

Proposing a new romanization system of Japanese

Kyouro ver. 4.1.0

Part 1

- For romanization, both Hepburn and Kyouro are used; Hepburn is *italicized* and Kyouro is **boldfaced** where clarification is necessary. In sections where only Kyouro is used, Inter is used as a dedicated font.

Purposes and principles

The problems

- Typing
- Pronunciation (especially tone and intonation)
- Parsing

Problem 1: Typing

こんにちは

How many N's?

konnnnitiha

- Today, the primary way of writing Japanese is to type it in on digital devices.
- Students do not know exactly how Japanese is typed in on their computers and smartphones.
- Rarely taught in class (AFAIK).
- The typed spelling is not the same as the Hepburn system or any of the major systems.

Problem 2: Pronunciation



こんにちは

How do you pronounce it?

konnichiwa

ame

“rain” or “candies”?

Iff the “*me*” part is pronounced
in the falling tone, “rain.”

- The *kana* spelling system is complex and has irregularities e.g., おう pronounced ō and は pronounced as wa.
- Tones and intonations are completely unmarked.
- Some students are unaware of lexical accents until they get to some advanced level (or ever).
- Some schools/teachers/textbooks teach them, but none of the pre-existing notations or theories are good enough and accessible to everyone.

Problem 3: Parsing



「今朝目に入って後で返そうと
思って放置してました」

「今週集荷依頼出せるところか
ら住所教えて」

Where are the word
boundaries?

- The Japanese orthography does not mark syntactic boundaries. (A space in English does not always correspond to a syntactic boundary, but much better than Japanese.)
- Some syntactic boundaries—not necessarily a word boundary—needs decomposing a single *kana* letter, e.g., *kak-u* (“write”) vs. *kak-anai* (“not write”).
- Many fossilized function words in idiomatic, seemingly complex, expressions, e.g., *とはいえ*, *それでは*, *こんばんは*. They are holistic and if broken down lose meaning.

The solutions



konniti'hå

'ka'er·u

kaèru'

'ame

ame'

- Typing: written as it is typed.
- Pronunciation (especially tone and intonation): mark them with diacritics and special symbols.
- Parsing: define syntactic boundaries and mark them with spaces and special symbols.

Prosodic units and components

Prosodic units and components

- The syllable
- The mora
- The AP
- The IP
- Contour tones (R, Lv, F; nF = R or Lv)
- Predicative intonations (R%, nR%)
- The boundary fall

The syllable



1 2 3 4 5
ga.kkⁱten

1 23 45
h^utō.kō

1 23 4

gu.rī.n

1 2 3 4 5

shittō.i

- Maximally CCGVG, where C = consonant, G = glide, and V = nucleus.
- V is usually a vowel, *a*, *i*, *u*, *e* or *o*, each with a longer counterpart, but marginally *n*.
- The offglide (syllable-final G) is mostly *i* or *n*, but marginally *e* and *u* can appear.
- C (*k*, *g*, *s*, *z*, *t*, *d*, *n*, *h*, *p*, *b*, *m*, *y*, *r*, *w*, etc.) can be short or long and plain or palatalized.
- The only onglide is *y*, which can be inserted between a palatalized consonant and a following nonfront vowel.

The mora



otoko

anata

ochiba

suki datta n da kara

apointomento

samugaritagaritagaritai

- The mora is a unit of tone allocation.
- In a CCGVG structure with all segments short, the first C bears one mora, CGV bears one, and the last G bears one.
- A long segment (some written with two letters in Hepburn) adds one mora to where it is placed, e.g., *ga.kkiten* (2 syllables, 5 moras), *hutō.kō* (2 syllables, 5 moras)
- The GVG part of a syllable can be at most 2 mora long; G cannot follow a long vowel, e.g., *gu.rī.n* (3 syllables, 4 moras), *shittō.i* (2 syllables, 5 moras)

Contour tones



o.to.ko

a.na.ta

o.chi.ba

s^uki. da.tta n. da. ka.ra

a.po.in.to.men.to

sa.mu.ga.ri.ta.ga.ri.ta.ga.ri.tai

hi.kk'tai

sh'kai.sha

i.sh'ka.ri.hei.ya

- Primary distinction: nF (“non-falling”) vs. F (“falling”).
- Within a certain unit (AP), all F’s follows all nF’s.
- In the sequence of nF’s, the first few moras can be R (“rising”) as opposed to Lv (“level”).
- The sequence of R’s marks the beginning of an IP.
- The sequence of R’s extends from the beginning through *the second mora’s syllable’s nucleus* within the limits of the AP’s nF sequence.
- The prosodic hierarchy: Sentence > IP > AP > Syllable > Mora.

- The basic ideas of the units and the hierarchy is based on Kodama (2008) with some modifications:
 - What we call the IP is a prosodic unit that is in Kodama (2008) called the *p-phrase*.
 - Kodama (2008) assumes in the “default” pronunciation only the first mora of an IP receives the R tone; we care about the syllable structures of the first few moras and distinguish different lengths of the rising tone.

Predicative intonations



kinō nani tabeta (R%)

ashita yotei aru (R%)

kinō gomasaaba tabeta (nR%)

*ashita omatsuri iku ($R\%$) tte
kikareta kara ($nR\%$)*

watashi wa dōsureba ii n
deshōka (nR%)

*masaka dotakyan tte koto wa
nai yonee (nR%)*

*uketeru dake ja shōgi wa
katenai yo (R%)*

- Typically, a question is marked by R% (the “rising” intonation) on the predicate.
- But in certain types of questions the predicate receives nR% (the “non-rising” intonation).
- When the predicate comes in the middle of a sentence, the intonations stays on the predicate.

The boundary fall



umakatta

uma katta

uma katta

ashita omatsuri iku watashi

ashita omatsuri iku, watashi

ame wa ashitaniwa yamu darō

ame wa ashitaniwa yamu darōkedo

- Where two AP's abut and the left one ends with an nF mora, the right AP can begin with a perceptually lowered nF tone (the boundary fall).
- Kodama's (2008)] in the context of nF]nF.
- In Kodama (2008), lowering in level tone, marked by the symbol], is part of the AP templates; we redefine it as a boundary tone.
- The boundary fall is inserted in certain syntactic boundaries where additional conditions are met. (Will be discussed later.)

The distribution of R



*Moeru gomi wa Getsuyōbini
dashimasu.*

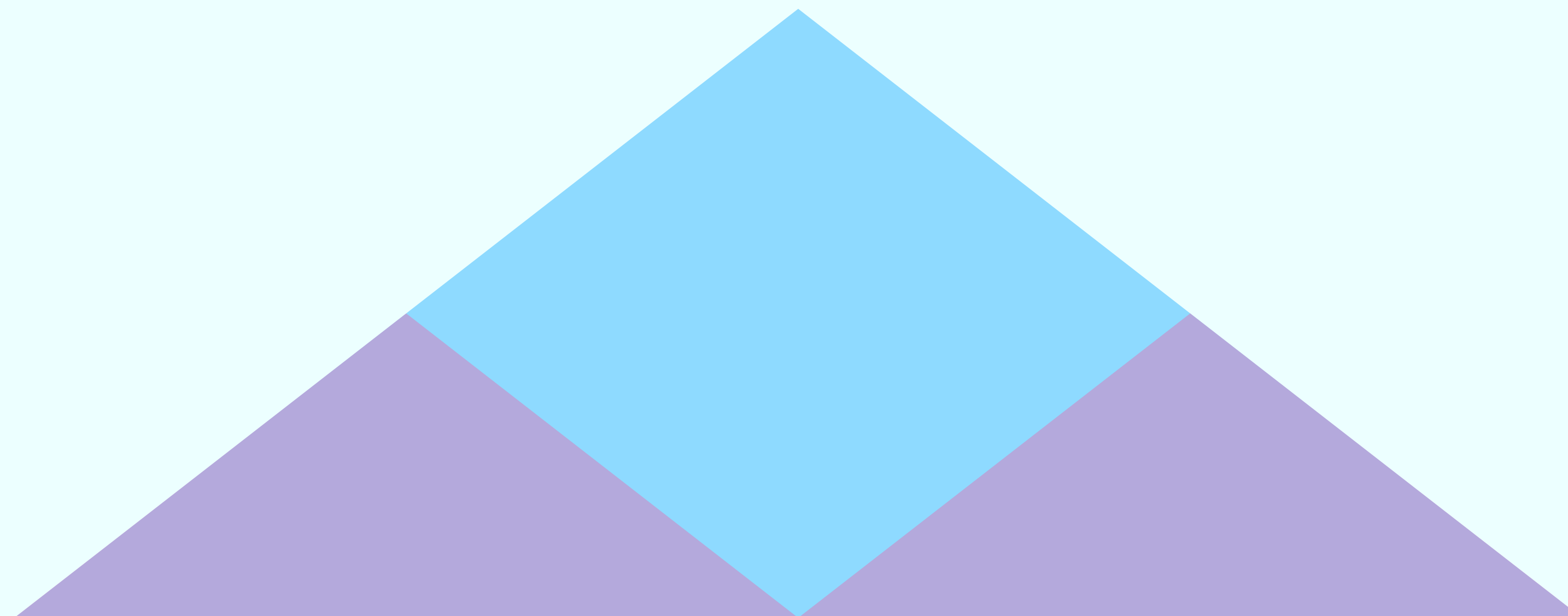
*Moeru gomi wa getsuyōbini
dashimasu.*

*Moeru Gomi wa Getsuyōbini
Dashimasu.*

**moeru Gomi wa getsuyōbini
dashimasu*

**Moeru gomi wa getsuyōbini
Dashimasu*

((*moeru gomi_wa*) (*getsuyōbini*
dashimasu)))



moeru

gomi wa

getsuyōbini

dashimasu

	+	-	-	-
*	+	-	-	+
	+	-	+	-
	+	-	+	+
	+	+	-	-
*	+	+	-	+
	+	+	+	-
	+	+	+	+

- An AP can have an R sequence. (= it can be the beginning of an IP)
- In the triangle structure, an AP that corresponds to the left corner of a triangle whose right corner corresponds to an AP with an R sequence must have an R sequence.
- The leftmost corner of the structure must have R. (= all sentences consist of a whole number of IP's.)

The lexical forms

The lexical forms

- The upper accent
- The lower accent
- The heavy rhyme rule
- Segmental alternation
- AP boundary
- The boundary fall

The upper accent



ke nara

te nara

kaki nara

yama nara

haru nara

karada nara

onna nara

tamago nara

remon nara

ke' nara	<u>ke</u> nara	0
'te nara	<u>te</u> nara	1
kaki' nara	<u>kaki</u> nara	0
ya'ma nara	<u>yama</u> nara	1
'haru nara	<u>haru</u> nara	2
karada' nara	<u>karada</u> nara	0
onn'na nara	<u>onna</u> nara	1
ta'mago nara	<u>tamago</u> nara	2
'remonn nara	<u>remon</u> nara	3

- The accent (') (modifier letter vertical line) is a special symbol used in Kyouro that is inserted immediately before the mora up to which the nF sequence extends within the AP. (First approximation; will be modified later.)
- The first and only the first word of an AP has an accent. (= iff it has an accent somewhere, it is the first word of the AP.)
- The term *accent* or, as opposed to the *lower accent*, the *upper accent* may refer to the symbol itself in the spelling or the corresponding phonological feature.

The lower accent



ke-kara

te-kara

kaki-kara

yama-kara

haru-kara

karada-kara

onna-kara

tamago-kara

remon-kara

ke-kara

ke nara

te-kara

te nara

kaki-kara

kaki nara

yama-kara

yama nara

haru-kara

haru nara

karada-kara

karada nara

onna-kara

onna nara

tamago-kara

tamago nara

remon-kara

remon nara

ke-kara nara

te-kara nara

kaki-kara nara

yama-kara nara

haru-kara nara

karada-kara nara

onna-kara nara

tamago-kara nara

remon-kara nara

ke-no

te-no

kaki-no

yama-no

haru-no

karada-no

onna-no

tamago-no

remon-no

ke-no

ke nara

te-no

te nara

kaki-no

kaki nara

yama-no

yama nara

haru-no

haru nara

karada-no

karada nara

onna-no

onna nara

tamago-no

tamago nara

remon-no

remon nara

ke nara

ke-no

ke-kara

te nara

te-no

te-kara

kaki nara

kaki-no

kaki-kara

yama nara

yama-no

yama-kara

haru nara

haru-no

haru-kara

karada nara

karada-no

karada-kara

onna nara

onna-no

onna-kara

tamago nara

tamago-no

tamago-kara

remon nara

remon-no

remon-kara

-ka₁ra

-₁no

karada' + -ka₁ra → karada₁-ka'ra

onn'na + -₁no → onn₁na-'no

ke' nara

ke,-'no

ke,-ka'ra

'te nara

'te-,no

'te-ka,ra

kaki' nara

kaki,-'no

kaki,-ka'ra

ya'ma nara

ya,ma-'no

ya'ma-ka,ra

'haru nara

'haru-,no

'haru-ka,ra

karada' nara

karada,-'no

karada'-ka,ra

on'na nara

onn,na-'no

onn'na-ka,ra

ta'mago nara

ta'mago-,no

ta'mago-ka,ra

'remon nara

'remon-,no

'remon-ka,ra

- The ablative *-kara*, the possessive *-no*, and some other items extend nF to a specified position when certain conditions are met.
- Observation: The ablative *-kara* extends nF to *ra* iff the item on the left's accent is at its left edge.
- Observation: The possessive *-no* extends nF to *no* iff the item on the left's accent is immediately before the leftmost mora and not at the initial position.
- (and other observations.)

- Generalization: *Certain items “move” the immediately preceding item W’s accent A to the lexically specified position P iff A is not W-initial and is either W-final or within one mora from P.*
- In the lexical form, P is marked by the special symbol (,) (modifier letter low vertical line) and is called the lower accent.
- When A is “moved” to P, P is then marked by ' to show that it is phonetically effective and A’s original position is then marked by , to show the trace. (P and A can be seen as a pair; when P shifts up, A shifts down.)
- Otherwise accents are shown in the lexical form.

karada' + -ka₁ra → karada₁-ka'ra

onn'na + -₁no → onn₁na-'no

ke' nara

ke,-'no

ke,-ka'ra

'te nara

'te-,no

'te-ka,ra

kaki' nara

kaki,-'no

kaki,-ka'ra

ya'ma nara

ya,ma-'no

ya'ma-ka,ra

'haru nara

'haru-,no

'haru-ka,ra

karada' nara

karada,-'no

karada'-ka,ra

on'na nara

onn,na-'no

onn'na-ka,ra

ta'mago nara

ta'mago-,no

ta'mago-ka,ra

'remon nara

'remon-,no

'remon-ka,ra

The heavy rhyme rule



hachiji nara

hachiji-no

hachiji-kara

kinō nara

kinō-no

kinō-kara

kō.ban. na.ra

jō.dan. na.ra

nē.san. na.ra

senshū nara

senshū-no

senshū-kara

ototsui nara

ototsui-no

ototui-kara

mokuyō nara

mokuyō-no

mokuyō-kara

sennsyuu' nara	sennsyuu,-'no	sennsyuu,-ka'ra
ototu'i nara	ototu,i-'no	ototu'i-ka,ra
moku'you nara	moku'you-,no	moku'you-ka,ra

hachiji nara

hachiji-no

hachiji-kara

kinō nara

kinō-no

kinō-kara

ha'tızı nara

kino'u nara

ha'tızı-,no

kino,u-'no

ha'tızı-ka,ra

kino'u-ka,ra

- The V or VG (the presence or absence the onglide is irrelevant) part of the syllable structure is called the *rhyme*.
- If the rhyme of a syllable is two mora long (“heavy”), an accent can intervene.
- If an accent is placed in the middle of a heavy rhyme, the rhym’s latter mora’s contour tone will be identical to the mora immediately following.
- A form whose last two moras are a heavy rhyme with an accent intervening it is said to belong to the *kinō group*.

Segmental alternation



taberu

taberare

tabereba

tabenai

tabeyō

tabeta

tabe

taberu

taberare

tabereba

tabenai

tabeyō

tabeta

tabe

asobu

asobare

asobeba

asobanai

asobō

asonda

asonde

asobu

asobare

asobeba

asobandai

asobō

asonda

asobi

taberu

asobu

ru - u

taberare

asobare

rare - are

tabereba

asobeba

reba - eba

tabenai

asobanai

nai - anai

tabeyō

asobō

yō - ō

tabeta

asonda

ta - (?)da

tabe

asobi

o - i

warau

waraware

waraeba

warawandai

waraō

waratta

warai

warau

waraware

waraeba

warawandai

waraō

wara()tta

wara()i

	<i>ru</i>	<i>ra</i>	<i>re</i>	<i>a</i>	<i>yō</i>	<i>t</i>	<i>i</i>
<i>k</i>	<i>ku</i>	<i>ka</i>	<i>ke</i>	<i>ka</i>	<i>kō</i>	<i>it</i>	<i>ki</i>
<i>g</i>	<i>gu</i>	<i>ga</i>	<i>ge</i>	<i>ga</i>	<i>gō</i>	<i>id</i>	<i>gi</i>
<i>s</i>	<i>su</i>	<i>sa</i>	<i>se</i>	<i>sa</i>	<i>sō</i>	<i>shit</i>	<i>shi</i>
<i>t</i>	<i>tsu</i>	<i>ta</i>	<i>te</i>	<i>ta</i>	<i>tō</i>	<i>tt</i>	<i>chi</i>
<i>n</i>	<i>nu</i>	<i>na</i>	<i>ne</i>	<i>na</i>	<i>nō</i>	<i>nd</i>	<i>ni</i>
<i>b</i>	<i>bu</i>	<i>ba</i>	<i>be</i>	<i>ba</i>	<i>bō</i>	<i>nd</i>	<i>bi</i>
<i>m</i>	<i>mu</i>	<i>ma</i>	<i>me</i>	<i>ma</i>	<i>mō</i>	<i>nd</i>	<i>mi</i>
<i>r</i>	<i>ru</i>	<i>ra</i>	<i>re</i>	<i>ra</i>	<i>rō</i>	<i>tt</i>	<i>ri</i>
<i>w</i>	<i>u</i>	<i>wa</i>	<i>e</i>	<i>wa</i>	<i>ō</i>	<i>tt</i>	<i>i</i>
<i>i</i>	<i>iru</i>	<i>ira</i>	<i>ire</i>	<i>i</i>	<i>iyō</i>	<i>it</i>	<i>i</i>
<i>e</i>	<i>eru</i>	<i>era</i>	<i>ere</i>	<i>e</i>	<i>eyō</i>	<i>et</i>	<i>e</i>

- *r* and *y* are deleted after a consonant.
- *t* becomes *d* after *g*, *n*, *b* or *m*.
- A vowel is deleted after a vowel.
- *k* and *g* become a nucleus *i* between a preceding *e* and a following consonant.
- *k* and *g* become an offglide *i* between a preceding *a*, *i*, *u* or *o* and a following consonant.
- *m*, *b*, and *m* become an offglide *n* before *C*.
- *t* becomes long (*tt*) after *t*, *r* or *w*.
- *t*, *r* and *w* is deleted before *tt*.
- *w* is deleted before a nucleus or offglide *i*.
- *i* is inserted between a preceding *s* and a following *t*.

- $\{r, y\} \rightarrow \emptyset / C_.$
- $t \rightarrow d / \{g, n, b, m\}_.$
- $V \rightarrow \emptyset / V_.$
- $\{k, g\} \rightarrow i \text{ (nucleus)} / e_C.$
- $\{k, g\} \rightarrow i \text{ (offglide)} / \{a, i, u, o\}_C.$
- $\{n, b, m\} \rightarrow n \text{ (offglide)} / _C.$
- $t \rightarrow tt / \{t, r, w\}_.$
- $\{t, r, w\} \rightarrow \emptyset / _tt.$
- $w \rightarrow \emptyset / _i \text{ (nucleus or offglide)}$
- $\emptyset \rightarrow i / s_t.$

AP boundaries



|*moeru* |*gomi wa* |*getsuyōbini* |
dashimasu

|yopparau |tabini |mukashino |dōkyūseini |
puropōzu |suru no wa |mō |yameru|beki da.

|*mainichi* |*nattōkimuchitamago*
o |*taberu*|*bekida*

yamerubeki

**yametabeki*

**yamenaibeki*

taberubeki

**tabetabeki*

**tabenaibeki*

- Some lexical forms have AP boundaries in the middle.

The boundary fall



ame wa ashitaniwa yamu darō

ame wa ashitaniwa yamu darōkedo

otokoippiki

tsugi kuru toki wa wasurenaidekudasai ne

ienakiko

- Some lexical forms have boundary falls.

Conclusion of Part 1



- A language must be described before it can be applied to spelling rules.
- Therefore Kyouro includes a grammatical description as a major component. (Developing Kyouro is similar to describing the target language.)
- The spelling rules must satisfy the typeability principle; it has to be spelled as it's typed. Phonological components that are otherwise unmarked therefore must be marked by diacritics and special symbols.

References



- Kodama, N. (2008). 曲線声調と日本語韻律構造. 『ありあけ』 熊本大学言語学論集 7 1-40.