Core Syllables vs. Moraic Writing

Eugene Buckley
University of Pennsylvania

Understanding writing systems:
From core issues to implications for written language acquisition

TENTH INTERNATIONAL WORKSHOP ON WRITING SYSTEMS AND LITERACY

Nijmegen :: 12–13 May 2016
Outline of talk

• Units of writing systems represent categories of spoken language
  – word, syllable, consonant, ...

• In some ways, “syllabaries” appear to represent moras
  – originally a unit of syllable weight

• Closer analysis confirms that apparent moraic systems are built on core CV syllabic signs
  – supports view that moras are not constituents in the prosodic hierarchy
Traditional Typology of Writing

- **Ideographic**
  - not truly LINGUISTIC: semasiographic

- **Logographic**
  - most often accurately “morphographic”

- **Syllabic**
  - or is it moraic?
  - this is the main question here

- **Alphabetic**
  - segments, potentially just consonants
Moraic Writing

• Poser (1992) made the claim most forcefully
  – a number of scholars have followed him

• Syllabaries are really *moraic* systems
  – “In addition to the syllable and the segment, writing systems exist that make use of the constituents *mora* and *rhyme*, and the notion *head*."
  – “[Pure] syllabaries are *exceedingly rare*."

• So is the traditional analysis mistaken?
Direction of Generalization

• Poser’s “wrong question”
  – *To what phonological constituents do the individual graphs correspond?*
  – i.e. what is the graph ⇒ constituent mapping

• Poser’s “right question”
  – *On what sort of phonological analysis must this writing system be based?*
  – i.e. what is the constituent ⇒ graph mapping

• Definition of systems depends on this assertion
  – but the “phonological analysis” could still determine the basic value of the independent graphs
Poser’s Categories

- **Syllabic**: each syllable ⇒ distinct sign
  - Akkadian, Loma
- **Moraic**: each mora ⇒ distinct sign
  - Japanese, Cree, Vai
- **Proto-Moraic**: nonhead mora written same as head
  - Cherokee
    - uses CV signs to write coda consonants
- **Head Moraic**: only head mora is written
  - Linear B
    - writes onset clusters, but not coda consonants
More Moras

- Ratcliffe (2001); cf. also Kurzon (2013)
  - Arabic and *kana* have moraic “representational goal”
- Coulmas (2003) has mostly syllabaries, but:
  - “in the *kana* syllabaries, each basic sign is interpreted as one mora”
- Rogers (2005) calls many systems moraic:
  - Japanese, Akkadian, Cherokee, Mayan, Linear B
- Gnanadesikan (2011) includes syllabaries, but:
  - “some writing systems that are traditionally called syllabaries actually are moraic” : Japanese, Vai
Syllabaries

- **Traditional** analysis is still widely used
  - for example, Sampson (2015)
- The moraic approach is often **not addressed**
  - i.e. **arguing actively** for the syllable interpretation
  - that is my goal in this presentation
- Gnanadesikan (2011) does argue for syllabaries
  - but also accepts moraic analysis in certain cases
  - more on her claims later
Primacy of Syllables

• Much research on syllables as more accessible to conscious awareness
  – claimed basis for predominance of early syllabaries (Daniels 1992)
• Why then would moras be the basic unit?
  – they don’t seem to be primary or salient in this way
• Fundamental ambiguity
  – CV is a syllable that consists of one mora
  – which fact is most relevant?
  – how can we test the respective predictions?
Moras and Vowel Length

- In moraic phonology, a second (nonhead) mora is the sole representation of segment length.
- Long vowel correlates with a heavy syllable.

<table>
<thead>
<tr>
<th>Short /a/, one mora, light</th>
<th>Long /aː/, two moras, heavy</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ</td>
<td>σ</td>
</tr>
<tr>
<td>μ</td>
<td>μ</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
</tr>
</tbody>
</table>
Moras and Coda Consonants

• In many or perhaps most languages, a coda consonant also makes the syllable heavy.
• This requires a representation with two moras.

- short V, one mora, light
- coda C, two moras, heavy

\[ \sigma \rightarrow \mu \rightarrow \text{a} \rightarrow \sigma \rightarrow \mu \rightarrow \text{a} \rightarrow \text{m} \]
Moras and Onset Consonants

- Theories differ about where the onset links
- Related to the status of the mora
  - prosodic constituent like the syllable, or a measure of weight
Constituency of CV

- If onset links to first mora, CV is a unit:
  \[
  \left[ \left[ \text{ta} \right]_\mu \left[ \text{n} \right]_\mu \right]_\sigma
  \]

- If onset links directly to syllable, CV is **not** a unit:
  \[
  \left[ \text{t} \left[ \text{a} \right]_\mu \left[ \text{n} \right]_\mu \right]_\sigma
  \]

- **Onset-linking** is more typical in modern phonology
  - if that is correct, then moraic writing is problematic
  - but we’ll proceed on the assumption that it’s still possible
Core syllables

• Every language has CV syllables
  – many also have V syllables, if onset is optional

• (C)V is the most basic syllable type
  – “core”
  – “minimal”
  – “universal”
  – “maximally unmarked”

• phonological constraints express this idea
  – Onset, NoCoda, *Complex
“Core”, not “Light”

• Tagalog RECENT PERFECTIVE includes prefix ka- and a CV reduplicant

  ka-ta-trabaho  ‘work’
  ka-bo-bloaut  ‘give special treat’

• If relevant notion were light syllable, we should expect CCV instead
  – so reliance on weight (one mora) is insufficient
Vai Syllabary

- Aligns well with moraic analysis
- Invented around 1833 in Liberia
  - revised by committee in 1960s
- Signs of the shape CV and V
  - also final nasal sign, similar to Japanese
- Spoken syllables fairly limited
  - CV
  - CVV
  - CVN
<table>
<thead>
<tr>
<th>Vai Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>ɓa lo kìɓálò</td>
</tr>
<tr>
<td>la</td>
</tr>
<tr>
<td>la ha</td>
</tr>
<tr>
<td>ku wu</td>
</tr>
<tr>
<td>ke ŋ</td>
</tr>
</tbody>
</table>
### Vai Complex Syllables

<table>
<thead>
<tr>
<th></th>
<th>Spoken</th>
<th>Written</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVN</td>
<td>+ mora</td>
<td>+ sign</td>
</tr>
<tr>
<td>CVV</td>
<td>+ mora</td>
<td>+ sign</td>
</tr>
</tbody>
</table>

- Signs added to derive from basic CV
- Perfect correlation between moras and signs
  - but based on very limited syllable types
Vai Signs and Moras

- Vai spellings have **one sign for each mora**
- But each syllable presents just a **few options**
  - adding length or final N also adds a mora
  - lacks onset clusters to examine a more complex syllable that does not add a mora
- No way to **test difference** between analyses
  - one sign per mora; or
  - core CV plus supplemental sign
- Moraic analysis is possible, but **not required**
### Ethiopic Alphasyllabary

<table>
<thead>
<tr>
<th>ä</th>
<th>u</th>
<th>i</th>
<th>a</th>
<th>e</th>
<th>(i)</th>
<th>o</th>
</tr>
</thead>
<tbody>
<tr>
<td>ኢ</td>
<td>ኢ</td>
<td>ኢ</td>
<td>ኢ</td>
<td>ኢ</td>
<td>ኢ</td>
<td>ኢ</td>
</tr>
<tr>
<td>በ</td>
<td>በ</td>
<td>በ</td>
<td>በ</td>
<td>በ</td>
<td>በ</td>
<td>በ</td>
</tr>
<tr>
<td>ጥ</td>
<td>ጥ</td>
<td>ጥ</td>
<td>ጥ</td>
<td>ጥ</td>
<td>ጥ</td>
<td>ጥ</td>
</tr>
</tbody>
</table>

- Shape modified for vowel; some irregularity
  - treat synchronically as CV unit

---

20
Tigrinya CV and CVC

- Perfect correlation between moras and signs
  - sign for CV ↔ mora for light syllable
  - sign for C ↔ mora for coda
Tigrinya Geminates

- Gemination not written, yet purely moraic
  - it would be easy to do: እሆኔ ዋ-ብ-ቦ
  - no sign for C ≠ mora for coda
Moras and Gemination

\[\text{ʔabo} = \text{ʔㄢ} \]

\[\text{ʔabbo} = \text{ʔㄢ} \]

\[\sigma \quad \sigma \]

\[\mu \quad \mu \]

\[\text{ʔ a b o} \]

\[\sigma \quad \sigma \]

\[\mu \quad \mu \quad \mu \]

\[\text{ʔ a b o} \]

distinction is exactly a mora
**Tigrinya Complex Syllables**

<table>
<thead>
<tr>
<th></th>
<th>Spoken</th>
<th>Written</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coda</strong></td>
<td>+ mora</td>
<td>+ sign</td>
</tr>
<tr>
<td><strong>Geminate</strong></td>
<td>+ mora</td>
<td><strong>no sign</strong></td>
</tr>
</tbody>
</table>

- Poor correlation between moras and signs
  - real principle seems to be writing every C(V)
  - ignores length (i.e. moras)
Japanese *kana* as moraic

“Although the *kana* scripts are often called syllabaries, they are in fact moraic systems. Each symbol in the *kana* scripts represents one mora. Most of these are CV sequences, but final /N/ or /Q/ count as separate morae, and vowel length adds a mora.”

Rogers (2005)
Japanese Syllables

• Basic CV **ka** with five deviations:
  – Long vowel : **ka:**
  – Coda N : **kan**
  – Coda C as part of geminate : **kat.ta**
  – Complex onset : **kja**
  – Novel CV in borrowings : **fa**

• Two **kana** writing systems
  – **hiragana** : (mainly) native words and affixes
  – **katakana** : foreign words, italics, etc.
CVV and CVN

おかあさん  o-ka-a-sa-N  oka:san  ‘mother’
          o_ μ ka_ μ μ sa_ μ n_ μ

ボールペン  bo:-:ru-pe-N  bo:rupen  ‘ballpoint pen’
          bo_ μ μ ru_ μ pe_ μ n_ μ

• Perfect correlation between moras and signs
  – sign for length ↔ mora as unit of length
  – sign for N ↔ mora for coda /n/
CVC with Geminate

あっさり  
a-tsu-sa-ri  assari  ‘easily’
   aᵣ sᵣ saᵣ riᵣ

ロケット  
ro-ke-tsu-to  roketto  ‘rocket’
   roᵣ keᵣ tᵣ toᵣ

• Gemination written with small tsu sign
  – historically, final /t/ surfaced as [tsu] or as geminate
  – (small) sign for C ↔ mora for coda
Complex Onset

じゅみょう
zi-ju-mi-jo-u
dʒumjo:
‘life span’

キャベツ
ki-ja-be-tsu
kjabetsu
‘cabbage’

• C plus /j/ written Ci plus small jV sign
  – “absorbed” into coronal obstruents: /zj/ → [dʒ]
  – (small) sign for glide ≠ mora
Borrowed CV Sequences

ファイト
fu-a-i-to  faito  ‘fight’
fa\_i\_ to\_μ

ティッシュ
te-i-tsu-si-yu  tissu  ‘tissue’
ti\_i\_ \_s\_i\_ yu\_μ

• $C_1V_1$ plus small $V_2$ yields $C_1V_2$
  – for sequences not normally found in Japanese
  – (small) sign for vowel ≠ mora
## Japanese Complex Syllables

<table>
<thead>
<tr>
<th></th>
<th>Spoken</th>
<th>Written</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVV</td>
<td>+ mora</td>
<td>+ kana</td>
</tr>
<tr>
<td>CVN</td>
<td>+ mora</td>
<td>+ kana</td>
</tr>
<tr>
<td>CVC</td>
<td>+ mora</td>
<td>+ small kana</td>
</tr>
<tr>
<td>CjV</td>
<td>no new mora</td>
<td>+ small kana</td>
</tr>
<tr>
<td>novel CV</td>
<td>no new mora</td>
<td>+ small kana</td>
</tr>
</tbody>
</table>

Small kana used to write length, etc. (Nishimura poster)
Japanese *kana* and Moras

- By the nature of spoken *Japanese phonology*, nearly every deviation from CV yields a new mora
  - this gives the impression that signs correspond to moras
  - as long as small *tsu* for gemination is counted
- But CjV and borrowed CV sequences are also written with two *kana*
  - *weakens correlation* between sign and mora
- Evidence for moraic writing is not actually strong
A More nuanced View

• Gnanadesikan (2012)
  – “Japanese kana and Maldivian Thaana encode moraic information: one can count the number of moras by the number of signs used”
  – “in both languages mora count plays a key role in the phonological system, and this is captured by the native scripts”

• Two new types of writing system
  – moraic syllabary (kana)
  – moraic alphabet (Thaana)
And Yet ...

- This still attributes **too much causation** to the moraic element of the writing
  - I maintain that the correlation is secondary
  - not a design feature (or “goal”) of these systems
- Japanese **kana**
  - we’ve already seen that kana count does not ultimately line up well with mora count
- Dhivehi **Thaana**
  - strong mora-count evidence, but still not persuasive
Dhivehi

• Language of the Maldives
  – Indo-Aryan
  – closely related to Sinhala (of Sri Lanka)

• Thaana is the writing system for the language
  – originated between 16th to late 17th centuries
  – structurally fairly similar to Arabic (including R→L)

• Unusual graphical basis
  – Arabic and Indic numerals 1–9
  – some Arabic letters for borrowed sounds
Thaana Consonants

- First nine (clearly) from Arabic numerals 1–9
  
- Next nine (possibly) from Indic 1–9
  
- Six more used mainly for loanwords
  - most are clearly adapted from other letters
Thaana Vowel Diacritics

• Short vowels based on Arabic pointing

\[ \acute{\cdot} \quad \bar{\cdot} \quad \ddot{\cdot} \quad \breve{\cdot} \quad \grave{\cdot} \]

a  i  u  e  o

• Long vowels by doubling short signs

\[ \acute{\cdot} \quad \bar{\cdot} \quad \ddot{\cdot} \quad \breve{\cdot} \quad \grave{\cdot} \]

a:  i:  u:  e:  o:

• Irregularity of \( \acute{o} \) (o:)
  – Gnanadesikan: abbreviation of doubled \( \acute{o} \) (o)
Open Syllables

<table>
<thead>
<tr>
<th>Syllable</th>
<th>Pronunciation</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>di, ve, hi</td>
<td>divehi</td>
<td>‘Dhivehi’</td>
</tr>
<tr>
<td>a, li, fu</td>
<td>alifu</td>
<td>‘alifu (letter)’</td>
</tr>
<tr>
<td>ka, ku</td>
<td>ka:ku</td>
<td>‘who’</td>
</tr>
<tr>
<td>ka, ku</td>
<td>kaku:</td>
<td>‘knee’</td>
</tr>
<tr>
<td>ri, zo, tu</td>
<td>rizo:tu</td>
<td>‘resort’</td>
</tr>
</tbody>
</table>
Thaana non-CV

• V without preceding C rests on *alifu*

  o\textsubscript{\( \mu \)} i\textsubscript{\( \mu \)} oi  ‘(ocean) current’

• C without following V is marked by *sukun*
  – whether preceding C or end of word
  – in this case *alifu* is /ʔ/ or obstruent gemination

\textbullet  \textbullet  no V  ?
### Closed Syllables

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Pronunciation</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>مُسْرَى</td>
<td>baום sום</td>
<td>bas</td>
</tr>
<tr>
<td>مُسْرَى</td>
<td>d3aום lום saום</td>
<td>d3alsa:</td>
</tr>
<tr>
<td>مُسْرَى</td>
<td>aום nום giום</td>
<td>angi:</td>
</tr>
<tr>
<td>مُسْرَى</td>
<td>koום ؟ום koום</td>
<td>kokko</td>
</tr>
<tr>
<td>مُسْرَى</td>
<td>uום nום miום duום</td>
<td>ummi:du</td>
</tr>
</tbody>
</table>
Thaana and Moras

• Number of diacritics equals number of moras
  – CV has one vocalic diacritic
  – coda C has sukun (for lack of vowel)

• Long vowel signs must be counted as 2
  – graphically they are doubling of short sign
  – but irregularity of <o:> casts doubt on this as an active (synchronous) generalization
Prenasalized Stops

- Normal NC cluster includes *sukun*
  - two segments, two moras (with following V)
- Prenasalized stop has NC but no *sukun*
  - single segment, single mora (with following V)
  - confirmed by the stress pattern of the language

\[
\text{handa} \quad \text{‘sea snail’}
\]
\[
\text{handu} \quad \text{‘moon’}
\]
Evidence for Moraic View?

- Gnandesikan says that the writing of prenasalized stops supports the mora count.
- But this is a relatively new practice.
  - Previously (and sometimes still) written as simple voiced stop, i.e. one sign.
- I suggest the added `<n>` modifies the stop.
  - As a diacritic, or part of a digraph.
  - One vowel sign for the pair, like Indic conjuncts.
- Not designed to serve a moraic count.
Borrowed Clusters

- sku:\l ~ 'school'
- film ~ 'film'
- progra:m ~ purogura:m ~ 'program'
- hausin(g) ~ 'housing'
Lack of Vowel

- Simple native syllable structure
  - hard to test the predictions of the moraic analysis
- Novel borrowed structures provide a test
  - note the use of sukun simply meaning “no vowel”
  - easily could have been written unadorned if mora count was important to the users of the system
- English <ng> with no final sukun
  - perhaps another digraph, or etymological spelling in which the C is not pronounced, so no diacritic
What Would Be Persuasive?

• Signs for **CCV** alongside **CV**
  – both are one mora
  – parsimony argument applies to core syllables also

• Writing of coda **C** only when it is moraic
  – this might be sonorants but not obstruents
  – yet non-writing of codas is rather common for homorganic nasals, which are high in sonority

• Writing of **vowel length** but not the coda in a language with non-moraic codas
  – but if anything, the opposite is more common
Conclusion

• CV is not established as a moraic constituent
  – so not parallel to segment and syllable
• CV systems may write non-moraic elements
  – onset clusters, final extrametrical consonants (e.g. Arabic)
• CV systems may fail to write some moraic elements
  – vowel length, gemination, coda consonants
• CV is really just a core syllable
  – supplemental distinctions sometimes correlate with moras, but this is not a design feature
Selected References


