Evidence for base-driven alternation in Tgdaya Seediq

Jennifer Kuo, *University of California, Los Angeles*
jenniferkuo2018@ucla.edu

1 Overview

UR discovery: Two approaches

- **"Cobbled" URs (Kenstowicz and Kisseberth, 1977):** maximally informative URs
  - → Determine which slots in paradigm reveal underlying contrast(s), **'cobble'** these together to set up UR.
  - → UR discovery is harder, but resulting grammar is simple.
- **Surface bases (Albright, 2002, et seq.):** input to morphophonology is a **single** surface form.
  - → Pick a slot in the paradigm to be the base, and project other slots using this base.
  - → UR discovery is easier, but resulting grammar is more complex, requires exceptions.

Current study: Tgdaya Seediq

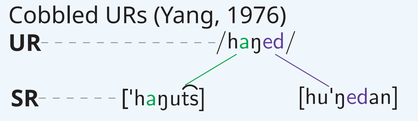
- Seediq (iso:trv) is an Austronesian language spoken in Taiwan.
- Extensive alternations in verb paradigms make it a good test case for comparing theories of morphophonology.
- **Finding:** Asymmetries in Seediq lexicon support the surface base approach.

3 Two solutions

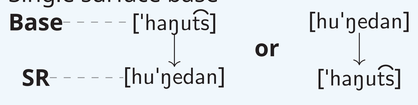
Given a paradigm of this sort…

STEM	SUFFIXED	
'haŋuts	'huŋedan	'to cook'

Cobbled URs (Yang, 1976)



Single surface base



Predictions about errors/ reanalyses:

- Cobbled URs: reanalyses from both stem and suffixed forms.
- Surface base: reanalyses will **always** be from the base.

References

https://tinyurl.com/y25t4gyn

Acknowledgements

Thanks to Bruce Hayes, Kie Zuraw, Claire Moore-Cantwell, members of the UCLA Phonology Seminarm for their helpful comments; Huang mei-yu and other Seediq consultants for their invaluable time and assistance.

2 Sources of alternation in Seediq

Neutralization from vowel reduction:

- Stress is always penultimate; suffixation shifts stress rightwards.
- Pretonically:

STEM	SUFFIXED	DESCRIPTION
'atik,'utik,'etik…	'tikan	Onsetless vowels delete
'pahik, 'puhik,'pehik…	piˈhikan	Assimilate if separated by /h,ʔ/
'patik, 'petik, 'putik…	puˈtikan	Else, reduce to [u]

  → Result: Neutralization of contrast in **suffixed forms.**

- Post-tonically:

STEM	SUFFIXED	DESCRIPTION
'patuk	puˈtekan,puˈtokan,puˈtukan	/e,o,u/ → [u] in closed syl.

  → Result: neutralization of contrast in **isolation stems**

Final consonant neutralization:

- Many processes of word-final consonant neutralization, some examples listed:

STEM	SUFFIXED	DESCRIPTION
'patik	puˈtikan,puˈtipan	/p/, /b/, /k/ → [k]
'patic	puˈtitan, puˈtidan, puˈtican	/t/, /d/, /t͡s/ → [t͡s]
'patiŋ	puˈtiŋan,puˈtiman	/m/, /ŋ/ → [ŋ]

  → Result: neutralization of contrast of **isolation stems**

Overall: All forms of a paradigm to suffer from neutralization

4 Predictability from stem

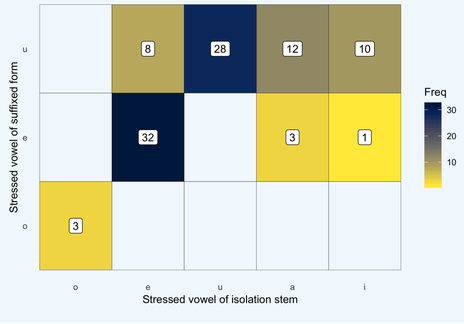
Despite apparent ambiguity, patterns in lexicon make it so that **suffixed forms are highly predictable from stem (non-suffixed) forms**

Data: 340 paradigms, taken from both online dictionary data and personal fieldwork.

Predicting vowel alternations

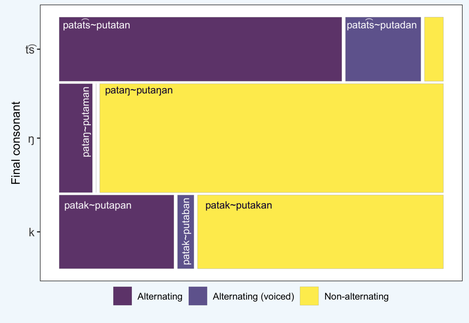
- Due to post-tonic vowel reduction…
  CVCuC∼{CuCeCan, CuCoCan, CuCuCan}
- But, identity of vowel in suffixed form is predictable via **"vowel matching":**

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



Predicting consonant alternations

- Most final alternations either:
  - almost always occur (t͡s∼t)
  - almost never occur (ŋ∼m)
- Result: a speaker can predict with almost perfect accuracy whether or not a final consonant will alternate.



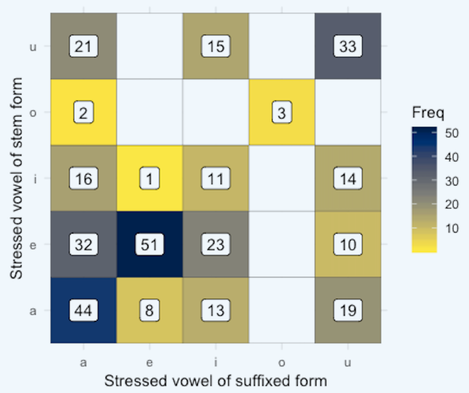
6 Conclusion

- Seediq suffixed forms are highly predictable from their stems.
- Asymmetries in Seediq lexicon suggest that **reanalysis from a stem base** has occurred.
  - Unexpected under the cobbled UR approach.
  - Natural result of Albright's surface-base approach, assuming that speakers have designated the stem form as base.
- Ongoing: wug-testing

5 Predictability from suffixed forms

Notably, **stems are not as predictable from suffixed forms** (i.e. suffixed forms are **less informative**

- In the suffixed forms of a verb, the **penultimate vowel of the stem** is always neutralized due to **pretonic VR.**



- Patterns of predictability for 'undoing' pretonic VR are relatively weak.
  - e.g. [puˈtasan] most likely has the stem ['patas]. However, this is correct only 38% of the time (44/115)
  - Overall, picking the 'most likely' option correctly predicts **181/316 forms (49%)**.
- pretonic VR also **affects more forms** than the neutralizing processes which affected the stem.

6 More evidence from modeling

Rule-based models confirm **stem-suffix asymmetry**, which can be better explained under the surface-base approach.

Implementation: a model for surface-base learning

- Rule-based model (cf. Minimal Generalization Learner, Albright and Hayes, 2003)
- Takes a surface form as base, derive other forms of the paradigm with a series of **rules**.

Model Evaluation

- **Rules** evaluated using adjusted confidence:
  - **Confidence:** % of forms where rule application results in correct output (≈ accuracy)
  - **Adjusted confidence** (Mikheev, 1997): penalizes rules that have less evidence
- **Lexical items** are given a **'score'** (≈ well-formedness) based on the adjusted confidence of the rules applied to them.
- **"Better"** model assigns **higher scores** to the lexical data.

Data

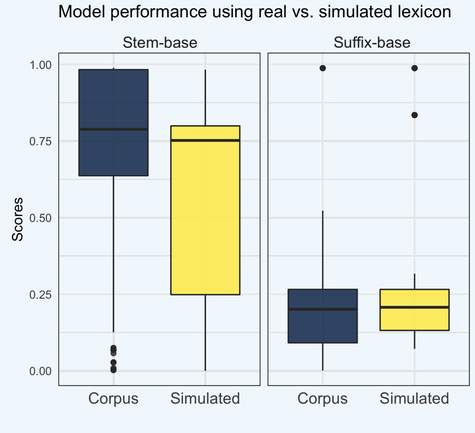
Compared **two models**: Stem-base vs. Suffix-base
Tested **two "lexicons"**:

- REAL: 342 existing Seediq paradigms
- SIMULATED: 700 paradigms, where rates of alternation are determined by baseline frequencies of sounds in Seediq lexicon.

Model Results

- **Comparing models:** 'Stem to Suffixed' model (where **stem** is the base) performs much better than the 'Suffixed to Stem' model.
- **Comparing "lexicons":** The 'Stem to Suffixed' model does much worse on the SIMULATED set.

  ⇒ Asymmetry suggests that Seediq speakers have reanalyzed verb paradigms to be **predictable from stem**.



References

Albright, A. (2002). *The identification of bases in morphological paradigms*. PhD thesis, University of California, Los Angeles.

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

Kenstowicz, M. and Kisseberth, L. M. (1977). Topics in phonological theory.

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

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