

Santa Clara University

Scholar Commons

Mutsun Text Collection: mutsun riicakma hummen

Archives and Special Collections

9-21-2022

Mutsun Text Collection - Introduction

Natasha Warner
nwarner@arizona.edu

Lynnika Butler

Heather van Volkinburg

Quirina Geary
Tamien Nation, qgeary@tamien.org

Follow this and additional works at: <https://scholarcommons.scu.edu/mutsun>

Recommended Citation



Warner, Natasha; Butler, Lynnika; van Volkinburg, Heather; and Geary, Quirina, "Mutsun Text Collection - Introduction" (2022). *Mutsun Text Collection: mutsun riicakma hummen*. 5.
<https://scholarcommons.scu.edu/mutsun/5>

This Front matter is brought to you for free and open access by the Archives and Special Collections at Scholar Commons. It has been accepted for inclusion in Mutsun Text Collection: mutsun riicakma hummen by an authorized administrator of Scholar Commons. For more information, please contact rscroggin@scu.edu.

Introduction: Mutsun Text Collection¹

Quick-start guide

If you want to just be able to flip through the text collection and read some things, the line of Mutsun to read out loud and pay attention to is the second line of each numbered text entry (the bold one in red). The line of English to look at is the Trans. line (first line below the Mutsun words, in red). Ignore * marks. Do not start with sentences where part of the word has hyphens on both sides of it, like -w- or -s-. These are explained below. Here's an example:

<p>24 awesnakma awesna -kma awesna -mak₁ star plural N N:Any</p> <p>Translation: stars Orig. spell: 'awes/'nakma Source: 41/0059b-2 Compare: rehearing Merriam Source trans.: Research notes: Stamped Sep 6, 1929.</p>		<p>awesnakma This is the <u>Mutsun</u> word. Ignore the (-) hyphen or (=) equal sign and read together. It's separated to show the parts of the word</p>
	<p>This is the English translation</p>	

Pronunciation guide:

a, e, i, o, u: Spanish (**taco**, **pero**, **mi**, **usted**)

c: English 'ch' (**chip**)

L: English 'ly' (**tell you**)

r: Spanish single 'r' (**pero**)

S: English 'sh' (**shop**)

T: Curl tongue tip up and back to make a noisy 't'²

tY: English 'ty' (**Katya**)³: little gap in English uh-oh

double letters: same but longer.

Capitals are special sounds. Letters not listed here are pronounced like in English.

The Mutsun language

Mutsun is a Southern Costanoan language, and part of the Utian language family (Callaghan 2012). It is from the area of the Pajaro River drainage, where the modern towns of San Juan Bautista, Gilroy, and Hollister are now located (Levy 1978). In mission times, the Mutsun people were the primary group at the San Juan Bautista Mission. Within a short time after the founding of the Mission, the Mutsun language became threatened and then endangered.

Harrington includes some mention of the use of Mutsun being forbidden in the mission and use of Spanish required by as early as 1813 (Reel 57/0468).³ The last historical speaker of Mutsun, Doña Ascension Solarsano, died in 1930, and the last fluent speaker of any Costanoan language,

¹ Throughout this introduction, special terms linguists use appear in the footnotes.

² Voiceless retroflex stop.

³ The organization by "reels" is explained below.

Isabelle Meadows (Rumsen) died in 1939 (Mills 1981). The Mutsun community as of the current writing in 2022 is very active, and language revitalization work has been underway since 1995.

What is in this text collection?

The text collection includes very nearly all of the original source material about the Mutsun language from Father Felipe Arroyo de la Cuesta (collected 1808-1815), C. Hart Merriam (collected 1902), J. Alden Mason (collected 1916), and J.P. Harrington (collected 1922 and 1929-30). It is our intention to include all of the language material on Mutsun. However, many of these early authors made multiple copies of their fieldnotes, sometimes changing something in one copy, or they re-elicited similar but not identical versions of the same sentences multiple times. It is often not clear what is just an extra copy of previous notes (sometimes in neater handwriting or transferred to a multi-language list for comparison) and what might have a few bits of new information in it. We have attempted not to include exact duplicate copies of notes.

Harrington also included occasional words of Mutsun language material within other portions of his work about cultural and historical information or in portions that are primarily about other languages and are filed under those languages. It is extremely difficult to locate all the language material within his very large collection of notes, despite the Smithsonian Institution's efforts toward organizing his papers (Mills 1981). We have probably missed a small proportion of the language material. However, the text collection includes very nearly everything that was written down about how to speak Mutsun, and any additional missing material is likely to represent mostly different pronunciations of known words, not additional unknown words.

A few entries in some chapters have been removed from the publicly available copy of this document because they contain culturally protected information. Community members can access the Community version, which contains those entries. Please consult with the archive for how to access that document if you are currently reading the Public version and would like to access the Community version.

What can you use this book for?

If you are Mutsun and you are interested in learning the Mutsun language, you can use this set of texts to see how your ancestors actually talked: to read sentences, words, and a few stories that they themselves said. This is not a language textbook, and you should use other language learning materials as your main way to learn how to talk and make sentences in Mutsun. The sentences in these texts are not constructed to show you how the grammar works. The purpose of these texts is to give you the original sentences native Mutsun speakers said. You should keep in mind, though, that sometimes the speakers were struggling to figure out how to say something, or Harrington pushed them to try saying it a different way, so there are some sentences that they probably would not have said in normal conversation. Still, this book gives you a lot of examples of how native Mutsun speakers expressed various things.

If you are a linguist, this text collection makes it possible to find the usages of a given morpheme, or to see how we have analyzed particular utterances. It also provides examples produced by native speakers that illustrate features of Mutsun phonology, morphology, and syntax, going beyond the examples given in Okrand (1977). Furthermore, the inclusion of the

original source spelling, original source translation, and notes from both the original source and our own analysis allow additional future analyses of Mutsun. For example, one could use the diacritics and vowel length in the original spellings to attempt a new analysis of Mutsun stress and length patterns.

Chapters of the text collection:

The text collection is divided into chapters based on which source documents the language information came from. The chapters are:

- Reels 58-67 and 69-71: The Mutsun language portions from each microfilm reel in Part II Reel 58-Reel 71 of the Harrington Papers in the Smithsonian's microfilm collection. Each reel has its own chapter. (Before the Harrington Papers were put online, they were organized by microfilm "reels," and we continue to use that numbering system because it makes it easier to find a given page in the materials. The "reel" numbers are also used in the Smithsonian's online copy, with slightly different page numbers.) Some parts of this include Harrington rehearing older sources. For example, the last approximately 50 entries of Reel 61 contains Harrington's rehearing of the Rumsen Torres vocabulary, Reel 67 is the Mutsun words that came up while rehearing a Rumsen dictionary Harrington was working on with Doña Solarsano, and Reel 69 begins with rehearing of Comelias' Santa Cruz vocabulary. A few parts are Harrington's own elicitation of Mutsun, not rehearing of other sources.
- Reels 36-40: The Mutsun language portions from each microfilm reel in Part II Reel 36-Reel 40 of the same collection. Each reel has its own chapter. Some parts of these reels were probably from Harrington's earliest work with Mutsun, before he knew the language as well as he did later.
- Harrington letter to Smithsonian: The Mutsun parts of a letter Harrington sent to the Smithsonian, because the text here is slightly different from the text he gives in the fieldnotes for the same story.
- Arroyo's *Vocabulario de Mutsun* Corpus: Arroyo's collection of Mutsun sentences
- Harrington's Rehearings of Arroyo's *Vocabulario* (abbreviated Ha Ar. sentence rehearings): Harrington's re-checking of Arroyo's sentences, including anything else he and Doña Solarsano discussed during that process (Reels 41-43)
- Reel 57 Arroyo's material: Arroyo's Mutsun material that Harrington included as hand-recopying, on Reel 57. These entries are translations of Christian prayers into Mutsun, a translation of a catechism, etc., and they are not included in Arroyo's sentence collection. They also include Arroyo's work *El Oro Molido*. Harrington included a copy of the original page of Arroyo's work for some of these texts, but the hand-recopied versions are much more legible, so those are what we included in the text collection. The hand copies are probably copied by Marta Herrera, the granddaughter of Doña Solarsano, who worked with Harrington. The texts in this chapter were not Mutsun speech spontaneously said this way by native Mutsun speakers, since they are translations of Christian documents from Spanish. We do not know how often Arroyo asked native Mutsun speakers for help in making good translations, vs. deciding how to express the ideas in Mutsun himself.
- Reel 57 (Harrington material): Harrington's Mutsun language material on Part II Reel 57, which is mostly Harrington's re-elicitation of words here and there from the Reel 57 Arroyo work. As far as we know, Harrington did not do a thorough re-elicitation of that material.

- Arroyo's *Gramática Mutsun*: The Mutsun language material from Arroyo's grammar of Mutsun. Suffixes or ends of words given without a word context are omitted; all other Mutsun examples are included.
- Mason's Vocabulary: J. Alden Mason collected a very small amount of Mutsun language information himself, and this is included here. He also analyzed Arroyo's sentence collection to produce a dictionary, but that is not included here, because it does not include any new information that is not in the other texts.
- Merriam Vocabulary, Merriam Vocabulary Part II, and Merriam II: Barbara: These three chapters include C. Hart Merriam's Mutsun language material.
- Re-hearing of Merriam and Henshaw: Harrington re-elicited small parts of the Merriam material, as well as of an earlier text by Henshaw that was on a Northern Costanoan language rather than Mutsun. These re-elicitations appear on Reel 41 of the microfilm.
- Pinart and Hewitt: These are two short word lists which almost certainly contain Pinart's and Hewitt's recopying of Mutsun words from Arroyo materials, not additional fieldwork material.⁴ The spelling generally matches Arroyo's except where they may have made errors copying Arroyo's words over.

The chapters are ordered by ease of readability, reliability of the Mutsun in a given source, usefulness of the content, and relationship among the chapters, not chronologically. For example, the Arroyo sentence collection, Arroyo grammar, and Reel 57 Arroyo material (translations of Christian documents) are the earliest sources of information about Mutsun, but they contain many words and phrases that are probably incorrect transcriptions or misunderstandings. Reel 57 in particular is not representative of native Mutsun culture or native language use, since the purpose of those materials was to translate Christian documents into Mutsun. Reel 57 would be very difficult and misleading to use in language learning. The early Harrington microfilm reels (especially 36, 37) include material Harrington collected in 1922, when he first began working on Mutsun. He gained greater understanding of the language as he continued working with Mutsun, so his later transcriptions seem more reliable.

We suggest that readers begin with Reel 66, so we have chosen that as the first chapter. Reel 66 contains sentences that are not too complex, but that are more than just a word list, and its content is a good starting place. If you would like to read something that is relevant to Mutsun culture, we suggest reading Reel 59 starting at entry 217. This is the only traditional story we know of that was written down in Mutsun. The chapter of Harrington's letter to the Smithsonian is also very culturally relevant material. Reel 58 starting at entry 1261 contains a discussion of use of tobacco, and the Mason vocabulary starting at entry 40 contains the only letter from one person to another that was ever written down in Mutsun. Reel 66 entries 37-45 represent the speaker thinking about his home. All of these are very culturally relevant.

We did not include Henshaw's vocabulary list (1884), because although Henshaw labeled it as being Mutsun, the speaker was not from a Mutsun area, and the vocabulary seems to be Tamyen

⁴ The Pinart document has "Arroyo" written on one page in a way that suggests the material came from an Arroyo document. Daisy Njoku of the Smithsonian (personal communication) suggests that Hewitt never conducted fieldwork in the western part of the U.S., and Hewitt's and Pinart's spelling for Mutsun words both match Arroyo's too closely to be independent sources. The title of the Pinart document states that the material is copied from other sources.

or some other Costanoan language instead of Mutsun. Within the Harrington material, we did not include anything labeled as coming from his work with Isabelle Meadows (notes headed “Iz.”), because she was a native speaker of Rumsen, not Mutsun (Mills 1981). Harrington might occasionally have been asking Isabelle Meadows how she thought one might say something in Mutsun, but mostly he was clearly asking her to translate Mutsun sentences into Rumsen, and she was responding with Rumsen words and sentences.⁵ We also did not include non-language material that is in Harrington’s papers on Mutsun, such as discussion of the Spanish version of the Catholic catechism used in the Missions, or discussion of cultural topics or historical information given in English or Spanish without anything in the Mutsun language. Cultural comments about a particular Mutsun word or sentence are included in the Research notes field, but not other information that was not about a Mutsun word.

Each of the source documents brings with it its own history, its own mistakes, and its own contribution to the understanding of the Mutsun language. The characteristics of each source are described in the Introduction to the Dictionary (Warner et al. 2016), as well as in Okrand (1977) and in Warner et al. (2006). Details of how the author of each source transcribed the Mutsun sounds are explained below.

What information is in each text entry?

Each text entry is numbered consecutively, starting at 1 at the beginning of each chapter, so you can keep track of the order. Figure 1 shows an example, and numbers in the text below refer to it:

19	waak-was	yete	kuwa	men-appase	1	.
	wak	=was	yete	kuwa	men=	appa -se 2
3	wak	=was	yete	kuwa	men=	appa -se
	he/she/it	=him/her/it	will, later	say, tell	your=	father objective 4
	Pro	Pro	Adv	V	Pro	N N:Any 5
6	Translation: He will say it to your father.					
	Orig. spell: wá̄k was yete kuwa men’appase’ 7					
	Source: 58/0006b-1 8					
9	Compare: Cap. 1					
	Source trans.: el va decira tu padre 10					
	Research notes: There is a '9' circled in the top right corner. 11					

Figure 1. Example of a text collection entry, entry 19 from Reel 58.

After the entry number (19 in this example), there are five lines, with words aligned under each other:

- 1st line of each text entry (1): This is how the original source’s author thought that the sentence sounded. It is spelled in the writing system we use in the Mutsun community (the practical orthography). This reflects what the linguist (Harrington, Arroyo, etc.) thought they heard one particular time that someone said the sentence, not necessarily a

⁵ Rumsen has word-final consonant clusters and an /x/ (transcribed by Harrington as x). Mutsun has neither of these, so it is very easy to see that the Isabelle Meadows material, marked “Iz.,” is in Rumsen.

stable consistent pronunciation. Each of these linguists also made various kinds of mistakes, described below. Therefore, if the first and second lines are different, the second line is the better way to pronounce the sentence.

- 2nd line of each text (2), in bold and red: This is how we think the sentence should be pronounced. It's our best estimate of how Mutsun speakers would have said the sentence--our best understanding of the correct way to say it. This is the line to read out loud. (There is one exception, explained below: anything written with hyphens on both sides of it, like -tY- or -s-.)
- 3rd line of each text (3): This line shows you what form to look up in the Mutsun-English dictionary if you want to find one of the words or parts of words in the sentence. Notice that for a lot of words (all the words in this example), the 1st, 2nd, and 3rd lines are all the same. We explain below why you need all three lines and what the difference is.
- 4th line of each text (4): short English meaning of each word or part of word. You could look these up in the English-Mutsun dictionary if you wanted to know more about how to say that English word in Mutsun. They are also here to show you how the parts of the sentence fit together to make the meaning of the whole sentence, and you can figure out a lot of Mutsun grammar by looking at these.
- 5th line (5): Part of speech (like noun, verb, etc.) of each word or part of a word. See the Introduction to the dictionary (Warner et al. 2016) for explanations of what the various parts of speech are.

Right below these five lines there are about 5-6 more lines, but these are not lined up word by word under the texts. These are:

- Translation (6): The free translation, which is what this sentence really meant in Mutsun, or our best understanding of what it meant if we're not sure. This is based on what the Mutsun sentence means, not on what the original source's author said it meant. We chose the free translation by comparing to other sentences that use some of the same words, to what the source's author gave as a translation, and to Okrand's (1977) explanations of Mutsun grammar. If you are trying to learn Mutsun, this is the line to use to see what the sentence really means.
- Orig. spell (7): The original spelling the source author used to write the sentence, in whatever system that author used to write the sounds of Mutsun. See below for more explanation of transcription systems. Do not use this line for learning Mutsun, because each author wrote sounds differently.
- Source (8): Information about where to find this sentence in the original source document. Details about page numbers are below.
- Compare (9): If this sentence came from Harrington, and he was trying to find out how to say a sentence that was in the earlier sources (for example Arroyo), this tells you where in the earlier sources to look to find out what earlier sentence he was asking about. This comes up because Harrington took the much older and less accurate earlier source documents, especially Arroyo's *Vocabulario*, and asked Doña Solarsano how to say the sentences in them in order to clear up confusions in the old Arroyo material. So this line tells you which Arroyo sentence Harrington was comparing to. Sometimes Harrington and Doña Solarsano then started discussing other sentences or followed up on a tangent, so not all entries are directly related to the "Compare" entry.

- Source trans. (10): This shows exactly what the original source author gave as the meaning of the sentence, in whatever language the original author used for the translation, usually Spanish. These meanings are often not very exact, or not quite right. Sometimes the original author didn't give a translation at all. Whatever the source author did give as a translation is here, but do not assume it's meant to be an exact translation, or that the source linguist always understood what the Mutsun speaker was saying.
- Research notes (11): This includes any notes the original source made about the sentence and information about what else is on the same page, as well as information we want to give you about the sentence, for example saying that the sentence is very hard to interpret and we're not sure what it means, or that a word in the sentence is used in an unusual way.

What do -, =, and * mean, and how are they written in the regular Mutsun spelling system?

When we write words and sentences in the text collection, we separate out all the parts of a word that have meanings (the suffixes and other similar things). The English word “cats” is made up out of “cat” and “-s,” and the “-s” means plural, more than one. In the same way, many Mutsun words are made up of several parts. For example, entry 19 of Reel 58, shown in Figure 1 above, is **wak-was yete kuwa men-appase** ‘He will say it to your father.’ That is how we would write it in the normal Mutsun spelling system, and how you should write it when you write in Mutsun. In the normal Mutsun spelling system, we use a hyphen to show where we have attached a “tack-on pronoun”⁶ to a word. The **-was** in this sentence is a tack-on pronoun meaning “it,” and the **men-** is a tack-on pronoun meaning “your.” (You can read about these in the introduction to the dictionary, or look them up in the dictionary, to learn more about them. These are also included in Mutsun language learning materials.)

But in the second and third lines for this entry in the text collection (labeled 2 and 3 in the figure above), you'll see that this sentence is written as **wak =was yete kuwa men= appa -se**. In these lines of the text collection, we use = to show the tack-on pronouns, and we use hyphens to show other parts within a word (suffixes). **appase** is made up of **appa** ‘father’ and **-se**, which shows that the word “father” is the object of the sentence (the person who is being told). In the 11th entry in Reel 58, you see the word **appakma** ‘fathers,’ with the second and third lines written as **appa -kma**. This shows that **-kma** is a separate part of the word, the part that means ‘-s’ (plural, more than one). There are also spaces around the - and = marks, separating the parts of the word. We use this different way of writing only for the second and third lines of each entry in the text collection, because in the text collection it's important to be able to see what pieces make up each word. The spaces in these lines of the text collection are there so that the meaning of each

⁶ “Tack-on pronoun” refers to clitic pronouns. Mutsun has enclitic singular subject and object pronouns and proclitic singular possessive pronouns, as well as a few less common plural clitic pronouns. The practical orthography in use in the community uses a hyphen to separate pronoun clitics (both pro- and enclitics) from the rest of the word, and does not write any separation between roots and affixes. There are a few non-pronoun enclitics, such as the =s question particle, and these are not marked with a hyphen in the practical orthography because doing so makes words difficult to read. However, these non-pronoun clitics are relatively rare, and were going out of use by Doña Solarsano's time, and it is not clear that their status as clitics rather than suffixes is important for language learners. The interlinear parse lines of the text collection use = to separate all clitics (pronoun or not) and - to separate affixes.

part of the word can be given right below it, to help readers learn how parts of words go together to build Mutsun words.

When you read the sentences in the text collection out loud, don't make a pause at the - or = marks. Just read all the parts together as a word. When you write Mutsun yourself, put in - to separate tack-on pronouns like **-ka**, **-ak**, **-me**, **kan-**, **men-**, etc. from the rest of the word, but don't put any mark in to separate suffixes (just write **appase** or **appakma** with no separation in the word). Here's a summary:

	How you should write it (practical orthography)	How it's written in lines 2 and 3 of the text collection (linguistic notation)
Piece of word (suffix)	No separation: appakma 'fathers' appase 'father (obj. of sent.)'	With spaces and hyphen: appa -kma 'fathers' appa -se 'father (obj. of sentence)'
Tack-on pronoun	Separated by hyphen: men-appa 'your father' kuwa-ka 'I tell'	With spaces and = sign: men= appa 'your father' kuwa =ka 'I tell'

Tack-on pronouns are a little bit separate from the rest of the word, unlike suffixes, which really become completely part of the word they attach to.⁷

In some entries, you'll see a * mark in the second and third lines. This is a kind of information the database keeps about some words, and you can just ignore it.⁸

Parts of words with hyphens on both sides of them (for example -w-, -tYe-) (Advanced topic: do not start with this)

Some parts of words in Mutsun attach into the middle of a word, not onto the beginning or end. In the second and third lines in the text collection, these are written with hyphens on both sides of them, to show that they need to attach into the middle of something, not onto the edge. For example, **-w-** in Mutsun means that something gets undone, opened, or done outward. **maahi** means "to close, to cover" something, but if you add **-w-** in the middle, it makes a word **mahwi** that means "to open, to uncover." (For another example, **kicca** 'to lock' becomes **kicwa** 'to unlock.')

These parts of words that combine into the middle of a word like **-w-** are called infixes. There are only four infixes in Mutsun: **-w-**, **-tY(e)-**, and two both pronounced **-s-** but with different meanings. **-tYe-**, which is sometimes pronounced **-tY-**, makes the meaning of the verb it goes into stronger, meaning that the person really, extremely does that action (**ricca** 'to talk,' **rictYa** 'to talk a lot'). One of the **-s-** infixes goes into verbs and means that one does the

⁷ Clitic pronouns count as part of the word for some phonological processes (e.g. vowel lengthening/shortening), but they do not for others. See Okrand (1977) and footnote 10 of the Introduction to the dictionary for further information.

⁸ FLEEx uses * to mark bound morphemes. * does **not** indicate ungrammaticality or historical reconstruction. Morphemes are marked as bound if there was substantial evidence in the corpus that the morpheme could only occur with certain suffixes. The evidence is either that Doña Ascension stated the word could not be said in its unsuffixed form, or that there are quite a lot of tokens of it in the database and it never occurs without a suffix.

action repeatedly, or many people do it (**tere** ‘to cut,’ **terse** ‘to cut a bunch of things,’ for example you use **terse** for cutting fingernails because there are ten of them). The other **-s-** infix only goes into words for family members (kinship terms), and means “my” (**appa** ‘father,’ **apsa** ‘my father’).

Just like with the suffixes, when we write Mutsun normally, we don’t use any symbol to separate infixes from the word they go into, we just write **mahwi**, **rictYa**, **terse**, **apsa** etc. But in the second and third line of the text entries, they’re written with hyphens on both sides, and they’re also separated from the word by putting them after the word they go into. In Reel 58 in entry number 1186, you see **terses men-urihse** ‘Did you cut your hair?,’ written in the second and third lines as **tere -s- -s men= urih -se**. This means the **-s-** needs to be put into the middle of the word before it, **tere**. If there is an infix in the sentence (a piece with hyphens on both sides of it, one of **-w-**, **-tYe-**, or **-s-**), you can’t just read the second line of the text entry out loud as it is, you have to put the infix into the middle of the word when you pronounce it.

You see from the examples above that the pronunciation of a word sometimes changes a little bit when an infix is put in the middle. For example, the long **aa** in **maahi** shortens when the **-w-** is inserted, making **mahwi**, and the long **pp** in **appa** shortens when the **-s-** is inserted, making **apsa**. Usually the change is just to shorten whichever segment is long (spelled as a double letter). The best way to find out for sure how to pronounce a word when you’ve put the infix into the middle is to look up the word in the Mutsun-English dictionary. For example if you know you need to put **-s-** into the middle of **appa**, you could start looking under either **appa** or **apsa**, and you will find the word. But you can also look at the first line in the text collection, and it will usually give you a good hint of how to pronounce the word with the infix inserted into it. (Because the first line shows how the original source’s author heard the sentence that particular time, the first line is usually not the correct way to pronounce the sentence, but it can often show you how to insert the infix.)

This means that when you run into a word with an infix (piece with hyphens on both sides) in it, it takes more work to read the sentence out loud. Since these don’t come up very often, though, we did not add an extra line to show how to pronounce the sentence with the infix.

Differences between the 1st, 2nd, and 3rd lines of a text entry

Remember, if you just want to get started reading quickly, the second line of each text entry (the bold red one) is the one to pronounce. This section gives advanced information about the differences between lines, but you do not need to know this in order to read the second lines.

1st vs. 2nd lines⁹: As mentioned above, the first line of each text entry shows how the original source’s author thought the sentence sounded, at least that particular time that they heard it. When a linguist is writing down how a sentence is said while listening to a speaker, it is very

⁹ The first line encodes the original source’s recording of a given token, converted from the source’s transcription system to practical orthography but not otherwise altered. This means that the first line includes variation stemming from dialect differences, sound change between source time periods, attrition, free variation, neutralizations imposed by a given source linguist’s transcription system, and probably most often error by the original linguist. The second line (bold) encodes our best understanding of the most likely surface form, which is the recommended form for language learning and revitalization purposes. See the introduction to the dictionary for discussion of factors considered in choosing the preferred form.

hard to get the pronunciation exactly right every time. All speakers pronounce words a little bit differently from one time to another. Sometimes a speaker says something slower, and it might sound like one of the vowels is long, but another time that vowel sounds short. Sometimes the Mutsun sounds **c** and **T** sounded very similar to the source linguist's ears¹⁰, so they confused those sounds in some words. Other times, the speaker wasn't completely sure about a word and said it several ways, sometimes to see which way sounded right to them. In some other cases, the typesetter who worked on publishing Arroyo's handwritten documents couldn't read the handwriting clearly, and we can't be sure what letters Arroyo meant. Sometimes, there are just two ways to pronounce a word (like "Caribbean" in English, which you can say as CA-rib-BE-an or as ca-RIB-be-an, without changing the meaning). All of these things lead to inconsistent spellings in the first line of a text entry. By comparing across all of the sentences that contain each word, we have decided what we think the most likely pronunciation for each word is, and this is what is given in the second line of the text entry. For example, entry number 22 in Reel 58 has the first line as **men-aananTu**, but the second line as **men= aanan -Tuk** (in the normal writing system without the = sign, this would be **men-aananTuk** 'with your mother'). This is because the suffix **-Tuk** was often pronounced as **-Tu** in Arroyo's time, but was almost always pronounced as **-Tuk** by Doña Solarsano's time. To summarize: when the first and second lines are different, the second line is the one we think is the more likely or better pronunciation. The introduction to the dictionary explains how we chose which pronunciation to use (Warner et al. 2016).

2nd vs. 3rd lines¹¹: If the 2nd and 3rd lines of the text entry are different, it shows you that part of one of the words has two pronunciations, and you use one in certain contexts, and the other in other contexts. For example in entry number 11 of Reel 58, the second line shows **appa -kma** (normal writing system: **appakma** 'fathers'), but the third line shows **appa -mak**. This is because the plural suffix is **-kma** if the word it attaches to ends in a vowel, like **appa** does, but it's **-mak** if the word ends in a consonant. You can see this in **aananmak** 'mothers' in entry 24 of Reel 58. Because **aanan** ends in a consonant, both the second and third lines show **-mak**. The main form of the suffix is listed as **-mak** in the dictionary, so in a word like **appakma** that uses the form **-kma**, the difference between the second and third lines shows you that a different pronunciation is used here. Suffixes and words like this that have different pronunciations depending on the sounds around them are explained in the dictionary introduction under "Other pronunciation."

Entry 9 in Reel 58 shows both of the kinds of pronunciation differences in one word. The first line shows **appaTu**, the second line shows **appa -hTuk**, and the third line shows **appa -Tuk** for 'with one's father.' The difference between the second and third lines is a regular, predictable difference: **-hTuk** is used if the word it attaches to ends in a vowel, and **-Tuk** is used if it ends in a consonant. Since **appa** ends in a vowel, **-hTuk** is the right pronunciation for this word. But the difference between the first and second lines is a different kind: the pronunciation seems to

¹⁰ We are considering all of the early authors who documented Mutsun as having been linguists, whether they had formal training in linguistics or not.

¹¹ In the case of allomorphy, the second line (bold) encodes the correct allomorph for the environment it is in. The third line encodes the allomorph listed as a headword in the dictionary. See the introduction to the dictionary for explanation of which allomorph is chosen as the headword listing. In cases where the original source uses an unexpected allomorph for a given environment (e.g. uses **-mak** or **-Tuk** after a vowel), the second line "corrects" it to the expected allomorph, maintaining the prevalent pattern of allomorphy.

have changed from Arroyo’s time to Doña Solarsano’s time. We have decided based on comparing all the words with this suffix that **-Tuk** (after consonants) and **-hTuk** (after vowels) is the best pronunciation of the word, not **-Tu**. The second line gives you the pronunciation that we think is most likely or most correct, so you should read **appahTuk** from the second line.

Numbers in the 3rd line of the text entry

Some words and parts of words in the third line have a small subscript number at the end. For example, entry 6 in Reel 58, **horkos** ‘throat’ has a subscript 2 with **horko** and a subscript 1 with **-s**. This means that there is more than one entry in the dictionary that sounds the same, and it shows you which one is meant here¹². If you look in the Mutsun-English dictionary (Warner et al. 2016), you will find that there are two separate words written as **horko**. The one with a number 1 means “to hang, to strangle,” and the one with a number 2 means “to swallow, to gulp.” Under **-s** you see that there are two separate suffixes that sound like **-s**. The one with a number 1 is a nominalizer; you add it to a verb and it makes it into a noun that means the thing you use to do the verb. The one numbered 2 is past tense. (There is also a tack-on **=s** that makes the sentence into a question, and two infixes **-s-**.) In entry 6 in Reel 58, **horkos**, the numbers tell us that we have “to swallow, gulp” and the **-s** suffix that means “thing you use to do the verb,” so this word **horkos** means “thing you use to swallow,” which is the throat. If it were **-s** number 2 instead, this would mean “swallowed, gulped” instead of “throat.”

Abbreviations

Common abbreviations Harrington uses, which you may see especially in the Research Notes field:

Harrington’s abbreviation	Meaning
ch., Ch., etc.	“Clearly heard” or “clearly heard forever,” meaning Harrington thought he was sure how the word was pronounced (but he may have changed his mind later)
Nesc, nesc., N.	“Nescit,” Latin for “Does not know,” meaning that Doña Solarsano did not know the word
Adiv.	“Adivina,” meaning that Doña Solarsano guessed the word would probably be pronounced that way, but was not sure
rhd.	“Reheard” meaning Harrington re-checked the word again later, or had checked it another time earlier, with Doña Solarsano
Inf., Asc.	Informant or Ascension, both referring to Doña Solarsano
Iz., Is.	Isabelle Meadows, a Rumsen speaker Harrington worked with
Ms., ms.	Manuscript (referring to Arroyo’s handwritten manuscript rather than the published version)
sic	Latin for “thus,” for example in “Ms. sic” to mean the manuscript spells a word a particular way
ev.	“evidently”
Prob.	“probably”

¹² These are homophone/homograph numbers.

Sp.	“Spanish”
□	ungrammatical (incorrect in Mutsun, meaning he tried asking if you could say it this way, and Doña Solarsano said that you can’t)
*	a guess, similar to “Adivina”
SJ, SJB	San Juan Bautista, can also refer to the Mutsun language

For a longer list showing Harrington’s less common abbreviations, see Mills (1981: 161-176).

We also use some abbreviations in our own notes in the Research Notes line, especially “Asc., Iz., Ms.” for the same meanings as Harrington uses them. We abbreviate Harrington as “Ha” or “H,” Merriam as “Me,” Arroyo as “Ar,” and Mason as “Ma.”

Arroyo often abbreviates Spanish words by writing the beginning of the word, and then writing the last letter or two of the word as a superscript letter. He does this most often in the Christian documents, where the words are common and easy to guess, such as “J^{to}” for “Jesu Christo” or “q^e” for “que.”

We use “{entry}” or “[entry]” to mean that the entire entry (the Original Spelling line) appears at that point. For example, Reel 58 entry number 38 has a research note: ‘H notes "But {entry}”,’ this indicates that Harrington’s entry for this line, “há̄wu,” is preceding by the word “But.” We use “{next entry}” or “[next entry]” to show the next entry down appears at this point in a note. Occasionally we also use { } or [] to enclose other words of explanation about an entry, such as “{faint text}” to indicate that there is illegible text on the page that we could not include.

Page numbers

The “Source” line tells you where each text entry came from, including what page or section of the source document it came from. The format of this information is different for the various source documents.

The Harrington material has been put on microfilm by the Smithsonian. All of Harrington’s Mutsun language information is in Volume II of his papers (Northern and Central California). Each page in the microfilm shows two of Harrington’s original sheets of paper, one on the left and one on the right. We give the Harrington source information as microfilm reel number followed by a slash mark, then the page number within the microfilm reel, “a” or “b” to indicate the left or right page of the microfilm page, and then a hyphen and the number of the sentence on that page. For example, 41/0319b-3 means that you could find that sentence on Volume II, reel 41, page 319, right half of the microfilm shot, third Mutsun sentence (or line, see below about sentence boundaries) on that page. You can also find the Harrington notes online through the Smithsonian. Unfortunately the online version uses slightly different page numbers than the original microfilm page numbers we used throughout most of our project. We added the entries for several reels and parts of a few other reels after the Smithsonian had put them online, using the online page numbers. We have included a note at the start of each chapter of Harrington material giving the page number conversions between what we list and the online version of the Smithsonian microfilms for that chapter.

The various Merriam documents are separate wordlists or partially overlapping copies of wordlists that are available through the Bancroft Library at the University of California, Berkeley, either on microfilm or online. It is difficult to determine which copies of wordlists come from which section of the Merriam papers, and we do not have exact microfilm reel and page numbers for all of the wordlists. The chapter of the text collection titled Merriam Vocabulary comes from the C. Hart Merriam Papers, Banc MSS 80/18c, Film 1022, Reel 74, BNeg Box 1556. A similar copy also appears on Reel 30, frames 80-163. The page numbers in the Source line for this document are formatted as in the example “Me74a1-2,” where a/b/c refer to three separate wordlist documents, 1 refers to the page number within each wordlist, and the number after the hyphen is the line/item number on the page (in this example, the second line of Mutsun on the first page). The other Merriam chapters are separate documents Merriam assembled that include Mutsun material, also numbered with page number and item number on the page.

For Arroyo’s *Vocabulario*, numbering refers to the sentence number in the published copy (that is, Ar1 is the first numbered sentence of the *Vocabulario* and Ar2884 the last). For Arroyo’s *Gramática*, we listed the chapter (Capítulo, abbreviated “Cap.”) the item occurred in, and entries are in order within each chapter. For Arroyo’s translations of Christian documents, we used the copy Marta Herrera made for Harrington, that appears in Harrington’s Reel 57, rather than using the much less legible originals. Page numbers for the Arroyo material in Reel 57 Arroyo follow the page numbering convention for Harrington material explained above. The very short Mason document is numbered by page number and sentence number within the page for the few pages of this document. The Pinart document is from a folder labeled “Native American vocabularies, WA MSS S-285, Box 1, Folder #6,” and the page numbers we list within it are the page numbers of the pdf copy, where page 1 is a picture of the folder itself. The Hewitt is from the Smithsonian (NAA) and is labeled NAA.MS3950, and the page numbers we use are the image numbers in a scan of that document.

How this text collection was made

This section has information for linguists about decisions we had to make while preparing the text collection. You do not need to read this section in order to read the texts.

Overall process:

This text collection was finalized using Fieldworks Language Explorer (FLEX) v. 9.0.17.455. However, much of the entering and parsing of Mutsun language material was originally conducted in Shoebox and then in LinguaLinks, followed by earlier versions of FLEX. To create each text entry, we entered the words or sentences from the original source document in the transcription system of the source, along with the translation given by the source linguist, page number information, and any notes or other information in the original document. For Harrington’s rehearsals, we also entered information about what item from another source document he was rehearsing. We then converted the original transcription of the item to a rendition of the same sounds in the practical orthography. For example, if Arroyo wrote “aisa,” we converted this to **aysa** because Arroyo’s transcription system uses **i**, **u** in place of **y**, **w**, and Mutsun rarely has two adjacent vowels. However, we did not convert it to **haysa** even though **aysa** is known to be a common Arroyo variant for **haysa** ‘they,’ because Arroyo’s transcription gives no indication of an /h/. (Arroyo almost always omits word-initial /h/, although he could

have used Spanish “j” to express /h/, and sometimes did in other positions.) That is, we do not rely on comparison with other sources to “correct” a source linguist’s transcriptions in entering them into the first line of each entry, even if the word is well known, since we want this first line to reflect the original linguist’s impression of how the utterance sounded. (However, when faced with a sequence of high vowels and glides, as in words like **hiwsen** ‘to want, like’ or **huyni** ‘to fish,’ which Arroyo usually writes as “iu” or “ui,” we did consult the other sources to determine which segment should be a vowel and which a glide, in order to disambiguate /wi/ from /uy/, etc.)

The published version of Arroyo’s works on Mutsun (1861, 1862) contains very frequent errors made by the typesetter, as Harrington also noted. This is not surprising, since Arroyo’s older handwriting style makes many letters look similar (e.g. m, n, u; c, e), and these works were published after his death. For Arroyo’s *Vocabulario*, we had access to a copy of the manuscript version, and we used this along with the typeset version to correct typesetter errors. Harrington also often noted typesetter errors in his rehearsals, since he also had access to a copy of the manuscript version. We include a note in each entry where we have corrected a typesetter error based on our own or Harrington’s examination of the manuscript. For Arroyo’s *Gramática*, we did not have access to a copy of the manuscript version, but the paradigmatic nature of most of the Mutsun words in the *Gramática* makes it easier to spot typical typesetter errors. We noted corrections based on assumed typesetter errors in these entries as well.

We then analyzed each utterance into morphemes, and parsed each entry by choosing the morphemes from the lexicon through the software and approving each parse by hand. That is, we did not use any automatic parser, but instead examined each entry separately. This process created the second line of each entry (expected surface pronunciation of the entry, except for infixes), third line (dictionary headword listing of each morpheme), fourth line (morpheme glosses), and fifth line (parts of speech). We also added information about anomalous, unexpected grammatical usages or words for which we were unsure of the parse in the Research Notes field where relevant.

For a few words in the Arroyo material, especially his translations of Christian materials, there was simply too little information about a word to parse it as anything at all. If Arroyo gave no translation or the translation seems unrelated to a given word of the sentence, there are no phonologically similar words in the dictionary, and Harrington did not rehear the sentence or Doña Solarsano was unable to give any information about the word at all, then we did not parse the unexplained word. For example, an Arroyo text beginning on 57/0592b urging people to participate in confession includes an unknown word **amteam** that does not closely resemble any known word. No specific translation for the text is given, Harrington did not re-elicite this, and it does not re-occur elsewhere. Any parse we could give would be a guess, so we have not parsed this word. We suspect this was especially common in the translations of Christian materials because Arroyo was trying to express abstract theological concepts in Mutsun, so he may have been using low frequency words that later underwent attrition. We also left a very small number of words in other materials unparsed, for example words to native songs that appear to be chanted syllables rather than lexical items or words where the text was too faint to recover missing parts of the word.

Arroyo sometimes uses a word that is rather phonologically distant from the most similar word in the dictionary, but that is a plausible match for meaning. For example, in sentence 905,

Arroyo includes two words he spells as “hujus speciei.” In the practical orthography, this would be **uhus spekyey** (including a phonotactically impossible onset **sp** cluster, furthermore the string **kyey** occurs nowhere else in the database). Arroyo’s translation covers only the earlier part of the sentence regarding spinning a wheel, and gives only “etc.” for these words. Harrington was not able to re-elicite either word. We have interpreted these as **uhwis cekle**, ‘took out, to be uneven,’ which are the nearest phonological forms in the database that could have something to do with spinning a wheel. That is, if Arroyo’s word in such unclear cases is the only evidence in the database for that form, we did not enter it into the dictionary as an additional word, but rather parsed it as the better-attested but phonologically distant match. The Arroyo *Vocabulario* in particular contains so many typesetter errors and unexplained forms that even after consulting the manuscript copy, a mistake of some sort seems more likely than a large number of additional otherwise unattested words. We included a note in the Research Notes line in such cases to note that the parse we have assigned may not be correct. However, if Arroyo uses a form consistently and gives a clear translation for it, we included it as a dictionary entry, even if Doña Solarsano did not recognize the word. This was the case for a few suffixes or adverbs that are used consistently in the Arroyo *Gramática*, such as **munna** ‘very distant in time (past or future),’ as well as a few words for religious terms in Reel 57 Arroyo translations of Christian documents. We assume these morphemes had undergone attrition by Doña Solarsano’s time.

After completing data entry and parsing for nearly all of the language material, the first author performed a further step of analysis for each headword in the lexical database by producing a concordance of all utterances parsed as containing that headword. This allowed for standardizing the presentation of information in the lexical database and thus the dictionary, and it also provided a further check of the parses in the text collection. See the introduction to the dictionary for more information.

Transcription conversions for the various sources

Each source linguist uses a different transcription convention. As mentioned above, all but Harrington’s transcription neutralize some phonemic distinctions, and some sources use a wide variety of inconsistent transcriptions. The conversions we have used are below. Okrand (1977) provides a thorough discussion of the process of interpreting Harrington’s transcription, especially for sibilants, along with arguments for what is or is not a phonemic distinction within Harrington’s transcriptions. We have followed Okrand’s careful analysis of phonemic distinctions on all issues but one, so we will not repeat his explanation here.

The one point where we depart from Okrand’s analysis of phonemic distinctions is that we do not attempt to distinguish between the single segment /ts/ and a cluster /ts/ across a syllable boundary. Although Harrington often notes a syllable boundary between word-medial /t/ and /s/ (**yat.sen** ‘is hurried,’ **pet.sen** ‘sugar,’ as given by Okrand (1977: 31), we do not think this necessarily indicates a distinction between the single segment /ts/ affricate at a syllable boundary and the consonant cluster /ts/ in the same environment. That is, we do not think there is strong evidence for a distinction between **pet.sen** with a /ts/ cluster and a hypothetical word **pe.tsen** with /ts/ as an affricate. For purposes of checking conformity to the maximal CVC syllable structure, we assume that /ts/ can be a single segment, meaning that word-medial consonant-/ts/ and /ts/-consonant clusters are possible, as in **huytsuhte** ‘narrow.’

We also do not mark initial glottal stop for otherwise vowel-initial words (e.g. we write **ama** ‘person’ where Okrand, Harrington, and most other sources would write **’ama** or some other transcription for initial glottal stop). Initial vowels may be produced with a glottal closure at onset, as often in English, but we do not include this in the practical orthography.

Transcription conversions used to map from the original source transcription to the practical orthography (first line of each text entry) are as follows (see also Warner et al. 2006):

Practical Orthography	IPA	Arroyo, Pinart, Hewitt	Merriam	Harrington
p	p	p	p	p
t	t	t	t	t
T	t̥	thr, ths, sth, tthr, occasionally others, almost always ts within the <i>Gramática</i>	tr, t’r	tr (distinct from t.r, which is a sequence of t and r); character similar to t̥ only if the notes are typed
tʸ	tʲ	ti	ty, t if before /i/	tʸ
k	k	c, qu	k	k
’	ʔ	h or omitted	’ or omitted (ambiguous with stress mark)	’
c	tʃ	ch	ch, tch	tc, sometimes tʃ
ts	ts	ts, tz, zs, or variable	ts	ts, t.s ¹³
s	s	s, sometimes zs	s	s, s̥
S	ʃ	si, is	sh	s̥ʲ, s̥, ʃ, sʲ; c or character similar to ɸʲ only if the notes are typed
h	h	j, g, or omitted	h	h; x only if the notes are typed
m	m	m	m	m
n	n	n	n	n
N	nʲ	ñ	ny	nʲ
l	l	l	l	l
L	lʲ	ll	l (or possibly also ly)	lʲ
r	r	r	r, rd, rr	r
y	j	ï, y	y	y, rarely j in early materials
w	w	u, gu, gü, etc.	w	w
i	i	i	ē, ĩ, ee (with no hyphen between), ēē (with no	i

¹³ As explained above, t.s is distinct from ts for Harrington, but not in the practical orthography.

			hyphen between), i in closed syllable, e in open syllable, e if before w	
e	e	e	ā, ě, a if in open syllable, e if in closed syllable	e
a	a	a	ah, ō, ǎ, a if in open syllable, o if in closed syllable	a
o	o	o	ō, o in open syllable	o
u	u	u	u, oo	u

Geminates and long vowels: Harrington writes geminate consonants using a double letter or partial double letter, e.g. “tt” for /tt/, “ttr” for /TT/, or in some portions of his notes with an underline mark under the consonant, e.g. “n” for /nn/. He usually writes long vowels with a macron above the vowel. Arroyo and Merriam most often neutralize length distinctions, but Arroyo sometimes writes “h” between two identical vowels to indicate a long vowel (e.g. “ehē” in Arroyo is sometimes /ee/ but sometimes /e’e/), and he sometimes does write double letters for geminate consonants. Merriam occasionally writes long vowels or geminate consonants by putting a hyphen, which he uses typically as a syllable break mark, between two symbols.

In addition to these conversions, Arroyo and especially Merriam use other transcriptions motivated by the spelling of their native languages (Spanish and English respectively). For example, although Merriam usually writes /a/ as “ah” or “ō” he writes /aw/ as “ow” or even “ough.” He uses “ī” or (if in an open syllable) just “i” for Mutsun /ay/, and sometimes uses “u” for /yu/. Merriam also uses “x” for /ks/. Merriam in fact notes in his explanation of his transcription system (included in one of the documents) that he sometimes spells a Mutsun word using English spelling if it sounds similar to an existing English word.

Merriam’s transcriptions of vowels are especially variable, to the point of being highly unreliable. For example he usually uses “a” in an open syllable to mean /e/ (following the English “long” and “short” vowel pronunciations used to teach English reading, in which “long a” is English /e^l/), but sometimes comparison to other sources shows that Merriam used “a” for /a/ even in an open syllable. It is more likely that Merriam used his own transcription system inconsistently than that he could not hear the distinction between /a/ and /e/. We did not attempt to “correct” the first line of text entries based on comparison to other sources, since it is impossible to tell when Merriam used an inconsistent transcription and when he misheard a vowel. This matches our policy of not “correcting” Arroyo’s omission of /h/, discussed above.

Arroyo’s use of “si” or “is” for /S/ along with his use of “i” for /y/ creates considerable ambiguity, especially in words where /S/ actually is adjacent to /i/ or /y/. This is even more problematic after /t/, where a cluster /tS/ could be confused with /c/. For example, Arroyo has four tokens of the root **atSa** ‘girl,’ which he transcribes as “atsia, atzia, atshsia, atchai.” Mutsun has maximally CVC syllable structure, and almost all Mutsun noun and verb roots consist of two syllables (Okrand 1977). We have used this as a guide in determining how to convert these sequences. For example we do not convert “si” to /S/ if this would make a syllable larger than CVC. Since Mutsun rarely has strings of consecutive vowels (other than long vowels), we

convert Arroyo's "i" and "u" to glides where that will avoid strings of vowels. It should be assumed that some of the variation shown in the first line of the text entries represents the inadequacy of Arroyo's and Merriam's transcription systems for recording all of the distinctions, rather than that Arroyo or Merriam always misheard the sounds. Arroyo simply had no way to distinguish among /S, si, sy/, for example.

Harrington uses a large number of diacritics, including various accent marks that may mark some unknown property of stress or intonation, as well as marks for non-distinctive aspiration on consonants or breathy offset of voicing. Okrand (1977) provides an excellent discussion of this situation and of what information might be recoverable. We have used the system developed by the Harrington Project at the University of California, Davis for converting Harrington's transcriptions to typed format in the Original Spelling and Research Notes lines (<https://nas.ucdavis.edu/sites/g/files/dgvnsk7031/files/media/documents/jphkeybd1.pdf>). The most common Harrington symbols for Mutsun that require a conversion to type them are "š" (entered as "s/°") and "šy" (entered as "s/_y"). In general, / indicates that the symbol after the backslash is written below the preceding letter and \ that the symbol after the slash is written above or superscripted to the previous letter. Long marks that Harrington placed above vowels are entered just after the vowel, without a \ mark, as in "ú̄" or "ē." The relatively rare usage of ŋ in Harrington's Mutsun notes is entered as "«n" or similar symbols. "\$" represents Harrington's rare use of a long cursive f with looped ends in his early Mutsun notes. In the small portion of Harrington's Mutsun notes that is typed, he seems to have used an unusual symbol that looks almost like an open-ɔ with a hook on the bottom. He used this as "tɔ" for /T/ and as ɔ^y for /S/. We were not able to find a conversion for this in the Davis Harrington Project system, so we entered them as typed here. Harrington's many accent marks on vowels are entered simply as á, é, etc., with additional accent or long marks after the vowel if Harrington used more than one on a single vowel. For all sources, any diacritics that do not map onto Mutsun segmental distinctions, e.g. probable stress or intonation marks or syllable boundaries, are included in the Original Spelling line. Because the language material was entered into the database over a span of more than 20 years by a large team on various computers, we were not able to be completely consistent in which symbol was used for curved apostrophes, for example, but we attempted to maintain Harrington's distinction between right-curving vs. left-curving apostrophes. Where we failed to maintain that distinction and used just a straight apostrophe, Harrington generally uses right-curving apostrophe at the beginning of an otherwise vowel-initial word for glottal stop, and left-curving apostrophe at the end of a word for a breathy off-breath similar to aspiration, so the type of apostrophe is largely predictable.

Although Okrand (1977) attempted a thorough analysis of Mutsun stress, he acknowledges that his conclusions about stress are very unsure, and that a large number of tokens in the Harrington material conflict with the partial generalizations he was able to draw. The patterns Okrand identified may also be typologically unlikely as a stress system.¹⁴ Therefore, we have decided that stress placement is too unsure to encode in the practical orthography.

Because there is very little original material from Mason, we have not included his transcription system in the table above. He usually uses "š" for practical orthography /S/ and "t" with a

¹⁴ This is based on brief personal communication with a colleague with expertise in stress (Michael Hammond), and is not a definitive conclusion. Further discussion is beyond the scope of this Introduction.

comma or circle under it for /T/. We record this symbol as “t/,” or “t/°” in the original transcription field, consistent with our keyboarding system for Harrington’s diacritics.

Sentence boundaries, word boundaries, hyphens, and punctuation in the source documents

None of the source authors is consistent in their use of punctuation, placement of sentence boundaries, etc. Harrington’s handwritten notes can have one or a few Mutsun words followed by translations and commentary in Spanish and/or English, and then continue with more Mutsun words, and it is not clear whether the Mutsun words are meant to form a single utterance together or not. Harrington frequently writes notes and additional words or sentences in margins or sideways in between, above, or below other material, or draws lines connecting words. Harrington’s handwriting is not neat, and there may be no visible difference between a period, a comma, and a scratch on the microfilm, if Harrington even used period vs. comma to mean distinct things. It is clear that information about sentence boundaries is not usually recoverable from the Harrington material.

The Arroyo *Vocabulario* contains mostly sentences, but some entries appear to be several related words listed in a row that do not form a sentence (especially if separated by colons). The Arroyo *Gramática* primarily lists word paradigms or example words rather than sentences. Arroyo’s use of punctuation, especially in the *Gramática* and the translations of Christian texts (Reel 57 Arroyo) departs considerably from modern norms. Merriam’s language information consists almost entirely of single words, but he gives some entries as single words that are either compounds or entire sentences, where he probably was not aware that what he heard was more than one word.

Because none of the sources provide reliable information on sentence boundaries, we have entered each line in an original source as an utterance, combining lines in the Harrington material where they appear to form a single sentence. However, different choices could have been made about what to group together into a single sentence.

For word boundaries, we have maintained the original source’s word boundaries in the Original Spelling line, but have altered them to our understanding of the structure of the words in the first line of the text entry. This is necessary in order to be able to parse the morphemes correctly. For example, Arroyo writes **men-koro** ‘your foot’ variably as “men coro” and “mencoro” in his *Gramática*. We have entered both as written in the Original Spelling line, but as **men-koro** in the first line of the text entry. In the Harrington handwritten material, it is often difficult to be sure whether Harrington intended a space between two symbols or not because of his erratic handwriting, and this is a further reason to base the line used for analysis on our understanding of the morphological structure.

For the possessive proclitics such as **kan-** ‘my,’ it is not completely sure that these are distinct from the independent subject pronouns such as **kan** ‘I.’ Mutsun can express possessive through a sequence of two nouns (e.g. **purcu rukka** ‘Purcu’s house,’ **kaknuh rukka** ‘chicken hawk’s house’), so **kan-rukka** ‘my house’ could be analyzed as **kan rukka**. However, Okrand analyzes these as proclitic possessive pronouns, and Harrington and other sources often write them as part of the word, suggesting they may form a phonological word with the noun. We have therefore maintained Okrand’s (1977) analysis of the clitic pronouns.

For words containing an apparent suffix or clitic **-t**, there is some ambiguity. We have analyzed these as one of 1) a clitic **=t** that adds a future meaning, and can be added to a wide range of words, 2) a suffix **-t** that adds a subjunctive meaning and can only be added to a few pronouns and demonstratives, as demonstrated in Arroyo's *Gramática*, to make words such as **kat** 'if I were to,' **makset** 'if we were to,' **nuniSSat** 'if that one were to,' or 3) part of a variant form of question words, since Arroyo often seemed to use alternative forms of question words that had a final /t/ (for example **hannit** instead of **hanni** 'where' with no apparent change in meaning, even where future tense is impossible). Additionally, there is 4) a suffix **-t** that marks a verb as being an imperative with a first-person object (e.g. **harat** 'give me'). This /t/ is easy to distinguish from the other three, and is documented by Okrand (1977) as well. Usages 1-3 are not completely sure. Given the wide range of meanings a subjunctive might cover, with "if you do verb" shading into "you will verb," and the rather approximate source authors' translations, it is possible that all the tokens we have parsed as usages 1-3 are actually the subjunctive. Historically, /t/ could be a shortened form of either **yete** 'later' (used to mark future, sometimes reduced to **et**) or **tukne** 'would that, if only.' However, since Arroyo's *Gramática* demonstrates this claimed subjunctive only on a very narrow range of words (a few pronouns), we have chosen not to analyze all ambiguous tokens as instances of it, and to maintain these three usages.

The suffix **-n(i)** is glossed as "mediopassive" in this text collection. This is the same term Okrand (1977) used for it. In the dictionary (Warner et al. 2016), we glossed this suffix as "inherent reflexive," but we have since decided that "mediopassive" is a better term.¹⁵ This suffix is often lexically specified. While it can attach to many verbs, there seem to be patterns of which verbs it can and cannot attach to, even within a single semantic type. As Okrand discusses, it most often attaches to verbs with meanings of change of state/inchoative, emotion, and motion along a path, and it can make a verb root intransitive. Analysis of **-n(i)** is beyond the scope of this introduction.

These and other points in the language material are amenable to alternative analyses. We do not present our analyses as the definitive answer about the morphological structures of Mutsun, and we do not think we have access to the necessary information to reach a definitive analysis on some points. Furthermore, we know there are errors in this document because it proved impossible to keep all inputting of all of such detailed material consistent throughout the project. We present the texts as accurately as possible with our analyses in order to allow for future research, and most of all to facilitate language learning for the Mutsun people.

Acknowledgements

A very large number of people contributed to the making of this set of documents. They are thanked in the Acknowledgments section of the introduction to the dictionary (Warner et al. 2016). Additionally, since the publication of the dictionary, we wish to thank Lillian Camarena, Zoe Tribur, Cheyenne Wing, Christina Newhall, Sam Prins, Kristine Chow, and Keith Alcock for their work on the project. We also wish to thank members of the Mutsun community for the feedback they have given on formatting for community usability.

¹⁵ Thanks to Heidi Harley and Marianne Mithun for discussion of usage of **-n(i)**.

This work was funded by a National Endowment for Humanities Preservation and Access grant (#PA-51356-05), and by a community scholarship grant from the Woodrow Wilson Foundation. We are very grateful for these programs.

References

Arroyo de la Cuesta, Felipe, d. 1842. Alfab[eticu]s rivulus obeundus exprimatio[nu]m causa horum indorum Mutsun misionis [sic] sancti Joann[is] B[aptist]ae exqui[si]tarum a Fr. Philipp[o] ab Ar[ro]yo de la Cuesta supradictae misionis indior[um] minist[ro], BANC MSS C-C 19, The Bancroft Library, University of California, Berkeley.

Arroyo de la Cuesta, Felipe. 1861. *Grammar of the Mutsun language*. *Shea's Library of American Linguistics* Vol. 4. New York: Cramoisy Press. Accessed at <http://www.gutenberg.org/ebooks/11081>, March 2022.

Arroyo de La Cuesta, Felipe. 1862. *A vocabulary or phrase book of the Mutsun language of Alta California*. *Shea's Library of American Linguistics* Vol. 8. New York: Cramoisy Press. Accessed at <https://archive.org/details/vocabularyorphra00arro>, March 2022.

Callaghan, Catherine. 2012. *Proto Utian Grammar and Dictionary*. (*Trends in Linguistics Documentation*.) Mouton de Gruyter.

Harrington, John Peabody. 1922, 1929-30. The Papers of John Peabody Harrington. National Anthropological Archives, Smithsonian Institution. Accessed at <https://sova.si.edu/details/NAA.1976-95#ref17247>, <https://sova.si.edu/details/NAA.1976-95#ref17248>, and <https://sova.si.edu/details/NAA.1976-95#ref17249>, March 2022.

Hewitt, John. Undated. MS 3950 Lists of Costanoan and Moquelumnan numerals collected by John N. Hewitt. NAA.MS3950. National Anthropological Archives, Smithsonian Institution. Accessed at <https://www.si.edu/object/archives/sova-naa-ms3950>, March 2022.

Levy, Richard. "Costanoan." *Handbook of North American Indians* 8 (1978): 485-495.

Mason, John Alden. 1916. John Alden Mason report on trips made to San Juan Bautista and Watsonville, California, CU-23.1 no.23, University Archives, The Bancroft Library, University of California, Berkeley.

Merriam, C. Hart. 1902. C. Hart Merriam Papers, BANC FILM 1022, The Bancroft Library, University of California, Berkeley.

Merriam, C. Hart. 1902. C. Hart Merriam Papers. Microfilm, Bancroft Library, University of California, Berkeley. Collection number: BANC MSS 80/18 c. Negative number: BNEG Box 1556 (Reels 69, 74).

Mills, Elaine L. 1981. *The Papers of John Peabody Harrington in the Smithsonian Institution 1907-1957, A Guide to the Field Notes*, vol. 2 (Northern and Central California). White Plains, NY: Kraus International Publications.

Okrand, Marc. 1977. *Mutsun Grammar*. Unpublished Ph.D. dissertation, University of California, Berkeley. Accessed at <https://www.proquest.com/docview/302868490>, March 2022.

Pinart, Alphonse Louis. Undated. *Lenguas Indigenas de la Baja California. Documentos y vocabularios extratos de varios autores publicados nuevamente.* WA MSS S-285, Box 1, Folder #6. Yale University, Beinecke Library.

Warner, Natasha, Lynnika Butler, and Quirina Luna-Costillas. 2006. "Making a dictionary for community use in language revitalization: The case of Mutsun." *International Journal of Lexicography* 19.3: 257-285.

Warner, Natasha, Butler, Lynnika, and Geary, Quirina. 2016. *Mutsun-English English-Mutsun Dictionary: mutsun-inkiS inkiS-mutsun riica pappel.* Language Documentation & Conservation Special Publication no. 11. Honolulu: University of Hawai'i Press. 672 pages.
<http://hdl.handle.net/10125/24679>, accessed March 2022.