**Sundanese [ar~al] allomorphy as aggressive reduplication**

*Juliet Stanton, NYU (stanton@nyu.edu)*

(§i) **Overview.** Dissimilatory processes typically exhibit locality effects: if two segments \(a\) and \(b\) are forbidden from co-occurring at some distance \(x\), they are forbidden from co-occurring at all distances less than \(x\). This generalization has both typological and experimental support: non-local dissimilation asymmetrically implies local dissimilation (e.g. Suzuki 1999, Zymet 2012), and participants in artificial grammar learning tasks are consistently unable to learn systems that violate this implication (Hansson & McMullin 2014). These results suggest that the locality generalization is a robust universal that ought to be captured by any successful theory of dissimilation.

This paper focuses on the sole apparent exception to the locality generalization: a pattern of phonotactically conditioned allomorphy in Sundanese (Cohn 1992, Bennett 2015), which suggests that \([r]\)s and \([l]\)s can co-occur only in adjacent syllable onsets (§ii). Developing a suggestion by Zuraw (2002:433), I show that the pattern can be analyzed by appealing to two unbounded co-occurrence constraints, \(*[r]...[r]*\) and \(*[l]...[l]*\), whose effects in local contexts are obscured by a general desire for identity among adjacent syllables (for a related analysis see Suzuki 1999; §iii). Additional evidence for the proposal comes from statistical trends in the Sundanese lexicon (§iv).

The proposed analysis lets us uphold the claim that dissimilation obeys a locality restriction: if a co-occurrence restriction targets \((a,b)\) at some distance \(x\), it targets \((a,b)\) at all distances less than \(x\). This conclusion serves as an argument for theories that derive the locality generalization (e.g. Suzuki’s 1999 GOCP) and a challenge for theories that do not (e.g. Bennett’s 2015 SCTD).

(§ii) **Data.** In the general case, the Sundanese plural infix is realized as [ar]. But when infixed to a root containing an [r], the plural infix generally maps to [al] (1). This alternation suggests a dispreference for the co-occurrence of [r]s in Sundanese. (Data from Cohn 1992, Bennett 2015.)

(1) a. [k-ar-usut], [g-ar-ilis] ‘messy (pl.), beautiful (pl.)’ default allomorph is [ar]
   b. [c-al-ombrek], [ŋ-al-umbara] ‘cold (pl.), go abroad (pl.)’ [al] when stem has an [r]

There are two kinds of exception to this generalization. The first: in certain local configurations, [r]s are permitted to co-occur. Specifically, if the syllable begun by the infixed consonant is preceded or followed by a syllable whose onset is [r], [ar] surfaces (2).

(2) [r-ar-ahit], [c-ar-uriga] ‘wounded (pl.),’ suspicious (pl.)’ neighboring onset is [r]

The second: in certain segmental contexts, the [al] allomorph appears unexpectedly. Specifically, if the word-initial onset is [l], [al] surfaces, whether or not another [r] exists (3).

(3) [l-al-itik], [l-al-irena] ‘little (pl.), take a break (pl.)’ initial onset is [l]

(§iii) **Analysis.** The data in (1) suggest that a preference to use the [ar] allomorph can be overridden by a constraint on co-occurring [r]s. I encode the preference for [ar] with USE[ar], which assigns a penalty if the exponent of /ar/ is not [ar] (see e.g. MacBride 2004 on allomorph preference constraints). This constraint is dominated by *[r]...[r]*, which assigns a violation for each pair of [r]s. This derives the facts in (1): [ar] is used unless *[r]...[r]* is violated (4–5).

(4) \([ar]\) is default allomorph

<table>
<thead>
<tr>
<th>([ar,al])+kusut</th>
<th>*[r]...[r]</th>
<th>USE[ar]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\not\in) a. k-ar-usut</td>
<td>*[r]...[r]</td>
<td>USE[ar]</td>
</tr>
<tr>
<td>b. k-al-usut</td>
<td>(*!)</td>
<td></td>
</tr>
</tbody>
</table>

(5) \([al]\) used to satisfy *[r]...[r]*

<table>
<thead>
<tr>
<th>([ar,al])+ormat</th>
<th>*[r]...[r]</th>
<th>USE[ar]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\not\in) a. h-ar-ormat</td>
<td>*[r]...[r]</td>
<td>USE[ar]</td>
</tr>
<tr>
<td>b. h-al-ormat</td>
<td>(*!)</td>
<td></td>
</tr>
</tbody>
</table>

To account for the fact that [r]s are permitted to occur when they occupy adjacent syllable onsets (2), I propose, following Zuraw (2002:433), that there is a more general drive in Sundanese for adjacent syllables to be coupled in a reduplication-like structure. I assume that the constraint promoting coupling is REDUP-σσ, which assigns a violation for every pair of adjacent syllables that is not coupled (see Zuraw 2002:405 for a discussion of adjacency). Coupled substrings are
subject to additional similarity requirements: in Sundanese, coupled substrings must have identical onsets (in other words, IDENT-κκ[onset], where each κ is a coupled substring, is undominated). The fact that [ar] surfaces in (2) shows that REDUP-σσ dominates *[r]...[r]: the drive for adjacent syllables to be coupled takes priority over avoidance of co-occurring [r]s (7). (The fact that the infixal and not the stem consonant alternates to satisfy IDENT-κκ[onset] can be attributed to high-ranked IO-IDENT; I assume that [ar] and [al] are both faithful realizations of the infixal morpheme.)

(7) \( \text{REDUP-σσ >> } [\star[r]...[r]] \)

\[
\begin{array}{|c|c|c|c|}
\hline
\text{ar,al}+\text{curiga} & \text{ID-κκ[ons]} & \text{REDUP-σσ} & *[r]...[r] \\
\hline
\varphi \ a. \ c-[\text{ar-u}]_{k}[\text{ri}]_{k}\text{ga} & \** & * \\
\hline
\varphi \ b. \ c-[\text{ar-uriga}] & \*** & * \\
\hline
\varphi \ c. \ c-[\text{al-u}]_{k}[\text{ri}]_{k}\text{ga} & ![ & ** \\
\hline
\varphi \ d. \ c-[\text{al-uriga}] & ![ & ** \\
\hline
\end{array}
\]

[1]s, dominates REDUP-σσ. The fact that [1]s are permitted to co-occur when they occupy the first two syllable onsets can be explained if REDUP-σ1σ2, a context-specific version of REDUP-σσ that requires the first two syllables of the word to be coupled, dominates *[1]...[1] (8).

(8) \( \text{REDUP-σ1σ2 >> } [\star[1]...[1]] \)

In sum, Sundanese disfavors co-occurrence of [r]s and [1]s, unless this allows for adjacent syllables to be coupled, either generally (in the case of [r]) or in the first two syllables (in the case of [1]). Under this analysis, there is no need to claim that the constraints motivating dissimilation target non-local contexts only: it’s just that the local effects of unbounded *[r]...[r] and *[1]...[1] are obscured by a more general desire for adjacent syllables to resemble one another.

(§iv) Further evidence. As noted by Zuraw (2002), trends in the Sundanese lexicon suggest a preference for adjacent syllable identity. For example, Cohn (1992) observes that in 76/87 \( r_{1}V(C_{1})r_{2}V(C_{2}) \) roots, \( r_{1}V(C_{1}) = r_{2}V(C_{2}) \). An initial investigation of 1537 (non-morphologically reduplicated) \( C_{1}VC_{2}VC_{3}V \)-initial words, from Rigg (1862), suggests that the preference for adjacent syllable identity is more general than previously recognized. For adjacent syllables (i.e. \( C_{1}VC_{2}V \) and \( C_{2}VC_{3}V \)), onset identity is significantly correlated with nucleus identity (9), but for non-adjacent syllables (i.e. \( C_{1}VC_{3}V \)), it is not (10). (Both \( p \)-values are from \( \chi^{2} \) tests.)

(9) Matching in \( \sigma_{1}\sigma_{2}/\sigma_{2}\sigma_{3} \) (\( p < .001 \))

\[
\begin{array}{|c|c|c|}
\hline
\text{V Match} & \text{V Mismatch} \\
\hline
C \text{ Match} & 188 & 74 \\
C \text{ Mismatch} & 1157 & 1527 \\
\hline
\end{array}
\]

(10) No matching in \( \sigma_{1}\sigma_{3} \) (\( p = .72 \))

\[
\begin{array}{|c|c|c|}
\hline
\text{V Match} & \text{V Mismatch} \\
\hline
C \text{ match} & 30 & 50 \\
C \text{ Mismatch} & 558 & 830 \\
\hline
\end{array}
\]

Data from the corpus also confirm that [r]s and [1]s can co-occur in a local configuration but not elsewhere. Of the 24 forms with co-occurring [r]s in my corpus, in 23 they are adjacent onsets (\( C_{1}VC_{2}V \) or \( C_{2}VC_{3}V \); see also Cohn 1992:214); of the 34 forms with co-occurring [1]s, all are adjacent. Such a dispreference for non-local co-occurrence is not found for other consonants: of the 56 forms with co-occurring [s]s, for example, in 42 the [s]s are local and in 14 they are not.

These lexical trends support the above analysis of the [ar–al] allomorphy: they confirm that Sundanese exhibits a desire for adjacent syllable identity which masks a dispreference for identical liquid co-occurrence. Possible ways to unify the analysis of the [ar–al] alternations with the statistical trends identified above will be discussed.