Substantive bias in paradigm reanalysis: the case of Malagasy weak-stem alternations

Jennifer Kuo

UCLA

July 1, 2022
TripleAFLA (Manchester)

Outline

Paradigm reanalysis

- Paradigms which contain neutralizing alternations are difficult to learn, and prone to reanalysis.
- Example of reanalysis in Latin:
  - *Before*

NOM SG.	GEN SG.		*After* NOM SG.	GEN SG.
hono:**s**	hono:**r**is	→	hono**r**	hono:**r**is (**s**→**r**)
soro**r**	soro:**r**is		soro**r**	soro:**r**is

Paradigm reanalysis

- Kiparsky (1978): studying patterns of reanalysis can give us insight into the factors that drive phonological learning.
- Recent work suggests the following factors:
  - Probabilistic distributions (frequency matching)
  - biases towards less marked outputs.

Paradigm reanalysis

- Kiparsky (1978): studying patterns of reanalysis can give us insight into the factors that drive phonological learning.
- Recent work suggests the following factors:
  - Probabilistic distributions (frequency matching)
  - biases towards less marked outputs.
- Very few quantitative models of reanalysis.
- Existing ones predict that reanalysis will always be based on probabilistic distributions.
  - Albright's Minimal Generalization Learner (MGL 2002; 2003, etc.)
  - Nosofsky's (2011) Generalized Context Model (exemplar-based analogical models).

Goals of the current talk

- Show that in Malagasy, reanalysis is driven by **both**:
  - distributional information
  - a substantive bias towards phonetically natural outputs.
- Outline a constraint-based model of reanalysis which captures both effects.

Basics of Malagasy phonology

- Official Malagasy (OM), based largely on the Merina dialect
- Inventory (Howe, 2021):
  - 5 vowels (/a e i o u/)

	bilabial	labiodental	dental	alveolar	retroflex	velar	glottal
plosives*	p, b		t, d			k, g	
	ᵐp, ᵐb		ⁿt, ⁿd			ᵑk, ᵑg	
affricates*				ts, dz	ʦ, ʣ		
				ⁿts, ⁿdz	ⁿʦ, ⁿʣ		
nasals	m		n			(ŋ)	
trills/flaps				r∼ɾ			
fricatives		f, v		s z			h
lat. approximants				l			

*Pre-nasalized stops/affricates will be written as nasal-consonant sequences (e.g. *mp* [ᵐp]).

- (C)V syllable structure.

Weak stems (Keenan and Polinsky, 2017)

- always end in one of the weak syllables ‘ka’, ‘ʈʂa’, ‘na’
- When weak stems are suffixed, the consonant in the weak syllable (ʈʂ/k/n) may alternate with another consonant.

pattern		**active (m + stem)**	**passive (stem + ana)**	
na ∼	n	mandʐávi**n**a	andʐavín**n**ana	‘to bear leaves’
	m	manándʐa**n**a	andʐá**m**ana	‘to try’
ka ∼	h	mangáta**k**a	angatá**h**ana	’to ask for’
	f	manáha**k**a	anaháfana	’to scatter’
ʈʂa ∼	r	miána**ʈʂ**a	ianárana	‘to learn’
	t	manándʐa**ʈʂ**a	anandʐá**t**ana	‘to promote’
	f	mandʐáku**ʈʂ**a	andʐakú**f**ana	‘to cover’

Table 1: Patterns of consonant alternation in Malagasy weak stems

Historical basis of weak stem alternations

- final consonant neutralizations, followed by vowel epenthesis, resulted in weak stems (Adelaar, 2012).
- ∼400AD: contact with Bantu resulted in a strict CV syllable structure.
  - a subset of consonant-final forms underwent final vowel epenthesis.

Historical basis of weak stem alternations

Example adapted from Mahdi (1988)

(1) **ʈʂa∼t alternation**

PMP	*Rabut	*z-abut-a	
	↓	↓	
PSEB	*ˈawut	*piaˈwutan	
	↓	↓	
PMlg	*ˈavuʈʂ	*fiaˈvutan	Final affrication: *-t > -ʈʂ
	↓	↓	
	ˈavuʈʂa	fiaˈvutana	ʈʂa∼t alternation
Mlg	ˈavuʈʂa	fiaˈvutana	‘to uproot’

PMP Proto-Austronesian
PMP Proto-Malayo-Polynesian
PSEB Proto-South East Barito

Reanalysis in weak stems

- In general, observed alternants are supposed to match specific historical final consonants.
- In cases where there is a mismatch, reanalysis has likely occured.
- Examples of possible reanalyses:

DIRECTION	ACTIVE	PASSIVE
r→t	miánaʦa	ianárana → ianátana
t→r	manándʐaʦa	anandʐátana → anandʐárana

Reanalysis in weak stems

- In general, observed alternants are supposed to match specific historical final consonants.
- In cases where there is a mismatch, reanalysis has likely occured.
- Examples of possible reanalyses:

DIRECTION	ACTIVE	PASSIVE
r→t	miánaʦa	ianárana → ianátana
t→r	manándʐaʦa	anandʐátana → anandʐárana

Question: What factors influence the direction of reanalysis in Malagasy?

Outline

Reanalysis in Malagasy

- In Malagasy, the observed alternant often does *not* match the historical PMP one, suggesting that extensive reanalysis has occurred.
- As a preview, reanalysis appears to have largely happened in the following directions:

Ending	**Direction**	
ka	f→h	Predicted by lexical statistics
na	m→n	Predicted by lexical statistics
ʈʂa	f, t → r	**Not** predicted by lexical statistics

Reanalysis in Malagasy

- In Malagasy, the observed alternant often does *not* match the historical PMP one, suggesting that extensive reanalysis has occurred.
- As a preview, reanalysis appears to have largely happened in the following directions:

Ending	**Direction**	
ka	f→h	Predicted by lexical statistics
na	m→n	Predicted by lexical statistics
ʈʂa	f, t → r	**Not** predicted by lexical statistics

- **Note: in the rest of this talk, I will focus on reanalysis in ʈʂa-final stems.**

Method and data

- Approach: compare modern Malagasy forms with their historical variants, and look for cases where the observed alternant doesn't match the expected one.
- Data:
  - Protoforms from Austronesian Comparative Dictionary (ACD; Blust and Trussel, 2010)
    - supplemented with loanwords from the World Loanword Database (Adelaar, 2009).
  - Modern Malagasy forms from the Malagasy Dictionary and Encyclopedia of Madagascar (MDEM; de La Beaujardière 2004), confirmed with a native speaker consultant.
  - 73 Malagasy ʈʂa-final weak stems with known historical forms.

Predicting directions of reanalysis

- Strong preference for t (67%)
- A slight dissimilatory tendency (*r...r), which has also been noted by Mahdi (1988)

expected	**count**	**preceding r**
t (<*t,*C)	49	13 (27%)
r (<*j,*d,*ḏ)	16	2 (13%)
f (<*b,*p)	8	1 (13%)

Table 2: Expected distribution of alternants for ṯṣa-stems, given final consonants of PMP

Predicting directions of reanalysis

- Strong preference for t (67%)
- A slight dissimilatory tendency (*r...r), which has also been noted by Mahdi (1988)

expected	**count**	**preceding r**
t (<*t,*C)	49	13 (27%)
r (<*j,*d,*ḏ)	16	2 (13%)
f (<*b,*p)	8	1 (13%)

Table 2: Expected distribution of alternants for ṯs-stems, given final consonants of PMP

- Suppose that reanalysis was purely driven by distributional information...
- then, we expect it to result in **more t-alternation** (i.e. r→t), and potentially **more dissimilation**.

Results: directions of reanalysis

- Expected vs. actual alternant of ʦa-final stems in modern Malagasy, based on known protoforms/loanwords

	alternant				
match?	expected	modern Mal.	change	count	preceding r
yes	r	r		15	0
yes	t	t		21	14
yes	f	f		6	1
no	r	t	r→t	1	1
no	t	r	t→r	28	0
no	f	r	f→r	2	0

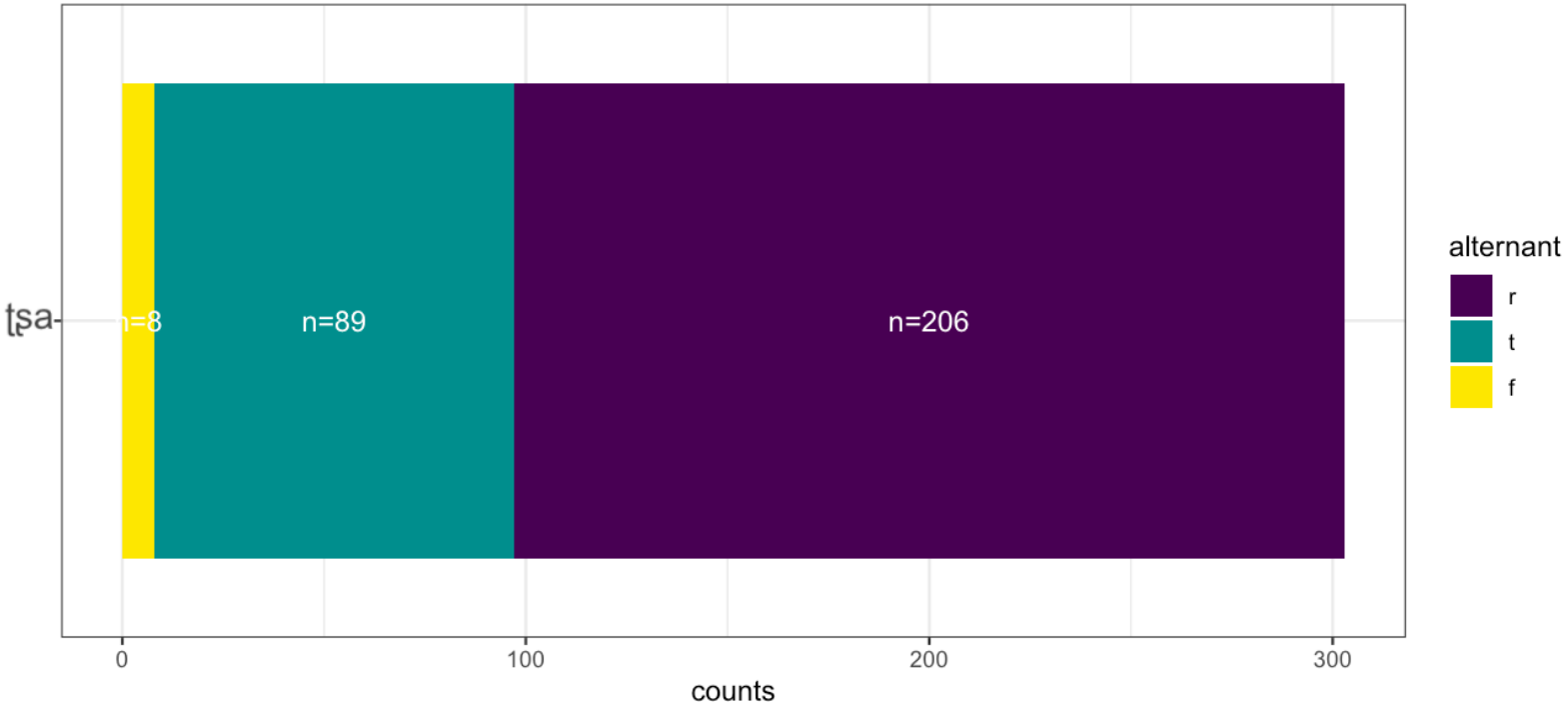
- Overwhelmingly, reanalysis is in the direction of t→r
  - ...except when there a preceding r
- We observe dissimilation, but also change in the opposite direction of what is predicted by distributional information.
- note: t>r is **not** a regular sound change, and VtV sequences are found in stems.

Results: ʈʂa-alternation in modern Malagasy

- What does ʈʂa-alternation look like in modern Malagasy?
- data: 305 ʈʂa-final weak stems (from the MDEM).

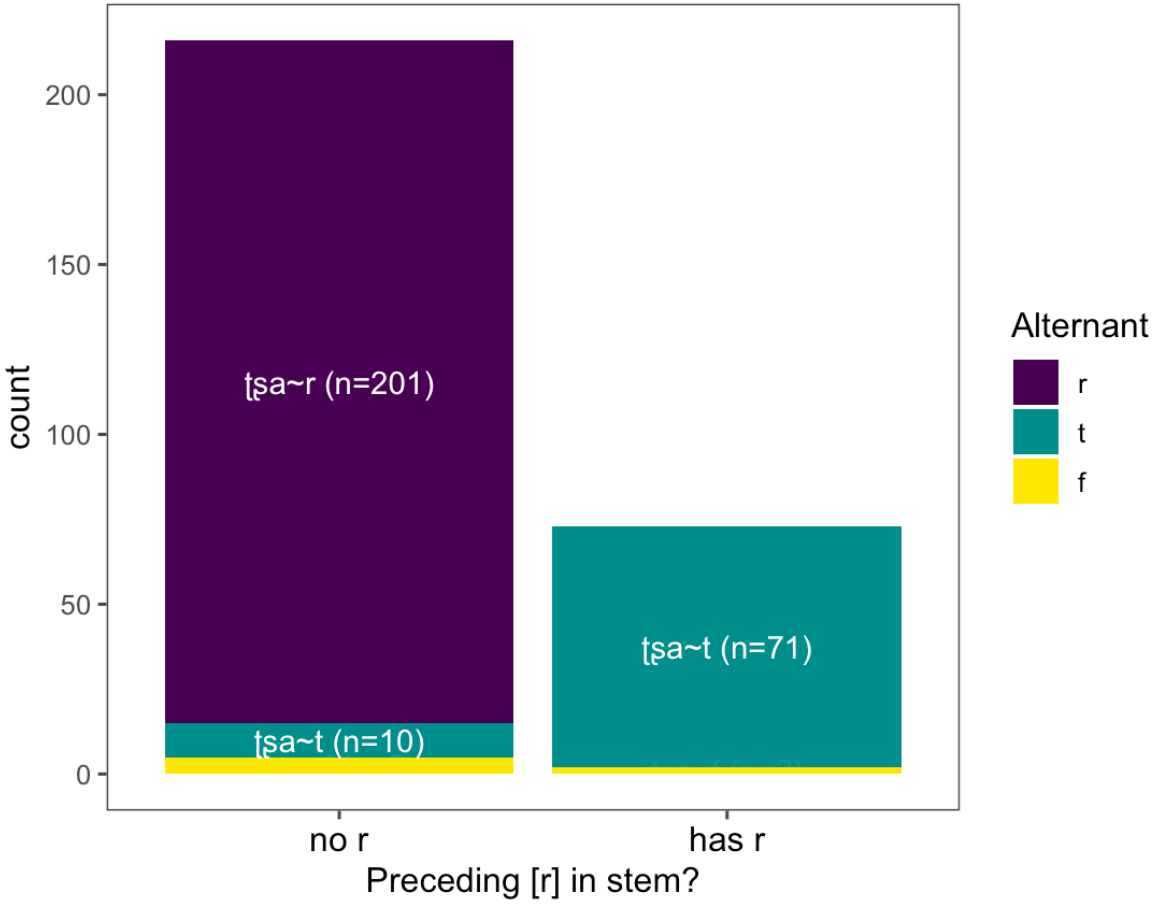
Results: ʈʂa-alternation in modern Malagasy

- What does ʈʂa-alternation look like in modern Malagasy?
- data: 305 ʈʂa-final weak stems (from the MDEM).
- general preference for r as alternant



State of ʈʂa-alternation in modern Malagasy

- strong dissimilatory pattern
- r is the default, **except** when stem already has an r



Summary of pattern

historical Malagasy

preference for t-alternation; weak dissimilatory patterns (*r...r)

current Malagasy

preference for r-alternation; strong dissimilatory patterns

Summary of pattern

historical Malagasy

preference for t-alternation; weak dissimilatory patterns (*r...r)

current Malagasy

preference for r-alternation; strong dissimilatory patterns

Why was there reanalysis towards r-alternation?

- not predicted by distributional probabilities; problem for existing quantitative models of analogical change
- Cause: markedness constraint against intervocalic stops?
  - Evidence from Malagasy
  - Typological evidence

Markedness constraints against intervocalic stops

- Evidence from Malagasy
  - historical lenition/spirantization: *b>v; *p>f; *d,*ɖ>r; *k,*g>h (Adelaar, 2012)
  - ...resulted in there being fewer intervocalic stops at some point in historical Malagasy
  - ka-final weak stems always alternate with fricatives (h, f) e.g. ˈalo**k**a∼aˈlo**h**ana, ˈhiri**k**a ∼hiri**f**ana

- Typology
  - phonetically natural, from both an articulatory (Kirchner, 1998) and perceptual (Kaplan, 2010; Katz, 2016) point of view.
  - Examples of lenition of stops specifically at morpheme boundaries: English tapping (Hayes, 2011, p. 143-144), Korean lenis stop voicing (Jun, 1994).

Outline

Modeling

- Reanalysis of ʈʂa-final stems is qualitatively influenced by markedness constraints.
- **Next step:** explicitly demonstrate how this works in a quantitative model
- Model should predict reanalysis of *t>r, despite the distribution of final stops in historical Malagasy, which favors t over r.

Modeling

- Reanalysis of ʦa-final stems is qualitatively influenced by markedness constraints.
- **Next step:** explicitly demonstrate how this works in a quantitative model
- Model should predict reanalysis of *t>r, despite the distribution of final stops in historical Malagasy, which favors t over r.
- Model is based in Maximum Entropy Harmonic Grammar (Maxent; Goldwater and Johnson, 2003; Prince and Smolensky, 1993), a stochastic variant of OT which assigns candidates probabilities.
- Faithfulness constraints are biased to have lower weights.
  - Captures the intuition that learners may value markedness over faithfulness constraints in the beginning stages of learning (e.g. Tesar and Smolensky, 2000; Jusczyk et al., 2002).

Input

- tra-, ka-final,na- final weak stems
- Frequency counts based on the PMP protoforms.
- rather than URs, the input is **surface forms**, mapping from stem to suffixed.

  e.g. INPUT → CANDIDATES
  ˈvuliʈʂa → vuˈlir-ana
  vuˈlit-ana
  vuˈliʈʂ-ana

  - Reason: Empirically, all reanalyses are from the stem→suffixed forms, and this happens when learners have access to the surface stem, but not the suffixed forms.
  - Similar to approach to Albright (2002)
- Note: for simplicity, I'm ignoring ʈʂa∼f alternating forms, which are rare and do not influence model outcomes.

Constraint set: faithfulness

- IDENT-OO constraints (output-output identity constraints, by feature)

Constraint set: faithfulness

- IDENT-OO constraints (output-output identity constraints, by feature)
- Following Wilson (2006), biases are implemented as Gaussian priors, with a preferred weight (μ) for each constraint.
  - Faithfulness constraints have $\mu = 0$, and are therefore penalized for having high weight.

Constraint set: Markedness

- *ʈʂ]V, *k]V, *n]V, which assess violations for every C]V, where C is **at a morpheme boundary**.
  - motivates alternation of the final consonant of weak stems.
  - Need to reference morpheme boundary because within stems, prevocalic ʈʂ, k, and n are allowed (e.g. beʈʂoka ‘to swell up’, ʈʂano ‘box’)
  - Example:

ˈvuliʈʂa	*ʈʂ]	ID-OO[anterior]	ID-OO[voice]
a. vuˈlit-ana		*	
b. vuˈlir-ana		*	*
c. vuˈliʈʂ-ana	*!		

Constraint set: Markedness

- ***V [-cont] V**: motivates re-analysis from t>r

ˈvuliʈʂa	*V[-cont]V	ID-OO[voice]	ID-OO[ant]
a. vuˈlit-ana	*!		*
☞ b. vuˈlir-ana		*	*

- ***r...r**: motivate r-dissimilation.

ˈvuriʈʂa	*r...r	*V[-cont]V	ID-OO[voice]	ID-OO[ant]
☞ a. vuˈrit-ana		*		*
b. vuˈrir-ana	*!		*	*

Markedness constraints

- Notably, *V[-cont]V and *r...r do not trigger alternations in the lexicon.
- Examples of stems which violate these constraints:
  - *akándzo* ‘coat, dress’, *áto* ‘close at hand’
  - *ráraka* ‘spilled’, *boréra* ‘weak, limp’
- However, they appear to be present as weak phonotactic constraints in the Malagasy lexicon.
  - Based on a phonotactic grammar built using the UCLA Phonotactic Learner (Hayes and Wilson, 2008).

Model results: one iteration

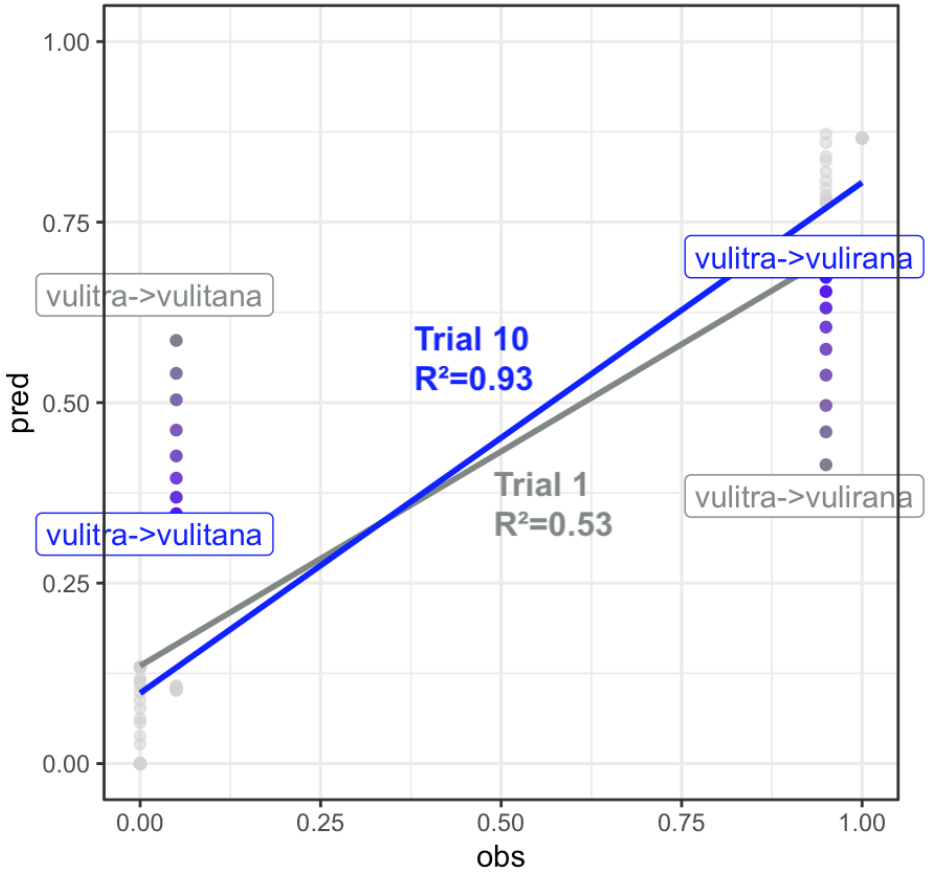
Results after one iteration of the model:

- Bias in the correct direction, but with a small magnitude.

input	output	type	obs (PMP)	P (biased)	
vuliʈʂa	vulir-ana	ʈʂ∼r	0.28	**0.33**	**+5%**
	vulit-ana	ʈʂ∼t	0.72	0.67	
	vuliʈʂ-ana	non-alt	0	0	
vuriʈʂa	vurir-ana	ʈʂ∼r	0.13	0.13	
	vurit-ana	ʈʂ∼t	0.87	0.87	
	vuriʈʂ-ana	non-alt	0	0	
vulika	vulih-ana	k∼h	0.90	0.88	
	vulif-ana	k∼f	0.10	0.11	
	vulik-ana	non-alt	0	0.01	
vulina	vulim-ana	n∼m	0.04	0.04	
	vulin-ana	non-alt	0.96	0.96	

Model results: multigenerational

- Running the model multiple times could simulate the effect of a bias over multiple generations.
- Method: run the model 10 times, with the results of the previous trial as the input to the next.



- Plot: observed rates of alternation in modern Malagasy vs. model predictions.
- A biased model improves fit to modern Malagasy lexicon, suggesting that it better predicts directions of reanalysis.

Conclusion

- In Malagasy, the direction of reanalysis in ʈʂa-final weak stems is not predicted by lexical statistics.
- Instead, it is drive by two markedness constraints, both arguably phonetically natural.
  - *V[-cont]V
  - *r...r
- Reanalysis can be modeled in MaxEnt, where frequency matching is modulated by a substantive bias.

Special thanks to...

- Vololona, for generously using her time to provide native speaker judgements.
- Bruce Hayes, Kie Zuraw, Claire Moore-Cantwell, David Goldstein, and members of the UCLA phonology seminar for their invaluable feedback.

References I

Alexander Adelaar. Malagasy. In Martin Haspelmath and Uri Tadmor, editors, *World loanword database (WOLD)*. Max Planck digital library, 2009.

Alexander Adelaar. Malagasy phonological history and bantu influence. *Oceanic Linguistics*, pages 123–159, 2012.

Adam Albright and Bruce Hayes. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90(2):119–161, 2003.

Adam C Albright. *The identification of bases in morphological paradigms*. PhD thesis, University of California, Los Angeles, 2002.

Robert Blust and Stephen Trussel. Austronesian comparative dictionary, web edition. *Blust's Austronesian Comparative Dictionary Website*, 2010.

Jean-Marie de La Beaujardière. Malagasy dictionary and encyclopedia of Madagascar, 2004.

Sharon Goldwater and Mark Johnson. Learning ot constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm workshop on variation within Optimality Theory*, volume 111120, 2003.

Bruce Hayes. *Introductory phonology*, volume 32. John Wiley & Sons, 2011.

Bruce Hayes and Colin Wilson. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3):379–440, 2008.

References II

Penelope Howe. Central Malagasy. *Journal of the International Phonetic Association*, 51 (1):103–136, 2021.

Sun-Ah Jun. The status of the lenis stop voicing rule in Korean. *Theoretical issues in Korean linguistics*, pages 101–114, 1994.

Peter W Jusczyk, Paul Smolensky, and Theresa Allocco. How english-learning infants respond to markedness and faithfulness constraints. *Language Acquisition*, 10(1): 31–73, 2002.

Abby Kaplan. *Phonology shaped by phonetics: The case of intervocalic lenition*. PhD thesis, University of California, Santa Cruz, 2010.

Jonah Katz. Lenition, perception and neutralisation. *Phonology*, 33(1):43–85, 2016.

Edward L Keenan and Maria Polinsky. Malagasy (austronesian). *The handbook of morphology*, pages 563–623, 2017.

Paul Kiparsky. Analogical change as a problem for linguistic theory. *Studies in the Linguistic Sciences Urbana, Ill*, 8(2):77–96, 1978.

Robert M. Kirchner. *An effort-based approach to consonant lenition*. PhD thesis, UCLA, 1998.

Waruno Mahdi. *Morphophonologische Besonderheiten und historische phonologie des Malagasy*, volume 20. D. Reimer, 1988.

References III

Robert M Nosofsky. The generalized context model: An exemplar model of classification. *Formal approaches in categorization*, pages 18–39, 2011.

Alan Prince and Paul Smolensky. Optimality theory: Constraint interaction in generative grammar. *Optimality Theory in Phonology*, page 3, 1993.

Bruce Tesar and Paul Smolensky. *Learnability in optimality theory*. Mit Press, 2000.

Colin Wilson. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive science*, 30(5):945–982, 2006.