

Daniel Campos

12/07/2019

LING 550

### Project Part B: KZ-62(Mambay) and Spanish

#### Language Background and information

#### Mystery Language KZ-62(Mambay)

Mambay is a language spoken primarily in Cameroon and Chad. The language is spoken by approximately 15,000 people (Anonby, 2006). 80% of Mambay speakers live in Northern Cameroon and the remaining 20% live across the border in southern Chad. Within the language there are no known dialects, so the language does not feature much variation in its phonetic structure. Mambay is considered a 6a language; meaning all children in the community actively learn the language (Eberhard, 2019). Despite historical drops in population due to ethnic conflicts and expansion of larger regional languages the vitality of Mambay is strong and it is considered a vigorous language. Like many North Central African languages, speakers of Mambay tend to be multilingual especially men. Since Mambay is not a broadly spoken language but is also central to the identity of community that speaks it, the majority of speakers are L1 speakers. Many Speakers of Mambay learn other L2 languages such as: Arabic, French, or Hausa. Mambay is part of the Kebi-Benue(Group 6) group withing the Adamawa branch of Niger-Congo (Anonby, 2006). Like many North Central African languages, there is no highly accurate ASR engine which can transcribe or understand spoken Mambay. There are not many L2 learners of Mambay as many L1 speakers are multilingual but there are small groups of L2 speakers both in the form of researchers and people in Chad and Cameroon.

## Mexico City Spanish

Spanish is the second most common L1 language after Mandarin (F. Moreno-Fernández, 2007).

Around the world over 534 million (Eberhard, 2019) people speak Spanish. Spanish is the primary language of many countries like Mexico, Spain, and Cuba. Spanish is a Gallo-Iberian language which is part of the wider more commonly known Romance languages (Eberhard, 2019). Mexican Spanish (MS) is a dialect of the broader Castilian Spanish. Mexican Spanish is spoken by over 127 million people, serves as the first language for 88% of the population (Avelino, 2017) and is the de facto national language of Mexico. Mexico City Spanish is a specified dialect of Spanish representing some of the language variations present in the > 20 million people who live in Mexico City. Spanish, and by inheritance Mexico City Spanish (MCS) is considered a fully developed language and is the primary form of media publication in Mexico City. Spanish is considered a global language and as such it is both a common L1 and L2 language. Spanish is such a popular L2 language that in many countries, like the United States of America where roughly 10% of students are L2 speakers of Spanish to some degree (Mitchell, 2017). Unlike Mambay there are many ASR engines that work on Spanish moreover there are even ASR systems developed specifically for MS (Herrera-Camacho, 2017). Broad ASR for general Spanish are commonly used everywhere thanks to popular software products like Google and Siri. For MS over 10 dialect specific corpora exist which have been combined with cutting edge systems such as HTK and Kaladi to produce a Word Error rate of 44% (Herrera-Camacho, 2017). Unfortunately, it is hard to ascertain how many L1 or L2 Spanish speakers are L1 or L2 Mambay speakers but one could posit that because of the distance, speaker size and linguistic differences there are few to none.

## Spanish Consonants

	Bilabial	Labio-dental	Denti-alveolar	Post-alveolar	Velar
Plosive	p b		t d		k g
Affricate			(ts) (tʃ)	(tʃ)	
Nasal	m		n	ɲ	
Fricative		f	s	(ʃ) ʝ	x
Lateral			l		
Flap			ɾ		
Trill			r		
Approximant				j	w

Figure 1:Linguistic Literature's Mexico City Spanish

	Bilabial	Labiodental	Dental	Alveolar	Palatal	Velar
Plosive	p b		t d			k g
Affricate					tʃ ʝ	
Nasal	m			n	ɲ	
Tap or flap				ɾ		
Trill				r		
Fricative		f	θ	s		x
Lateral approximant				l	ʎ	

Figure 2:Linguistic Literature's Castilian Spanish

	Bilabial	LabioDental	Dental	Alveolar	Palato-alveolar	Palatal	Velar
Plosive/Stop	p b			t d			k g
Nasal	m			n	ɲ		
Flap				ɾ			
Trill/Tap				r			
Fricative		f		s, z	ʃ		
Lateral Fricative							
Approximate					j		ɥ
Lateral Approximate				l	ʎ		

Figure 3: My Mexico City Spanish

consonant	example word	Transcription	English meaning
p	peso	peso	weight
b	beso	beso	kiss
t	tos	tos	cough
d	dos	dos	two
s	seso	seso	brain
k	casa	kasa	house
g	gafa	gáfa	glasses
m	mexico	me:jico	mexico
n	nunca	nunka	never
ɲ	ni	nijo	boy
r	carro	caro	car
ɾ	caro	caro	expensive
f	freno	freno	brake
s	suero	suero	I.V.
j	ya	ja	yes
ɥ	oaxaca	ɥuahaca	Oaxaca
l	lago	lago	lake
ʎ	llegue	ʎegue	arrived

Figure 4: My Examples of consonants

Consonants Mambay

	Bilabial		Labio-dental	Alveolar		Retroflex	Palatal	Pre-glottalized palatal	Velar	Labial-velar	Pre-glottalized labial-velar	Glottal
Plosive	p	b		t	d				k	g	kp	gb
Implosive	ɓ					ɗ						
Nasal	m			n					ŋ			
Flap	ɣ					ɾ						
Fricative			f	v	s	z						h
Approximant							j	ɟ		w	ʋ	
Lateral approximant				l								

Figure 5: Linguistic Literature's Mambay

	Bilabial	LabioDental	Dental	Alveolar	Palato-alv.	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive/Stop	p b			t d		c	k g			
Nasal	m			n						
Trill/Tap				r						
Fricative		f v		s z					ħ	h
Lateral Fricative										
Approximate				ɹ						
Lateral Approximate						ʎ				

Figure 6: My Mambay

ID	Word	Transcript	ID	Word	Transcript	ID	Word	Transcript	ID	Word	Transcript
1	Field	pa.ra'	21	Pressure	Aahana	41	Child	biga	61	Nose	wo
2	Elephant	ba.la'	22	Skin	wa.ga	42	My	m	62	Split	wo
3	Fish-sp	bau.bau	23	Crack-n	wa.gã	43	My child	bigm	63	Burn	la
4	Flour	ta.bau	24	Mud shelt	la.ga	44	Female in-	azi	64	Lift	na
5	Mouth	dada	25	Gift	maba	45	Female in-t	azi	65	Fish-sp	nateyɔgy
6	Squirrel	nan.ga	26	Filth	paga	46	Bean	azi	66	Cat-sp	nakeiyɔgy
7	Hare	gabla	27	Ibis-sp	bana	47	Bean leaf	na:a	67	Abundant	bazam
8	Pair of	ba	28	Truth	mana	48	Bean eate	la:a	68	Yellow	ba
9	Axe	cala	29	Fish-sp	maza	49	Frog-sp	pu.a:ani	69	Scorpion	zaba
10	Distant	batga	30	Daughter	fa.zi	50	Meet	bak	70	My scorpion	zakm
11	Trap-n	maba	31	To hide	vuana	51	Respect	ma	71	Child	bi.ga
12	Cloud	nana	32	Circumcisi	ca:da	52	Find	dada	72	My child	bi.gm
13	Current	alba	33	Crab	halaga	53	Leak	na	73	Smallness	baga
14	Spread	rana	34	News	p <sup>h</sup> amuda	54	Hug	rak	74	Milk	pap.ba
15	Back	fala	35	Circumcisi	ca:da	55	Be blind	ara	75	My milk	pap.m
16	Grass-	svala	36	Squirrel	nan.ga	56	Existential	to	76	squirrel	ganga
17	Tail	saba	37	Fog	hoboa	57	Negation	niai	77	Food	huua
18	Scorpion	zaba	38	Hurt	kiak	58	Press	iaha	78	My food	hualm
19	Wasp-	hamza	39	Body	inu	59	Call	taha	79	To carry	rink
20	Life	ian.ga	40	Hurt the b	kiakinu	60	Fig	wa	80	To go up	he.na
									81	To heal	bar.na

Figure 7: Mambay Transcriptions

Vowels Spanish

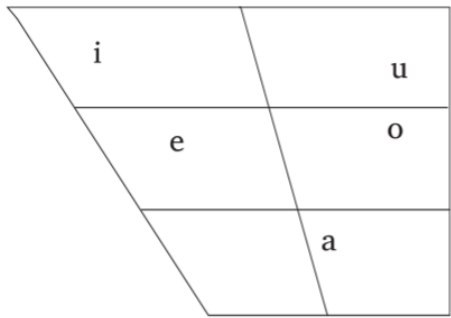


Figure 8: Linguistic Literature's Vowels

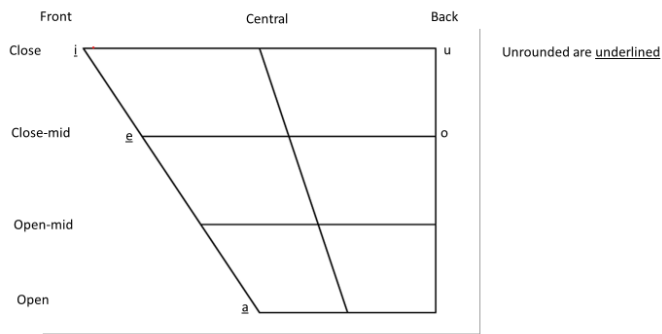


Figure 9: My Vowels

Vowels Mambay

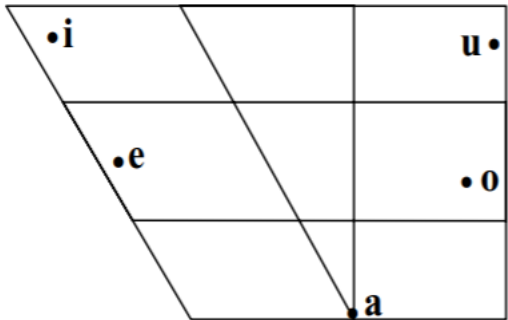


Figure 10: Linguistic Literature's Vowels

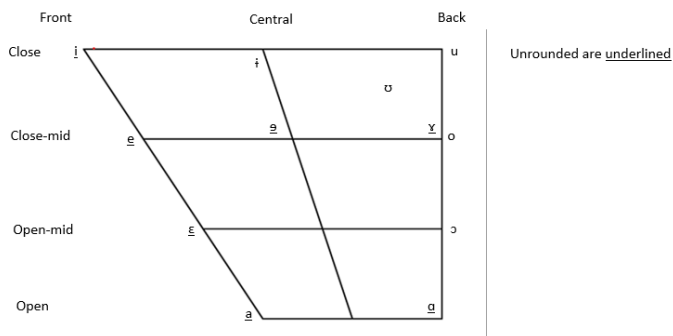


Figure 11: My Vowels

ID	Word	Transcript ID	Word	Transcript ID	Word	Transcript ID	Word	Transcript ID	Word	Transcript				
1	palm or sc	di	11	bottom	ni.nu	21	melt	re	31	hand of	se	41	polish	ro.no
2	burn	de	12	slave	be.ge	22	spread ou	ra	32	cow of	za	42	glue	dni
3	mouth of	da	13	pangolin	be.la	23	amuse	ro	33	snake of	ðä	43	rot-v	dna
4	drink	do	14	elephant	bala	24	leave with	ru	34	dip	di	44	lean	gnu
5	them	du	15	captive	balo	25	cobra	dyno	35	spy	bɛ	45	name-n	ia
6	fish-sp	p'ai	16	money	co.bo	26	manioc-sp	bu.ni	36	fil	bap	46	name-of	ee
7	virginity	huai	17	relative	coo.la	27	millet-sp	bvo.na	37	ugliness	bɔ	47	your snar	eeam
8	frog-sp	kau	18	magic	sula	28	clean out	(ri	38	gather	bu			
9	grass-sp	deu	19	herd	zulɔ	29	blind or si	ra	39	slither	re			
10	eve	ni.nu	20	carry	ri	30	slip somet	dɔ	40	tree-sp	rega			

Figure 12: Mambay Vowel Transcription

## Tone

ID	Phrase	Transcript ID	Phrase	Transcript ID	Phrase	Transcript ID	Phrase	Transcript	
1	head	ka.lâ	6	my breast	nù.nũ	11	you move	moia	
2	my head	ká.li.m	7	drumstick	ga.gam	12	you finish	moíâ	
3	axe	ka.lă	8	my drums	ga.gaŋ	13	you move	moiâ	
4	my axe	kă.li.m	9	to wash	sua.na	14	you finish	moîâ	
5	breast	nú.nũ	10	wash n or	súa.na	15	move!	môia	
							20	my brother	bíc.turutin

Figure 13: Mambay Tone Transcriptions

## My Analysis vs. Published

After relistening to the audio samples and thinking about what was presented in class since part

A of the project I have not provided updates to my transcriptions for either MCS or Mambay.

## Spanish

When my inventory is compared with the analysis presented in the literature there is a high

overlap. Our vowel inventories match absolutely but our consonants have marked differences

which I will explore in depth.

My analysis lacks the Affricative Palatal and Dental-alveolar as they were not items we covered much in class. Additionally my analysis features an additional fricative alveolar(z), lacks a velar fricative(x) and includes a Palatal Lateral approximate (ʎ). All of these difference have to do with how I perceive the language and are likely influenced by my high degree of exposure to English. In my understanding of MCS there is no x because in every word that may feature it, the pronunciation is instead of a J (e.g. Mexico/Mejico, Xavier/Javier). This is such a common substitution that in MS it is widely accepted that many words that is spelled with a x can be spelled with a j. The words that are the exception to this rule like maximo are usually pronounced with a alveolar fricative s (macsimo). My inventory features a fricative alveolar(z) which is likely due to me and my speaker group in Mexico city being multi lingual. In MCS it is common to replace the z altogether s or x. This alteration can be commonly viewed in how L1 MCS speakers use a z in their L2 language. Back in Mexico, most of my family pronounces my wife's name(Zoe) as either soe or xoe while MCS speakers where MCS is their L2 more commonly pronounce it as the Greek derived zoe. A similar phenomenon can be found with bilabial stops(b) and labiodental fricatives(v). Since MCS and MS do not include v in their inventory, when L1 speakers are interacting with other language with a marked use of vs native MCS speakers will pronounce vs as bs. This can be understood with pronunciation of words such as Volvo(Mexican Spanish pronunciation would be bolbo) or Venezuela(benezuela). The final minor detail in inventories is around the palatal lateral approximate ʎ. In the literature words that in my analysis of MCS include a ʎ are represented with a j(e.g. yeso(meaning cast) is head by the literature as jeso (Avelino, 2017), while I hear it as ʎeso).

## Mambay

Unlike my Spanish inventories, there is a high degree in variability in my inventory of Mambay and what is presented in literature. The root of this is in my background in hearing foreign sounds and understanding of tone. As is visible in my vowel charts I believe there is a large variety in vowels in Mambay which the literature does not support. In the literature, Mambay only includes 5 core vowels(a,e,i ,o,u) while to my ear there was a broader range featuring 13. Looking at waveforms did not drive me to unify my transcription with the literature because in spectrograms, the vowels in question matched the profile of the vowels in my inventory. I do not have a broad exposure to tonality so it is very possible that I am conflating tonal vowels with other vowels.

The biggest divergence between the literature and my assessment is around tonality. Tonality is something new to me which makes it hard for me to hear and thus my transcriptions of tonal phrases lack the specificity presented in the literature. When I was transcribing, having examples that clearly indicated differences cause by tonality(such as breast and my breast) helped me understand and convey the tonal structure of Mambay and with a larger sample I am certain I could have heard and understood the broader tonality with a near overlap to the literatures.

When I contrast the literature's consonant inventory with mind I find they have a higher degree of specificity likely gained with further experience in transcription. My transcriptions are lacking Labial-velar approximate(w), Palatal Approximate(j) and nasal velar(ŋ). When listening and analyzing the sample's spectrograms I was unable to hear where the literature gained these interpretations. The literature also features a bilabial flap (ɸ) which my inventory instead



hears as a palatal lateral approximate( $\lambda$ ) which is likely because of different L1 and L2 which the author and I may share.

## Conclusion

In studying the phonemic and phonetic properties of a mystery language (Mambay) and documenting and analyzing a language I speak (Mexico City Spanish) I have broadened my understanding of the underlying structure of spoken language. I really struggled with transcription because with a language as foreign as Mambay, my brain was constantly trying to project sounds into a semantic space I knew and based transcriptions on English/Spanish. The area I struggled with most was the tonal aspect of Mambay. Since both English and Spanish have minimal use of tonal features, tone is not a concept my brain or ear are familiar with. When audio samples had some form of tonality, I found myself listening to the samples 50-60 times, looking at spectrograms, and trying to reproduce the sounds all while still feeling I was not understanding the sound properly. In the transcriptions I grew very amazed by the tonality of Mambay because it seemed to denote some concept of ownership. The idea that changing how a word is said as a way to denote ownership and distinction between similar words was very foreign to me but once I experienced it I wondered why it's not present in English or Spanish. Tonality seems like an incredible way of compressing information while maintaining fidelity. During my transcriptions, I tried to leverage spectrograms but, in most cases, understanding the  $f_1$  and  $f_2$  of a word did not make it easier for me to transcribe a word. In this project it was satisfying once I had listened to the audio samples a few times because my transcriptions stabilized. After the 4<sup>th</sup> listen I had few transcriptions which I changed from listen to listen. When I reached this point, I was elated because I was producing transcriptions of an

unknown language with consistency! After reading through all my transcriptions I was able to assemble a heuristic which could describe the sound system of a language which would generalize to unseen samples. This project satisfied the goals of this project extremely well: listen to new sounds, describe a sound system, compare sound systems, and combine all this into linguistic literature. Finally, and most importantly, in transcribing a language which has so little genetic nature shared with my L1, and L2 I have been able to understand what sounds and combinations are common and shared across languages and which are unique. In thinking about shared noises, I generated my own hypothesis on why some sounds are common and others aren't.

#### References

- Anonby, E. J. (2006). Mambay. *Journal of the International Phonetic Association*, 36.
- Avelino, H. (2017). Mexico City Spanish. *Illustrationf of the IPA*, 223-230.
- Eberhard, D. M. (2019). *Ethnologue: Languages of the World. Twenty-second edition. Dallas, Texas: SIL International*. N/A: Ethnologue.
- Eugenio Mart'inez-Celdran, E. M.-C.-S. (n.d.). Castilian Spanish. *Illustrations of the IPA*.
- F. Moreno-Fernández, J. O. (2007). Atlas de la lengua espanola en el mundo. *Real Instituto Elcano-Instituto Cervantes-Fundación Telefónica*.
- Herrera-Camacho, C. D.-M.-R. (2017). Automatic speech recognizers for Mexican Spanish and its open resources. *Journal of Applied Research and Technology*, 259-270.
- Mitchell, C. (2017). *How Muich Foreign Language Is Being Taught in U.S. Schools*. American Academy of Arts and Sciences Report.