording their text, when this was contextually appropriate. This is the case in tx12 (bundle escuela), where Sebastián Fuentes Holguín recorded the history of the local school and gave advice to his children, who were present during the recording, about the importance of formal education.

A series of video recorded texts focus on descriptions in situ of agricultural practices and food preparation, made with both elders and younger speakers. Some representative examples include: (i) a recording with Mr. José María Fuentes describing the traditional structure used in the community to store corn (rikoa) (Figure 9); (ii) a recording with Bertha Fuentes Loya describing the agricultural cycle (Figure 10); and (iii) a recording with Mrs. María Dolores Holguín describing how to prepare pozole (oribisi), a traditional stew prepared with hominy.

Figure 9. Mr. José María Fuentes describes how corn is stored after the harvest (bundle “troja”)

Recorded texts are relatively homogenous in terms of the amount of planning speakers had when preparing their texts. In most cases, they would be asked to recount a particular topic or they would volunteer themselves a topic for a text, and we would proceed to record the monologue. In some instances, speakers also recorded a free Spanish translation of their texts, but for the most part, free translations, annotation, and analysis was produced in separate sessions with either the authors themselves or a native speaker project member (no single text transcription and annotation of this collection has been carried out without a native speaker project member).

\[11\] In some cases, however, speakers sought to have more planning involved. This was the case of a project participant who, at the beginning of our collaboration, would first write her text in Spanish and then record her translation of the text into Rarámuri.
These texts are crucial to linguistic analysis (including the reference grammar underway that includes examples from the assembled corpus) and are part of the developing record of the history, culture and language of the community as envisaged by community members. As such, they constitute a crucial component of the deposited collection.

Figure 10. Bertha Fuentes Loya describes the agricultural cycle (bundle “maizal”)

3.2.2 Interviews Several community member project participants carried out interviews with elders on topics such as the history of the community, agricultural practices, kinship terms, and ritual practices, as well as personal biographical information. Younger speakers also participated in these recordings, discussing knowledge about agricultural information, changing ritual practices, and food preparation. The interviews include accounts by interviewees of their experience growing up in Choguita and the differences brought about by recent changes in the community, such as the introduction of the government-run local school or the introduction of electricity.

Interviews were video recorded with high quality audio by community members (sometimes with my assistance), with additional, separate audio recordings made at the time of the recording with the solid-state recorder. These interviews had a special focus on endangered domains of knowledge still possessed by elders, but in attrition or non-existent for younger speakers (e.g., the complex kinship term system that at the time of recording was not known with its full complexity by speakers in their thirties and younger). From the perspective of community members, the interviews served the purpose of creating a record of the history, culture, and language of the community. From my perspective as a linguist, these documentary materials were
Mrs. María Dolores Holguín describes how to make the oribisi stew (bundle “oribisi”) aimed at diversifying the genres of speech represented in the corpus and at documenting inter-speaker variation.

This type of document is exemplified in “Entrevista a Don Federico León por Sebastián Fuentes, parte 1” and “Entrevista a Don Federico León por Sebastián Fuentes, parte 2 sobre parentesco,” two bundles of documents of a planned interview about agricultural practices and a follow-up semi-spontaneous interview about kinship terminology recorded in 2006. As described in the titles, project participant Sebastián Fuentes Holguín carried out the interview in Rarámuri with Mr. Federico León Pacheco, community elder and healer. These document bundles include video and audio files, plus metadata information and transcriptions in Word, a corresponding PDF file, and a time-aligned ELAN transcription. These media files, which amount to approximately 40 minutes, were fully transcribed and annotated by Sebastián Fuentes Holguín, Bertha Fuentes Loya, and by me. Original annotations were made in word and PDF documents and include a broad phonetic transcription, an underlying phonological representation, morphological glosses, Spanish free translations, notes about recording circumstances, and notes on grammatical analysis. These annotations also include sporadic time-alignment notes. These annotations were then transferred onto time-aligned ELAN documents. A screenshot of one of the ELAN transcription files is shown in Figure 12.

These recordings have been shown to community members at different gatherings. The video recording “jueves_proyeccion”, made by Sebastián Fuentes Holguín, documents a video projection session recorded in April of 2009 at the local church during the Easter celebration, a time of year where members of neighboring commu-
nities join Choguita community members for the ritual festivities. A screenshot of this recording is shown in Figure 13.

Figure 13. Screenshot of “jueves_proyeccion,” video projection of interviews with elders

Figure 12. ELAN annotation of “in63,” interview with Mr. Federico León Pacheco
3.3 Materials with a community-based focus

3.3.1 Music  Ritual music is one of the domains of local culture that is mainly mastered by elders, with few young people learning this form of art. A series of audio recordings and video recordings document traditional violin music, performed by Mr. Morales Fuentes Hernández. The recordings were made indoors and in private. Figure 14 provides a screenshot of a video recording of one of these sessions.

The collection includes other examples of ritual music performed in context (at healing ceremonies and other rituals). The recordings with Mr. Fuentes Hernández are unique since they reflect his initiative to document his own personal style of performance of this traditional form of art. There is no musical transcription yet available of these recordings, though it is hoped that a future collaboration will yield this kind of annotation for these documents.

Figure 14. Mr. Morales Fuentes Hernández plays violin next to his wife, Ms. María Dolores Holguín (“morales_music1.mpeg,” bundle “Morales musico”)

3.3.2 Video recordings of community events  A series of video recordings in the collection document important events through the lens of community members, who designed the recording sessions and decided what aspects of the events to document. The recorded events include rituals and traditional activities, as well as events that are connected to the greater contact this community has had with mestizo practices and institutions (such as the government-run local school and the Catholic church) since the early twentieth century.
Ritual and traditional events documented include healing ceremonies, races, rain and harvest ceremonies, ritual appointment of local authorities, and Easter celebrations. This kind of document is exemplified in two recordings of women’s traditional races (rowea). A screenshot of one of the racers and her pacers is shown in Figure 15.

![Figure 15](rowea2.mpeg) Screenshot of “rowea2.mpeg,” a video recording of a women’s race

In addition to the actual running, these recordings also document the traditional betting process that takes place before and during the races, which involves a complex process led by appointed representatives of each of the competitors (called chokéami) that set up the race, arrange its details, and are in charge of pairing up the objects that supporters of each competitor bet. A representative screenshot of this process is shown in Figure 16.

Another example of documentation of ritual and traditional events is a series of recordings of different activities related to the Holy Week and Easter celebration, a large event involving the people from Choguita and neighboring towns within the ejido. In addition to the main celebrations that take place on Easter Sunday, these recordings include documentation of: (i) preparations for the festivities by ritual dancers (fariseos and pintos) and local authorities; (ii) preparations by host families (tenanches) who are in charge of feeding all participants; and (iii) appointment of new local authorities. These files are organized in the collection in several bundles labeled

---

The Rarámuri are famous endurance runners, and races are an important aspect of community life across the sierras. This is linked in the ethnographic literature to persistence hunting, the practice of hunting without projectile weapons and through driving mammals to exhaustion through running long distances, which the Rarámuri have been documented to have practiced in the past (Bennett & Zingg 1935; Pennington 1963).
Choguita Rarámuri (Tarahumara) language description and documentation

Figure 16. Screenshot of “rowea.mpeg” with race competitor’s representatives (chokéami)

Noríruachi and exemplified in Figures 17 and 18. In the recordings exemplified in these figures, made by project participant Sebastián Fuentes Holguín, documentation focuses on a vigil ritual conducted by ritual pinto and fariseo dancers.

Figure 17 exemplifies an aspect of the vigil that is highly ritualized (with praying and ritual music inside the local church), while Figure 18 shows an example of an informal gathering where ritual dancers take a break outside the church. Crucially, this latter recording exemplifies documentation design made by community members that is inaccessible to community outsiders in terms of intimate knowledge of what the celebration and rituals involve, including these kinds of informal gatherings and the preparations that take place in the days previous to the festivities.

Finally, two representative examples of community events linked to mestizo institutions and practices adopted by the community more recently are recordings of a Mother’s Day celebration hosted by the local school (bundle chichi omawara) and a quinceañera celebration (bundles quinceañera and quinceañeras), a coming of age party for fifteen-year-old girls popular across Mexico and the rest of Latin America. Figure 19 provides a screenshot of piñata breaking during the Mother’s Day celebration, and Figure 20 provides a screenshot of a quinceañera celebration.

While events such as the ones exemplified in these last two cases are relatively recent in the community (the quinceañera recorded in the file exemplified here was the first one of its kind in the community), they nonetheless involve more traditional practices and language, such as oratory by local authorities, traditional music, and other rituals that have been practiced by community before the advent of more recent
customs. As expressed by community members, all of these events are part of the life of the community and, as such, they are worthy of documentation.
4. Access restrictions of materials  

Each person recorded for this project consented to participation and specified their preferences in terms of possible uses the recorded
materials could have inside and outside their community. All possible uses are non-commercial and intended for educational, academic and community-specific purposes. While some participants were comfortable with having their contributions be open access, other participants placed restrictions on possible audiences (e.g., for some speakers, their contributions can be accessed by community outsiders but cannot be mobilized within Choguita).

In terms of community-wide sensitivities expressed either individually to me or in local meetings, there is widespread consensus that cultural, historic, and linguistic legacy materials should be available to community members and special sensitivities are placed on video recordings of community-wide celebrations, as outsiders in the past have recorded them without consulting local authorities or the community at large. Thus, in accordance with the sensitivities expressed by community members and individual contributors, most video files plus a few other documents in the deposited collection (41 resources out of 966) are currently restricted to be accessed by subscribers of the deposit. The rest of the contents of the collection (925 resources out of 966) are accessible to registered ELAR users. Restrictions on sensitive materials allow for informing potential users about possible uses of particular materials as decided by individual contributors on a case-by-case basis and in terms of community-wide sensitivities. As described below, the development of open-access materials that respect these sensitivities is underway.


Users should also acknowledge individual contributors if citing specific resources or document bundles, as well as the support of the Endangered Languages Documentation Programme in the creation of the collection.

5. Work in progress and future directions  In addition to continuing research on the particular typologically and theoretically relevant phenomena of the language, ongoing work seeks to produce a comprehensive grammatical description of CR. This grammatical description will be made available in the form of a book, the published version of which will provide a number of examples that will be linked to a web-based user interface version of the CR corpus. The goal is to allow readers to access the larger contexts which the grammatical description is based upon, which enables wider dissemination of the results, and allows interested academics and community members to carefully examine the analyses and description set forth in the grammatical description. A crucial component of this project is also to enable a version of this corpus in a format that is tailored to community members.

CR documentation materials that contain ELAN annotations are now available through a web-based corpus,13 currently accessed through an interface that links WAV files, ELAN annotations, and document metadata (with speaker and specific doc-

---

13This tool was developed by Russell Horton (UCSD, Linguistics) and further developed by project member Lucien Carroll. Interested readers can find the documentation of this tool in the following website: https://github.com/ucsd-field-lab/kwaras.
ument information, as well as time stamp of every recording fragment annotated) (Caballero et al. in prep.). The corpus provides translations, source metadata information, a citation system for referencing and finding individual utterances, lexical forms, and linked audio. When available, morpheme-by-morpheme glosses are also provided, as well as any other annotations (grammatical or other) available in the original ELAN annotations. The web-based corpus also contains a link to an associated lexical database generated through FLEx.

This web-based version of the corpus is in the process of being turned into a curated, open-access user interface (Caballero et al. 2017). This user interface will contain only the texts and other materials that individual contributors authorize as open access and that are consistent with community-wide sensitivities. These materials will have views designed for both community members and academic users, with links to audio of texts, linguistically annotated data, and transcriptions in a practical Rarámuri orthography, as well as Spanish and English translations. Texts and other materials will be listed by title or by contributing authors, as community members may be interested in accessing legacy materials of particular community members. Materials will be searchable through a function enabling global searches, as well as searches by field (Rarámuri orthographic representations, broad IPA transcriptions, and translations into Spanish and English). Each text will include its unique identifier that references the deposited collection in ELAR. This has the purpose of linking the most recent annotations to their corresponding archived versions that are maintained for long-term preservation.

These web-based corpora are currently used as a tool in grammar writing, since they allow finding examples of specific constructions from a variety of speech genres and from several speakers through the search function of the website. A curated, open-access version of these materials will enable greater transparency of the analyses proposed, as closer inspection of the data will be available for interested users. This will provide opportunities to modify or replace existing analyses or identify new linguistic phenomena in CR, and for community members to enrich or modify the existing annotations as they deem necessary for their own language planning purposes.

6. Conclusion The deposited collection and analytical products stemming from the linguistic research carried out over the past fourteen years has multiple stakeholders: community members interested in language and culture preservation and the history of their community, as well as linguists and other academics interested in the CR language and culture. With the rapidly changing political context of the Rarámuri speaking area, increased migration has brought about accelerated attrition of the language in the last decade. At the same time, access to new resources, such as smartphones and social media, allows CR speakers to communicate and use their language in new contexts. Thus, community members may now have changing needs for language resources and an increased ability to access the documented collection and any other associated materials. Making the language documentation collection accessible and usable to all interested users therefore involves standing challenges, which also
includes anticipating changing needs and possible future agendas, such as language reclamation efforts in diaspora communities.

This paper responds to calls for guides that allow accessing, using, and understanding the larger context of language documentation collections (Woodbury 2014), which in turn inform associated analytical materials, such as corpus-based reference grammars or other publications that make reference to materials produced through language documentation (Mosel 2014). This paper thus provides contextualization of the CR documentation collection and associated materials, ongoing developments, and expected results. The CR reference grammar and other publications are intended for language typologists and arealists, and provide a static picture of a complex linguistic system. The documentary corpus, on the other hand, is intended for a wider audience, and provides a hint to the true dimensions of complexity of the language, a dynamic system with a large amount of variation and change in progress. It is the hope that the current guide provides a first step into understanding and accessing this particular collection, as it continues to be developed for the evolving needs of community members and our academic fields.

References


Choguita Rarámuri (Tarahumara) language description and documentation


Pennington, Campbell W. 1963. The Tarahumara of Mexico: Their environment and material culture. Salt Lake City: University of Utah Press.


Gabriela Caballero
gcaballero@ucsd.edu