

A Typologically Odd Phonological Reconstruction
for Proto-Sapotekan:
Stem-final *k

T e r r e n c e K a u f m a n

University of Pittsburgh

Aries 2000

revised Scorpius 2015

Copyright 2015 by Terrence Kaufman

The Sapotekan languages.

The Sapotekan language family has two branches, Sapoteko and Chatino. Sapotekan belongs to the Oto-Mangean stock, and its closest relative within Oto-Mangean is Masatekan (AKA Popolokan).

There are many Sapoteko languages: the number that would be recognized on purely structural linguistic criteria is not yet known, since not enough kinds of Sapoteko are well enough documented to answer this question.

Note: between 2007 and 2010 the PDLMA carried out a dialect survey of Sapoteko and Chatino, interviewing a speaker from every Sapoteko-speaking and Chatino-speaking community with a questionnaire of 1300 numbered entries, some with sub-entries. The results of this dialect survey are still being analyzed, but the question of the number of different Sapoteko languages will be answered on the basis of the data collected.

An extensive lexicon and a grammatical study were compiled for Central Sapoteko by Juan de Córdova during the second half of the 16th century [pub. 1578]. The lexicon was reworked during the colonial period. During the colonial and republican periods the documentation of Sapoteko was relatively sparse until the 1940s, when SIL workers began working on Atepec, Rincón, Villa Alta, Mitla, Juchitán, Coatlán, and Cuixtla Sapoteko. Since the 1940s more SIL linguists have worked on a large number of types of Sapoteko and Chatino. Some of their work has been published; more of it has not been published. Since the 1970s a number of academic linguists not affiliated with the SIL have worked on several Sapotekan languages. Since 1995 the PDLMA has been working on 11 Sapotekan languages with the aim of producing dictionaries of each of them, and to use the results of this research to reconstruct the lexicon of proto-Sapotekan, and to apply the results of this reconstruction to the decipherment of Sapoteko (Oaxacan) hieroglyphic writing.

Classification.

The following classification, based on published and unpublished materials, displays the languages being documented by the PDLMA. (Abbreviations in square brackets are from data gathered outside the PDLMA). All the major types of Sapotekoan language are believed to be represented.

Sapotekan family

Sapoteko branch/sub-family

Northern & Central Sapoteko

Northern Sapoteko

Ixtlán or Sierra de Juárez Sapoteko: Atepec
(ATP)

Rincón Sapoteko [RIN]

Villa Alta Sapoteko: Yatzachi (YTZ)
and Zoogocho (ZOO)

Choapan Sapoteko: Choapan (CHO)

Central Sapoteko

Isthmus Sapoteko: Juchitán (JCH), La Ventosa,
Xadani

Chichicapan Sapoteko: Chichicapan (CHI)

Córdova's Sapoteko (COR)

Mitla Sapoteko [MTL]

Ayoquesco Sapoteko [AYO]

Southern Sapoteko

Coatlán Sapoteko: Coatlán (COA), Loxicha

Cuixtla Sapoteko (CUI)

San Juan Mixtepec Sapoteko [SJ]

Papabuco Sapoteko

Zaniza Sapoteko (ZAN)

Texmelucan Sapoteko [TEX]

Western Sapoteko

Laxichío (LCH)

Chatino branch/sub-family

Zenzontepec Chatino (ZEN)

Yaitepec Chatino (YAI)

Structural characteristics of Sapotekan languages.

Phonology

Sapotekan

TAM markers are proclitics.

The vowel nucleus may be simple, long, checked, or broken.

Every syllable bears a tone.

Stems can have one syllable or two (clitics are usually monosyllabic).

Sapoteko

Words can be stressable or clitics.

Stressable stems are stressed on the penult or unique syllable.

Some languages have lost the vowels of proclitics, others have lost posttonic vowels, and others have lost both.

Syllables begin with a single or geminate consonant.

Except for those forms of Sapoteko that have lost posttonic vowels, every syllable ends in a vowel.

These seem to be the basic typological traits of Sapoteko phonology.

Chatino

The last syllable of a disyllabic stem is the most prominent.

Some languages have lost some of the vowels of non-final syllables.

Syllables begin with a single consonant.

Vowels may be oral or nasal.

Every syllable ends in a vowel.

These seem to be the basic typological traits of Chatino phonology.

Morphology

There are no suffixes.

There are frozen prefixes.

Verbs receive proclitics marking

valency: causative/more active and mediopassive/less active

movement: andative, venitive, and voltative

iterative

aspect and mood

Some nouns receive proclitics marking

possessed state

animacy

Numerals receive proclitics marking
cardinal
ordinal

There is a number of additional clitics (proclitics and especially enclitics) in any Sapotekan language.

\$

Comparative work achieved to date.

Sapoteko

Morris Swadesh 1947 (IJAL 13) [ca 100 reconstructions]

María Teresa Fernández de Miranda 1962? (1998?) [ca 500 reconstructions]

Jorge Suárez 1973 (IJAL 39) [ca 70 reconstructions]

Joe Benton 1986,1988 [ca 350 reconstructions]

Terrence Kaufman 1965-now [ca 600 reconstructions]

Chatino

Upson & Longacre 1964? [ca 200 reconstructions]

Sapotekan

Calvin Rensch 1966 (1976)

Terrence Kaufman (1965-now) [ca 300 => 700 reconstructions]

Oto-Mangean

Terrence Kaufman (1983-now) [comparative phonology, ca 400 reconstructions, the verb complex, Woerter und Sachen]

Proto-Sapotekan phonology.

[before the revision to be offered in this article]

c o n s o n a n t s					v o w e l s	
t	c	ty	k	kw	i	u?
	s	s^			e	o
l					a	
n						
		y		w		

It is not clear whether both /o/ and /u/ need to be reconstructed to proto-Sapotekan: most Sapotekan languages lack such a contrast, but a few have it.

s y l l a b l e c a n o n

\$(C)CV(:)(n)(7)T\$

/:/ vowel length

/n/ vowel nasality

/V7/ checked vowel

/V:7/ or /V"/ broken vowel

T tone [not yet reconstructed]

CC is a geminate consonant: whenever the underlying identity of the first C can be established, it is always and only /k/. /w/ and /y/ do not occur geminate: /k/ + /y/ yields /cc/

Some proclitics have the shape (C)V+, others (C)Vk+, others (C)Vy+

The problem.

Villa Alta Sapoteko has some words of the shape CVR and CVRw where other forms of Sapoteko have words of the shape CV; it has some words of the shape CVCR and CVCRw where other Sapoteko languages have words of the shape CVCV.

/R/ is a postvelar spirant, optionally voiced; /Rw/ is a labialized postvelar spirant, optionally voiced. In a handful of cases it occurs non-word-finally: its basic position is word-final.

Villa Alta words of shape CVC correspond to words of shape CVCV in other Sapoteko languages. Villa Alta words of shape CVCR and CVCRw also correspond to words of shape CVCV in other Sapoteko languages but these words have a distinctive vowel correspondence. For the moment, this set is labeled *CVCVx.

Correspondences

pSap	*CVCi	*CVCe	*CVCa	*CVCo	*CVCVk
ATP	CVCi	CVCe	CVCa	CVCu	CVC(i)a
Cs]	[ATP has -a after *ty, *c, and *s^, -ia after all other				
VA	CVC	CVC	CVC	CVC	CVCR, CVCRw
CHO	CVCi	CVCe	CVCa	CVCo	CVCo(g)
LCH	CVCi	CVCe	CVCa	CVCo	CVCa, CVCo ??
JCH	CVCi	CVCi	CVCa	CVCu	CVCe
CHI	CVCi	CVCe	CVCa	CVCu	CVCa
"COR"	CVCi	CVCi	CVCa	CVCo	CVCe

Swadesh reconstructed *CVCVx as *CVCia, because he had ATP but no VA data. FdM reconstructed *CVCVx as *CVCV-RV and *CVCV-Ru, depending on the Villa Alta reflexes, but in general she could not determine the vowel that preceded her *-RV and *-Ru.

I originally reconstructed enclitics *+xa and *+xu (but considered *..ax and *..ux) -- this would correlate with stem-final vowels in those languages that keep them, and with /R/ and /Rw/ in VA, which

drops them -- but three kinds of data forced me to reconsider this.

[1] CHO shows *..o* word finally, but *..og* whenever a vowel-initial enclitic follows. Since CHO generally does not lose stem-final vowels, these words must have an underlying shape with *../og/*.

[2] Manuscripts of almanacs embodying the precolumbian calendar and written in the latter part of the 17th century were confiscated by the Inquisition from several Northern Sapoteko towns. These almanacs are in a dialect akin to that of both CHO and VA, symbolized as "1700". It does not lose posttonic vowels, but it has words ending in <ig>, <eg>, <ag>, and <og>.

[3] Corresponding to Villa Alta words of shape CVCR and CVCRw, Chatino shows words of shape C(V)CV, with all the possible vowels in final position.

Already in the 80s, point [3] led me to consider reconstructing **..ix*, **..ex*, **..ax*, and **..ox*.

pSap	*CVCik	*CVCek	*CVCak	*CVCok
ATP	CVC(i)a	=>	=>	=>
	[ATP has -a after *ty, *c, and *\$, -ia after all other Cs]			
VA	CVCR	=>	=>	CVCRw
"1700"	CVCig	CVCeg	CVCag	CVCog
CHO	CVCo(g)	=>	=>	=>
LCH	CVCa	=>	=>	CVCo
JCH	CVCe	=>	=>	=>
CHI	CVCa	=>	=>	=>
"COR"	CVCe	=>	=>	=>
CHT	CVCi	CVCe	CVCa	CVCu

All these data indicate that earlier Northern Sapoteko had words ending in a lenis (voiced) velar consonant that could be preceded by any vowel. All other forms of Sapoteko, and CHO and ATP as well,

merged all vowels before this word final velar, which conditioned the resultant vowel quality, but differently for each language, then the final velar consonant was deleted.

The consonant that is responsible for these correspondences is no doubt *k. (In Sapoteko languages non-geminate consonants are lenis, and may become voiced or spirantized). Whenever a consonant can be identified before another consonant, it is /k/. *k is thus the only consonant that has an unusual distribution. Rules can be formulated

[1] to delete *k (or [g], [g̃], or [x]) after it has affected a previous vowel (CHO, JCH, COR, etc.);

[2] to change to [g] to [R] (or [Rw] after [o]) in absolute word-final position, and then remain distinct from word-final [g] that arises out of word-medial single *k (VA).

NOTE: it is possible that the fuller data being collected by the PDLMA may show that CHO and LCH have more than one reflex of stem final pSapoteko *..Vk. (Not all the lexical databases have been completed, and not all the available data have been canvassed.)

The correspondences that reflect proto-Sapotekan shapes like *CVk are not fully attested: there were apparently few such items in the language.

Monosyllables and apparent monosyllables ending in *k.

Monosyllables

*yak 'ir//to go': ATP /yia/, RIN /yeR/, CHO /yo/, VA /yeR/; CHI /ya/, MTL /y3/, AYO /ya7a/, JCH /ye/, COA /ya/; SJ /ye/; TEX /ya/; LCH /ya/; CHT /ya/

*e7k 'beber//to drink': ATP /7ya/, RIN /e7R/, CHO /7og/, VA /e7eR/; MTL /37/, JCH /e7/, COR <ye>

Apparent monosyllables with medial *y

*kweya7k 'moho//mold, huitlacoche//cornsmut': ATP /bia/, RIN /b@aR/, VA /beRw/; CHI /be/, AYO /b3w/ 'vapor//steam', MTL /kwa-b3h/ 'enmohecer//to get moldy', JCH /be/; TEX /me7e&/, SJ /mey/; CHT /kwaya7/ [NOTE: this set may be identical with the following: only in CHT are the two etyma distinct]

*kweyok 'nube//cloud, niebla//mist, neblina//fog' => 'nublado//cloudy': ATP /bia/, RIN /b@aR/, VA /beRw/, CHO /beo/; COR <pe(y)e>, CHI /be/, AYO /be/, MTL /b3h/, JCH /be/ 'nube en el ojo//cataract'; CUI /bew/, COA /m-bee7/, /gi be(e)7/; CHT /koo/

*kiyok 'lluvia//rain': ATP /iyya/, RIN /gy@R/, VA /yeRw/; AYO /yew/, CHI /jya/, MTL /nis gih/, JCH /nisa gye/, COR <nicza-quiie>; COA /gii7/, CUI /yuu/; SJ /gye7l/; TEX /gyey/; LCH /kiyu/; CHT /kyoo/

*kwe+ ciyok 'flauta de carrizo//reed flute': CHO /bjyo/, VA /bz^eRo/; CHI /bi-je:/ ~ /bi-c^iw/; LCH /c^i/

*keyek 'piedra//stone': ATP /iyya/, RIN /gy@R/, CHO /gyo/, VA /yeR/; MTL /gih/, AYO /g@/, JCH /gye/, COR <quiie>; CUI /kee/, COA /gee/; CHT /kee/

*ke7yek 'flor//flower': ATP /iyya/, RIN /y@R/, CHO /gyo/, VA /yeR/; CHI /jya7(a)/, MTL /gi/, AYO /g@7/, JCH /gye7/ COR <quiie ~ guiie>; CUI /ye7/, COA /ge7/, SJ /gye7/; TEX /gye7e/; LCH /kye/, CHT /kee/

Disyllables in final *k (supporting forms not cited)

*kwe+ ttyu:7s^s^ik 'tomate//mater'

[This appears as WWasteko /tu8ay/, CWasteko /tu8ey/, possibly via Chochoan as an intermediary]

*yikkek 'cabeza//head'

*yekkenk 'torcer//to twist'

*o+ s^o:7nak 'correr//to run'

*yakkwak 'chayote//vegetable pear'

*ke:7sak 'harina//flour'

*nittak 'caña//cane'

*kwe7nak 'lagarto//gator'

*lattak 'vaciar//to empty'

*asak 'bañar//to bathe'

*alak 'nacer//to be born'

*assak 'negro//black'

*kallak 'veinte//twenty'

*wicak 'pasado mañana//day after tomorrow'

*necok 'dar//to give'

*kwennok 'lodo//mud'

*k~t-itok 'agujerear//to make a hole in'

*accok 'reventar//to burst'

*k~t-iccok 'estallar//to shatter'

*k~t-is^ok 'pagar//to pay'

*las^ok 'salario//wage'

*kes^s^ok 'red//net'

Evidence that final *k is sometimes optional.

Sometimes, within a single etymology, some languages point to a final *k in their ancestor, while other languages point to an absolutely final vowel, and this requires at least two reconstructions for such etymologies, one with and one without a final *k. This indicates that sometimes stem-final *k (or *Vk with undeterminable V) was added to an existing stem ending in a vowel. Sometimes *ok was added to such a stem. What function or meaning might these (undoubtedly enclitic) elements ending in or consisting of *k have had? I imagine that deixis might be the most plausible phenomenon, since these are meanings that can be attached to virtually any type of constituent, but have no evidence to support this. Another possibility [p.c. Bill Poser] is 'only, just' or 'also'. In any case proto-Sapotekan forms with final *k include nouns, numerals, adjectives, adverbs, and verbs, so there is no functional category that prefers or facilitates the presence or adding of final *k.

*li7 'cercado//fence': ATP /li7/, RIN /l37/, CUI /b-li7/; MTL /bili7i/; TEX /li7/

*li7k : VA /le7eR/; CHI /le:7/, MTL /lo-le7/, AYO /ro-le7e/ 'patio', JCH /le7/, COR <lee ~ leeya ~ leye>

*te 'ceniza//ashes, polvo//dust': ATP /te/, RIN /d@/, VA /de/

*ti: : CUI /dii/, COA /dii/; LCH /ti/; CHT /hi:/

*tek : COR <te(e)>, CHI /de/, JCH /de/, AYO /d@/, MTL /d3h/

*kwe7 'cangrejo//crab': ATP /be7 citta/, RIN /b@c^u b@7/, VA /c^io be7/, CHO /be7/; LCH /z^uka be/; CHT /kwe7/

*kwek : JCH /bic^u be/ 'caracol//snail', COR <mani pee>, CHI /b@/

=== **ADD *..ok**

*kwetya 'pavo//turkey': ATP /bera/, RIN /b@ra/

[This is borrowed as WHuastec /kwita7/, CHuastec /pi:ta7/]

*kwetyok : CHO /bero/, VA /beRw/; JCH /bere/; CHT /kwetu/

*ala 'abierto//open': CHT /la/

*alok : ATP /alia/, RIN /alR/, VA /alRw/; CHI /ala/, JCH /ale7/

*kwela 'estrella//star': CHT /kwela/

*kwelok : ATP /belia/, RIN /b@Rla/, CHO /belo/, VA /beLRw/; CHI /bela/; COR <pelle>; JCH /belle gi/; LCH /bela:/

*kas^s^a 'cerca//near': ATP /es^s^a/; JCH /gas^a/, AYO /gaz^a7a/

*kas^s^ok : CHI /gas^u/; LCH /as^u/

=== **ADD *..k to *..a**

*tilla `pelear//to fight': ATP /tilla/, RIN /dila/, VA /dill/
*tillak : JCH /dinde/

*a:7na `arar//to plow': ATP /a7na/, VA /a7ann/, JCH /aana/ `rozar'
*a:7nak `escarbar//to dig': JCH /g~d-aanye/

=== **ADD *..k to *..i**

*kwetti `zorrillo//skunk': ATP /bette/, RIN /b3te/, VA /bet/, CHO /b3ti/; LCH /s^ibiti/; CHT /kwih/;
*kwettik : JCH /be7te/, COR <pe(e)te>; CHI /be7(e)ta/

*kwe+ ya7kki `cuervo//crow': JCH /bya7ki/, CHI /byac^i/
*kwekkik : ATP /bekkia/, VA /bec^R/; LCH /biyeka/

*k~t-issi `distribuir//to deal out': RIN /g~d-isi/, VA /y~d-is/; JCH /giizi/
*k~t-issik : ATP /g~t-i88ia/

*latti `lugar//place': ATP /lati/
*la(t)tik : RIN /latRa/, VA /latR@/; JCH /lade/

*kwe+ s^s^osi `padre//father': RIN /bi-s^uz/, VA /b-s^oz/, CHO /ps^uzi/; LCH /z^uci/; CHT /suti/
*kwe+ s^s^osik : ATP /be-s^u6ia/; JCH /bi-s^oze/, COR <pixoze>, CHI /bi-s^oza/

*lassi `delgado//thin': ATP /la88i/, RIN /lasi/, CHO /lasi/, VA /las/; LCH /lece/; CHT /lati/
*lassik : JCH /na-lasse7/, CHI /lasa/

With the reconstruction of stem-final *k to proto-Sapotekan, it is necessary to revise the statement of its syllable structure.

The following formulas represent shapes that can be directly reconstructed: that is, they are surface representations in the proto-language.

monosyllable:

\$(k)CV(:)(n)(7)T(k)\$

disyllable:

\$(k)CV(:)(n)(7)T\$(k)CV(:)(n)(7)T(k)\$

Conceivably, there might have been underlying combinations of **\$(k)CV(:)(n)(7)T(k)\$** with **\$(k)CV(:)(n)(7)T(k)\$**. These would presumably yield the "disyllable" formula given above, by simplification of //VkkCV// to /VkCV/

\$

Certain sound changes from Proto-Sapotekan to Chatino.

[1] *t => *8 => *h

[2a] *s => *8 => *t

[2b] *ty => t

[3] *s^ => *s

loss of *k in geminates and at end of syllable

Sound changes from proto-Sapotekan to proto-Sapoteko.

[these are all due to foreign "civilized" influence, from Mije-Sokean and/or Matlatzinkan]

shift of stress to penult syllable of disyllabic stems [M-S]

loss of vowel nasality (after *an => *ǎ) [M-S, Mtl]

drift of single *kw to [p] [M-S]

addition of *mm through loans [M-S,Mtl]

The Typological Oddity and its Possible Explanation.

Most Sapotekan languages do not have syllables that end in consonants. Nor do Masatekan (aka Popolokan), Mistekan, Amusgo, Tlapanekan, or Chorotegan. Reconstructed proto-Oto-Mangean has syllables of the shape

\$(H) (n) (y) CV(n) (H) T\$

T stands for "tone"

H stands for *ʔ or *h

preconsonantal *ʔ yields glottalized segments in some daughters

preconsonantal *h yields aspirated segments in some daughters

preconsonantal *n prenasalizes

preconsonantal *y palatalizes

postvocalic *n nasalizes

There is something that might explain the typologically odd presence of preconsonantal and syllable-final *k in proto-Sapotekan. Prevoalcalic pOM *x and *h both change to *k in proto-Sapotekan. If preconsonantal and postvocalic *h survived in Sapotekan, instead of being deleted or changed to *ʔ or vowel length, then they would show up as *k. This regular sound change yielded a typological oddity which has persisted from 1500 BCE (when the various Oto-Manguen branch ancestors had differentiated themselves) until 1700 CE in some areas, but started being eliminated as soon as Chatino separated from Sapoteko around 500 BCE.

\$

ASCII transcription

- /c/ [ts]
- /c^/ "che"
- /s^/ "esh"
- /z^/ "zhe"
- /@/ "barred i"
- /6/ "edh"
- /8/ "theta"
- /kw/ "labialized k"
- /ty/ "palatalized t"
- /R/ "uvular r"
- /Rw/ "labialized uvular r"
- /3/ "aesc"
- /j/ "jay"
- /7/ glottal stop

- /+/ marks off a clitic

ACKNOWLEDGEMENTS.

Conversations with Paula Radetzky have helped me to improve the organization and content of this article.

The data in CHO that show that some stems end in a final velar consonant that is deleted before silence were brought to my attention by Paula Radetzky in June 1999.

This paper was delivered at the Eighth Spring Workshop on Theory and Method in Linguistic Reconstruction on Friday 31 March 2000.

Ives Goddard suggested that final /k/ may have become [x] before influencing the pronunciation of adjacent vowels, since velar and postvelar spirants can more readily influence the quality of adjacent vowels than a velar stop can. This is conceivable, but it would entail /k/ acquiring the articulation [ç] or [x] in final position even in languages that did not voice single /k/ intervocalically by 1500, since, for example, final *.Vk became /e/ by 1500 in Córdoba's Valley Sapoteko, while single medial /k/ remained [k].

Bill Poser suggested that additional meanings of clitic *(o)k may have been 'only, just' or 'also'.