

Advancing the diachronic & typological study of tone with comparative lexical databases:

A case study of 5 Mixtec varieties

SANDRA AUDERSET^{1,2}, CHRISTIAN D. BRENDL¹,
ERIC W. CAMPBELL¹, SIMON L. PETERS¹,
KEVIN SCHAEFER¹ & ALBERT VENTAYOL-BOADA¹

1 UNIVERSITY OF CALIFORNIA, SANTA BARBARA

2 MPI FOR THE SCIENCE OF HUMAN HISTORY, JENA

Introduction

Motivations for this study:

- studies of tone are underrepresented in linguistic research
- tone change is poorly understood
- computer-assisted methods can help, provided that the data to analyze includes robust representation of tones

Goals of this study:

- adaptions to conventional database design to handle tone
 - to advance studies of tone
 - to advance understanding of tone change
 - support needs of multiple audiences

Roadmap

1. Studying tone: issues and challenges
2. An introduction to Mixtec languages in Mexico and in the California diaspora
3. Data collection
4. Database uses: documentation, community, language history
5. Database design

Thanks

We would like to thank our collaborators in Ventura County, California, for their continued support and for allowing us to share their data here today:

- ❖ Griselda Reyes Basurto
- ❖ Carmen Hernández Martínez
- ❖ Gabriel Mendoza
- ❖ Juvenal Solano
- ❖ Yésica Ramírez

The study of tone

- approx. 60-70% of the world's languages (Yip 2002: 1; cf. Maddieson 2013) are tonal (i.e. make use of pitch difference for distinguishing lexical and/or grammatical categories)
- but these languages are still underrepresented in linguistic research, especially in historical linguistics (Campbell in press; Janda & Joseph 2003: 173)
- if we want to understand human language and how it works, we cannot exclude tone (Hyman 2018)
- computer-assisted approaches in historical linguistics have gained traction in recent years
 - integration of tonal data into the database design requires theoretical and practical consideration of several key issues in tonal analysis

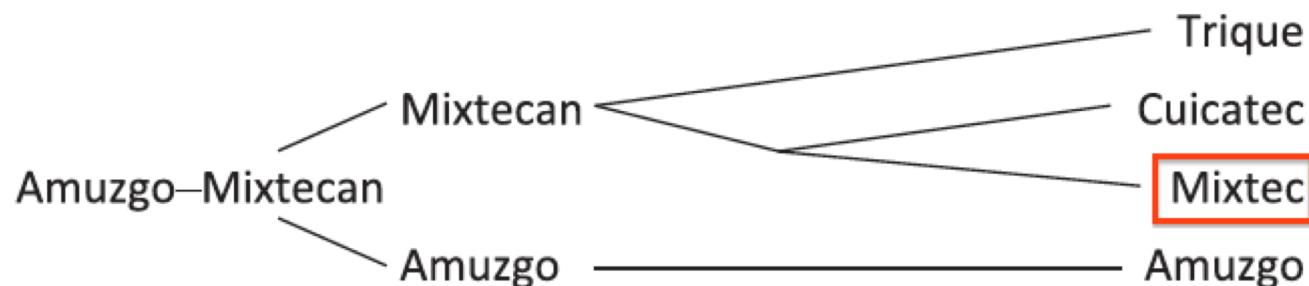
Challenges in the study of tone

- analytical challenges:
 - tonemes (Strecker 1979; Greenberg 1948) vs. tone melodies (Dürr 1987; Cruz & Woodbury 2014a)
 - comparability of tone correspondences across varieties
 - grammatical tone
 - comparability of tone and segments
 - floating tones and tone sandhi

- representation:
 - number notation vs. diacritics vs. IPA bar notation
 - consistent Unicode representation of tone diacritics

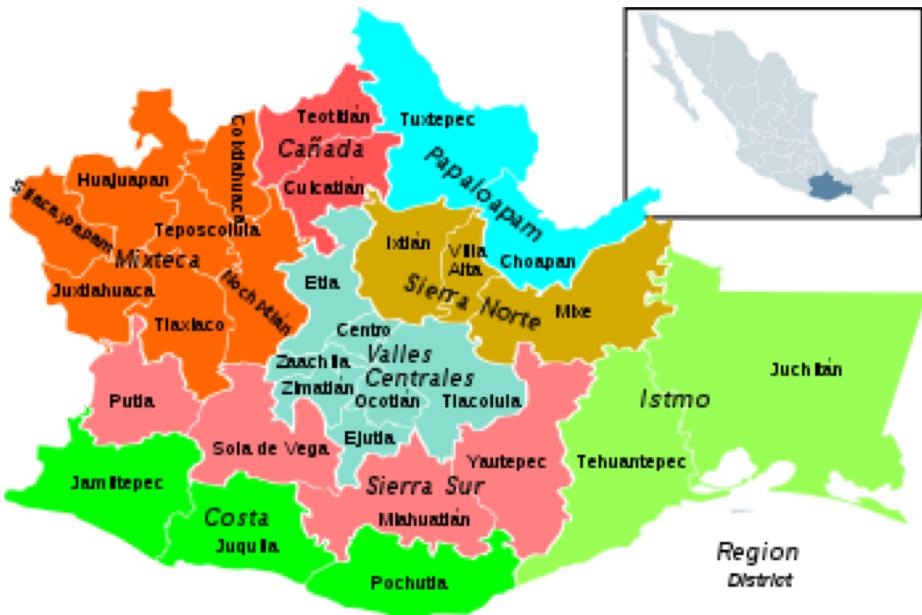
A brief introduction to Mixtec

- Otomanguean > Eastern Otomanguean > Amuzgo-Mixtecan > Mixtecan
- Mixtec language group
 - 12 primary subgroups (Josserand 1983)
 - 52 languages (Egland 1983; Simons & Fennig 2017)
 - 81 varieties (INALI 2008)
- complex systems of grammatical and lexical tone
- tone-bearing unit is the mora



Amuzgo-Mixtecan branch of the Otomanguean stock (adapted from Campbell 2017:3)

The speech communities



Oaxaca, Southern Mexico

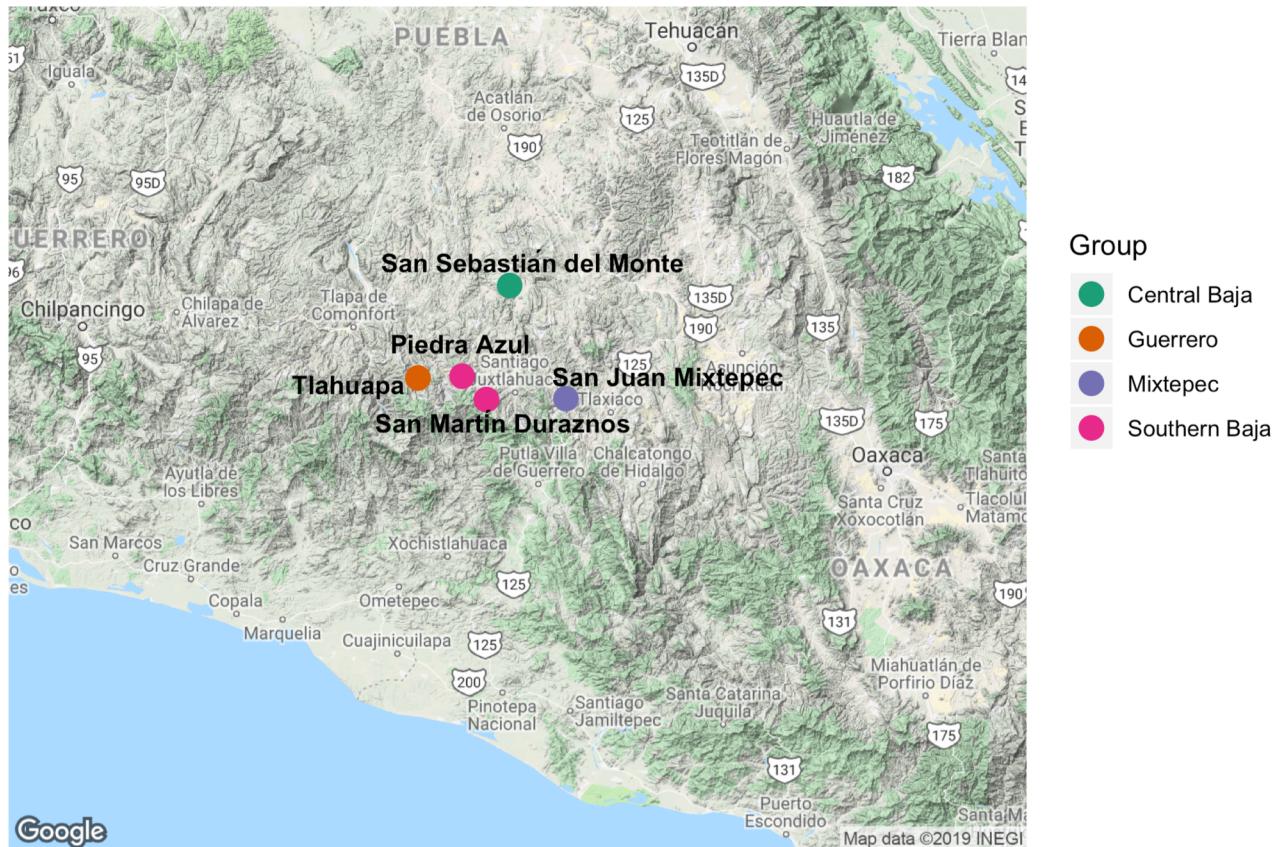


Ventura County, California,
USA

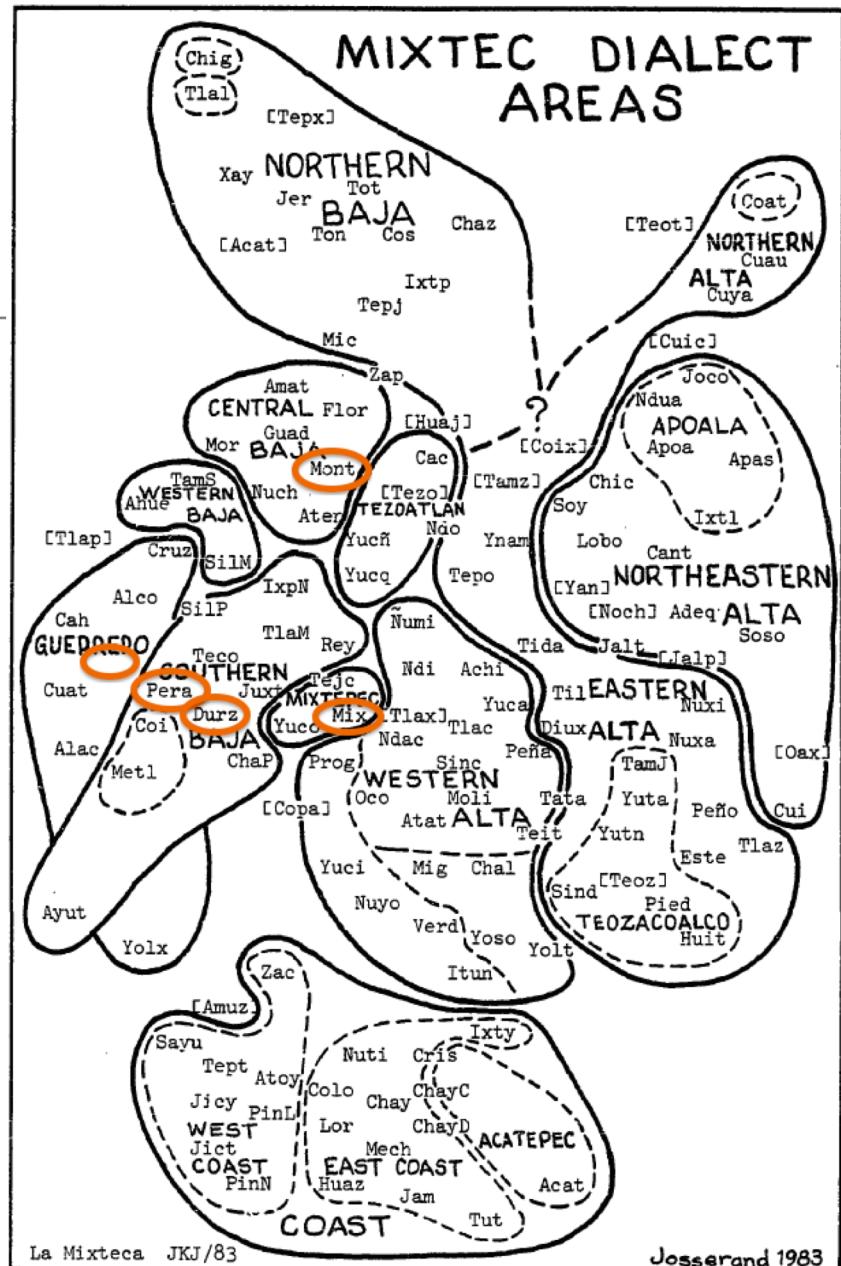
The 5 varieties in this study

Team member	Location in MX Village (Municipality, State)	Dialect area (Josserand 1983)	Endonym
Griselda Reyes Basurto	Tlahuapa (Alcozauca, Guerrero)	Guerrero	Tù'un Sàví
Gabriel Mendoza	Piedra Azul (San Martín Peras, Oaxaca)	Southern Baja	Tù'un Ntá'ví
Carmen Hernández Martínez	San Martín Duraznos (San Sebastián Tecomaxtlahuaca, Oaxaca)	Southern Baja	Tù'un Ntá'ví
Juvenal Solano	San Sebastián del Monte (Santo Domingo Tonalá, Oaxaca)	Central Baja	Tù'un Nta'ví
Yésica Ramírez	San Juan Mixtepec (San Juan Mixtepec, Oaxaca)	Mixtepec	Tù'un Ndá'ví

Geographical location of the 5 varieties



Subgroup membership



Manual data collection

- community-based documentation (Cruz & Woodbury 2014b, Czaykowska-Higgins 2009) in Oxnard, California through partnership with the Mixteco/Indígena Community Organizing Project (MICOP)
- entries in practical orthographies, which were developed together with speakers



Manual data collection

For each variety:

1. build up a basic vocabulary list
2. add a column in the DB
3. segmental phonemic analysis
4. tone analysis: tone groups (Cruz & Woodbury 2014a)
5. practical orthography design, update database
6. verb inflection database

Example lexical sets

	Tlahuapa (Guerrero)	Piedra Azul, SMP (S. Baja)	S.M. Duraznos (S. Baja)	S.S.d. Monte (C. Baja)	S.J. Mixtepec (Mixtepec)
'sun'	<i>tkandìí</i>	<i>tsìjkantsijȋ</i>	<i>kuàntxií</i>	<i>ndikàndiì</i>	<i>nikanchîí</i>
'water'	<i>tikwíí</i>	<i>tsìkuíí</i>	<i>txìkuíí</i>	<i>tikuií</i>	<i>chukwíí</i>
'thread'	<i>yu'và</i>	<i>yi'và</i>	<i>yò'và</i>	<i>yìvà</i>	<i>yùvà</i>
'smoke'	<i>yù'mà</i>	<i>i'màn</i>	<i>ñù'mà</i>	<i>ñìmà</i>	<i>ñu'mà</i>
'neck'	<i>sùkú</i>	<i>sojkún</i>	<i>sùkùn</i>	<i>jikò</i>	<i>sùkú</i>
'tripe'	<i>siti</i>	<i>xitsâ̄n</i>	<i>xitxi</i>	<i>sutì</i>	<i>tsitì</i>
'green bean'	<i>ndìchi</i>	<i>nchìchî</i>	<i>nchìchi</i>	<i>ndìchi</i>	<i>nchichî</i>
'corn'	<i>nùní</i>	<i>nùjnî̑</i>	<i>nùnì</i>	<i>nònì</i>	<i>nùnî̑</i>
'hard'	<i>te'é</i>	<i>tsié'é</i>	<i>txé'é</i>	<i>kásí</i>	<i>kátsi</i>

Computer-generated data

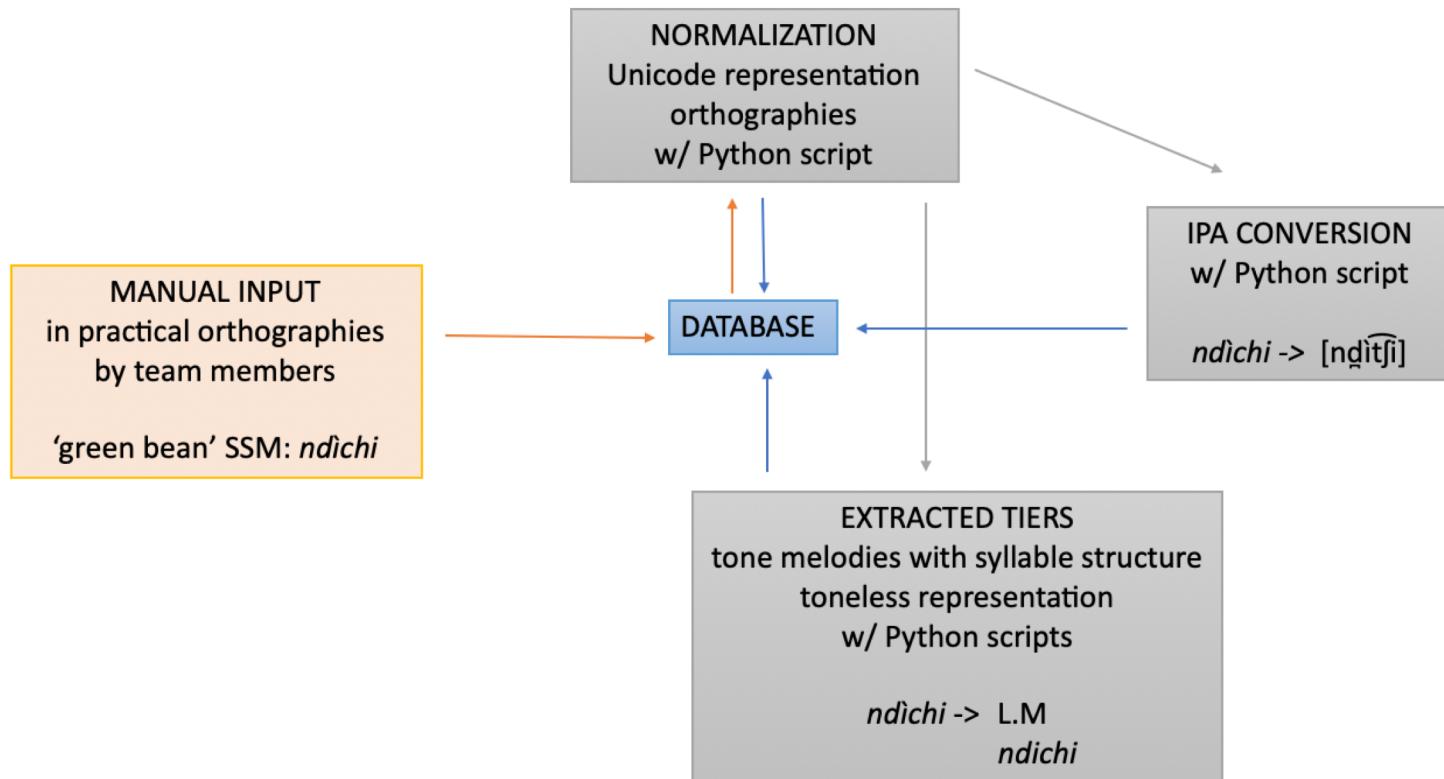
Steps automated with Python scripts

- regularization of the orthography
- generation of toneless representations
- extraction of tone melodies (tone representations w/o segments, but w/ syllable boundaries)
- conversion of orthographies to IPA

Advantages:

- recognizes orthography mixing
- speeds documentation
- makes the database multipurpose

Workflow



Uses of the database: Comparative method

1. The data entered serves as the basis for establishing sound correspondences and cognate sets and refining/adding reconstructions
2. These cognate sets are then used to further investigate subgrouping (both based on sound changes and lexical differences/similarities)
3. These results are important for the development of community-oriented materials:
 - interpretation services
 - multi-variety resources

Scorpion -- Alacrán								
TLH	<i>t</i>	í	<i>s</i>	ò	?	<i>m</i>	<i>a</i>	
PA	<i>ts̪</i>	ì	<i>s</i>	<i>o</i>	?	<i>m</i>	à	
SMD	<i>tç̪</i>	ì	<i>s</i>	<i>u</i>	?	<i>m</i>	à	
SSM	<i>t</i>	ì	<i>j</i>	<i>i</i>		<i>m</i>	<i>a</i>	
SJM	<i>t</i>	ì	<i>s</i>	<i>u</i>	?	<i>m</i>	à	

Uses of the database: Community

efficient documentation:

- searchability
- different character compositions
- everything is derived from the practical orthography
 - the linguist is not the gate-keeper
 - changes in the orthography as the documentation advances are not a problem
- expandable to other varieties

online dictionary:

- the database can easily be embedded on a website and serve as a multilingual, multi-variety dictionary

Database design considerations

- for tone languages:
 - toneless representation necessary (user-friendly, searchable)
- for multi-variety projects:
 - files that can be worked on at the same time by multiple people in different locations (cf. The increasing number of geographically disjoint communities)
 - i.e. a live spreadsheet
 - multi-variety information retrievable at the same time
- for community-based projects in general:
 - low-tech solutions (e.g. spreadsheets) for inputting and modifying data
 - general data management principle: accessible, general-purpose tools with long-term support

Importantly: this also improves research!

13th Meeting of the ALT - Università degli Studi di Pavia – September 4-6 2019

An empirical reconsideration of tone volatility and segment stability in Mixtec languages

SANDRA AUDERSET^{1,2}, CHRISTIAN D. BRENDL¹,
ERIC W. CAMPBELL¹, SIMON L. PETERS¹,
KEVIN SCHAEFER¹ & ALBERT VENTAYOL-BOADA¹

1 UNIVERSITY OF CALIFORNIA, SANTA BARBARA

2 MPI FOR THE SCIENCE OF HUMAN HISTORY, JENA

THANKS FOR LISTENING!

Acknowledgements



Carmen Hernández
Martínez, Arcenio López,
Javier Martínez, Yesica
Ramirez, Juvenal Solano,
Vanessa Terán

UCSB Field Methods
students, Mary Bucholtz,
Alexia Fawcett, Adrienne
Tsikewa

Grant numbers 1650114,
1660355

References

- Campbell, Eric W. (in press). Why is tone change still poorly understood, and how might documentation of less-studied tone languages help? In Patience L. Epps; Na'ama Pat-El & Danny Law (eds.), *Historical linguistics and endangered languages: exploring diversity in language change*. New York: Routledge.
- Campbell, Eric W. 2017. Otomanguean historical linguistics: past, present and prospects for the future. *Language y Linguistics Compass* 11: e12240.
- Cruz, Emiliana & Woodbury, Anthony C. 2014a. Finding a way into a family of tone languages: The story and methods of the Chatino Language Documentation Project. *Language Documentation & Conservation* 8: 490–524.
- Cruz, Emiliana & Woodbury, Anthony C. 2014b. Collaboration in the context of teaching, scholarship, and language revitalization: experience from the Chatino Language Documentation Project. *Language Documentation & Conservation* 8: 262–286.

References

- Czaykowska-Higgins, Ewa. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian Indigenous communities. *Language Documentation & Conservation* 3(1):15-50.
- Dürr, Michael. 1987. A preliminary reconstruction of the proto-Mixtec tonal system. *Indiana* 11: 19–61.
- Egland, Steven. 1983. La inteligibilidad interdialectal en México: Resultados de algunos sondeos. México, D.F.: Instituto Lingüístico de Verano, A.C.
- Greenberg, Joseph. 1948. The tonal system of Proto-Bantu. *Word* 4(3): 196–208.
- Hyman, Larry M. 2018. What tone teaches us about language. *Language* 94(3): 698–709.
- INALI. 2008. Catalogo de las lenguas indígenas nacionales: Variantes lingüísticas de México con sus autodenominaciones y referencias geoestadisticas. Diario oficial del Instituto Nacional de Lenguas Indígenas.

References

- Janda, Richard D. & Brian D. Joseph. 2003. On language, change, and language change—or, of history, linguistics, and historical linguistics. In Brian D. Joseph and Richard D. Janda (eds.), *The Handbook of historical linguistics*, 3–180. Oxford: Blackwell.
- Josserand, Judy Kathryn. 1983. Mixtec dialect history. PhD dissertation, Tulane University.
- Maddieson, Ian. 2013. Tone. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*.
- Simons, Gary F. & Charles D. Fennig (eds.). 2017. *Ethnologue: Languages of the world*, twentieth edition. Dallas: SIL International.
- Strecker, David. 1979. A preliminary typology of tone shapes and tonal sound changes in Tai: the lān nā a-tones. In Therapan L. Thongkum, Pranee Kullavanijaya, Vichin Panupong and M.R. Kalaya Tingsabadh (eds.), *Studies in Tai and Mon-Khmer phonetics and phonology*. Bangkok: Chulalongkorn University Press. 171-240.
- Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.