Exploring the relation between alternation and phonotactics of Sundanese liquid assimilation and dissimilation

SWELL Workshop

Dec. 23rd

Yoona Yee

Outline

1. Introduction
2. Data
   - Alternations
   - Phonotactics
3. Previous Researches
4. Phonotactic Learning Simulation
5. Discussions
6. Conclusions

1. Introduction

- Sundanese, a Malayo-Polynesian language, shows both assimilation and dissimilation at the same time regarding the same segments, namely liquids /l/ and /r/ (Bennett 2015, Cohn 1992, Stanton 2019, a.o.).
- **A plural affix /ar/** alternates between two surface forms: [ar] and [al].
- It is a prefix before a vowel initial stem (1a), and an infix when it attaches to a consonant initial stem (1b). ('=' denotes a morpheme boundary).

(1) Plural affix /ar/ in Sundanese (from Bennett 2015[1], citing Cohn 1992)

a. ar-VCVC	[**ar**=aj̵m]	'patient (pl.)
b. C-ar-VCVC(V)	[k=**ar**=usut]	'messy (pl.)'
	[g=**ar**=ilis]	'beautiful (pl.)'

1. Introduction

- This plural affix /ar/ shows [l]~[r] alternation which arises by assimilation or dissimilation.
  - When the affix is preceded by /l/, it is realized as [al] as shown in (2), while it becomes [al] when another /r/ follows in the stem (3).

    (2) L-assimilation: /r/ assimilates to /l/, **after a preceding /l/**.
    - /ar/+/lɨtik/ [**l**=a**l**=ɨtik] 'little (pl.)'

    (3) R-dissimilation: /r/ dissimilates to /l/, **before another /r/**.
    - /ar/+/ŋumbara/ [ŋ=a**l**=umba**r**a] 'go abroad (pl.)'

- Sundanese lexicon is also known to mirror the alternation patterns (Cohn 1992, Stanton 2019, a.o.).
  - Liquids tend to not co-occur within a word.
  - If they do, it is only when they are onsets of adjacent syllables, in which case they are both /r/s or /l/s.

1. Introduction

- Aims of the paper:
  - ✓ To explore whether the observation made about the Sundanese lexicon in the previous literature is true.
  - ✓ To investigate whether Sundanese phonotactics regarding liquid segments is captured by a phonotactic learner UCLA phonotactic learner (Hayes and Wilson 2008).
  - ✓ To explore the relation between alternation and phonotactics of Sundanese liquid assimilation and dissimilation.

2. Data

- The plural affix /-ar-/ alternates between [ar] and [al], which is the only known context that creates a condition for liquid alternation.
- The affix /ar/ acts either as a prefix or an infix depending on whether the stem begins with a vowel or a consonant.
  - A prefix before V-initial stems: /ar=VCVC/ -> [ar=aj**ɨ**m] 'patient (pl.)'
  - Infix before C-initial stems: /C=ar=VCVC(V)/ -> [k=ar=usut] 'messy (pl.)'

➢ When this affix combines with a stem that also contains liquid(s), it shows either assimilation or dissimilation.

(6) L-assimilation: /r/ assimilates to /l/, **after a preceding /l/**.

a.	/ar/+/l**ɨ**tik/	[**l**=a**l**=**ɨ**tik]	'little (pl.)'

Cf. only root-initial /l/s result in assimilation:

b.	/ar/+/gilis/	[g=**ar**=ilis]/*[g=a**l**=i**l**is]	'beautiful (pl.)'
c.	/ar/+/ŋabedol/	[ŋa=b=a**r**=edo**l**]/*[ŋa-b=a**l**=edo**l**]	'pull in (pl.)'

(7) R-dissimilation: /r/ dissimilates to /l/, **before another /r/**.

a.	/ar/+/ŋumbara/	[ŋ=a**l**=umba**r**a]	'go abroad (pl.)'
b.	/ar/+/hormat/	[h=a**l**=o**r**mat]	'respect (pl.)'
c.	/ar/+/combrek/	[c=a**l**=omb**r**ek]	'cold (pl.)'
d.	/ʔ/+/ar/+/ulur/	[ʔ=a**l**=ulu**r**]	'lower on a rope (pl.)'

2. Data: alternations

- However, there is an exception to R-dissimilation in (7). When the two /r/s are onsets of adjacent syllables, dissimilation does not occur.

(8) When two /r/s are onsets of adjacent syllables: No R-dissimilation

a. /r=ar=ahɨt/ [**r**=a**r**=ahɨt] 'wounded (pl.)'
b. /c=ar=uriga/ [c=a**r**=u**r**iga] 'suspicious (pl.)'

- When there are two /r/s being onsets in adjacent syllables, i.e., so R-dissimilation is not supposed to occur, but at the same time /l/ precedes the affix, creating the environment for L-assimilation, L-assimilation takes place.

(9) When word initial onset is /l/ with the following /r/ in the stem: L-assimilation

a. /l=ar=itik/ [**l**-a**l**-itik] 'little (pl.)'
b. /l=ar=iren/ [**l**-a**l**-i**r**en] 'take a break (pl.)'

➢ Summary of alternations

- **R-dissimilation**: /r…r/ is dispreferred and is repaired by rendering it /l…r/, except when both /r/s are onsets of adjacent syllables.
- **L-assimilation**: /r/ in /ar/ becomes /l/ after a stem-initial (i.e. word initial) /l/. Even when there is another /r/ in the stem and that /r/ is in the adjacent to the syllable containing the suffix /r/, /l/-assimilation still takes place.

2. Data: phonotactics

- /l/~/r/ alternation does not occur within a stem.

(10) No /l/~/r/ alternation within a stem

a.	/liren/	[**li**r**e**n]	'take a break'
b.	/radar/	[**r**ada**r**]	'radar'
c.	/restoran/	[**r**esto**r**an]	'restaurant'

➢ Several previous researches state that **static distributions of liquids in Sundanese lexicon parallel the alternation patterns** laid out in the previous subsection (Cohn 1992, Stanton 2019).

3. Previous research

- Stanton (2019): Sundanese liquid alternation is a type of aggressive reduplication such as suggested by Zuraw (2000).
  - Aggressive reduplication: Phonologically similar adjacent syllables tend to get more similar.
  - Liquid dissimilation occurs, which is though masked by aggressive reduplication in adjacent syllables.
    - When the adjacent syllables both have liquid onsets, they become identical liquids, which is why L-assimilation occurs and R-dissimilation does not occur when the /r/s are the adjacent syllable onsets.
  - To further support her claim, she probed into the tendency in Sundanese lexicon to have identical onsets when they are in adjacent syllables.
    - Dictionary data: Lembaga Basa & Sastra Sunda (1985) - approximately 16,000 words
    - A log linear regression on adjacent onsets in disyllabic and tri-syllabic words.

3. Previous research

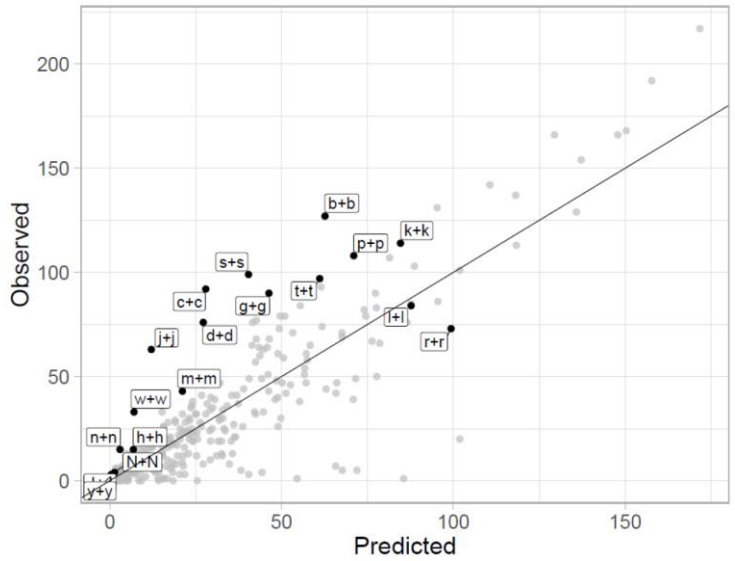

Figure 1. Predicted vs. observed frequencies of $\sigma_1\sigma_2$ onset pairs (from Stanton 2019)

- Adjacent identical onsets in the first and second syllables are over-attested, while /r+r/ and /l+l/ sequences are under-attested.
  - ➢ Adjacent syllable identity, i.e., patterns that conform to aggressive reduplication, is indeed observed throughout the Sundanese lexicon.
- It also supports the constraints on the /l/~/r/ alternation, namely */l…l/ and */r…r/.
  - ➢ Underattested /l…l/ and /r…r/ sequences in the lexicon.

3. Previous research

- Summary of Stanton (2019)
  - Her main interest lies in capturing the tendency of aggressive reduplication of adjacent syllables.
  - It shows that /l…l/ and /r…r/ are under-attested, which according to her parallels dispreference in alternations for identical liquids in a word, except for when they are onsets of adjacent syllables.

3. Previous research

- Cohn (1992) investigated all the /l/-initial and /r/-initial words in the dictionary.

- Question: Are multiple liquids, including not only identical liquids within words but also different liquids such as /l…r/ and /r…l/, under-attested?

- Conclusions: roots with two /r/s in the native lexicon almost does not exist, unless they are a phonological copy of the first syllable. In addition, there are no two /r/s attested within a syllable (e.g. */rer/).

3. Previous research

- Of approximately 960 /r/-initial entries that she searched, 105 have more than one /r/, and those 105 cases fall into the two patterns.

(12) Phonotactic distribution on /r/s (from Cohn 1992)

- Copying of first syllable (87 cases in total)
  a. rara ‘braid’
- Recent borrowings (18 cases in total)
  b. radar ‘radar’

- She also describes that there is also a restriction on /l/ and /r/, based on 960 /r/-initial forms and 990 /l/-initial forms that she searched.

(13) Phonotactic constraints on nonidentical liquids in a word

- rVlV (4 cases)
  a. rəlatip ‘relative’
- lVrV (14 cases in total with 4 recent borrowings)
  b. lori ‘small train’
- Words that have an alternate form (12 cases)
  c. loris ~ roris ‘check’

=> Forms containing non-identical liquids (/# [αlateral] V [-αlateral] V …/) are rare, and when they occur, they often have an alternate form of the shape /rVrV…/.

3. Previous research

- Summary of Cohn (1992)
  - Identical liquids are avoided within a word (both in mono-morphemic and morphologically complex forms), unless they are onsets of adjacent syllables.
  - Occurrences of multiple liquids in a word including non-identical liquid pairs (i.e., /l…r/ & /r…l/) are rare altogether.
  - This parallels the patterns observed in the allophonic alternation of the plural affix.

    ⇒ The constraints on the distributions of two /r/s and /r/ and /l/s hold much more generally than just in the context of the plural affix /ar/.

- My impression on Sundanese liquid phonotactics:
  - Quite many words contain multiple liquids, no matter whether they are identical or non-identical: /l…l/, /r…r/, /l…r/, /r…l/.

4. Phonotactic learning

- It would be worthwhile to investigate what constraints there are in the Sundanese lexicon, to see if liquids' phonotactic patterns match the alternation or not.
  - Using a phonotactic learner would be a good way to do this.
- I will try phonotactic learning using a UCLA phonotactic learner (Hayes and Wilson 2008), with the dictionary data of Lembaga Basa & Sastra Sunda (1985), which contains about 16,300 words.
- How is the phonotactic patterns different from the alternations?
  - Is co-occurrence of liquids – whether they are identical or not - rare within words?
  - When liquids do co-occur within words, do they tend to be non-identical, unless they are onsets of adjacent syllables?
    - Is there R-dissimilation in lexicon such as that in alternations?
  - Are these patterns captured by the learner?

4. Phonotactic learning

- Learning data
  - Data source: Lembaga Basa & Sastra Sunda (1985)-a dictionary containing 16,300 words, which is the most extensive data of Sundanese.
  - Scope of the data:
    - Reduplicated words have been excluded, as they may distort the learning results.
      - 16,328 words → 15,878 words (450 words deleted)
    - Morphological complexity of the word was not considered.
      - The data were not limited to mono-morphemic words to see if liquid dissimilation and assimilation holds across morpheme boundaries.

4. Phonotactic learning

- Learner Settings
  - Maximum number of constraints: 100, 180, or non-limited
  - Maximum Gram Size: 2
  - Maximum Gram Size of an Onset Tier: **4**
  - Max OE: 0.3
  - Allow complements: Yes
  - Onset tier setting: Onset +onset, +sonorant: lateral, word_boundary
- Learned constraints

Constraint	Tier	Weight
***[-word_boundary][-lateral][-word_boundary][-lateral]**	(tier=Onset)	0.1

- 175th out of 180 constraints
- Weak R-dissimilation among /r/s in non-adjacent syllables, which are in a non word-initial position.
- It matches the alternation data.

4. Phonotactic learning

Rank	Constraints	Tier	Weight
96	***[+lateral][-word_boundary][+lateral]**	(tier=Onset)	0.491
97	***[-lateral][-word_boundary][-lateral]**	(tier=Onset)	0.479

- 96th & 97th out of 100 constraints
- Liquid dissimilation (both /l/s and /r/s) in non-adjacent syllables
- */l…l/, */r…r/, which matches the alternation data

Rank	Constraints	Tier	Weight
96	***[+lateral][-lateral][-word_boundary]**	(tier=Onset)	0.488

- 96th out of 100 constraints
- L-assimilation or R-assimilation in adjacent syllables */l…r/, in a non word-final position

5. Discussions

- Currently, the study is looking at all words including morphologically complex words, but it would eventually be necessary to look at mono-morphemic words.

- Learning of native words vs. loanwords is necessary to see if the grammar varies according to lexical strata.

- Some additional tier(s) should be implemented in learning.
  - consonant tier: onset-coda relations

- If possible, a well-formedness judgement experiment will need to be conducted to see what the actual grammar of speakers looks like.

6. Conclusions

- Sundanese exhibits a very unique alternation regarding the plural affix /ar/, in that same segments show assimilation and dissimilation at the same time.

- It is an interesting issue to see whether alternation matches phonotactic patterns; whether the patterns observed in alternations are also attested in lexicon.

- To address the above question, phonotactic learning has been carried out with the UCLA Phonotactic Learner (Hayes and Wilson 2008), with the extensive dictionary data as an input.

6. Conclusions

- So far, similar patterns to the alternations are partly being learned.

- However, more revisions of the data and the learner settings are needed to obtain more accurate grammar.

References

- Anderson, S. R. 1972. On nasalization in Sundanese. *Linguistic Inquiry*, 3. 253-268.
- Bennett, W. G. 2015. Assimilation, dissimilation, and surface correspondence in Sundanese. *Natural Language & Linguistic Theory*, *33*(2), 371-415.
- Cohn, Abigail C. 1992. The consequences of dissimilation in Sundanese. Phonology 9: 199-220.
- Eringa, F.S., 1949. *Loetoeng kasaroeng* (Vol. 8). Brill.
- Hayes, B. and Wilson, C., 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry*, *39*(3), pp.379-440.
- Holton, David. 1995. Assimilation and Dissimilation of Sundanese Liquids. In Jill Beckman, Laura Walsh-Dickey & Suzanne Urbanczyk (eds.), Papers in Optimality Theory, vol. 18 University of Massachusetts Occasional Papers, University of Massachusetts Graduate Student Association.

References

- Lembaga Basa & Sastra Sunda. 1985. Kamus Umum Basa Sunda. Indonesia: Penerbit Tarate Bandung.
- Müller-Gotama, Franz. 2001. *Sundanese*. München: Lincom Europa.
- Rigg, Jonathan. 1862. *A dictionary of the Sunda language of Java*. Batavia (Jakarta): Lange.
- Robins, R. H. 1959. Nominal and verbal derivation in Sundanese. *Lingua* 8. 337-369.
- Shiohara, A., and M. Furihata. 2011. Plural infix *-ar-* in Sundanese. *Asian and African Languages and Linguistics* 6: 89–107.
- Stanton, Juliet. 2019. Aggressive reduplication and dissimilation in Sundanese. Ms.
- Van Syoc, W. B. 1959. *The phonology and morphology of the Sundanese language* (Doctoral dissertation, Univ. of Michigan [Ann Arbor]).
- Zuraw, Kie. 2002. Aggressive reduplication. *Phonology* 19. 395–439.