

# Brief Descriptions of Previous Research in Zapotec

Yukki Baldoria<sup>1,2</sup>, Seunghun J. Lee<sup>2,3</sup>, Julián Villegas<sup>4</sup>

<sup>1</sup>Yonsei University, <sup>2</sup>Internatinal Christian University, <sup>3</sup>University of Venda, <sup>4</sup>The University of Aizu

## 1 Introduction

This volume was compiled with data from eight sources on different varieties of Zapotec. These sources and their dialect variety are listed alphabetically according to source name as follows:

- a. Arellanes Arellanes (2010): San Pablo Güilá Zapotec
- b. Avelino Becerra (2004): Yalálag Zapotec
- c. Brazilai and Riestenberg (2021): San Pablo Macuiltianguis Zapotec
- d. Beam de Azcona (2004a): Coatlán-Loxicha Zapotec
- e. Bueno Holle (2016): Isthmus Zapotec
- f. Esposito (2003): Santa Ana del Valle Zapotec
- g. Sonnenschein (2004): San Bartolomé Zoogocho Zapotec
- h. Tejada (2012): Sierra Juárez-San Juan Atepec Zapotec

The volume has four chapters and a supplementary data file in Excel. The chapters sort the data in alphabetical order based on headwords in Zapotec (Chapter 2), Spanish (Chapter 3), or English (Chapter 4). The supplementary Excel file can be used for further data analysis.

## 1.2 Why Zapotec in the Context of the PSYPHON Project?

The project “Psychoacoustic features for phonation prediction” (PSYPHON) aims at creating psychoacoustic-based predictors of phonation by means of Machine Learning (ML) techniques. One of the main outcomes at the end of PSYPHON is a software tool to predict various types of phonation, (mainly modal, breathy, and creaky) based on psychoacoustic models (roughness, loudness, sharpness, etc.).

For machine learning, it is important to have samples of different types of phonation made by the same speaker. That way, the system can truly learn the differences between these types of phonation. For example, in a language in which creakiness is contrastive, but not so breathiness, an ML-system may learn the difference between creaky and non-creaky phonation, but not between modal and breathy phonation.

Few languages use modal, creaky, and breathy phonation for distinguishing the meaning between words. It has been reported that Zapotec is one of those languages. For that reason, we have decided to focus on this language to build a corpus that can be used for our machine learning purposes. The present volume summarizes the hitherto descriptions of different dialects of Zapotec, constituting a crucial document for planning further data collection.

## 2 Summary of Each Study

### 2.1 Arellanes Arellanes (2010): San Pablo Güilá Zapotec

The research in Arellanes Arellanes (2010) focuses on the phonological relevance of two ‘degrees’ of laryngealization in San Pablo Güilá Zapotec. Specifically, it refers to the contrast between modal vowels and two types of laryngialized vowels distinguished by their level of phonetical manifestation, with the research scope limited to the latter. In his research, the variation on the expression of non-modal vowels is attributed to two factors: (1) tonal context and (2) speaking style, and the paper focuses on the conditions in which the non-modal vowels are realized. In total, the research presents 84 individual tokens that were used in the present study for analysis.

## 2.2 Avelino Becerra (2004): Yalálag Zapotec

In Avelino Becerra (2004), the main research focus is on the phonetic structures of Yalálag Zapotec, divided into eight chapters across three sections. In the first section, he briefly introduced Yalálag Zapotec including the orthography, morphology, and sentence structure. In the second section, the phonetic structure, including vowels, consonants, and tone, is elaborated upon. Acoustic properties of different vowel phonations are also illustrated as well as an acoustic and palatographic study of coronals, denti-alveolars, post-alveolars, and alveo-palatals. In the third and final section, he delves into tone perception in Yalálag Zapotec. In total, 420 tokens were used for the present study.

## 2.3 Barzilai and Riestenberg (2021): San Pablo Macuiltianguis Zapotec

This paper focuses on phonetic enhancement of phonation contrast in the San Pablo Macuiltianguis Zapotec variety in relation to the context (Barzilai & Riestenberg, 2021). This research aims to fill the gap in Zapotec varieties by discussing the Sierra Juárez variety, of which San Pablo Macuiltianguis Zapotec belongs to. It is found that  $F_0$  cues the contrastive tone and that it also phonetically enhances the contrast. Three contrastive phonation types are presented in this variety: modal, checked, and rearticulated. In total, 143 tokens were taken from this study and analysed in the present study, and these tokens can be found throughout the original paper.

## 2.4 Beam de Azcona (2004a): Coatlán-Loxicha Zapotec

This dissertation touches on the Zapotec variety spoken in Southern Coatlans and Northern Loxichas. It describes the grammar of the language, making reference to the phonetics, phonology, morphology, and certain elements of the syntax. A total of 4,644 token entries were recorded in the Appendix section the dissertation, of which all were taken for analysis in the present study.

## 2.5 Bueno Holle (2016): Isthmus Zapotec

This dissertation uses data from the Isthmus variety of Zapotec, which belongs to the Central branch of the Zapotec language family. The main focus of the study is on the information structure—that is—the ways intonation, morphology, and syntax are organised by speakers and how they interact with each other to communicate information. A total of 222 individual tokens were taken from this dissertation and used in the present research.

## 2.6 Esposito (2003): Santa Ana del Valle Zapotec

This thesis discusses experimentally the phonation and tone of the Santa Ana del Valle variety of Zapotec. Phonation and tones are defined in the first section and a production experiment to measure the three way contrast—modal, breathy, and creaky vowels—was carried out. The second section consisted of  $F_0$  and position and their effects on phonation, and it was found that  $F_0$  influenced phonation. As for the data used in the present study, 44 monosyllabic words with [a], 20 disyllabic tokens, 20 words with [e], 20 words with [i], 20 words with [o], and 20 words with [u] were extracted and used. An additional 40 tokens were also taken from the second section of the study. In total, there were 183 individual tokens.

## 2.7 Sonnenschein (2004): San Bartolomé Zoogocho Zapotec

Sonnenschein (2004) presents a grammatical description of the San Bartolomé Zoogocho variety of Zapotec. The first section comprises of the ethnographic, sociolinguistic, sounds of the language, pronominal system, morphology, and syntax of the language, whereas the second of the lexical classes and word order of this variety of Zapotec. A total of 157 tokens were taken from this thesis and used in the current study.

## 2.8 Tejada (2012): Sierra Juárez-San Juan Atepec Zapotec

The aim of this dissertation is to examine floating tones and tone gesture deactivation in the Sierra Juárez variety of Zapotec. Most of the data collected in this study was elicited from a single consultant

who lived in Los Angeles, USA. In total, 71 tokens from this research were extrapolated and analysed in our present study.

## References

- Arellanes Arellanes, F. (2010). Dos ‘Grados’ de Laringización con Pertinencia Fonológica en el Zapoteco de San Pablo Güilá (“Two ‘Degrees’ of Laryngealization in the Phonology of San Pablo Güilá Zapotec”), in *Entre Cuerdas y Velo—Estudios Fonológicos de Lenguas Otomanguas* (“Between Strings and Veil—Phonological Studies of Otomanguan Languages”), 85-121.
- Avelino Becerra, H. A. (2004). *Topics in Yalálag Zapotec, with Particular Reference to its Phonetic Structures* (Doctoral dissertation, University of California, Los Angeles).
- Barzilai, M. L., & Riestenberg, K. J. (2021). Context-dependent phonetic enhancement of a phonation contrast in San Pablo Macuiltianguis Zapotec. *Glossa: a journal of general linguistics*, 6(1): 18. 1–36. DOI: <https://doi.org/10.5334/gjgl.959>
- Beam de Azcona, R. G. (2004a). *A Coatlán-Loxicha Zapotec Grammar* (Doctoral dissertation, University of California, Berkeley).
- Bueno Holle, J. J. (2016). *Information Structure in Isthmus Zapotec* (Doctoral dissertation, The University of Chicago).
- Esposito, C. M. (2003). *Santa Ana del Valle Zapotec phonation* (Doctoral dissertation, University of California, Los Angeles).
- Sonnenschein, A. H. (2004). *A descriptive grammar of San Bartolomé Zoogocho Zapotec*. (Doctoral dissertation, University of Southern California).
- Tejada, L. (2012). *Tone gestures and constraint interaction in Sierra Juárez Zapotec* (Doctoral dissertation, University of Southern California).