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

Jennifer Kuo

UCLA

July 1, 2022
TripleAFLA (Manchester)
Outline

Background

Directions of reanalysis

Modeling
Paradigm reanalysis

- Paradigms which contain neutralizing alternations are difficult to learn, and prone to reanalysis.
- Example of reanalysis in Latin:
  - *Before*       *After*
    NOM SG.   GEN SG.    NOM SG.   GEN SG.
    hono:s    hono:ris  →  honor    hono:ris (s→r)
    soror     soro:ris  soror     soro:ris
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.)
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/)

<table>
<thead>
<tr>
<th></th>
<th>bilabial</th>
<th>labiodental</th>
<th>dental</th>
<th>alveolar</th>
<th>retroflex</th>
<th>velar</th>
<th>glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>plosives*</td>
<td>p, b</td>
<td>t, d</td>
<td></td>
<td></td>
<td>k, g</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>m(^p), m(^b)</td>
<td>n(^t), n(^d)</td>
<td></td>
<td></td>
<td>n(^k), n(^g)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>affricates*</td>
<td>ts, dz</td>
<td>ʈʂ, ɖʐ</td>
<td></td>
<td></td>
<td>n(^t)ʂ, n(^d)ʐ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n(^t)s, n(^d)ʐ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nasals</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(ŋ)</td>
</tr>
<tr>
<td>trills/flaps</td>
<td></td>
<td></td>
<td>r~ɾ</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fricatives</td>
<td>f, v</td>
<td>s, z</td>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lat. approximants</td>
<td></td>
<td></td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Pre-nasalized stops/affricates will be written as nasal-consonant sequences (e.g. \(m)p \[^{\text{m}}\]p).*

- (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.

<table>
<thead>
<tr>
<th>pattern</th>
<th>active (m + stem)</th>
<th>passive (stem + ana)</th>
</tr>
</thead>
<tbody>
<tr>
<td>na ~ n</td>
<td>mandʒávina</td>
<td>andʒavínana</td>
</tr>
<tr>
<td>m</td>
<td>manándʒana</td>
<td>andʒámana</td>
</tr>
<tr>
<td>ka ~ h</td>
<td>mangátaka</td>
<td>angátahana</td>
</tr>
<tr>
<td>f</td>
<td>manáhaka</td>
<td>anaháfana</td>
</tr>
<tr>
<td>tʃa ~ r</td>
<td>miánatʃa</td>
<td>ianárana</td>
</tr>
<tr>
<td>t</td>
<td>manándʒatʃa</td>
<td>anandʒátana</td>
</tr>
<tr>
<td>f</td>
<td>mandʒáku tʃa</td>
<td>andʒakúfana</td>
</tr>
</tbody>
</table>

‘to bear leaves’
‘to try’
’to ask for’
’to scatter’
‘to learn’
‘to promote’
‘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).
- \(\sim\)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

<table>
<thead>
<tr>
<th>Stage</th>
<th>Language</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMP</td>
<td>*Rabut</td>
<td>*z-abut-a</td>
</tr>
<tr>
<td></td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>PSEB</td>
<td>*'awut</td>
<td>*pia'wutan</td>
</tr>
<tr>
<td></td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>PMlg</td>
<td>*'avuʈʂ</td>
<td>*fia'vutan</td>
</tr>
<tr>
<td></td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td></td>
<td>'avuʈʂa</td>
<td>fia'vutana</td>
</tr>
</tbody>
</table>

Final affrication: *-t > -ʈʂ

Mlg 'avuʈʂa fia'vutana ‘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 occurred.
- Examples of possible reanalyses:
  
<table>
<thead>
<tr>
<th>DIRECTION</th>
<th>ACTIVE</th>
<th>PASSIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>r → t</td>
<td>miánaʈʂa</td>
<td>ianárana → ianátana</td>
</tr>
<tr>
<td>t → r</td>
<td>manándʐaʈʂa</td>
<td>anandʐátana → anandʐárana</td>
</tr>
</tbody>
</table>
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 occurred.
- Examples of possible reanalyses:

<table>
<thead>
<tr>
<th>DIRECTION</th>
<th>ACTIVE</th>
<th>PASSIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>r → t</td>
<td>miánaʈʂa</td>
<td>ianárana → ianátana</td>
</tr>
<tr>
<td>t → r</td>
<td>manándʒzaʈʂa</td>
<td>anandʒátana → anandʒárana</td>
</tr>
</tbody>
</table>

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

Background

Directions of reanalysis

Modeling
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:

<table>
<thead>
<tr>
<th>Ending</th>
<th>Direction</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>ka</td>
<td>f→h</td>
<td>Predicted by lexical statistics</td>
</tr>
<tr>
<td>na</td>
<td>m→n</td>
<td>Predicted by lexical statistics</td>
</tr>
<tr>
<td>tʃa</td>
<td>f, t→r</td>
<td>Not predicted by lexical statistics</td>
</tr>
</tbody>
</table>
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:

<table>
<thead>
<tr>
<th>Ending</th>
<th>Direction</th>
<th>Predicted by lexical statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>ka</td>
<td>f→h</td>
<td></td>
</tr>
<tr>
<td>na</td>
<td>m→n</td>
<td></td>
</tr>
<tr>
<td>tsa</td>
<td>f, t → r</td>
<td>Not predicted by lexical statistics</td>
</tr>
</tbody>
</table>

- **Note:** in the rest of this talk, I will focus on reanalysis in tsa-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)

<table>
<thead>
<tr>
<th>expected</th>
<th>count</th>
<th>preceding r</th>
</tr>
</thead>
<tbody>
<tr>
<td>t (&lt;*t,*C)</td>
<td>49</td>
<td>13 (27%)</td>
</tr>
<tr>
<td>r (&lt;*j,*d,*ɖ)</td>
<td>16</td>
<td>2 (13%)</td>
</tr>
<tr>
<td>f (&lt;*b,*p)</td>
<td>8</td>
<td>1 (13%)</td>
</tr>
</tbody>
</table>

Table 2: Expected distribution of alternants for ʈša-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)

<table>
<thead>
<tr>
<th>expected</th>
<th>count</th>
<th>preceding r</th>
</tr>
</thead>
<tbody>
<tr>
<td>t (&lt;*t,*C)</td>
<td>49</td>
<td>13 (27%)</td>
</tr>
<tr>
<td>r (&lt;*j,*d,*d)</td>
<td>16</td>
<td>2 (13%)</td>
</tr>
<tr>
<td>f (&lt;*b,*p)</td>
<td>8</td>
<td>1 (13%)</td>
</tr>
</tbody>
</table>

Table 2: Expected distribution of alternants for ʈsa-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

<table>
<thead>
<tr>
<th>match?</th>
<th>alternant</th>
<th>modern Mal.</th>
<th>change</th>
<th>count</th>
<th>preceding r</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>r</td>
<td>r</td>
<td></td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>yes</td>
<td>t</td>
<td>t</td>
<td></td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>yes</td>
<td>f</td>
<td>f</td>
<td></td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>no</td>
<td>r</td>
<td>t</td>
<td>r→t</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>no</td>
<td>t</td>
<td>r</td>
<td>t→r</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>no</td>
<td>f</td>
<td>r</td>
<td>f→r</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

- 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: тša-alternation in modern Malagasy

- What does тša-alternation look like in modern Malagasy?
- data: 305 тša-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 ṭsa-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

<table>
<thead>
<tr>
<th><strong>historical Malagasy</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>preference for t-alternation; weak dissimilatory patterns (*r...r)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>current Malagasy</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>preference for r-alternation; strong dissimilatory patterns</td>
</tr>
</tbody>
</table>

### 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: \( *b > v; *p > f; *d, *d > r; *k, *g > h \) (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. 'alo\(k\)a ~ a'lo\(h\)ana, 'hir\(i\)ka ~ hir\(i\)fana

- 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

Background

Directions of reanalysis

Modeling
Modeling

• Reanalysis of ʈʃa-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 ṭṣa-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**
  
  'vuliṭša → vu'liŗ-ana
  vu'liŗ-ana
  vu'lišt-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 źa~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 ($\mu$) 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. ɓetşoka ‘to swell up’, ʈşano ‘box’)
- Example:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. vu'lit-ana</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Constraint set: Markedness

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. vu'lit-ana</td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. vu'lit-ana</td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

• *r...r: motivate r-dissimilation.

<table>
<thead>
<tr>
<th>'vuritša</th>
<th>*r...r</th>
<th>*V[-cont]V</th>
<th>ID-OO[voice]</th>
<th>ID-OO[ant]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. vu'rit-ana</td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. vu'rir-ana</td>
<td></td>
<td>!</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
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’
  - ráráka ‘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.

<table>
<thead>
<tr>
<th>input</th>
<th>output</th>
<th>type</th>
<th>obs (PMP)</th>
<th>P (biased)</th>
<th>P (biased)</th>
</tr>
</thead>
<tbody>
<tr>
<td>vuliṣa</td>
<td>vulir-ana</td>
<td>ʈʂ̕~r</td>
<td>0.28</td>
<td>0.33</td>
<td>+5%</td>
</tr>
<tr>
<td>vulit-ana</td>
<td>ʈʂ̕~t</td>
<td>0.72</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vuliṣ-ana</td>
<td>non-alt</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vuriṣa</td>
<td>vurir-ana</td>
<td>ʈʂ̕~r</td>
<td>0.13</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>vurt-ana</td>
<td>ʈʂ̕~t</td>
<td>0.87</td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vurtṣ-ana</td>
<td>non-alt</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vulika</td>
<td>vulih-ana</td>
<td>ʳ̕~h</td>
<td>0.90</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>vulif-ana</td>
<td>ʳ̕~f</td>
<td>0.10</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vulik-ana</td>
<td>non-alt</td>
<td>0</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vulina</td>
<td>vulim-ana</td>
<td>ⁿ̕~m</td>
<td>0.04</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>vulin-ana</td>
<td>non-alt</td>
<td>0.96</td>
<td>0.96</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 ṭsa-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


References II


References III


