Proposing a new romanization system of Japanese

Kyouro ver. 4.1.1

Ver. 4.1.0, October 28, 2023. @awesomenewways

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?

konnnitiha

- 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思うか
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

konnnɪtɪˈhȧ

'ka'er·u

kaėruꞋ

'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.kk¹ten

1 23 45

hᵘtō.kō

1   23 4

gu.ri.n

1 2 3 4 5

shᶦttō.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

_oto_ko

_ana_ta

_o_chiba

suki datta n da kara

*apo*intomento

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*

sᵘki. da.tta n. da. ka.ra

a.po.in.to.men.to

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

hi.kkⁱtai

shⁱkai.sha

i.shⁱka.ri.hei.ya

- Primary distinction: nF ("non-falling") vs. F ("falling").

- Within a certain unit (AP), all F's fallows all nF's.

- In the sequence of nF's, the first few moras <u>can</u> 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" pronounciation 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 gomasaba 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

*uma*katta

uma *katta*

uma *ka*tta

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.)
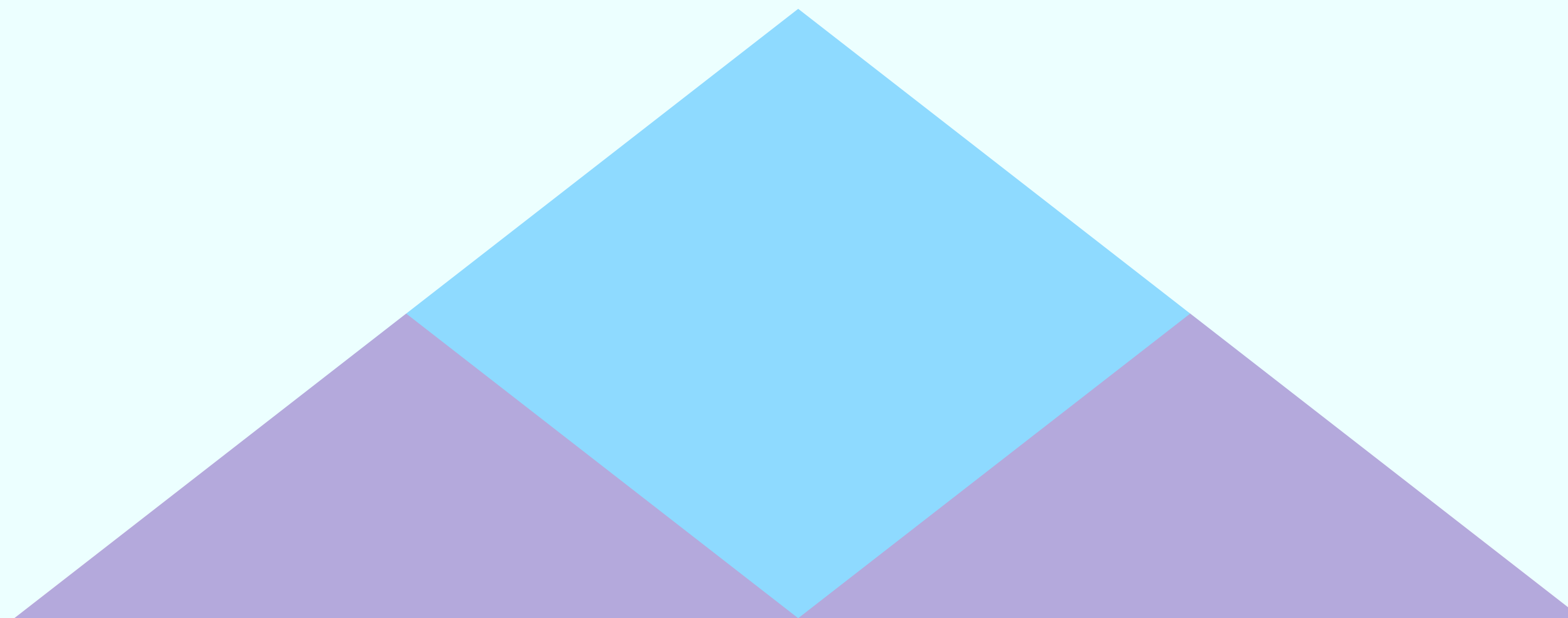
The distribution of R

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
+	−	−	−
+	−	−	+
+	−	+	−
+	−	+	+
+	+	−	−
+	+	−	+
+	+	+	−
+	+	+	+

(rows 2 and 6 marked with *)

- 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 bimoraic 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	_ke nara_	0
'te nara	_te nara_	1
kakı' nara	_kakı nara_	0
ya'ma nara	_yama nara_	1
'haru nara	_haru nara_	2
karada' nara	_karada nara_	0
onn'na nara	_onna nara_	1
ta'mago nara	_tamago nara_	2
'remonn nara	_remon 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

*har*u-kara nara

karada-kara nara

onna-kara nara

*tamag*o-kara nara

*re*mon-kara nara

ke-no

te-no

kaki-no

yama-no

haru-no

karada-no

onna-no

*tama*go-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*

-kaˌra

-ˌno

karadaˈ + -kaˌra → karadaˌ-kaˈra

onnˈna + -ˌno → onnˌna-ˈno

keˈ nara	keˌ-ˈno	keˌ-kaˈra
ˈte nara	ˈte-ˌno	ˈte-kaˌra
kakɪˈ nara	kakɪˌ-ˈno	kakɪˌ-kaˈra
yaˈma nara	yaˌma-ˈno	yaˈma-kaˌra
ˈharu nara	ˈharu-ˌno	ˈharu-kaˌra
karadaˈ nara	karadaˌ-ˈno	karadaˈ-kaˌra
onˈna nara	onnˌna-ˈno	onnˈna-kaˌra
taˈmago nara	taˈmago-ˌno	taˈmago-kaˌra
ˈremon nara	ˈremon-ˌno	ˈremon-kaˌra

- 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 (ˌ) (modifier letter low vertical line) and is called the lower accent.

- When A is "moved" to P, P is then marked by ˈ to show that it is phonetically effective and A's original position is then marked by ˌ to show the trace. (P and A can be seen as a pair; when ˌ shifts up, ˈ shifts down.)

- Otherwise accents are shown in the lexical form.

karadaˈ + -kaˌra → karadaˌ-kaˈra

onnˈna + -ˌno → onnˌna-ˈno

ke' nara	keˌ-'no	keˌ-ka'ra
'te nara	'te-ˌno	'te-kaˌra
kakɪ' nara	kakɪˌ-'no	kakɪˌ-ka'ra
ya'ma nara	yaˌma-'no	ya'ma-kaˌra
'haru nara	'haru-ˌno	'haru-kaˌra
karada' nara	karadaˌ-'no	karada'-kaˌra
on'na nara	onnˌna-'no	onn'na-kaˌra
ta'mago nara	ta'mago-ˌno	ta'mago-kaˌra
'remon nara	'remon-ˌno	'remon-kaˌra

The bimoraic rhyme rule

hachiji nara   *hachiji*-no   *hachiji*-kara

kinō nara   *kinō*-no   *kinō*-kara

kō.ban. _na_.ra

jō.dan. na.ra

_n_ē.san. na.ra

senshū nara	_senshū-no_	_senshū-kara_
ototsui nara	_ototsui-no_	_ototui-kara_
mokuyō nara	_mokuyō-no_	_mokuyō-kara_

sennsyuuˈ nara   sennsyuuˌ-ˈno   sennsyuuˌ-kaˈra

ototuˈɪ nara   ototuˌɪ-ˈno   ototuˈɪ-kaˌra

mokuˈyou nara   mokuˈyou-ˌno   mokuˈyou-kaˌra

hachiji nara   *hachiji*-no   *hachiji*-kara

kinō nara   *kinō-no*   *kinō*-kara

ha'tɪzɪ nara ha'tɪzɪ-ˌno ha'tɪzɪ-kaˌra

kɪno'u nara kɪnoˌu-'no kɪno'u-kaˌra

- 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, an accent can intervene.

- If an accent is placed in the middle of a bimoraic 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 bimoraic rhyme with an accent intervening it is said to belong to the *kinō group*.

Segmental alternation

taberu

taberare

tabereba

tabenai

tabeyō

tabeta

tabe

*tabe*ru

*tabe*rare

*tabe*reba

*tabe*nai

*tabe*yō

*tabe*ta

tabe

asobu

asobare

asobeba

asobanai

asobō

asonda

asonde

*asob*u

*asob*are

*asob*eba

*asob*anai

*asob*ō

aso<u>nd</u>a

aso<u>b</u>i

taberu	*asobu*	*ru - u*
taberare	*asobare*	*rare - are*
tabereba	*asobeba*	*reba - eba*
*tabe*nai	*asoba*nai	*nai - anai*
tabeyō	*asobō*	*yō - ō*
tabeta	*aso<u>nd</u>a*	*ta - (?)da*
tabe	*aso<u>b</u>i*	*0 - i*

warau

waraware

waraeba

warawanai

waraō

waratta

warai

*wara*u

*warawa*re

*wara*eba

*warawa*nai

*wara*ō

wara(_)tta

wara(_)i

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

- *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} —> 0 /C_.

- t —> d /{g, n, b, m}_.

- V —> 0 /V_.

- {k, g} —> i (nucleus) /e_C.

- {k, g} —> i (offglide) /{a, i, u, o}_C.

- {n, b, m} —> n (offglide) /_C.

- t —> tt /{t, r, w}_.

- {t, r, w} —> 0 /_tt.

- w —> 0 /_i (nucleus or offglide)

- 0 —> i /s_t.

AP boundaries

|moeru |gomi wa |getsuyōbini | dashimasu

|yopparau |tabini |mukashino |dōkyūseini | puropōzu |suru no wa |mō |yame<u>ru|beki</u> da.

|mainichi |nattōkimuchitamago o |taberu|bekida

yamerubeki

*yametabeki

*yamenaibeki

taberubeki

tabetabeki

tabenaibeki

- Some lexical forms have AP boundaries in the middle.

The boundary fall

_a_me wa _ashita_niwa _yamu_ darō

_a_me wa _ashita_niwa _yamu_ _dar_ōkedo

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.