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 (Chomsky and Halle, 1968): Derive surface contrasts from underlying distinctions.
  → Determine which slots in paradigm reveal underlying contrast(s), (cobbled) these together to set up UR.
  → UR discovery is harder, but resulting grammar is simpler.

• Surface bases (Albright, 2002, et seq.): Learners base UR on 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 Northeastern Taiwan.
• Extensive alternations in verbal paradigms make it a good test case for comparing theories of UR learning.

• Finding: Asymmetries in Seediq lexicon support the Albrightian surface base approach.

2 Neutralization in Seediq
Neutralization from vowel reduction:
• Stress is always penultimate; suffixation shifts stress rightwards.
• Pretonically:
  STEM  SUFFIXED  DESCRIPTION
  'pahik', 'pahik', 'pahik'... 'pahikan'  Assemble if separated by /h/.
  'patik', 'petik', 'petic'... 'punitkan'  Else, reduce to [u]
  → Result: Neutralization of contrast in suffixed forms.

• Post-tonically:
  STEM  SUFFIXED  DESCRIPTION
  'patik'  pu'tikan, pu'tikan, pu'tikan /e, o, u/ → [u] in closed syl.
  'pato'  pu'tawan, pu'toan /aw/ → [o]
  → Result: neutralization of contrast in isolation stems

Final consonant neutralization:
• Many processes of word-final consonant neutralization, some examples listed:
  STEM  SUFFIXED  DESCRIPTION
  'patik'  pu'tikan, pu'tikan  /p/, /b/, /k/ → [k]
  'patic'  pu'titan, pu'tidan, pu'tican  /d/, /t/, /c/ → [ts]
  'patiN'  pu'tinan, pu'timan  /m/, /n/ → [ŋ]
  → Result: neutralization of contrast of isolation stems

Overall: All forms of a paradigm to suffer from neutralization

3 Two solutions

• Given a paradigm of this sort...
  STEM  SUFFIXED
  'huNgardan' 'huNgardan' 'to cook'

• Cobbled URs (Yang, 1976)

UR: \\

SR: \\

• Albrightian surface base

Base  [huNgardan]  ['huNgardan']

SR: [huNgardan]  ['huNgardan']

4 Predictability from stem

Despite apparent ambiguity, statistical regularities in lexicon make it so that suffixed forms are highly predictable from non-suffixed forms (e.g. stem)

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 potus then putusan

  p(u, a, i) tus

Predicting consonant alternations:
• Most final alternations either:
  − always occur (c→t)
  − almost never occur (n→m)

  Result: a speaker can predict with almost perfect accuracy whether or not a final consonant will alternate.

5 More evidence from modeling
Models of surface-base learning reveal asymmetries (in stem vs. suffixed forms) which can be better explained under the Albrightian model.

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: proportion of forms where rule applies to give 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 to Suffix (stem is the base) vs. Suffix to Stem (suffix is 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 Suffix' model (where stem is the base) performs much better than the 'Suffix to Stem' model.

• Comparing lexicons: "Stem to Suffix" model does much worse on the SIMULATED set.

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

6 Conclusion

• Seediq suffixed forms are highly predictable from their stems.
• Asymmetries in Seediq lexicon suggest reanalysis towards the stem form of paradigms.
  - Unexpected under the cobbled UR approach.
  - Natural result of Albrightian approach, assuming that speakers have designated the stem form as base.

• Ongoing: wug-testing

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

PDF with references:
References


