Geminates and syllabification in Pattani Malay and Marshallese

Nina Topintzi, UCL
i.topintzi@ucl.ac.uk

1 Intro and aims

✦ Main claim: if the characteristic of geminates is to be underlyingly moraic (Ham 2001) and if moraic onsets exist (Topintzi 2006), then there should be geminates which are tautosyllabic in an onset position and carry a mora
✦ Empirically verify this with data from Pattani Malay and Marshallese
✦ This proposal not only fills in a logical possibility, but also accounts for initial geminates, which so far have resisted an adequate treatment in moraic theory

2 What is a geminate?

✦ Singleton vs. geminate: Singleton has no underlying moraicity, geminate has (Hayes 1989)

(1) Singleton vs. geminates in the input (Hayes 1989)

✦ Standard representation of a medial geminate:

(2) Flopped structure of geminates word-medially

✦ Advantages:
   ✦ increased length of geminate mirrored in double linking
   ✦ moraicity of coda and effects on the first syllable’s weight
   ✦ distinguishes [CVC_1] singleton from [CVC_2] geminate

3 Geminates as moraic onsets?

✦ One important observation (Ham 2001):
   The usual ‘flopped’ structure of a medial geminate (2) is not necessitated by moraic theory but by syllable theory which prefers onsetful syllables. Moraic theory only asks that the geminate bears a mora
(3) **Geminates word-medially**

a. Avoidance of onsetless σs

\* \( \sigma_1 \) \( \sigma_2 \)

\( \mu \) \( \mu \) \( \mu \)

V C: V

b. Preferred structure

\( \sigma_1 \) \( \sigma_2 \)

\( \mu \) \( \mu \)

V C: V

+ Double linking (3b) is thus seen as the result of avoiding onsetless syllables (3a) wherever possible

+ But there is an additional representation that satisfies the moraicity requirement of the geminate and presents desirable syllabification

(4) **Geminate as moraic onset word-medially**

\( \sigma_1 \) \( \sigma_2 \)

\( \mu \) \( \mu \)

V C: V

(5) **Peculiarities with (4):**

+ a) it assigns a mora to the onset
  - fine, in the light of other supporting languages, e.g. Pirahã, Karo, Arabela stress or Bella Coola word minimality (Topintzi 2006)

+ b) it predicts that \( \sigma_2 \) becomes heavy rather than \( \sigma_1 \) (cf. (3b))
  - logically possible. See Marshallese below

+ c) it does not present double linking; how about length and weight correlation?
  - Ladefoged and Maddieson (1996: 91-92): depending on the language, geminate stops can be anywhere between 1½ - 3 longer than singleton counterparts \( \rightarrow \) so no absolute 1:1 match between geminate length and weight

  - Moraic primacy in the implementation of timing, e.g. Hubbard (1994), Ham (2001). Geminates have to be moraic. Their exact length is regulated on a language-specific basis. The geminate mora is allocated a minimum target duration, whose implementation in terms of timing takes precedence over any other segment-specific effects, such as place of articulation and voicing. If the main characteristic of geminates is to be moraic, then it is anticipated that the priority is to achieve the mora’s minimum duration target, thus leaving smaller space for other durational segment-related differences. Ham (2001) tests this prediction and finds that indeed such durational differences are larger in moraic singletons than geminates

+ Prediction: if some geminates can be represented as moraic onsets word-medially, they should be able to do so word-initially too!

+ **Correct:** see geminates in Pattani Malay (below), Trukese (Davis and Torretta 1998, Muller 1999) and others [possibly Sa’ban, Nhaeun (Muller 2001), Logbara, Oneida (Hajek and Goedemans 2003; henceforth H&G) which only have initial geminates]
4 Initial geminates and their representations

- Word-initially, the flopped structure of (2) is impossible, since there is no preceding coda available!!!
- Problem for standard moraic theory
- Some proposals:

\[ \begin{array}{c}
\sigma \\
\mu \\
C: V \\
\end{array} \quad \begin{array}{c}
\text{PrWd} \\
\mu \\
C: V \\
\end{array} \]

- Problem for (6a): \( \mu \) is unlinked, but needs to associate to higher prosodic structure (cf. Kiparsky 2002) so that it counts for weight purposes, such as Word Minimality in Trukese or stress in P. Malay
- Problem for (6b): seems to partially identify initial geminates with unsyllabified consonants, cf. Bella Coola (Bagemihl 1991), Piro (Lin 1997), Arabic (Kiparsky 2002)

- BUT: Initial geminate as a moraic onset avoids these problems

\[ \begin{array}{c}
\sigma \\
\mu \\
\# p: o \\
\end{array} \]

5 Pattani Malay initial onset geminates

5.1 Pattani Malay facts

- Inventory: consonants /p t c k ? b d j g s h z y m n ñ m b ñ p l r l w y/; vowels /i e ë å u o ë e å y 2/
- Vowel length is predictable: Vs are long in open syllables and short in closed ones. Only the schwa-like /ë/ is relatively short even in open syllables
- Consonant length only contrasts word-initially

(8) Initial geminates vs. singletons in Pattani Malay (Abramson 1999: 592)

<table>
<thead>
<tr>
<th>Singletons</th>
<th>Geminates</th>
</tr>
</thead>
<tbody>
<tr>
<td>make</td>
<td>‘to eat’</td>
</tr>
<tr>
<td>lamaʔ</td>
<td>‘late’</td>
</tr>
<tr>
<td>ɣatɔ</td>
<td>‘comprehensive’</td>
</tr>
<tr>
<td>sepaʔ</td>
<td>‘to kick’</td>
</tr>
<tr>
<td>cabe</td>
<td>‘branch’</td>
</tr>
</tbody>
</table>

3
buŋo ‘flower’  buŋo ‘to bloom’
kukoh ‘to be stable’  kukoh ‘to render stable’

Stress facts:
- Non-geminated words (9a): primary stress on final σ; secondary stress on remaining syllables (9a.i, unless they include /i/ in which case they are stressless (9a.ii-iii)
- Geminated words (9b): primary stress always on first σ (9b), even if it includes /i/ (9b.iii); secondary stress on remaining syllables

(9) Geminates and stress
   a. Non-geminated words
   i. [buŋɔ] ‘fruit’ ~ i. [buŋɔ] ‘to bear fruit’ [from /b+buŋɔ/]
      [jɔlɛ] ‘road, path’ ~ [jɔlɛ] ‘to walk’ [from /b+jale/]
   ii. [pimɔtɔ] ~ ii. [mɔtɔ] ‘jewellery’
      [sidɔdɔ] ~ [diɔdɔ] ‘police’
   iii. [kidiɔ] ‘shop’ ~ iii. [kiɗa] ‘to the shop’ [from /k+i+kida/]

5.2 Hajek’s and Goedemans’ analysis (2003)

- Crucial components of H&G’s analysis:
  - Initial geminates are moraic onsets like the one in (7)
  - The moraic composition of syllables in P. Malay is:
    i) C1=1μ
    ii) CVV=2μ, where V is not /i/
    iii) CVC=2μ
    iv) C:V=2μ
    v) C:VV=3μ, where V is not /i/
  - The following constraint dominates ALIGN-HEAD-R which prefers final stress: The left edge of the main stress foot must be aligned to a moraic consonant

- How does it work?

(10) Moraic composition for H&G
    a. Non-geminated words
       [buŋɔ] (ɔ)(u) (ɔ)(u)
       [pimɔtɔ] (ɔ)(u) (ɔ)(u)
    b. Geminated words
       [buŋɔ] (ɔ)(u)(u) (ɔ)(u)
       [mɔtɔ] (ɔ)(u)(u) (ɔ)(u)
    c. Geminated and /i/-inclusive words
       [kiɗa] (ɔ)(u)(u) (ɔ)(u) predicted: *[k{id}a]

Roughly: primary stress is word-final, heavy σs get secondary stress (due to WSP), but monomoraic ones, i.e. those having /i/, remain stressless, e.g. pimɔtɔ. If there is a geminate, then the first syllable gets primary stress, as it is the heaviest in the word (trimoraic). Secondary stress is as usual
Prediction (given the above): *kùda* should be stressed as *[kːdá]*, because both syllables are bimoraic, thus right alignment should favour final primary stress. Wrong!!! Actually, it is: *[kːdà]*. Why?

H&G’s solution: Utilizing the constraint: *The left edge of the main stress foot must be aligned to a moraic consonant.* This forces stress to dock on a syllable with a geminate no matter what the following vowel is.

Problems with this analysis

- Use of such constraint is ungrounded. It’s completely *ad hoc*
- Proposes prioritization of onset weight, but if this is possible then we can also expect prioritization of coda weight yielding systems with: CVC > CVV > CV → unattested or the simultaneous combination of the two in one language, yielding: C^\mu V^\mu C^\mu > C^\mu V^\mu V^\mu, V^\mu V^\mu V^\mu > V^\mu → unattested

5.3 Present analysis

- Accepts moraicity of onsets, but assumes different moraification of syllables. Stress is partially quality-sensitive, thus it avoids the problems above
- Crucial components of analysis
  - Initial geminates are moraic onsets like the one in (7)
  - The moraic composition of syllables in P. Malay (based on the absence of phonemic vowel length) is:
    i) CV(V)=1\mu for all vowels including /h/
    ii) CVC=1\mu
    iii) C:V=2\mu for all vowels including /h/
- WSP >> ALIGN-HD-R so that heavy ot, i.e. those with geminates, attract stress
- Quality-sensitive stress: avoidance of stressed central vowels, i.e. use of *P/i* (cf. sonority-driven stress scales like *P/.onerror* > *P/i, u > *P/e, o* > *P/a, supported by languages like Mordwin, Kobon, Chukchee, Aljutor and Mari in Kenstowicz (1994), or Yil and Nganasan in de Lacy (to appear))
- However, WSP >> *P/i, due to examples like *[kːdà]*

How does it work?

(11) Moraic composition for present analysis
a. Non-geminated words
   [bùwś] (鸵)\mu (鸵)\mu
   [pimət] α\mu (鸵)\mu (鸵)\mu
b. Geminated words
   [bùwò] (鸵)\mu (鸵)\mu
   [máts] (鸵)\mu (鸵)\mu
   c. Geminated and /i/-inclusive words
   *[kːdà]* (鸵)\mu (鸵)\mu correct prediction
Analysis in more detail

- All syllables (except some of those including /i/) need to be parsed (high-ranking PARSE-σ) into single monosyllabic feet → (12) and (13)
- ALIGN-R (PrWd, HdFt) [abbreviated as ALIGN-HD-R] ensures that primary stress appears at the right edge → (14)
- Unless there is an initial geminate which attracts primary stress due to WSP_{PrWd} >> ALIGN-HD-R (cf. McGarrity 2003 for constraints specific to primary stress; WSP_{PrWd} = heavy syllables receive primary stress) → (15)
- A syllable that includes /i/ is normally left unstressed: *P/i >> FtFORM → (16)
- Unless it is also preceded by a geminate, in which case it receives stress due to WSP_{PrWd} >> *P/i → (17)

P. Malay favours stress on each σ (in violation of *CLASH) and formation of monosyllabic feet. How can we get that? By simultaneous satisfaction of the constraints in (12) (cf. Prince 1997, Green 2002)

(12) ALIGN-L (Ft, FtHd): Align the L edge of every foot with the L edge of a foot head
(13) FtHd-L, FtHd-R, PARSE-σ >> FtBIN, *CLASH

<table>
<thead>
<tr>
<th>buwɔh</th>
<th>Ft-Hd-L</th>
<th>Ft-Hd-R</th>
<th>PARSE-σ</th>
<th>FtBIN</th>
<th>*CLASH</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (buwɔh)</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. (búwɔh)</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. bu(wɔh)</td>
<td>*!</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. (bù)(wɔh)</td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>*</td>
</tr>
</tbody>
</table>

N.B: FtFORM abbreviates Ft-Hd-L, Ft-Hd-R, PARSE-σ

(14) [bùwɔh]: FtFORM, ALIGN-HD-R

<table>
<thead>
<tr>
<th>buwɔh</th>
<th>FtFORM</th>
<th>ALIGN-HD-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (bù)(wɔh)</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. (bù)(wɔh)</td>
<td></td>
<td>*!</td>
</tr>
</tbody>
</table>

(15) [bùwɔh]: WSP_{PrWd} >> ALIGN-HD-R

<table>
<thead>
<tr>
<th>buwɔh</th>
<th>WSP_{PrWd}</th>
<th>ALIGN-HD-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (bù)(wɔh)</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. (bù)(wɔh)</td>
<td>*!</td>
<td></td>
</tr>
</tbody>
</table>

(16) [pimɔtɔ]: *P/i >> FtFORM, ALIGN-HD-R

<table>
<thead>
<tr>
<th>pimɔtɔ</th>
<th>*P/i</th>
<th>FtFORM</th>
<th>ALIGN-HD-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (pì)(mà)(t5)</td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. pi(mɔtɔ)</td>
<td>*! (Ft-Hd-L), *! (PARSE-σ)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. pi(mà)(t5)</td>
<td>* (PARSE-σ)</td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>d. (pimɔ)(t5)</td>
<td>* (Ft-Hd-L)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. pi(mà)(t5)</td>
<td>* (PARSE-σ)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(17) \[ \text{\texttt{[k\textipa{\textdagger}\textipa{\textdagger}]}: WSP}_{\text{PwD}} \gg \ast P/1} \\
\begin{tabular}{|c|c|c|}
\hline
 & \text{\texttt{k\textipa{\textdagger}\textipa{\textdagger}a}} & \text{WSP}_{\text{PwD}} & \ast P/1 \\
\hline
\text{a} & \text{\texttt{(k\textipa{i})(\textipa{\textdagger}a)}} &  & \ast \\
\hline
\text{b} & \text{\texttt{k\textipa{\textdagger}}(\textipa{\textdagger}a)} &  & \ast ! \\
\hline
\end{tabular}
\]

\[ \text{\textdagger} \quad \text{Final ranking for Pattani Malay stress} \]

(18) \textit{Pattani Malay stress}  \\
\text{WSP}_{\text{PwD}} \gg \ast P/1 \gg \text{Ft-Hd-L/R, PARSE-\textbf{\textsigma}, ALIGN-HD-R} \gg \text{FtBin}, \ast \text{CLASH}

\[ \text{\textdagger} \quad \text{Advantages over H\&G} \]
\[ \quad \diamond \quad \text{No implausible prioritization of onset weight} \]
\[ \quad \diamond \quad \text{Use of well-attested generalizations on vowel-quality-sensitive stress} \]
\[ \quad \diamond \quad \text{No added mechanisms or assumptions. Maintains insight about onset weight} \]

6 Marshallese medial onset geminates

\[ \text{\textdagger} \quad \text{Marshallese (Ralik dialect) consonant doubling and stress patterns support a moraic onset analysis of medial geminates} \]

6.1 Distributive reduplication

\[ \quad \text{Most distributives are formed by consonant doubling or final syllable reduplication, or more usually by both, e.g. koto \rightarrow kkototo \text{‘windy \rightarrow always be windy’}. Here I only address gemination} \]

(19) \textit{Marshallese distributive - Ralik dialect (Abo et al. 1976)}

\begin{tabular}{|c|c|c|}
\hline
Root & Ralik & Gloss \\
\hline
korap & yokkoraprap & ‘gecko’ \\
tumej & yuttumejmej & ‘open eyes under water’ \\
panuk & yeppanuknu & ‘pile up, gather’ \\
bale & yebbaile & ‘a type of fish’ \\
diylah & yiddiylahlah & ‘nail’ \\
jekapen & yejjekapenpen & ‘less than half full’ \\
mede & yemmedede & ‘young coconut meat’ \\
nib & yinnibnib & ‘preemptive’ \\
lomjak & yolomjakjak & ‘be busy with’ \\
reja & yerrejava & ‘shave (from Engl)’ \\
\hline
\end{tabular}

\[ \text{\textdagger} \quad \text{Explanation on the presence of epenthetic \textipa{\textgamma}V- (whose vowel quality is the same as the base vowel, unless this is \textipa{\textalpha}/ in which case the epenthetic V is \textipa{\textgamma}/) is unclear. See Hedricks (1999) for speculative ideas} \]

\[ \text{\textdagger} \quad \text{An insight: the Ratak dialect forms the distributive by simply copying the first CV, e.g. betah \rightarrow bebetahtah, diylah \rightarrow didiylahlah. Perhaps both dialects require reduplication of 1\textmu. In Ralik, this can be satisfied by the absolutely minimal reduplication of a moraic onset} \]
geminate, i.e. diylah → ddiylahlah, whereas in Ratak, geminates are avoided, in which case the distributive is realized with the supraminimal CV

A geminate reduplicant wholly syllabified in the same σ captures this straightforwardly

6.2 Stress

All data and transcriptions from Zewen (1977: 40-41)

Stress is assigned within a trisyllabic window at the right edge. If the final three syllables are light, the antepenult gets stress. Codas - at least final ones - do not count for stress purposes (20)

(20) *Trisyllabic words LLL: Antepenultimate stress*

<table>
<thead>
<tr>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ékajet</td>
<td>*ekajét</td>
<td>‘to judge’</td>
</tr>
<tr>
<td>nûkileb</td>
<td>*nukîlîb</td>
<td>‘to have a big family’</td>
</tr>
<tr>
<td>jêkaru</td>
<td></td>
<td>‘coconut syrup’</td>
</tr>
<tr>
<td>lákatib</td>
<td>*lakatîb</td>
<td>‘to make angry’</td>
</tr>
</tbody>
</table>

A heavy syllable gets stressed (21)

(21) *Trisyllabic words LHL or Disyllabic LH: Stress H*

<table>
<thead>
<tr>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>je.ú:rur</td>
<td></td>
<td>‘commotion, excitement’</td>
</tr>
<tr>
<td>je.ro.án</td>
<td></td>
<td>‘to waste’</td>
</tr>
<tr>
<td>körâ:</td>
<td></td>
<td>‘woman’</td>
</tr>
<tr>
<td>jelâ:</td>
<td></td>
<td>‘to know’</td>
</tr>
<tr>
<td>kijê:k</td>
<td></td>
<td>‘fire’</td>
</tr>
</tbody>
</table>

Penultimate stress appears on disyllabic words where both syllables are either heavy or light, i.e. the leftmost of the two is stressed (22)

(22) *Disyllabic words (either LL or HH): Penultimate stress*

<table>
<thead>
<tr>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>nêbar</td>
<td></td>
<td>‘to praise’</td>
</tr>
<tr>
<td>májajt</td>
<td></td>
<td>‘to be clear of underwood’</td>
</tr>
</tbody>
</table>

In sum: Default stress is leftward within the trisyllabic window, but can shift so that it docks onto a heavy syllable

How about stress in words having medial geminates?

(23) *Geminate stress (Zewen 1977: 27)*

<table>
<thead>
<tr>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>jîbbûj</td>
<td></td>
<td>‘morning’</td>
</tr>
<tr>
<td>(y)énmán</td>
<td></td>
<td>‘good’</td>
</tr>
<tr>
<td>e’mmér</td>
<td></td>
<td>‘to be ripe’</td>
</tr>
</tbody>
</table>
If syllabification was [jiːb.buŋ], stress would be [jiːb.buŋ] under any assumption, namely:
  ♦ if codas contribute no weight, syllables are light, hence leftmost stress (22)
  ♦ if medial codas are weightful (finals cannot be, cf. (20)), then again leftmost stress should occur as this is the heaviest syllable (21)

Conclusion: Only moraic onset syllabification of the geminate yields the correct result. If it comprises an onset and is moraic, then it will render its σ heavy and stress-attracting (24)

\[(24)\] jiːb.buŋ \quad *jiːb.buŋ
(y)e.m* m*=áň \quad *(y)ém*.m*=áň

NB: I found one exception to this pattern, namely laːrrik ‘boy’ (Zewen 1977: 27) with stress on the first syllable. There are also plenty of examples where stress is not shown at all; some of these are: jillüb ‘sound produced by something that falls into the water’ (1977: 56), killebleb ‘to be very big, corpulent’ (1977: 59), konimán ‘make’ (1977: 107), elloːloː ‘see’ (1977: 110)

While additional data are required, the Marshallese facts are suggestive of a moraic onset geminate analysis

7 Conclusion

If all a geminate is about is being underlyingly moraic (Ham 2001) and if moraic onsets exist (Topintzi 2006), then it is anticipated that some geminates are best analysed as moraic onsets both initially (Pattani Malay) and medially (Marshallese)

If a language syllabifies medial geminates as CVC.CV then it should have: *MORAIC ONSET >> NOCODA

If a language syllabifies medial geminates as CV.CV then it should have: NOCODA >> *MORAIC ONSET

Given that cross-linguistically *MORAIC ONSET is usually highly-ranked, it is anticipated that most languages will have heterosyllabic geminates. Tautosyllabic ones are rarer, but by no means impossible

This solution provides a neat account for cases like Marshallese, but also importantly for the long-standing problem of initial geminates

References


Hubbard, Kathleen (1994) *Duration in moraic theory*. Doctoral Dissertation, University of California, Berkeley


Dordrecht: Foris [reissued in 2003 by the University of Chicago Press]

Kenstowicz, Michael (1994) Sonority-Driven Stress. Available as ROA-33


Ladefoged, Peter and Ian Maddieson (1996) *The sounds of the world’s languages*. Oxford; Cambridge, MA: Blackwell


