Base-driven alternation in Tgdaya Seediq

AFLA 27

Jennifer Kuo

University of California, Los Angeles

Overview

Tgdaya Seediq, where verb paradigms show extensive alternations, is a good test case for comparing between theories of morphophonology.

Overview

Tgdaya Seediq, where verb paradigms show extensive alternations, is a good test case for comparing between theories of morphophonology.

Evidence from Tgdaya Seediq supports an approach where URs are based off a single surface allomorph.

Two approaches to morphophonological analysis

'Cobbled' URs (Kenstowicz and Kisseberth, 1977)

- URs preserve as many contrastive properties as possible.
- When all forms of a paradigm are affected by neutralizing processes, the resulting UR must **'cobble'** information from multiple forms of the paradigm.

Example: *Tonkawa* (Kenstowicz and Kisseberth, 1977, p.16)

A (/C-stem-V/)	B (/V-stem-C/)	gloss	UR
notx	ntoxo	'hoe'	/notoxo/
netl	ntale	'lick'	/netale/

Two approaches to morphophonological analysis

'Cobbled' URs (Kenstowicz and Kisseberth, 1977)

Result: surface forms are predictable, and derivable from **exceptionless rules/constraints**.

Example: *Tonkawa* (Kenstowicz and Kisseberth, 1977, p.16)

(/C-stem-V/)	(/V-stem-C/)	gloss	UR
netl	ntale	'lick'	/netale/

/C-netale-V/	Rule
C-netle-V	Delete V_2
C-netl-V	Delete V_3
[C-netl-V]	

Two approaches to morphophonological analysis

Single surface base hypothesis (Albright, 2002b, et seq.)

- learners designate **one slot** (surface allomorph) in the paradigm to be a 'privileged base'.
  - **input** for morphophonology.

Two approaches to morphophonological analysis

Single surface base hypothesis (Albright, 2002b, et seq.)

- learners designate **one slot** (surface allomorph) in the paradigm to be a 'privileged base'.
  - **input** for morphophonology.

Example: *Slot B is chosen as base*

A	B	
notx	ntoxo	'hoe'
netl	ntale	'lick'

Deriving slot A of paradigm

[ntoxo]	Rules
notoxo	∅ →[o]/__Co
[notoxo]	

Two approaches to morphophonological analysis

Single surface base hypothesis (Albright, 2002b, et seq.)

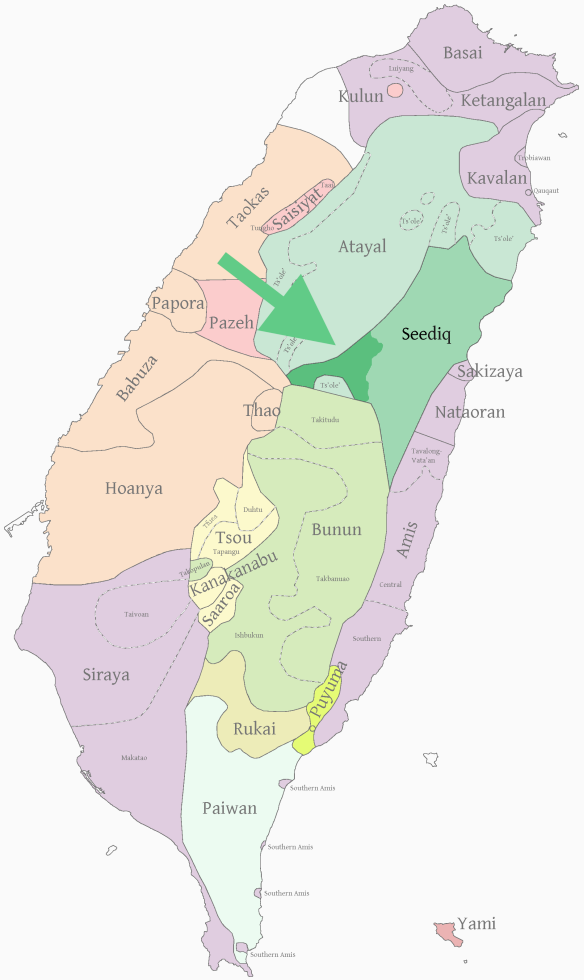
- If all forms of a paradigm have undergone some neutralization, **no base will work perfectly.**
  - rules/constraints **will have exceptions.**

Two approaches to morphophonological analysis

Single surface base hypothesis (Albright, 2002b, et seq.)

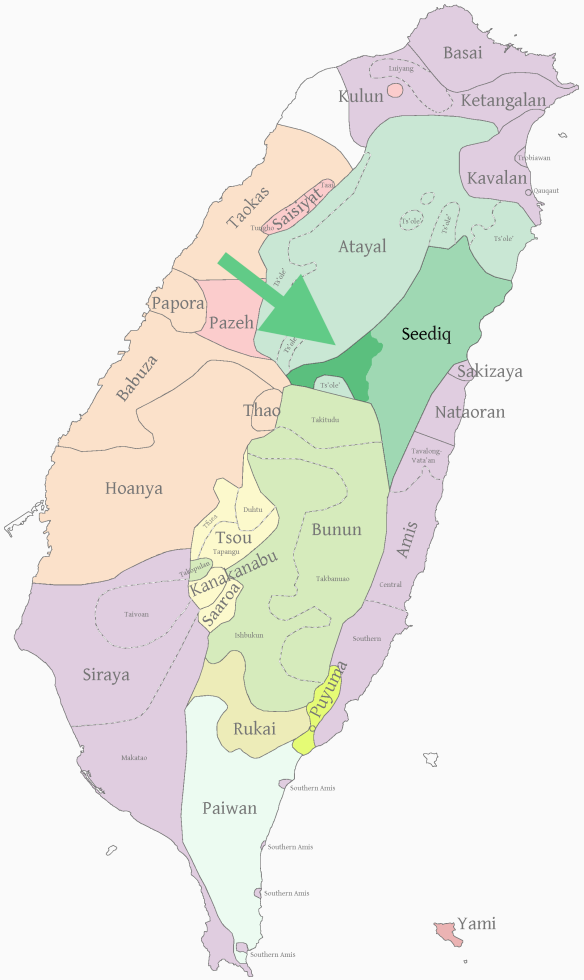
- If all forms of a paradigm have undergone some neutralization, **no base will work perfectly.**
  - rules/constraints **will have exceptions.**
- However, growing body of evidence from:
  - Historical change; e.g. Yiddish, Lakhota (Albright, 2010, 2002a)
  - Child learning errors; e.g. Korean (Kang, 2006)

Overview: Tgdaya Seediq



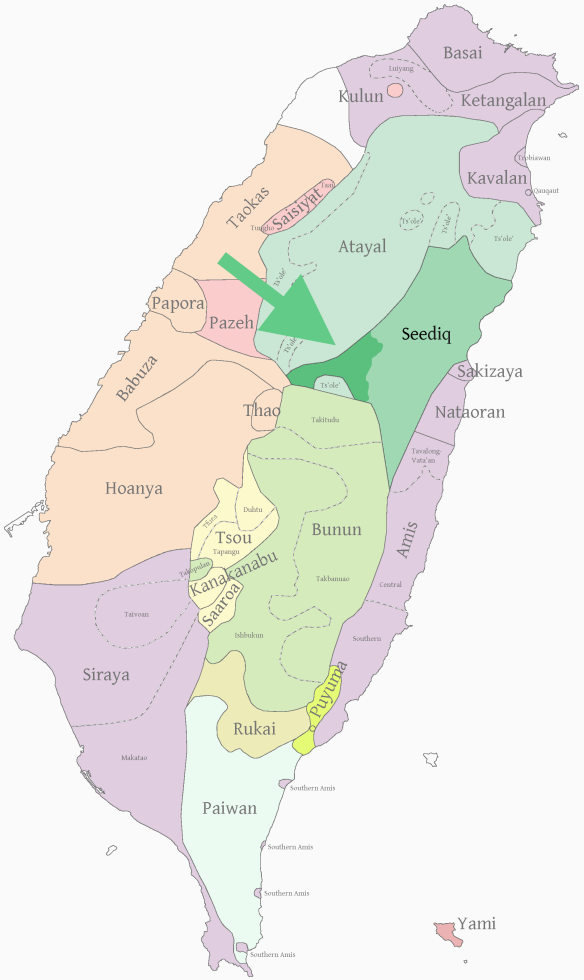
- Seediq is an Atayalic language, spoken in Taiwan.

Overview: Tgdaya Seediq



- Seediq is an Atayalic language, spoken in Taiwan.
- Tgdaya Seediq (德固達雅):
  - spoken primarily in Nantou
  - population ∼2500 (Tsukida, 2005)
  - Number of fluent speakers is thought to be much fewer.

Overview: Tgdaya Seediq



- Seediq is an Atayalic language, spoken in Taiwan.
- Tgdaya Seediq (德固達雅):
  - spoken primarily in Nantou
  - population ∼2500 (Tsukida, 2005)
  - Number of fluent speakers is thought to be much fewer.

All forms of a verb paradigm suffer from loss of contrasts ⇒ good test case for comparing the two theories of morphophonology.

Overview: Seediq morphophonology (Yang, 1976)

Verb inflection (Holmer, 1996).

	ACTOR FOC	LOC. FOC	PAT. FOC	INSTR. FOC
PRES	<m>/mu-	-an	-un	s-
PRET	<mun>	<n>-an	<un>	
FUT	mu(pu)-	RED-an	RED-un	
IMP	-i			

Significant alternations between suffixed and non-suffixed forms of verb paradigms

Overview: Seediq morphophonology (Yang, 1976)

Verb inflection (Holmer, 1996).

	ACTOR FOC	LOC. FOC	PAT. FOC	INSTR. FOC
PRES	<m>/mu-	-an	-un	s-
PRET	<mun>	<n>-an	<un>	
FUT	mu(pu)-	RED-an	RED-un	
IMP	-i			

Significant alternations between suffixed and non-suffixed forms of verb paradigms

Examples will compare bare stem vs. /-an/-suffixed forms

Sources of morphophonological alternation

Pretonic vowel reduction (VR)

Post-tonic VR

Consonant neutralization

Word-final monophthongization

Sources of morphophonological alternation

Pretonic vowel reduction (VR)

Post-tonic VR

Consonant neutralization

Word-final monophthongization

Pretonic vowel reduction (VR)

Five vowels /a i e o u/

Stress is penultimate; suffixation shifts stress rightwards.

Pretonically, all vowel contrasts are neutralized...

Pretonic vowel reduction (VR)

Five vowels /a i e o u/

Stress is penultimate; suffixation shifts stress rightwards.

Pretonically, **all** vowel contrasts are neutralized...

1. *Onsetless vowels delete* (36/36)
   a. ˈawak ∼ ˈwak-an ‘lead (by a leash)
   b. ˈeyah ∼ ˈyah-an ‘come’

Pretonic vowel reduction (VR)

Five vowels /a i e o u/

Stress is penultimate; suffixation shifts stress rightwards.

Pretonically, **all** vowel contrasts are neutralized...

1. *Onsetless vowels delete* (36/36)
   a. ˈawak ∼ ˈwak-an ‘lead (by a leash)
   b. ˈeyah ∼ ˈyah-an ‘come’
2. *Assimilation to stressed vowel if separated by [ʔ] or [h]* (35/35)
   a. ˈleʔiŋ ∼ liˈʔiŋ-an ‘hide (an object)’
   b. ˈsaʔis ∼ siˈʔis-an ‘sew’

Pretonic vowel reduction (VR)

Five vowels /a i e o u/

Stress is penultimate; suffixation shifts stress rightwards.

Pretonically, **all** vowel contrasts are neutralized...

1. *Onsetless vowels delete* (36/36)
   a. ˈawak ∼ ˈwak-an ‘lead (by a leash)
   b. ˈeyah ∼ ˈyah-an ‘come’
2. *Assimilation to stressed vowel if separated by [ʔ] or [h]* (35/35)
   a. ˈleʔiŋ ∼ liˈʔiŋ-an ‘hide (an object)’
   b. ˈsaʔis ∼ siˈʔis-an ‘sew’
3. *Vowel reduction to [u]* (201/201)
   a. ˈkesa ∼ kuˈsa-an ‘tell someone’
   b. ˈbarah ∼ buˈrah-an ‘rare’
   c. ˈbi͡tsiq ∼ buˈ͡tsiq-an ‘decrease’

Pretonic vowel reduction (VR)

Five vowels /a i e o u/

Stress is penultimate; suffixation shifts stress rightwards.

Pretonically, **all** vowel contrasts are neutralized...

1. *Onsetless vowels delete* (36/36)
   a. ˈawak ∼ ˈwak-an ‘lead (by a leash)
   b. ˈeyah ∼ ˈyah-an ‘come’
2. *Assimilation to stressed vowel if separated by [ʔ] or [h]* (35/35)
   a. ˈleʔiŋ ∼ liˈʔiŋ-an ‘hide (an object)’
   b. ˈsaʔis ∼ siˈʔis-an ‘sew’
3. *Vowel reduction to [u]* (201/201)
   a. ˈkesa ∼ kuˈsa-an ‘tell someone’
   b. ˈbarah ∼ buˈrah-an ‘rare’
   c. ˈbit͡siq ∼ buˈt͡siq-an ‘decrease’

⇒ Loss of contrasts in **suffixed** forms.

Stress-driven vowel alternations

- Post-tonically...
  1. /e,o/ reduce to [u] in closed syllables
     a. ˈremux ∼ ruˈmuxan ‘enter’ (u∼u, n=60)
     b. ˈpemux ∼ puˈmexan ‘hold’ (u∼e, n=36)
     c. ˈdoʔus ∼ doˈʔos-an ‘refine’ (metal)’ (u∼o, n=3)

Stress-driven vowel alternations

- Post-tonically...
  1. /e,o/ reduce to [u] in closed syllables
     a. ˈremux ∼ ruˈmuxan ‘enter’ (u∼u, n=60)
     b. ˈpemux ∼ puˈmexan ‘hold’ (u∼e, n=36)
     c. ˈdoʔus ∼ doˈʔos-an ‘refine’ (metal)’ (u∼o, n=3)

  ⇒ loss of contrasts in (non-suffixed) stem forms

Final consonant neutralization

- Various processes of final consonant neutralization, a subset of which are shown here:
  1. /p, b, k/ → [k]

	ALTERNATION		STEM	SUFFIXED	
(a)	[k∼k]	(n=19)	ˈtatak	tuˈtak-an	‘chop’
(b)	[k∼p]	(n=6)	ˈpatak	puˈtap-an	‘cut’
(c)	[k∼b]	(n=1)	ˈeluk	ˈleb-an	‘close’

  2. /t͡s, t, d/ → [t͡s]

(a)	[t͡s∼t͡s]	(n=1)	buˈt͡seba t͡s	bucuˈba t͡s-an	‘slice’
(b)	[t͡s∼t]	(n=16)	ˈdama t͡s	duˈmat-an	‘for eating’
(c)	[t͡s∼d]	(n=4)	ˈhara t͡s	huˈrad-an	‘build (a wall)’

Final cons. neutralization, continued

- Continued...
  3. /ŋ,m/ → [ŋ]
     (a) [ŋ∼ŋ] (n=32) ˈgilaŋ guˈlaŋ-an ‘mill (rice)’
     (b) [ŋ∼m] (n=3) ˈtalaŋ tuˈlam-an ‘run’
  4. /n,l/ → [n]
     (a) [n∼n] (n=3) ˈdurun duˈrun-an ‘entrust’
     (b) [n∼l] (n=19) ˈdudun duˈdul-an ‘lead’

  (alternations involving stem-final [g] are more complicated, and not discussed here)

Final cons. neutralization, continued

- Continued...
  3. /ŋ,m/ → [ŋ]
     (a) [ŋ∼ŋ] (n=32) ˈgilaŋ guˈlaŋ-an ‘mill (rice)’
     (b) [ŋ∼m] (n=3) ˈtalaŋ tuˈlam-an ‘run’
  4. /n,l/ → [n]
     (a) [n∼n] (n=3) ˈdurun duˈrun-an ‘entrust’
     (b) [n∼l] (n=19) ˈdudun duˈdul-an ‘lead’

     (alternations involving stem-final [g] are more complicated, and not discussed here)

⇒ loss of constrasts in stem forms.

Morphophonological learning in Seediq

All forms of a Seediq verbal paradigm suffer from some form of neutralization; some verbs undergo extensive alternations

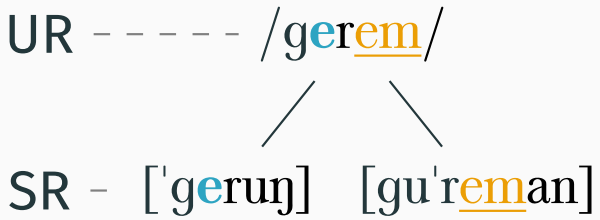
e.g. ˈ**gerɯŋ** ∼ **gu**ˈ**rem**an ‘to break’
ˈ**eluk** ∼ ˈ**leb**an ‘to close’

Morphophonological learning in Seediq

All forms of a Seediq verbal paradigm suffer from some form of neutralization; some verbs undergo extensive alternations

e.g. 'gerun ~ gu'reman ‘to break’
'eluk ~ 'leban ‘to close’

Cobbled UR approach

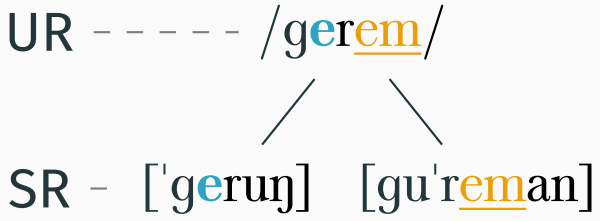


Morphophonological learning in Seediq

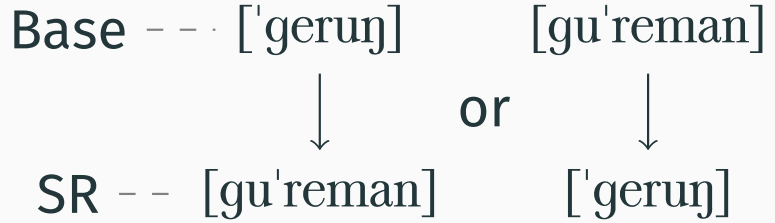
All forms of a Seediq verbal paradigm suffer from some form of neutralization; some verbs undergo extensive alternations

e.g. 'gerun ~ gu'reman 'to break'
'eluk ~ 'leban 'to close'

Cobbled UR approach



Single surface base approach

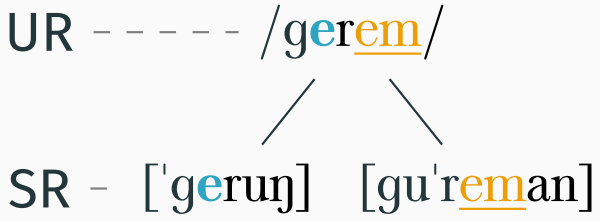


Morphophonological learning in Seediq

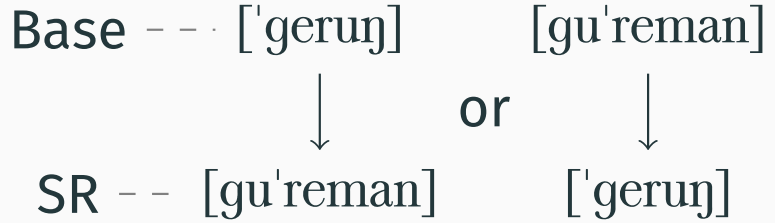
All forms of a Seediq verbal paradigm suffer from some form of neutralization; some verbs undergo extensive alternations

e.g. 'gerun ~ gu'reman 'to break'
'eluk ~ 'leban 'to close'

Cobbled UR approach



Single surface base approach



Comparing the two approaches

When the learner has incomplete data, what kind of renalysis/errors will take place?

Comparing the two approaches

When the learner has incomplete data, what kind of renalysis/errors will take place?

- **Cobbled UR:** the UR will be determined by whatever surface forms happen to be available.
  ⇒reanalyses in both directions are plausible.

Comparing the two approaches

When the learner has incomplete data, what kind of renalysis/errors will take place?

- **Cobbled UR:** the UR will be determined by whatever surface forms happen to be available.
  ⇒reanalyses in both directions are plausible.
- **Surface base:** Reanalyses will always be projected from the designated base (i.e. same slot in paradigm).
  ⇒resulting Seediq lexicon will have asymmetries in paradigm structure.

Quantitative Patterns

- Suffixed forms are highly predictable from stems, but not vice versa (i.e. stem forms are more informative)
- Suggests that Seediq speakers have identified the **isolation stem** as the base, per Albright's surface-base hypothesis.

Data collection

Results are based on a corpus of 340 verbal paradigms

- Taiwan Aboriginal e-Dictionary (n=184) (Mei-jin et al., 2014)
- fieldwork with three Seediq speakers (n=156)
  2F,1M; ages 69-78

Predictability from stems

Sources of contrast neutralization in stems/non-suffixed forms:

- Post-tonic vowel reduction
- Final consonant neutralization
- Final monophthongization

Predictability from stems

Sources of contrast neutralization in stems/non-suffixed forms:

- Post-tonic vowel reduction
- Final consonant neutralization
- Final monophthongization

Can these neutralizations be 'undone' in a principled way, based on statistical patterns of predictability?

Predictability from stems

Sources of contrast neutralization in stems/non-suffixed forms:

- Post-tonic vowel reduction
- Final consonant neutralization
- Final monophthongization

Can these neutralizations be 'undone' in a principled way, based on statistical patterns of predictability?

Predictability from stems: post-tonic vowel alternations

- Recall that due to post-tonic vowel reduction...

STEM		SUFFIXED
CVCuC	∼	{CuCeCan, CuCoCan, CuCuCan}

Predictability from stems: post-tonic vowel alternations

- Recall that due to post-tonic vowel reduction...

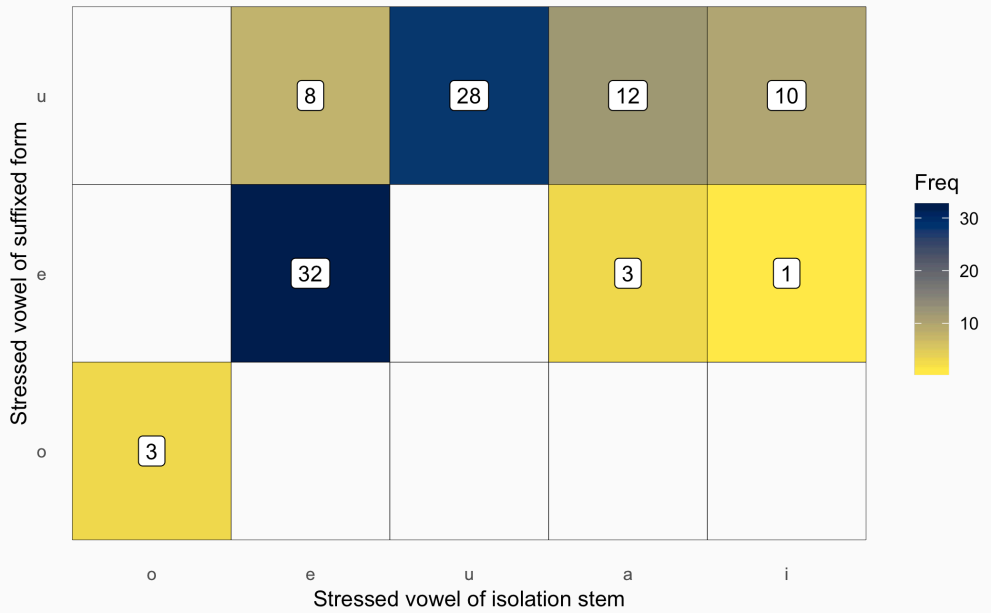
STEM		SUFFIXED
CVCuC	∼	{CuCeCan, CuCoCan, CuCuCan}

- But, identity of vowel is predictable via **VOWEL MATCHING**

if		*then*	
	potus		putosan
	petus		putesan
	p{u,a,i}tus		putusan

⇒a speaker can predict, with relatively high accuracy, what a post-tonic [u] will surface as in suffixed forms.

Predictability from stems: post-tonic vowel alternations

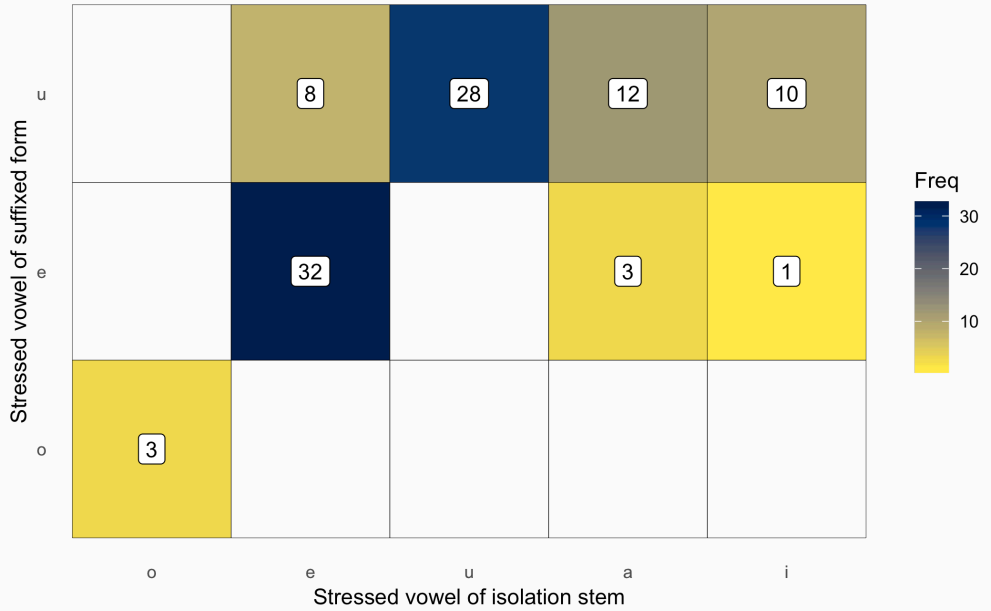


For example....

- ['putus] always surfaces as [pu'tusan] (∼28/28=100%)
- ['petus] likely surfaces as [pu'tesan] (∼32/40=80%)

Figure 1: How reduced [u] of non-suffixed CVCuC is realised when stressed under suffixation

Predictability from stems: post-tonic vowel alternations



For example....

- ['putus] always surfaces as [pu'tusan] (∼28/28=100%)
- ['petus] likely surfaces as [pu'tesan] (∼32/40=80%)

Figure 1: How reduced [u] of non-suffixed CVCuC is realised when stressed under suffixation

Note: [o] appears to be marginal in the lexicon.

Predictability from stems: final consonant alternations

- Due to final consonant neutralization, final [t͡s, k, n, ŋ] show the following alternations

STEM		SUFFIXED	STEM		SUFFIXED
[t͡s]	∼	[t, d, t͡s]	[n]	∼	[l, n]
[k]	∼	[p, b, k]	[ŋ]	∼	[m, ŋ]

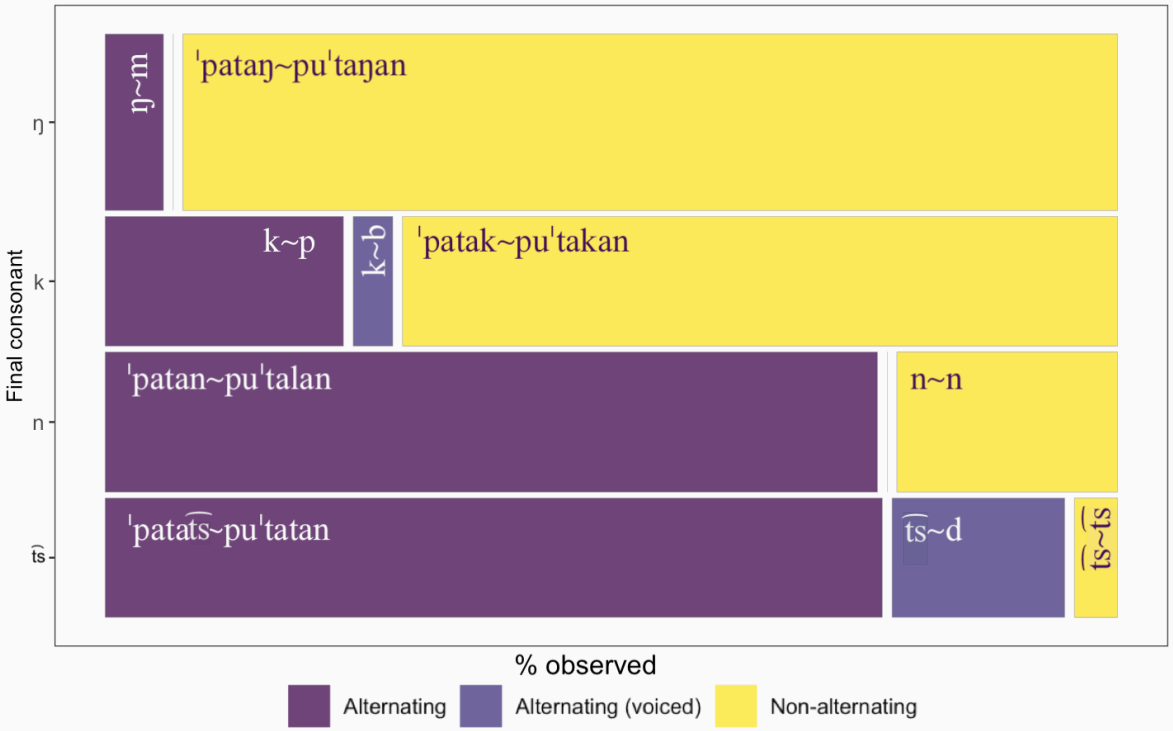
Predictability from stems: final consonant alternations

- Due to final consonant neutralization, final [t͡s, k, n, ŋ] show the following alternations

STEM		SUFFIXED	STEM		SUFFIXED
[t͡s]	∼	[t, d, t͡s]	[n]	∼	[l, n]
[k]	∼	[p, b, k]	[ŋ]	∼	[m, ŋ]

- Final consonants tend to **almost always** or **almost never** alternate
- Given a novel stem, (non-)alternation is relatively predictable.

Predictability from stems: final consonant alternations



For example...

[ˈpatiŋ]→[puˈtiŋan]
(32/35, 91%)

[ˈpati͡ts]→[puˈtitan]
(16/21, 76%)

Predictability from stems: summary

- Given a non-suffixed stem, it is impossible to perfectly predict the alternation of (i) [u] in post-tonic closed syllables, (ii) stem-final vowels and consonants ([t͡s, n, k, ŋ, g]).

Predictability from stems: summary

- Given a non-suffixed stem, it is impossible to perfectly predict the alternation of (i) [u] in post-tonic closed syllables, (ii) stem-final vowels and consonants ([t͡s, n, k, ŋ, g]).
- However, these alternations are **highly predictable from just the stem form** due to statistical regularities.

Predictability from stems: summary

- Given a non-suffixed stem, it is impossible to perfectly predict the alternation of (i) [u] in post-tonic closed syllables, (ii) stem-final vowels and consonants ([t͡s, n, k, ŋ, g]).
- However, these alternations are **highly predictable from just the stem form** due to statistical regularities.
- How about the other direction; will **suffixed forms** be a good base?

Predictability from suffixed forms

Given the suffixed form of a verb...

- Final consonants and vowels are completely predictable.

Predictability from suffixed forms

Given the suffixed form of a verb...

- Final consonants and vowels are completely predictable.
- **However**, the antipenultimate vowel of the stem is always neutralized due to pretonic VR

[pu'tim-an] → {'patiŋ, 'pitiŋ, 'petiŋ, 'potiŋ,'putiŋ}

Predictability from suffixed forms

Given the suffixed form of a verb...

- Final consonants and vowels are completely predictable.
- **However**, the antipenultimate vowel of the stem is always neutralized due to pretonic VR

[pu'tim-an] → {'patiŋ, 'pitiŋ, 'petiŋ, 'potiŋ,'putiŋ}

Compared to the neutralizing processes discussed so far, the patterns of predictability that would allow speakers to 'undo' pretonic VR are much weaker

Predictability from suffixed forms

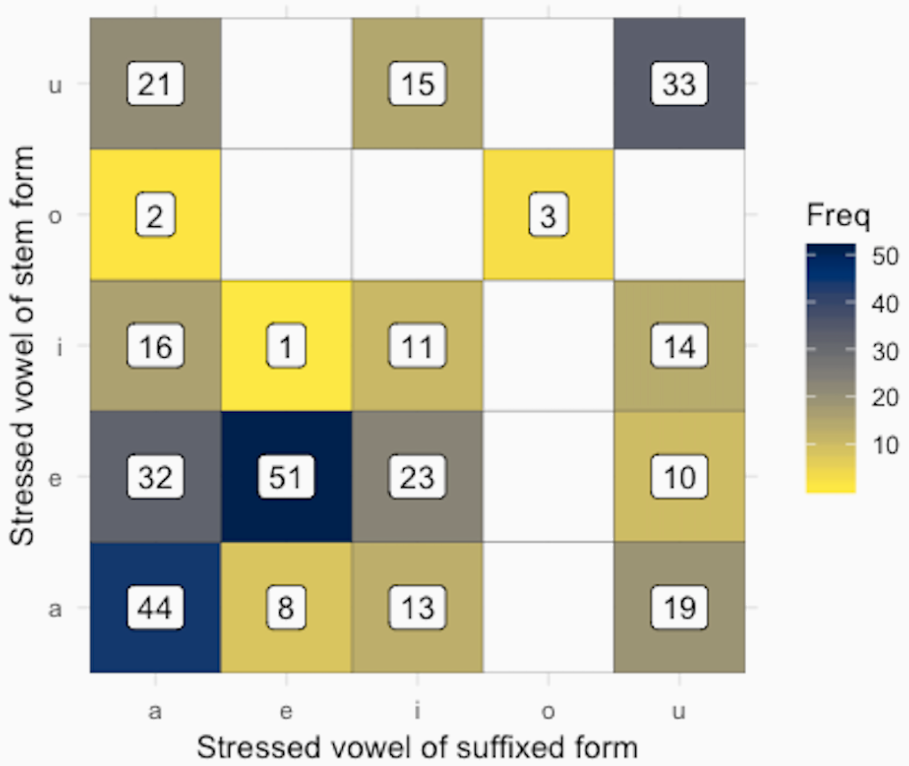


Figure 2: Distribution of stressed vowels in non-monosyllabic suffixed forms

Predictability from suffixed forms

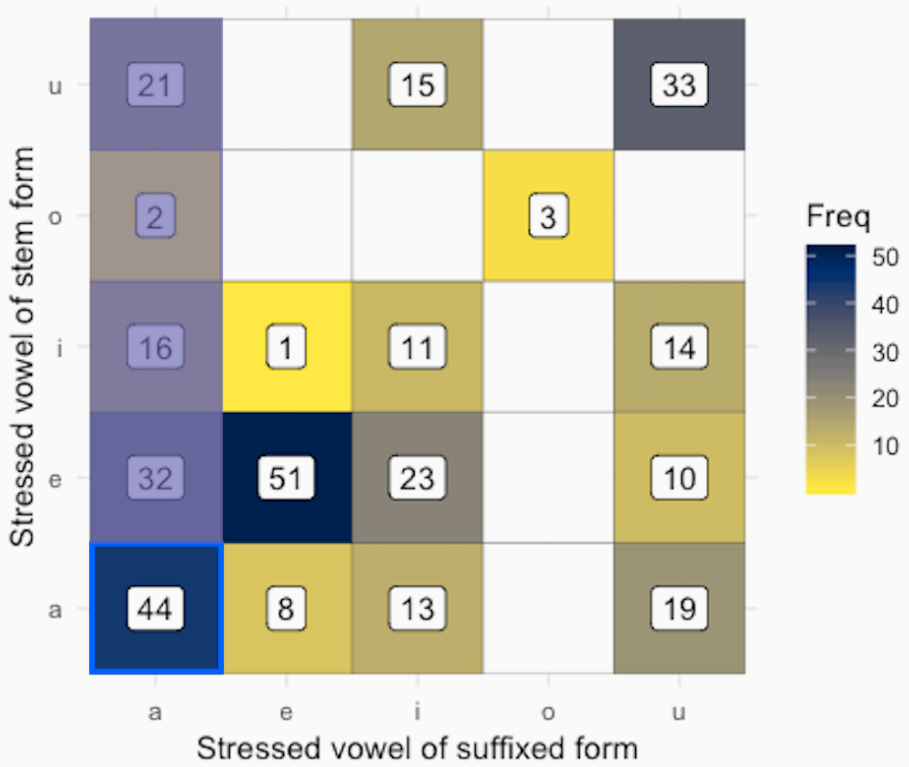


Figure 3: Distribution of stressed vowels in non-monosyllabic suffixed forms

For example...

- Given the form [pu'tasan], the most likely stem form is ['patas]. However, this is correct only 38% of the time (44/115)

Predictability from suffixed forms, cont.

- To undo pretonic VR, even picking the 'most likely' option based on statistical distributions would only result in correct predictions for 181/316 relevant forms (49%).

Predictability from suffixed forms, cont.

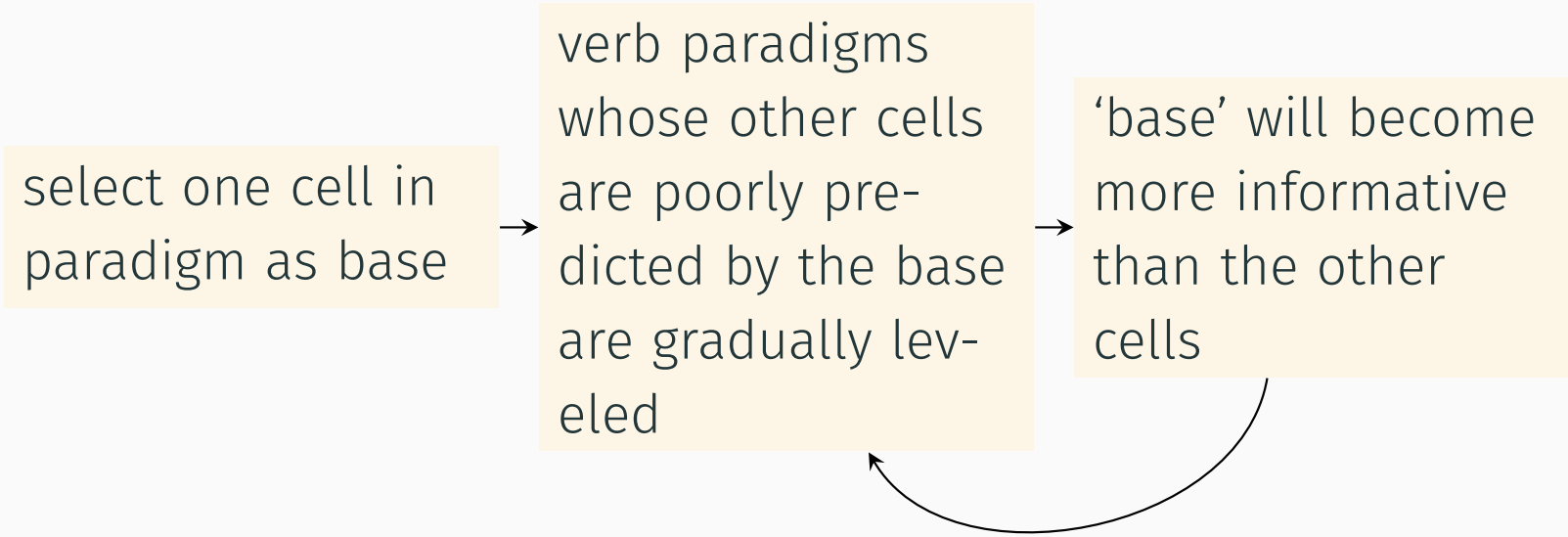
- To undo pretonic VR, even picking the 'most likely' option based on statistical distributions would only result in correct predictions for 181/316 relevant forms (49%).
- pretonic VR also affects more forms than the processes which cause loss of contrasts in the stem (336/340).

Interim summary

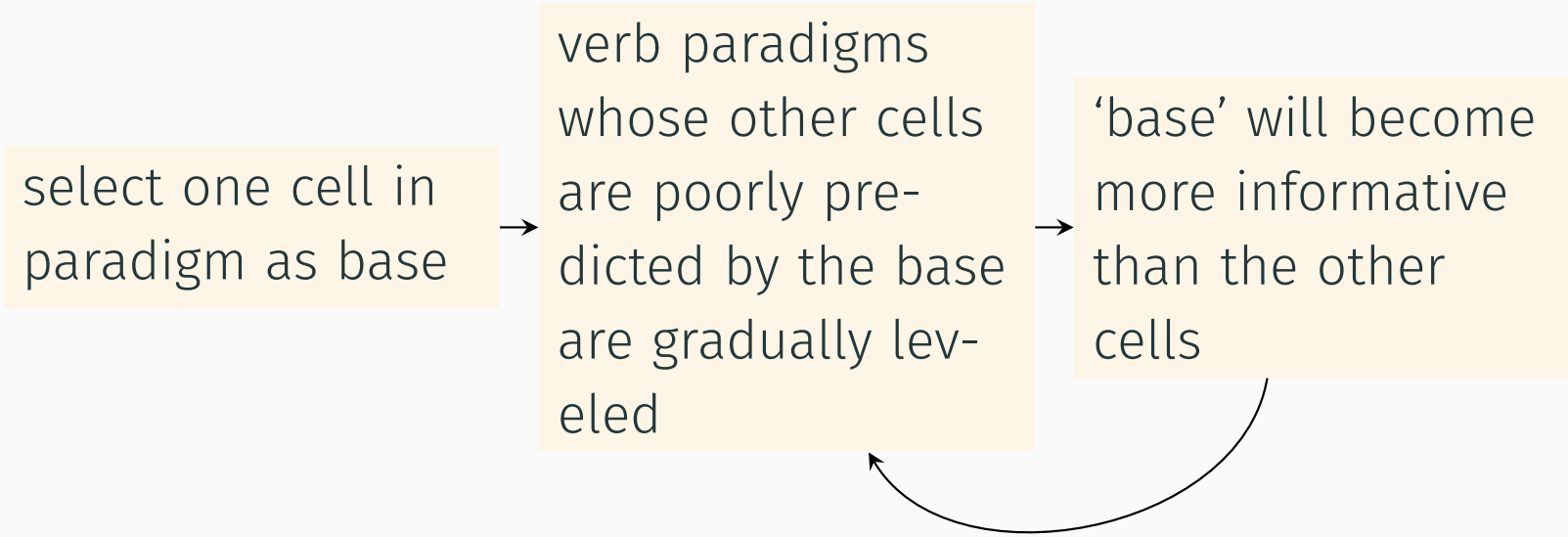
Asymmetry in informativeness of stem vs. suffixed forms, where stem forms are much more informative.

How does this asymmetry support the single surface base hypothesis?

Statistical asymmetries as evidence for base-driven reanalysis



Statistical asymmetries as evidence for base-driven reanalysis



⇒if one cell in a paradigm is much **more informative** than the others, and this asymmetry **cannot** be attributed just to phonological neutralization processes, restructuring from a single base form has likely happened.

Modeling

- Rule-based model confirms the stem/suffix asymmetry.
- Evaluation of models against a simulated lexicon provides more indirect evidence for base-driven restructuring of paradigms.

Model implementation

- Takes surface forms as input, and attempts to derive the other slots of the paradigm using phonological rules.
- based off of Minimal Generalization Learner (Albright and Hayes, 2003).

Model implementation

- Takes surface forms as input, and attempts to derive the other slots of the paradigm using phonological rules.
- based off of Minimal Generalization Learner (Albright and Hayes, 2003).
- Explicit algorithm for quantifying the informativeness of bases.
- Each stem-suffix pair in the lexicon is assigned a score, reflecting how well the model predicts the output.

Model evaluation

Compare how two models, stem-base vs. suffix-base, perform on the Seediq corpus.

The model (stem- vs. suffix-base) which assigns higher scores...

- better captures the lexicon.
- has the more informative base.

Stem vs. suffix base model

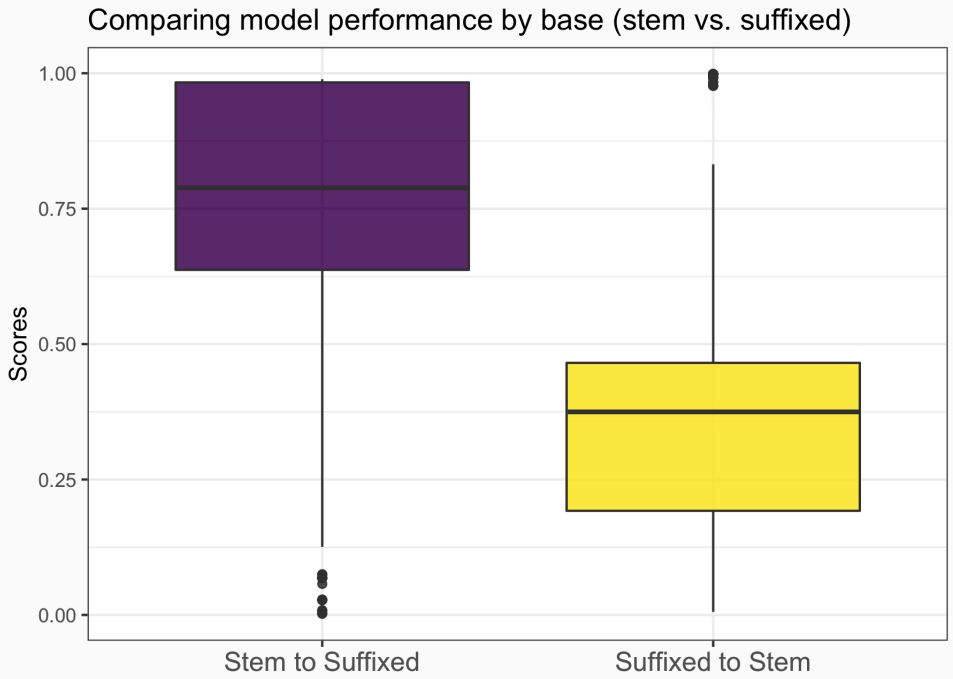


Figure 4: Performance of stem vs. suffix-base models

Indirect evidence for historical reanalysis

- Model results confirm stem-suffix asymmetry
- The stem form is a good base in part because neutralised segments either almost always or almost never alternate.

Indirect evidence for historical reanalysis

- Model results confirm stem-suffix asymmetry
- The stem form is a good base in part because neutralised segments either almost always or almost never alternate.
- Notably, this could be either due to
  - historical reanalysis exaggerating patterns of predictability
  - accidental effect of baseline phonotactic preferences
    e.g. final [t͡s] strongly prefers to alternate with [t]; this may be because there's a strong baseline phonotactic preference for [t] (relative to [t͡s]).

Indirect evidence for historical reanalysis

To account for this, test the two surface-base models were tested against a simulated lexicon

- rates of alternation are determined by relative frequencies of sounds in the Seediq lexicon.
- If the stem-base model performs equally well on the real and simulated data, then stem-suffix asymmetry can be attributed to phonotactic preferences.

Indirect evidence for historical reanalysis

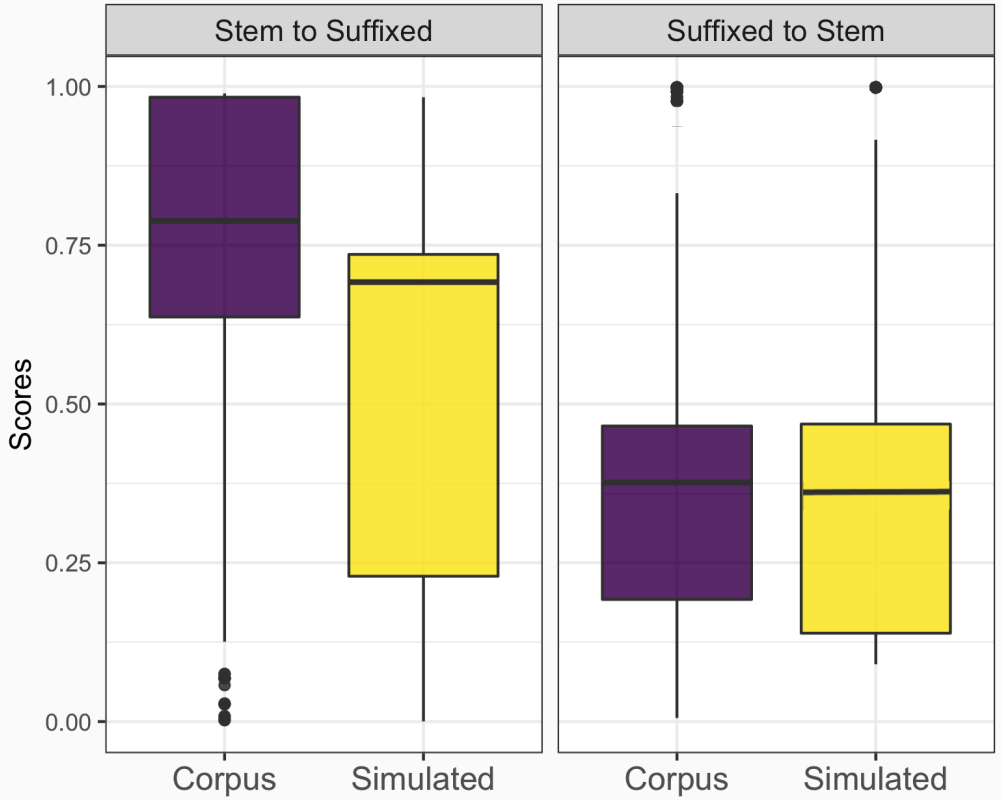


Figure 5: Model performance using real vs. simulated lexicon

Discussion

Conclusion

Non-suffixed forms of a Seediq paradigm are much more informative than the suffixed forms.

Conclusion

Non-suffixed forms of a Seediq paradigm are much more informative than the suffixed forms.

Modeling results suggest that this asymmetry cannot be explained by baseline phonotactic preferences.

Conclusion

Non-suffixed forms of a Seediq paradigm are much more informative than the suffixed forms.

Modeling results suggest that this asymmetry cannot be explained by baseline phonotactic preferences.

These results are ...

- puzzling under the cobbled UR approach, which makes no predictions about the direction of restructuring
- Expected under the single surface base approach, where restructuring from a base **exaggerates asymmetries in the data.**

Further testing

What other sources of evidence could there be for base-driven alternations in Seediq?

- Extensive historical evidence
- Productivity testing, to see if speakers apply (or don't apply) alternations as predicted by the surface base hypothesis.
  ⇒**Work in progress**

Thank you!

First, thank you to my three Seediq, consultants, 黃美玉, 陳玉妹, 謝芸薇, for their time and invaluable knowledge. Many thanks to Bruce Hayes, Kie Zuraw, and Claire Moore-Cantwell guidance on all aspects of this project. Thanks also to the UCLA Phonology seminar for much helpful discussion.

References i

Adam Albright. A restricted model of ur discovery: Evidence from lakhota. *Ms, University of California at Santa Cruz*, 2002a.

Adam Albright. Base-driven leveling in yiddish verb paradigms. *Natural Language & Linguistic Theory*, 28(3):475–537, 2010.

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, 2002b.

Simon J Greenhill, Robert Blust, and Russell D Gray. The austronesian basic vocabulary database: from bioinformatics to lexomics. *Evolutionary Bioinformatics*, 4:271–283, 2008.

References ii

Arthur Holmer. *A parametric grammar of Seediq*. PhD thesis, Lund University, 1996.

Yoonjung Kang. Neutralizations and variations in korean verbal paradigms. *Harvard Studies in Korean Linguistics*, 11:183–196, 2006.

Michael Kenstowicz and Larry M Kisseberth. Topics in phonological theory, 1977.

Paul Jen-kui Li. Reconstruction of proto-atayalic phonology. *Bulletin of the Institute of History and Philology*, 52, 1981.

Huang Mei-jin, Yu-yang Liu, and Xin-sheng Wu. Taiwan aboriginal language e-dictionary. *Taiwan Journal of Indigenous Studies*, 7(2):73–118, 2014.

Andrei Mikheev. Automatic rule induction for unknown-word guessing. *Computational Linguistics*, 23(3):405–423, 1997.

References iii

Naomi Tsukida. Seediq. the austronesian languages of asia and madagascar, ed. by alexander adelaar and nikolaus p. himmelmann, 291–325, 2005.

Hsiu-fang Yang. The phonological structure of the paran dialect of sediq. *Bulletin of the Institute of History and Philology Academia Sinica*, 47(4): 611–706, 1976.

Irregular alternations i

1. Irregular vowel alternations (n=11)

STEM	SUFFIXED	GLOSS	EXPECTED SUFFIXED
ˈhuruc	huˈridan	‘come to a stop’	(huˈrudan, huˈredan)
ˈtebas	tuˈbesan	‘sieve grains’	(tuˈbasan)

2. Irregular final vowel deletion (n=5)

ˈhado	ˈhadan	‘deliver’	(huˈdawan)
ˈqene	ˈqenan	‘extend’	(quˈneyan)

3. Non-alternating pairs (n=2)

ˈt͡saman	ˈt͡saman	‘pass the night’	(t͡suˈman-an, t͡suˈmalan)

4. [n]-insertion (n=3)

ˈqeya	quˈyan-an	‘hang’	(quˈya-an)

Comparing the two approaches

Cobbled UR approach

✓Phonotactically motivated markedness constraints or rules, which are (nearly) exceptionless.

✓Empirical predictions about range of possible alternations.

✗UR learning relatively difficult.

Surface-base approach

✗Some alternations can't be explained by general markedness; many exceptions

✓Evidence from historical change and child speech errors (e.g. Kang, 2006; Albright, 2010).

✓UR learning relatively easy.

Base-driven restructuring: a Seediq example

Statistical patterns in the modern Seediq lexicon reflect a strong dispreference for the stem-final **[ŋ]-[m] alternation**.

- Older system of Seediq with relatively more symmetrical distribution of segments
  - Dispreference for alternation → **weaker statistical tendency.**

Base-driven restructuring: a Seediq example

Statistical patterns in the modern Seediq lexicon reflect a strong dispreference for the stem-final **[ŋ]-[m] alternation**.

- Older system of Seediq with relatively more symmetrical distribution of segments
  - Dispreference for alternation → **weaker statistical tendency.**
- paradigms which showed the dispreferred [ŋ]-[m] alternation would gradually have been restructured, resulting in the **very skewed rates of alternation** that we see today.

Base-driven restructuring: a Seediq example

Statistical patterns in the modern Seediq lexicon reflect a strong dispreference for the stem-final **[ŋ]-[m] alternation**.

- Older system of Seediq with relatively more symmetrical distribution of segments
  - Dispreference for alternation → **weaker statistical tendency.**
- paradigms which showed the dispreferred [ŋ]-[m] alternation would gradually have been restructured, resulting in the **very skewed rates of alternation** that we see today.

One example (elicited) suggesting this type of reanalysis:

ˈlauŋ∼luˈuŋan (<*l-um-aum) ‘to burn’
(Li, 1981; Greenhill et al., 2008)

Model evaluation

Examples of rules in the stem-base model

	Name	Rule	Example	p (H/S)	$\hat{p}$
(a)	Pret. VR	$\begin{bmatrix} +\text{syl} \\ -\text{stress} \end{bmatrix} \rightarrow$ [u] / #C_	'patuk→pu'tukan	1.0 (265/265)	0.99
(b)	Pret. V-del.	$\begin{bmatrix} +\text{syl} \\ -\text{stress} \end{bmatrix} \rightarrow \emptyset$ / #_	'awak→'wakan	1.0 (36/36)	0.95
(c)	ŋ-to-m	[ŋ] → [m] / _]$_{stem}$V	'geruŋ→gu'reman	0.06 (2/34)	0.02
(d)	ruy-to-rig	[ruy] → [rig] / _]$_{stem}$V	'baruy→bu'rigan	1.0 (3/3)	0.6

- Rules vary in scope (number of input forms that meet structural description) and hits (forms where application results in correct output).
- **Confidence (p)** is Hits/Scope.
- Based on Mikheev (1997), rules are evaluated on adjusted confidence ($\hat{p}$), i.e. penalized for less evidence (AKA low scope).

Model evaluation

Each rule has a **confidence value**, reflecting how accurate it is.

Model assigns a score to each stem/suffix pair in the input data:

- **Score:** product of confidence of all the rules needed to derive the correct output form.

The model (stem- vs. suffix-base) which assigns higher scores...

- better captures the lexicon.
- has the more informative base.

Simulated lexicon

To account for this, test the two surface-base models were tested against a simulated lexicon

- 700 verb paradigms
- rates of alternation are determined by relative frequencies of sounds in the Seediq lexicon (regardless of which position in a word they occur in)
- e.g. across the corpus of 340 paradigms, [ŋ] (n=104) is around 2.1 times more frequent than [m] (n=49). Corresponding to this, the [ŋ]-final forms in the simulated lexicon are 2.1 times more likely to *not* alternate (than to alternate with [m]).

Selection of non-suffixed form as base

Why was the non-suffixed form, rather than the suffixed form, designated as the base form?

- Albright (2002b): the base should be the "most informative", that (i) has the fewest lexical items affected by neutralization, and (ii) suffers from the fewest neutralizations

Selection of non-suffixed form as base

(i) neutralizing processes affect the fewest lexical items

- **True: 336/340 suffixed forms** are affected by pretonic VR, while **287/340 non-suffixed forms** are affected by post-tonic VR and/or other final neutralization processes.

Selection of non-suffixed form as base

(ii) suffers from the fewest phonological neutralization processes

- **Not intuitively true;** non-suffixed forms are affected by more neutralizing processes (post-tonic VR and final consonant neutralization).
- Historical evidence suggests that pre-tonic VR occurred **prior to all of the post-tonic neutralization processes** (Li, 1981, 239).
- It is likely that at some point after pretonic neutralization, the non-suffixed forms of the Seediq verb paradigm had become much more informative than the suffixed forms.
- 'tipping point' for restructuring of paradigms.

Productivity of base-driven alternations

Results predict that speakers will be able to productively apply statistically preferred alternations when given novel stem forms.

NOVEL STEM	EXPECTED SUFFIX FORM
ˈpetus	puˈtesan (vowel matching)
ˈpatac	puˈtatan ([t͡s]-[t] alternation)
ˈpataŋ	puˈtaŋan (no [ŋ]-[m] alternation)

Productivity of base-driven alternations

Results predict that speakers will be able to productively apply statistically preferred alternations when given novel stem forms.

NOVEL STEM	EXPECTED SUFFIX FORM
ˈpetus	puˈtesan (vowel matching)
ˈpatac	puˈtatan ([t͡s]-[t] alternation)
ˈpataŋ	puˈtaŋan (**no** [ŋ]-[m] alternation)

Is this the case?

- tentative support from pilot 'paradigm-gap' tests.
- Current work in progress: more extensive testing.