

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** \Rightarrow **constituent** mapping
- Poser’s “right question”
 - *On what sort of phonological analysis must this writing system be based?*
 - i.e. what is the **constituent** \Rightarrow **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 \Rightarrow distinct sign
 - Akkadian, Loma
- **Moraic** : each mora \Rightarrow 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?

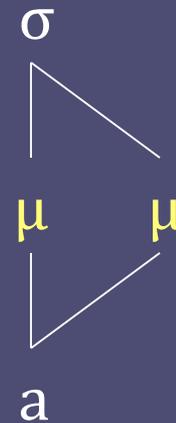
Moras and Vowel Length

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

short /a/,
one mora,
light

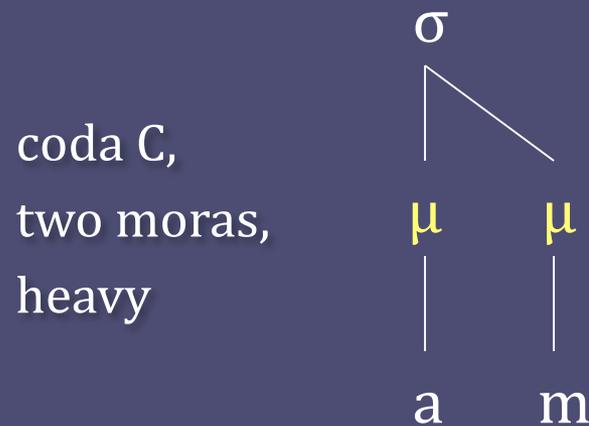
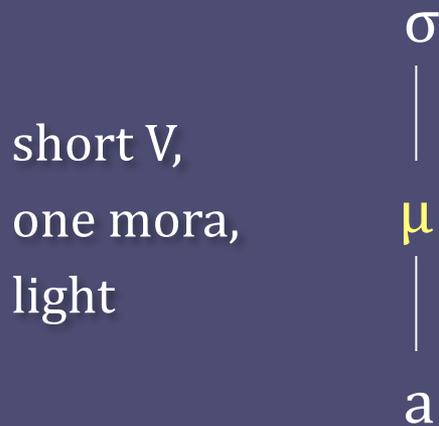


long /a:/,
two moras,
heavy



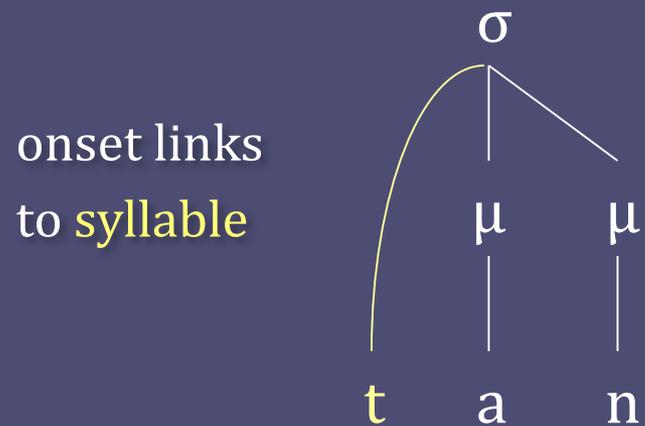
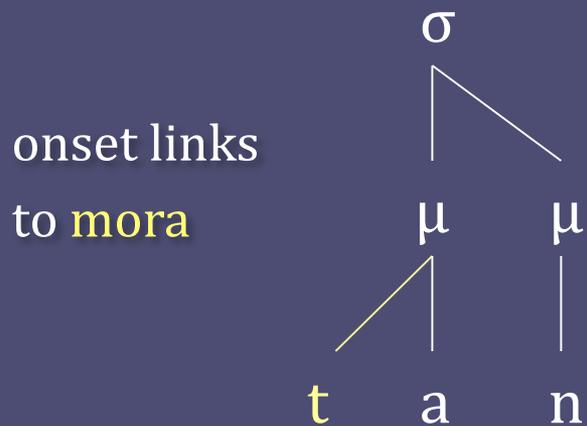
Moras and Coda Consonants

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



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:

$$[[\mathbf{ta}]_{\mu} [\mathbf{n}]_{\mu}]_{\sigma}$$

- If onset links directly to syllable, CV is **not** a unit:

$$[\mathbf{t} [\mathbf{a}]_{\mu} [\mathbf{n}]_{\mu}]_{\sigma}$$

- **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

Vai Examples

᠎ᠠ ᠪᠠ ᠎ᠠ

ki_μ ba_μ lo_μ

kìbálò

‘dream’

᠎ᠠ

la_μ

lá

‘tense marker’

᠎ᠠ ᠬᠠ

la_μ ha_μ

là:

‘paddle’

᠎ᠠ ᠠ

ku_μ wu_μ

kǔ:

‘compound’

᠎ᠠ ᠨᠠ

ke_μ ŋ_μ

kénj

‘house’

Vai Complex Syllables

	Spoken	Written
CVN	+ mora	+ sign
CVV	+ mora	+ sign

- 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

	ä	u	i	a	e	(ɨ)	o
ʔ	አ	ኡ	ኢ	ኣ	ኤ	ኦ	ኧ
b	በ	ቡ	ቢ	ባ	ቤ	ቦ	ቧ
g	ገ	ጉ	ጊ	ጋ	ጌ	ግ	ጘ
d	ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ
	<i>ʔa</i>	<i>bu</i>	<i>gi</i>	<i>da</i>			

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

Tigrinya CV and CVC

ሰላም

sä_μ la_μ m_μ

sälam

‘peace’

ሰልሚ

sä_μ l_μ mi_μ

sälmi

‘muddy soil’

ሰሊማ

sa_μ li_μ ma_μ

salima

‘lightning source’

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

Tigrinya Geminates

አቦ

ʔa_μ bo_μ

ʔabo

‘oleander’

ʔa_μ b_μ bo_μ

ʔabbo

‘father’

በለመ

sä_μ lä_μ mä_μ

sälämä

‘he was calm’

sä_μ l_μ lä_μ mä_μ

sällämä

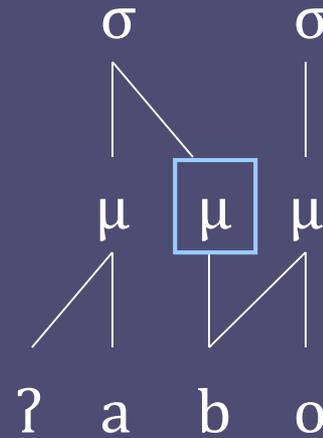
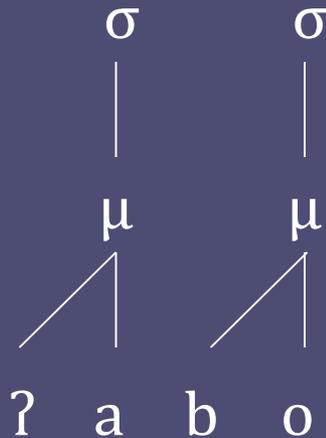
‘he adorned’

- Geminata not written, yet purely moraic
 - it would be easy to do: አብቦ ʔa-b-bo
 - no sign for C ≠ mora for coda

Moras and Gemination

$\text{ʔabo} = \text{ħʌ}$

$\text{ʔabbo} = \text{ħʌ}$



distinction is exactly a mora

Tigrinya Complex Syllables

	Spoken	Written
Coda	+ mora	+ sign
Geminate	+ mora	no sign

- 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
o_μ ka_{μμ} sa_μ n_μ

oka:sa

‘mother’

ボールペン

bo:-ru-pe-N
bo_{μμ} ru_μ pe_μ n_μ

bo:rupen

‘ballpoint pen’

- 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
a_μ s_μ sa_μ ri_μ

assari

‘easily’

ロケット

ro-ke-tsu-to
ro_μ ke_μ t_μ to_μ

roketto

‘rocket’

- 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
dzu_μ mjo_{μμ}

dzumjo:

‘life span’

キャベツ

ki-ja-be-tsu
kja_μ be_μ tsu_μ

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
fa_μ i_μ to_μ

faito

'fight'

ティッシュ

te-i-tsu-si-yu
ti_μ ∫_μ ∫u_μ

tiffu

'tissue'

- C₁V₁ plus small V₂ yields **C₁V₂**
 - for sequences not normally found in Japanese
 - (small) sign for vowel ≠ mora

Japanese Complex Syllables

	Spoken	Written
CVV	+ mora	+ kana
CVN	+ mora	+ kana
CVC	+ mora	+ small kana
CjV	no new mora	+ small kana
novel CV	no new mora	+ small kana

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

ح	س	ن	ر	ب	ل	ك	Ø	و
h	ʃ	n	r	b	l	k	∅	u

- Next nine (possibly) from Indic 1–9

م	ف	د	ت	ل	گ	ڄ	س	ڌ
m	f	d	t	l	g	ɟ	s	d

- Six more used mainly for loanwords
 - most are clearly adapted from other letters

Thaana Vowel Diacritics

- Short vowels based on Arabic pointing

◌̣ ◌̇ ◌̈ ◌̉ ◌̊
a i u e o

- Long vowels by doubling short signs

◌̣◌̣ ◌̇◌̇ ◌̈◌̈ ◌̉◌̉ ◌̊◌̊
a: i: u: e: o:

- Irregularity of ◌̊ (o:)

– Gnanadesikan: abbreviation of doubled ◌̊◌̊ (o)

Open Syllables

دِيْهِ

di_μ ve_μ hi_μ

divehi

‘Dhivehi’

أَلِفُو

a_μ li_μ fu_μ

alifu

‘*alifu* (letter)’

كَاكُو

ka_{μμ} ku_μ

ka:ku

‘who’

كَاكُو

ka_μ ku_{μμ}

kaku:

‘knee’

رِيْزُوْتُ

ri_μ zo_{μμ} tu_μ

rizo:tu

‘resort’

Thaana non-CV

- V without preceding C rests on *alifu*



o_μ i_μ

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



no V



ʔ

Closed Syllables

بَسْ

ba_μ s_μ

bas

‘word, language’

دِزَالَسَا

dza_μ l_μ sa_{μμ}

dzalsa:

‘assembly’ (Urdu)

اَنجِي

a_μ n_μ gi_{μμ}

angi:

‘glove’

كُوْكَوْ

ko_μ ʔ_μ ko_μ

kokko

‘younger sibling’

اُمِّيْدُو

u_μ n_μ mi_{μμ} du_μ

ummi:du

‘hope’

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 (synchronic) 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

رَسْرَدَا

ha_μ n_μ da_μ

handa

‘sea snail’

رَسْرَدُو

ha_μ n̄du_μ

han̄du

‘moon’

Evidence for Moraic View?

- Gnanadesikan 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

سکول

s ku_μ l_μ

sku:l

'school'

فیلم

fi_μ l_μ m_μ ?

film

'film'

پروگرام

p ro_μ g ra_μ m_μ

progra:m ~
purogura:m

'program'

رہائشی

ha_μ u_μ si_μ n_μ

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

- Coulmas, Florian. 2003. *Writing Systems: An Introduction to their Linguistic Analysis*. Cambridge.
- Daniels, Peter T. 1992. The syllabic origin of writing and the segmental origin of the alphabet. *Linguistics of Literacy*. Benjamins. 83-110.
- Gnanadesikan, Amalia E. 2011. Syllables and syllabaries: What writing systems tell us about syllable structure. *Handbook of the Syllable*. Brill. 397-414.
- Gnanadesikan, Amalia E. 2012. Maldivian Thaana, Japanese kana, and the representation of moras in writing. *Writing Systems Research* 4, 91-102.
- Kurzon, Dennis. 2013. Diacritics and the Perso-Arabic script. *Writing Systems Research* 5, 234-243.
- Poser, William J. 1992. The structural typology of phonological writing. Presentation to the Linguistic Society of America, Philadelphia, PA.
- Ratcliffe, Robert R. 2001. What do “phonemic” writing systems represent? *Written Language & Literacy* 4, 1-14.
- Rogers, Henry. 2005. *Writing Systems: A Linguistic Approach*. Blackwell.
- Sampson, Geoffrey. 2015. *Writing Systems*, 2nd ed. Equinox.