PROPOSING A NEW ROMANIZATION SYSTEM OF JAPANESE

Kyouro/教育ローマ字 ver. 5.2.9

PREFACE

About this PDF file

This document announces version 5 of the romanization system, tentatively called *Kyōiku Rōmaji* in Japanese. In this document, the said romanization system is called *Kyouro* (rhymes with Kyoto). Author: Junichi Iida, aka @awesomenewways on twitter.

The latest version is available under the following GitHub repository: https://github.com/NihongoTopics/Kyouro/.

If you wish to use Kyouro in your classroom, textbook, YouTube videos, blogs, etc., I'll be happy to help. Also if you're looking for a Japanese teacher, I can offer you personal lessons. Feel free to reach me on Twitter.

- This document has been released as incomplete (though effective) and will be updated at intervals. Errors are expected.
- Japanese sentences, phrases, and words are written in the Hepburn system and Kyouro. When clarification is necessary, Hepburn is italicized and Kyouro is boldfaced.
- Some sentences are taken from literary works such as novels and TV shows. Cited sentences are written in either Kyouro or Hepburn even if the original notation is in the standard orthography. The pronunciation is assumed where the source does not include sound.
- The reader is expected to speak and understand Japanese.
- As an experiment, alphabetically ordered English words are given in place of page numbers. Sometimes the symbol "»" is employed to mean "refer to" with the English word assigned to the designated page. ("»apple" reads "refer to the page apple.")
- Examples written in the Hepburn notation do not always reflect word boundaries as analyzed in Kyouro, but an effort was made to make the notation mostly consistent with Kyouro.
- The variant of the Japanese language described in Kyouro is intended to be the one that is conventionally called Tokyo Japanese in the literature or a version of it. In some parts of this document, the intended lect is referred to as the "reference dialect." Controversial claims about this dialect not supported by explicit references are based on my introspections.
- The person who writes Kyouro is referred to as the *speller*.
- The symbol # prefixed to a sentence indicates that the meaning differs from the intended one.
- This PDF file includes hyperlinks.
- This PDF file includes the table of contents accessible via the PDF menu.

Notable changes in 4.1

Ver. 4.1, in a separate file, released in October 2023.

In the linguistic theory:

- New syllable theory (the CCGVG theory) proposed.
- Binary predicative intonations: R% and nR%.

Notable changes in 5.0

Ver. 5.0, the original version of this document, released in December 2023.

In the linguistic theory:

- Added P5, G1, and G2 boundaries.
- C1 reinterpreted.
- “Adjunct” *ku* no longer considered an adjective suffix.
- Trimming proposed.
- Boundary generation proposed.
- Merging direction specification.
- Potential split of R2 proposed.
- Comment on syllabification and word boundary.
- Comment on *u* offglide.
- Comment on nasal before a long consonant.
- Comment on syntactic compositionality.
- Comment on the notions of word and inflection.
- Comment on the notion of morpheme.

In the spelling rules:

- Predicative intonation R% shown by ¿ prefixed to the affected word.
- ? no longer in charge of intonation.
- In-word comma rule.
- Ellipsis rule.
- Non-typed long vowel symbol: combining macron.
- Dot above reworked.
- Boundary fall reworked for verbal nouns.
- 1.x materials no longer effective.

Notable changes in 5.1

Ver. 5.1 released in January 2024.

In the linguistic theory:

- The role of the lexicon explained.

Notable changes in 5.2

Ver. 5.2 released in July 2024.

- $ theory now written “$-theory” with hyphenation.
- The subsection “subordinate clause” expanded with comments on some |C1 verb suffixes. »northwest-obeche
- The subsection stab “the te iru form” renamed to “progressive and perfect.” (The contents have been removed for easier editing but hopefully only temporarily.) »obeche
- Some C1| items now marked “provisional.” »jeans
- Several errors have been corrected.

Chapter One

PURPOSES AND PRINCIPLES

THE PROBLEMS

The lack of instruction on typing

Kyouro serves as a reasonable representation of how the language is typed in on digital devices. The input method of Japanese is rarely mentioned in Japanese education. Students spend a considerable amount of time practicing handwriting. But the instruction on the Japanese input method is often omitted. Since typing on a computer or smartphone is the primary way of writing Japanese today, this can be a serious problem.

For example, the moraic nasal of Japanese is written as a single *n* in the Hepburn romanization system. This may hinder the student from correctly guessing the keystrokes of common phrases like *konnichiwa*. In Kyouro the same phrase is written as **konnnıtı'hȧ**, which is exactly how it should be typed in on the computer or smartphone: k-o-n-n-n-i-t-i-h-a. The number of **n**'s is exactly how many times it has to be pressed down to type it. Similarly, small tsu in Kyouro is written either **xtu** or by doubling the following consonant letter, just like how it should be typed in.

The lack of instruction on tones and intonations

Kyouro shows lexical tones and intonations clearly. Lexical tones and intonations are important components of the phonological structure, but they are taught only briefly if at all, or taught through the problematic traditional level tone theory.

A separate section is in preparation to discuss why the level tone account is problematic. For the issue put into the context of Japanese education, Hasegawa (1995) is convincing and concise.

Kyouro allows teachers and textbook authors to formally represent the lexical tones and intonation components of a sentence in a sophisticated, more objective manner. Since it relies less on the learned "intuition" that the traditional account is based on, it can be more intuitive to the second language Japanese speakers.

Obscure syntactic boundaries

Kyouro represents the tokens and the structure of a sentence in terms of their syntactic and phonological properties. The Japanese orthography does not contain spaces between syntactic entities, and this makes it difficult for the student to parse a sentence correctly. This can be a major obstacle when learning Japanese.

In Kyouro, a sentence is spelled in such a way that it is clear what tokens it consists of. The syntactic status of each token is broadly marked by spaces and a few special symbols. With each token clearly represented with its phonological and syntactic properties, the meaning of a sentence can be more easily deduced from the properties of the tokens present.

DESIGN PRINCIPLES

The hierarchy of design principles

Beginner	How do I type this on the smartphone?
	How do I pronounce this?
	Why do I pronounce this this way?
	Why do I use this form?
Advanced	What forms are possible and what aren't?

The design principles of Kyouro can be summarized as the typeability principle, the pronounceability principle, and the parseability principle. These principles are prioritized in the order given: typing first, pronunciation second, and parsing last.

The typeability principle requires that the learner can easily guess from Kyouro spelling the keystrokes to get the correct orthographic form on screen. It has to be spelled as it is typed. An average layperson who speaks General American natively and knows nothing about Japanese should be able to guess and type in the sentence correctly (but not necessarily choose the conversion candidate correctly). By setting the bar to type Japanese so low, it intends to provide the potential learner with an easy way to write her choice of phrases before even starting to study Japanese. Since the Japanese input systems predict the correct conversion candidate most of the time, just knowing the keystrokes can allow the user to type Japanese correctly. It potentially lowers the initial burden of learning Japanese.

The pronounceability principle contends that a simple set of rules must deduce the phonetic form from Kyouro spelling. The student should be able to see how the entente is pronounced without memorizing ad-hoc rules. The spelled and phonetic forms however do not have a one-to-one correspondence. Many spelling forms can correspond to the same phonetic form. While the principle demands that the phonetic form can be deduced from the spelling, the reverse deduction is not expected. The same pronunciation can be obtained from multiple spelling forms.

The parseability principle is that the phonological, syntactic, and semantic properties of a unit form must be reasonably reflected in the spelling. The composition of a sentence must be to a reasonable degree transparent from the representation.

These major principles are implemented in the spelling system in the order of priority given above. That is, when two or more of the three major principles compete, the one with the higher priority has its way.

Broad design principles outside of the hierarchy

Kyouro shall not replace existing writing systems: It is mainly intended to be used alongside the full orthography of Japanese although it is also reasonable to occasionally use Kyouro without a parallel text. It is not intended to teach Japanese only with Kyouro or to use Kyouro as the everyday writing system. Kyouro is designed to help students learn how to write and read the standard orthography. Perhaps its role is somewhat analogous to yomigana. It tells you how to pronounce and how to type. Kyouro's spelling rules should be designed so that information that can be easily extracted from the usual orthography is not contained in the Kyouro spelling.

Change spelling as slowly as possible: As the theory develops, analyses often demand some of the previously accepted spelling forms be modified. Since as a spelling system stability is favored, when two analyses are nearly equally compelling, the one that makes smaller or fewer changes in spelling should be adopted.

Use fewer scary symbols: It is important to keep the spelling less scary for the student and the teacher. When two analyses are nearly equally compelling, the one that uses fewer scary symbols should be adopted. Also, new spelling rules reducing the use of scary symbols are more favored than the ones that increase it.

Don't run after two hares: If multiple phonetic realizations of the same written form are interchangeable, complicating the rules or the theory is necessary only if none of them can be sufficiently represented in the given version of Kyouro. If one is caught, let go of the others.

Put a burden of ten on the teacher if it means lifting a burden of one from the student: A possible criticism of Kyouro is that it is often too difficult to write. That it is difficult to write is partly intentional. It is designed to be difficult to write, but easy to read. After all, it is a spelling system to help learners.

It is intended that the written form reflects the best effort to analyze the language as formally and rationally as possible so that cultural background is not required to learn Japanese. The speller is responsible for providing the student with the best analysis of the language, as a direct component of the spelling or otherwise, to make learning easier and smoother, even if the effect is minimal. Kyouro is difficult precisely because human languages are difficult to objectively characterize.

LETTERS AND SYMBOLS

Typed symbols

Segments are written as it is typed. While in typical input systems, both the kunreishiki system and the Hepburn system are allowed for the most part, in Kyouro kunreishiki style is adopted where possible, with certain exceptions mentioned separately.

The primary reference system is currently the iPadOS's default Japanese input method, but it has been designed to be compatible with Windows as well. Other major systems may be added in the future. See also p. sweetheart.

Non-typed symbols

The following non-alphabetic symbols are used. These symbols are ignored when typing the orthographic Japanese. Inside the parentheses are the unicode names.

- ˈ (modifier letter vertical line): The upper accent. The phonetically ‘active’ accent responsible for the surface tones.
- ˌ (modifier letter low vertical line): The lower accent. The phonetically ‘inactive’ or ‘void’ accent not affecting the surface tones.
- ˋ (modifier letter grave accent): The boundary fall. Phonetically, a steep pitch fall between two adjacent nF moras.
- ˎ (modifier letter low grave accent): The lower boundary fall. 1. The trace of an unrealized boundary fall. 2. An AP boundary without a phonetic boundary fall within the limits of a tokenized form.
- ̇ (combining letter dot above): 1. Distinguishes a syllabic vowel from a nonsyllabic vowel (glide) after a vowel letter. 2. Given to the stem-final vowel of a vowel verb. 3. Added to a vowel **a** or **e** when the preceding **h** is not pronounced as is (e.g., **hȧ**)
- ̄ (combining macron): Vowel length that is not reflected in the orthographic form.
- · (middle dot): Used in suffixes of the verb group. For verb suffixes, it is inserted on the right side of a segment that is subject to the segmental alternation. For adjective suffixes, it is inserted between the stem and the suffix.
- **-** (hyphen minus): Used in suffixes of the noun group.
- **¿** (inverted question mark): The rising predicative intonation.

More detailed explanations are provided in Chapter Seven.

Alphabet

The following letters, including the dotted and dotless versions, are used:

Lowercase letters: **a ȧ b d e ė f g h ı i k m n o ȯ p r s t u v w x y z xtu**

Capital letters: **A Ȧ B D E Ė F G H I İ K M N O P R S T U V W X Y Z Xtu**

The long vowel symbol

The long vowel symbol is transcribed as the em dash: —.

Numbers

Numeric symbols (1, 2, 3, etc) are not used. Numbers are spelled out whenever possible. If it cannot be spelled out, it cannot be written in Kyouro.

Chapter Two

PROSODY

PROSODIC STRUCTURE

The prosodic hierarchy

Kyouro recognizes the following prosodic units: the mora, the syllable, the accent phrase (AP), the intonation phrase (IP), and the sentence. They constitute strict layers in the order given above (from the smallest to the largest). Each non-terminal prosodic unit contains a whole number of prosodic units that are exactly one level lower. In the hierarchy, the mora is the lowest level and is the terminal prosodic unit.

A sentence consists of a whole number of intonation phrases; an intonation phrase consists of a whole number of accent phrases; an accent phrase consists of a whole number of syllables; and a syllable consists of a whole number of moras.

The use of these symbols as well as the basic understanding of the tonal system and the prosodic structure of Japanese is largely based on Kodama (2008). The difference between the present Kyouro account and his original account will be noted where it matters. A subsection to discuss the details of Kodama's prosodic hierarchy is in preparation. To note some obvious changes, Kodama's p-phrase is called the IP and Kodama's p-word is called the AP; in Kodama's system, the mora is not explicitly mentioned as part of the hierarchy but we use the term *mora* to refer to a prosodic unit that belongs to and is exactly one level lower than a syllable in the prosodic hierarchy.

The mora

A mora bears one of the three contrastive contour tones: the rising tone (R), the level tone (Lv), and the falling tone (F). R and Lv can be grouped by the symbol nF, which stands for "non-falling."

The allocations of nF and F are lexical. A lexical item may have *accents*. The position of accents in an AP determines which moras of the AP receive nF and which receive F.

The occurrence of R as opposed to Lv is determined by the syntactic structure and the information structure of the sentence. A sentence however must have at least one R mora.

These tones, R, Lv, and F, play a major role in characterizing the prosodic units below the sentence level.

The syllable

A syllable bears a whole number of moras. It consists of the onset and the rhyme. The onset is made of zero to two consonants and does not bear a contrastive tone in the phonetic form. The rhyme consists of an optional onglide, an obligatory nucleus (usually a vowel, but marginally a syllabic nasal), and an optional offglide. The rhyme bears up to two moras and corresponding tones.

The accent phrase (AP)

An AP consists of a whole number of syllables. It begins with one or more consecutive nF moras and ends with zero or more consecutive F moras. An nF mora cannot intervene in a sequence of F moras and neither can an F mora intervene in a sequence of nF moras. Consequently, if an nF mora follows an F mora, there must be an AP boundary in between.

If an AP ends with an nF mora, the AP boundary can be phonetically not apparent. This is a direct consequence of the fact that an AP takes the shape nF1F0 (*"one or more nF's and following zero or more F's"*).

Between two APs, a boundary fall may be found. A boundary fall can appear at certain syntactic boundaries if the preceding AP ends with an nF mora.

The intonation phrase (IP)

An IP consists of a whole number of APs. It has an R mora on the nucleus of the first syllable if that is the last syllable containing an nF in the AP. Otherwise, it has R moras through the second mora's syllable's nucleus' first mora. No other moras can be R within the same IP.

Kyouro's IP corresponds to Kodama's (2008) *p-phrase*.

The sentence

A sentence consists of a whole number of IPs. It has at least one predicative intonation. The predicative intonations are relevant to whether the sentence is declarative, interrogative, etc.

The sentence as a prosodic unit is expected to correspond to the syntactic sentence.

One or more prosodic levels between the IP level and the sentence level may be confirmed as a result of more thorough phonetic and phonological analyses, but the current version of Kyouro does not recognize any intermediate levels in its spelling.

TONES AND INTONATIONS

R

The R tone is given to the sequence from the beginning through the second mora's syllable's nucleus' first mora within the initial AP's nF sequence of an IP. If the second mora belongs to the first syllable—this happens when the onset contains two consonants or at least one long consonant—or the IP's nF sequence is only one syllable long, the R tone is given to the first syllable's nucleus only.

The R tone at the beginning of a word may be difficult to pay attention to. The R tone is utterance-initially obligatory. When a native speaker introspects into the shape of a word, she inevitably introspects into what an utterance that contains it sounds like. Since an utterance imposes certain tonal or intonation components on words, including the R tone, a simple introspection fails to separate one component from the other components of the utterance and therefore fails to discern what components belong to the word shape.

The mechanisms determining the placement of R are explained later in this chapter and Chapter Six. A reader who speaks natively or fluently a variation of Japanese that shares the basic tonal and intonation system with our reference dialect can examine his own speech and test what tonal phenomena most resemble the R tone described here.

The R tone has been informally or implicitly known in the literature for several decades. In the level tone account, one aspect of the R tone has been known as the "initial lowering" (e.g., Selkirk, et. al. 2003), which refers to the fact that a phrase normally begins with a low-pitched mora followed by a high-pitched mora (LH) unless the first mora is accented. It has also been claimed by many that when the first syllable of the phrase is heavy the initial lowering is canceled or, as per Beckman & Pierrehumbert (1986), weakened and the pitch contour will be HH (a high-pitched mora followed by another high-pitched mora) or a weak L followed by an H.

Several phonetic studies (Sugito 1969, P&B 1986, Kori 2004, Maekawa 2018, among others) exhibit a rising pitch contour in the first one or two moras of a phrase. It seems that the traditional view interpreted this type of rising pitch contour as either LH or HH depending on its position and length relative to moras.

The phonological nature of R as opposed to Lv is described in Kodama (2008) and, in a different notation, Kori (2008a). Kori's (2008a) view is that the rising pitch contour is of the default word tone and it is suppressed under certain semantic contexts. This observation is reinterpreted in Kyouro and will be explained in Chapter Six.

blank

Lv

The Lv tone is characterized by its relatively flat pitch contour. When the starting pitch is high enough, typically, consecutive Lv moras are pronounced with a gradual lowering of the pitch.

It is important not to confuse the Lv tone with the high tone that is supposed in the traditional level tone account. In the level tone account, the intonation components and lexical tone components are conflated and both are treated in terms of high and low; the consecutive R tones at the top of an intonation phrase are treated as a sequence LH, and a single R tone as a single H.

The level tone account fails to capture certain phonological contrasts when multiple APs are present in the same IP.

F

F is the falling tone contrasted with nF. The placement of F and nF is lexical, and the F tone is observed at the end of words such as:

- *haru* (nF-F) ('spring')
- *aruku* (nF-nF-F) ('to walk')
- *tatsu* (nF-F) ('to stand up')

But it is not observed anywhere in words such as:

- *ashi* (nF-nF) ('foot')
- *uma* (nF-nF) ('horse')
- *yama* (nF-nF) ('mountain')
- *warau* (nF-nF-nF) ('to laugh')
- *suwaru* (nF-nF-nF) ('to sit')

The following are examples of words that have multiple F moras:

- *gengo* (nF-F-F) ('language')
- *minamiaoyama* (nF-nF-nF-nF-nF-F-F) ('Minami-Aoyama')
- *magunichūdo* (nF-F-F-F-F-F) ('magnitude')

The F tone can be understood as the default tone. If neither Lv nor R is given to a mora, the mora receives the F tone.

An F mora is characterized by its falling pitch contour. In a sequence of two or more F moras, the pitch will continue to fall smoothly throughout the sequence. If the pitch is already extremely low at the beginning of an F mora, the falling pitch contour may not be clearly realized. It seems that for some people an F after an nF is more noticeable than an F after another F.

Boundary fall

A boundary fall can be inserted among other places between the A1 and A4 syntactic boundaries.

The precise phonetic nature of the boundary fall is yet to be described. Although far from conclusive, Iida (2021) found a greater pitch fall in an F-F heavy syllable right after an accent nucleus compared to an nF-nF syllable in the same position. More thorough measurements and analyses are called for. Minimal triplets such as *umakatta* "it was tasty," *uma katta* "bought a horse," and *uma katta* "raised a horse" demand that the boundary fall be recognized as a phonological component.

The boundary fall as a phonological and phonetic phenomenon contrasted with contour tones in Tokyo Japanese has, to my knowledge, never been explicitly proposed in the literature before Kyouro. Implicit or indirect indications of the boundary fall, however, have been known for several decades.

Kawakami (1985) notes that *hana takashi,* where *hana* means "flower," accompanies a steep pitch fall over the word boundary with or without a following rising pitch contour, citing Kawakami (1953, 1963). This is a clear indication that, according to his introspection, a falling contour can occur where two adjacent nonfalling moras abut.

Hattori (1961) compares the surface forms of *buta kau* and *inu kau*. He transcribes them as the following:

- [LbuΓtakau]~[LbuHta'Γkau]
- [LiΓnuLkau]~[LiΓnuHkaLu]~[LiΛnu'˥kaLu]

(Hattori (1961), p. 195. Incompatible symbols are substituted with graphic equivalents.)

Again, these examples show that a falling contour can occur in the middle of or over adjacent nonfalling moras and it contrasts with a relatively flat pitch contour that occurs otherwise (e.g., "*taka"* in the first example's first possible realization).

It is not clear why he did not insert a low-tone symbol between *ka* and *u* in the first example. Since the first mora *ka* is in a high tone, the lowness of the second mora of the word should have no problem realizing phonetically. In the next page (p. 196), he writes the phonemic form of this example as /|buta|kaˀ'u/, where the would-be pitch fall is clearly indicated by the symbol "ˀ." The assumed phonemic form of *kau* is identical as the one in *inu kau*, which is /|'inuˀ|kaˀ'u/. Regardless, the point is he transcribed a falling tone in the second mora of *inu* whose corresponding position of *buta* is in a simple mid or high tone.

Note that this type of falling pitch contour cannot be equated with the F tone found after the *nucleus*. The nucleus of *inu* is at the second mora as indicated in Hattori's aforementioned phonemic notation. Since this form contrasts with forms with the nucleus at the first mora such as *kau* (/kaˀ'u/ in Hattori's notation), there must be some way to distinguish an nF mora pronounced with an allegedly falling pitch contour ([Λ˥] in Hattori's notation) from the usual F mora such as the second mora of *kau*. That is, nF-nF]nF (e.g., ***inu kau***) and nF]F-nF (e.g., ***kau inu***) are distinguished.

In Hattori (1961), the equivalent of the nucleus is called the *karnel* and is indicated by the symbol "ˀ".

In this subsection we use the contour tone notation as faithfully as possible to Kodama's (2008) original notation, but to aid visual parsing, hyphens are inserted.

Some may suspect that nF]nF is an equivalent of nF]F and therefore nF-nF]nF vs. nF]F-nF is a pseudo-problem. In fact, under the level tone account, nF]nF and nF]F are notationally equated: μ]μ.

If nF]nF could be equated with nF]F, the contrast between nF-nF]nF and nF]F-nF would become trivial as they could be rewritten as nF-nF]F and nF]F-nF. However it will be shown that they are not equivalents.

Kodama (2008) illustrates that nF]nF and nF]F contrast. He proposed that the shape of the AP can be generalized as the following:

- nF+]F+ (*yūkaku* type)
- nF+] (*odaka* type)
- nF+ (*heiban* type)

where "+" indicates succession (nF+ reads "one or more nF's in succession") and "]" indicates a lower tone on the right (nF]nF reads "a sequence of two nF moras where the latter is lower than the former") and is placed right after the nucleus. That is, every AP in Tokyo Japanese takes the shape of one of the three types (p. 28, p. 33).

Since an utterance is a sequence of APs (p. 5), both of these sequences are possible:

- nF]nF
- nF]F

To illustrate that they are contrastive, he puts forth three possible pronunciations of *tsukini sen'en* / *tsuki nisen'en* ("a thousand yen per month"/"two thousand yen per month") as examples:

- RLv](nF)(nF)+ "a thousand yen per month"
- RLv]F(nF)+ "a thousand yen per month"
- RLv](nF)+ "two thousand yen per month"

(Examples taken from p. 29.)

> Kodama (2008) transcribes IPs with R consistently placed only on the first mora for the "default" pronunciation. This is one of the differences between his original notation and mine.
>
> Under my introspection, the first one is acceptable only with a paralinguistic prominence on *ni*. I have not inquired of him whether this judgment of mine differs from his observation. Regardless, his claim is that in these examples nF and F contrast after].

The second and third pronunciations demonstrate the validity of the claim. nF and F contrast after] in these examples.

As we have now established that F and nF contrast after], it is readily that nF-**nF]**nF vs. nF]**F**-nF is a real contrast; nF] and F contrast before nF. (We do not worry about] before F as we will see later.)

In sum, before nF, F ≠ nF], nF] ≠ nF, and F ≠ nF; all three are contrastive. This necessitates contour tones and the boundary tone indicated by "]" simultaneously.

> That F ≠ nF before nF is assumed to be trivial as there would be no disputes here, e.g., μ]μ|μ vs. μμ|μ. (Kori (2008a) and Kori (2004) are relevant.)

Notice that in Kodama's (2008) formulation a sequence of an nF mora and the following F mora is always written as nF]F. nF-F never occurs. Hence,] here is redundant and can be safely ignored.

> This leads to another difference between Kodama's (2008) and my notations; I replace Kodama's nF]F with nF-F.

] is not redundant between nF moras as we have just established above that nF]nF and nF-nF contrast.

] never appears in other places—except utterance-finally, in which case, by definition, it cannot be phonetically realized as itself. No matter how *yūkaku*, *odaka*, and *heiban* are arranged, neither F]F nor F]nF can occur.

Thus, the only place where] matters is between nF's, which is where an instance of] is called the boundary fall. The boundary fall is a special case of Kodama's].

This subsection laid out some theoretical background of the boundary fall. To give some concrete examples, the following sentences include a boundary fall:

- *sōda, takara sagasanakya.*

(TRICK, ep. 10, 00:38:37)

- *kutsu yogoreru kara iya.*

(TRICK 3, ep. 2, 00:14:39)

Predicative intonations

Certain types of simple predicates are understood as part of a question if the final mora is pronounced with a rising tone at its later half. In the following example, the intonation associated with interrogation is marked by the symbol R% (R stands for "rising").

- *kinō nani tabeta* (R%) "What did you eat yesterday?"

R% can co-occur with the final mora's F tone without neutralizing it with nF.

- *ashita yotei aru* (R%) "Do you have any plans tomorrow?"

Aru is nF-F. Compare:

- *kyō mizugi iru* (R%) "Do we need a swimsuit today?"

Iru is nF-nF.

R% coexisting with F on the same mora after nF is found, among other times, when *no* follows an nF predicate (*iru* in the following example).

- *kyō mizugi iru no* (R%) "Because we need a swimsuit today?"

We will come back to this point later in this section.

The corresponding declarative predicate does not accompany a rising intonation. In this sense, the intonation at the end of the predicate is contrastive. In a declarative sentence, the last few syllables, especially the final one, receive additional steepness (marked by F%):

- *kinō gomasaba tabeta* (F%) "I ate gomasaba yesterday."

Kori (2008b, 2011) provides a detailed phonetic analysis of the added pitch drop at the end of a declarative sentence.

The F% intonation can be difficult to pay attention to. The following YouTube video includes an exemplary instance of how the absence of F% at the end of a plain declarative sentence makes the sentence unnatural. In the video, a man says, "*Boku wa hitono shi o jiyūni ayatsuru koto ga dekiru.*" without an F% intonation at the end. The sentence rather sounds like an attributive clause of a larger sentence: https://youtu.be/PIbCqmvAiNI?feature=shared

The R% intonation can appear in the middle of a sentence. The predicate of a question is typically positioned at the end, but when it is placed elsewhere, the R% intonation arises at the same place. In the following example, the predicate is placed in the middle due to inversion.

- *nani tabeta* (R%) *kinō* "What did you eat yesterday?"

There seems to be no accepted single orthographic standard as to where the question mark should be placed in the case of inversion. My impression as an everyday speaker is that the most natural position is after the predicate, but I would not be surprised if a significant minority places the question mark at the end and puts a comma after the predicate.

The typical pronunciation of the sentence above has an IP that consists of *tabeta* (R%) *kinō*. It includes the R% intonation in the middle. This would raise a question as to what component of the prosodic hierarchy the intonation belongs to. Some may speculate that the intonation belongs to AP. The following example, however, demonstrates that it can even intervene in an AP (unitalicized).

- *ashita omatsuri* iku (R%) tte *kikareta kara* “because (he) asked me, ‘Are you going to the festival tomorrow?’”

I speculate, without much confidence, that the other type of the predicative intonation, F%, can be subject to the same phenomenon where it intervenes in a prosodic unit smaller than or equal to the sentence when inversion or direct quotation is involved, but the effect, if any, is much less obvious than R%.

- *ashita omatsuri iku, watashi* “I’m going to the festival tomorrow.”

- *gomasaba tabeta, kinō* “I ate gomasaba yesterday.”

- *gomasaba tabeta tte itteta* “(he) said he had eaten gomasaba.”

When a predicative intonation intervenes in a sentence, the sentence can have multiple predicative intonation components. (This is most likely obligatory since there is no apparent reason—empirical or otherwise—to assume that having one in the middle lifts the one at the end.)

- *nani tabeta* (R%) *kinō* (F%) “What did you eat yesterday?”

  In this case, since *kinō* here is clearly not a predicate, the final F% may be better called a sentential intonation. The distinction, however, is inessential to our discussion.

- *ashita omatsuri iku* (R%) *tte kikareta kara* (F%) “because (he) asked me, ‘Are you going to the festival tomorrow?’”

The effect is easier to see if the final intonation is R%.

- *ashita omatsuri iku* (R%) *tte kikareta* (R%) “Did (he) ask you, ‘Are you going to the festival tomorrow?’?”

Summarizing our observations so far, the predicative intonations 1. can appear in the middle of a sentence, an IP, or even an AP; 2. can appear in multiple places of a sentence; and 3. are obligatory at least in one place per sentence. There seems to be no obvious, simple, strict rules governing the position and occurrence of these intonations with respect to the prosodic hierarchy, other than that a sentence must have at least one predicative intonation somewhere.

Certain types of interrogative predicates do not have R%.

- *watashi wa dō sureba ii n deshōka* (F%) “What would you think I should do?”

Some declarative sentences have R%, but it is limited to certain particles.

- *kuwashiku shirabenaito ikemasen ga, osoraku sotokara mottekita mono da to omoimasu yo* (R%)

(検察側の罪人.)

This *yo* (R%) is accented (nF) as we can infer from the following:

- *nande tabenai no? oishii* (-nF) *yo* (R%) “Why aren’t you eating? It’s delicious.”

- *yoru wa raito tsukenaito abunai* (-nF) *yo* (nF, R%) “At night, it’s unsafe if you don’t turn on the lights.”

In each, the predicate before *yo* (nF, R%) is a member of kino group (»harmonica). Hence, the fact that *oishii* and *abunai* are both nF for their final mora implies that the following item *yo* (R%) is accented (nF). This can be compared with another *yo* (often explained in terms of “emphasis”), unaccented (F):

- *yoru wa raito tsukenaito abunai* (-F) *yo* (F, F%) “At night, it’s unsafe if you don’t turn on the lights.”

- *gyōza wa shōyu tsukenakutemo oishii* (-F) *yo* (F, F%) “Gyoza is delicious even without soy sauce.”

And *no* (F):

- *yoru wa raito tsukenaito abunai* (-F) *no* (F, F%) “Because at night, it’s unsafe if you don’t turn on the lights.”

- *yoru wa raito tsukenaito abunai* (-F) *no* (F, R%) “Because at night, it’s unsafe if you don’t turn on the lights?”

As shown, the predicate *abunai* and *oishii* receives an F tone at the end when followed by F, regardless of whether the predicative intonation given to the unaccented item is F% or R%.

The fact that *no* is unaccented even with R% can be easily observed with examples such as the following:

- *ashita omatsuri iku no* (F, R%) “Because you are going to the festival tomorrow?”

Compare:

- *ashita omatsuri iku ko* (nF, R%) “Is (he) the kid who is going to the festival tomorrow?”

Clearly, nF and F are contrastive even with R%.

But the contrast becomes less clear after F:

- *wakaru* (-F) *yo* (nF, R%) “I understand that.”

- *wakaru* (-F) *no* (F, R%) “Because you understand that?”

The difference between them in the pitch contours is slim at best (my introspection, December 2023). Whether and how they differ must be investigated phonetically. But that *yo* (R%) is phonologically always nF can be inferred from the fact that *yo* (F, R%) is never allowed after nF:

- **iku* (-nF) *yo* (F, R%)

- **iru* (-nF) *yo* (F, R%)

- **asoko da* (nF) *yo* (F, R%)

> I initially overlooked this fact and in a privately shared earlier version of this document and potentially in other older documents recklessly stated that *yo* (nF, R%) and *yo* (F, R%) are contrastive. I hereby correct the error: they are not contrastive. The cause of the error was that I had introspected into two possible phonetic forms of *oishii yo*, one of which has *oishii* with the final mora F. I am not aware of evidence strong enough to claim that, despite the observations we have just made above, these two forms differ in the phonological tone of *yo*. Phonetically the clearest difference is at the final moras of *oishii*. The supposed difference in meaning (F R% for “to elicit confirmation” and nF R% for “to elicit attention”) could be paralinguistic, and there will be no explanation for why this form of *yo* can appear after *oishii* and not after *ikeru*, for example. I suppose if these two forms of *oishii* before *yo* in fact contrast, it will be necessary to account for the difference by some other means.

Yo (nF, Lv%) is related to making announcements. The vowel part of it tends to be long, in which case it is denoted here as *yō*. (Compare this with the case of *yonē* mentioned later in this section.) The intonation symbol Lv% refers to the fact that it is perceptually level.

- *mōsugu heiten desu yō* (nF, Lv%) “The store will be closing soon.”

- *iku yō* (nF, Lv%) “Here I go!” (throwing a ball)

Ne, sometimes translated as "right?" or "isn't that so?", is pronounced as a whole IP (indicated by capitalization):

- *Oishii Ne*. "It's delicious, isn't it?" (declarative)
- *Kawaii Ne*. "It's cute, isn't it?" (declarative)

Native speakers often broadly describe the intonation given to *ne* as a "rising" intonation, but the following sentences have the same phonetic shapes as above:

- *Oishii Ne*. "It's a delicious root."
- *Kawaii Ne*. "It's a cute root."

The perceived rises in the pitch contours of *ne* in these cases are just the IP-initial R tone. This is not considered as a predicative intonation. Some other ending particles including *ze* and *wa*, which are used in a masculine speech and a feminine speech, respectively, fall into the same class as *ne*.

One-mora IP's like *ne* ('root'), *te* ('hand'), *e* ('drawing'), etc. can have R%:

- *Oishii Ne* (R%) "Is it a delicious root?"
- *Ōkina Te* (R%) "Is it a big hand?"
- *Kireina E* (R%) "Is it a beautiful drawing?"

Some ending particles are prosodically complex. *Yone* (related to "confirmation") has the second mora in its own IP and accented and the first mora in the preceding IP is unaccented (which means *yo* belongs to the preceding AP). At first glance the fact that *yo* is part of the IP of the preceding word may seem unclear. If the beginning of *yo* coincides with the IP's left edge, *yo* must be nF by definition, which means the latter half of a long vowel at the end of *yone*'s preceding AP must be able to be nF without a problem. However, *oishii yone* is obligatorily pronounced with the latter half of *ii* in F. Hence, it must be that the mora *yo* belongs to the same AP as *oishii* and *yone* has an IP boundary in the middle.

When *yone* (not guaranteed to be the same item as the aforementioned *yone*) is used in a question, it receives a characteristic rising-falling intonation, tentatively represented here as RF%.

- *kono heyade atteru yone* (RF%) "This room is the correct one, right?" (interrogative)

Perceptually, the final syllable *ne* (RF%) is longer than it would be with F%. The perceptual length is sometimes represented in the orthography with an added vowel letter (よねえ？), often small (よねえ？), although it is probably more common to not mark the length directly, only leaving the question mark (よね？). In fact, if it is two moras, the impressionistic description of RF%—the first half is rising and the second half falling—will be understood as just F% with the first mora bearing the IP-initial R. If this assumption is correct, the above example can be rewritten as:

- *kono heyade atteru yonē* (F%) "This room is the correct one, right?" (interrogative)

For now, we assume that this intonation is just a phonetic realization of F% when the relevant syllable has a long vowel and is its own IP.

Kino group items at the end of a sentence allow multiple pronunciations for some speakers. For these speakers, including myself, *oishii,* for example, can be pronounced with the final mora in F or nF. Other speakers seem to consistently prefer only one of the two forms of the same item. It would not be unreasonable to suspect that these two forms contrast. For now, I refrain from suggesting one conclusion over the other. In this version of Kyouro, the said potential intonational contrasts are not indicated. *Oishii* at the end of a sentence is written as **oɪˈsɪˈ·ɪ** regardless of whether it is pronounced with or without an F tone at the end.

The utterance-final tone inflicted by the suffix *-yō* of the verb paradigm needs attention. In the following examples, the relevant items are unitalicized.

- *tsugi au toki wa chigau kachikande mina de* odorō.

(PUNPEE "Gizmo (Future Foundation).")

- *tsugino jikanni mo minna de* kangaeyō *ze*.

(丘の家のミッキー1.)

The predicates *odorō* and *kangaeyō* can be pronounced with R%, but the forms with F% and R% are pragmatically distinct. The latter—henceforth the *"suggestive" -yō*—is often written in the orthography with a question mark (?). When they are pronounced with R%, the pitch contour will be similar to that of a kino group item, such as *yasashii* and *oishii*.

At first glance, this may seem like the familiar tone alternation of kino group items, but the *-yō* form of a verb does not belong to kino group. One indication is that, in *kangaeyō ne*, the part *-yō* must be pronounced with F. This differs from *yasashii ne* and *oishii ne*, where the final moras of *yasashii* and *oishii* can be pronounced with nF. Consequently, the "suggestive" *-yō* is a separate phonological from the imperative *-yō*.

Additionally, potentially contrastive is a perceptually Lv% form of the "suggestive" *-yō*. While the phonological standing of this form remains unclear, orthographically, this form does not receive a question mark (?) where the corresponding R% form often does. Precise semantic descriptions aside, they are conventionally distinguished by laypeople.

Summarizing our investigation so far, the predicative intonations can be classed into the following three categories:

- F%: simple declarative sentences and interrogative sentences with certain predicate types, and potentially some others.
- Lv%: declarative sentences with certain ending particles, some of the *-yō* forms, and potentially some others.
- R%: simple interrogative sentences and declarative sentences with certain ending particles, and potentially some others.

The following chart summarizes the functions of these intonations (F%, Lv%, and R%) for each context.

	F%	Lv%	R%
simple predicate	declarative	declarative (contrast unclear)	interrogative
-shōka	interrogative	not found	not found
-yō (imperative)	imperative	imperative (contrast unclear)	not found
-yō (“suggestive”)	not found	“suggestive” (contrast unclear)	“suggestive”
yo (nF)	declarative	probably a variation of yō (nF)	declarative
yō (nF)	not found	declarative (announcement)	not found
yo (F)	declarative, imperative	undefined (necessity not observed)	not found
one-mora IP	declarative	undefined (necessity not observed)	interrogative
yonē (nē as its own IP)	interrogative	not found	not found

As shown, none of the predicate contexts clearly utilize a ternary (F%-Lv%-R%) contrast. Although further investigations may reveal a three-way (four-way including attributive) contrast somewhere in the system, for now, we assume that in each context, the predicative intonation is binary (ternary including attributive). Since as a native speaker I feel that the perceptually salient category is R%, here I assume the contrast is between nR% (non-rising, that is, either Lv% or F%) and R% and reorganize the chart accordingly.

	nR%	R%
simple predicate	declarative	interrogative
-shōka	interrogative	not found
-yō (imperative)	imperative	not found
-yō (“suggestive”)	not found	“suggestive”
yo (nF)	declarative	declarative
yō (nF)	declarative (announcement)	not found
yo (F)	declarative, imperative	not found
one-mora IP	declarative	interrogative
yonē (nē as its own IP)	interrogative	not found

SYLLABLE

Syllable structure

The syllable structure is maximally CCGVG. The onglide (the first G) can only be *y*. The offglide (the second G), when present, is *n* or *i* or marginally *e* or *u,* where *n* is a nasal glide which is realized as a stop before a stop. The vowel and at least one consonant of the onset can be long. The onset (the CC part) and the glides are optional, but the nucleus (V) is obligatory. Marginally, the nucleus can be a nasal stop.

The orthographic *y* is sometimes interpreted to be part of a palatalized consonant. See Okada (1999) for example.

The proposed syllable structure is substantially different from some earlier works including Okada (1999). The new approach is necessary for accurately describing the tone allocation rules in simple terms.

In a syllable of the (C)(G)VG structure, the onset bears zero moras, the nucleus one, and the glide one, and if the onset is complex (CC), the first C bears one mora, provided that all segments are short. Glides are always short and not allowed after a long vowel. The onglide bears zero moras. A long C or V adds exactly one mora to the corresponding spot. Denoting a long consonant as C̄ and a long vowel as V̄, both C̄C(G)VG and CC̄(G)V̄ are four moras long, and C̄(G)V is two moras long. A five-mora syllable is dubious, but I refrain from judging it impossible for now. An example of a four-mora syllable is the second syllable (*-kkitai*) of *hikkitai* [hiḵːd̥aɪ̯] 'the cursive form.'

Correction in 5.2.0: CC(G)V̄ is three moras. I thank @triethylamineq for pointing this out.

When the symbols C and V stand for consonants and vowels, the distinction between long and short segments are not always made. In most other places in this book (e.g., p. ditheism), long and short segments are denoted by C and V without a macron.

Tserdanelis, Beckman, Kong, Li, & Syrika (2005) as well as Hasegawa (1995) treat Japanese "palatalized *k*" as palatal. We do not discriminate between a palatalized or fronted velar and a true palatal in our IPA transcription.

The vowel *a* is usually transcribed as [a], [ä], or [ɐ]. In a natural speech, it can be as high as [ɜ] and occasionally it feels even as high as [ɘ] (my introspection). I use [a] unless the precise quality of the vowel is at issue.

It is customary to transcribe *hi* as [çi]. This style of transcription fails to capture the fact that the ostensive palatal fricative [ç] is only about as constrictive as the following vowel [i] (Kuroki p.c. and my personal observation). (Tronnier & Dantsuji 1993 provide a comparison of Japanese /h/ and German /h/ and /ç/, for both languages in the context of a following [i], and show that the F1 value of the Japanese /h/ is greater than the German /h/ (p. 220).) The vowel-like nature of this segment will be correctly captured if [h] is used instead. In fact, [h] often resembles the neighboring vowels. It is appropriately interpreted as a glide, something that is vowel-like and nonvocalic, in Chomsky & Halle (1968), p. 303. Japanese *h* does not seem to be an exception of this usual behavior of the so-called glottal fricative. The sequence *ho* in a word such as *toho* 'on foot' is typically pronounced as [ho], in which [h] is another way of representing [o̞̯] (informal observation). Accordingly, [hi] as used here is a notational variant of [i̞̯i] or [ji̞]. By the same token, *fu* will be transcribed as [hu] if necessity arises.

despair

The offglide *i* can follow nonfront short vowels (*u, o, a*) but not front vowels (*i, e*) and is often found in kanji words and katakana words (e.g., *gaikoku* 'foreign country', *sairiumu* 'psyllium') and in adjectives (e.g., *usui* 'is thin' (-*i* is a suffix.)). When it is expected to follow the vowel *i* in an adjective stem, a phonetic long vowel [iː] appears (e.g., *yasashii* 'kind'). The phonological behavior suggests that the offglide *i* is prohibited after the nuclei *i*.

In a natural speech, whether a given vocoid is a nucleus or a glide is often unclear and sometimes they are practically neutralized. As is usually the case, the phonetic specifications are given as abstractions that relate to physical representations, which are affected by uncontrolled real-world factors. If two sentences receive distinct phonetic specifications, the claim is that they systematically differ in how they relate to physical sounds, but not necessarily that they always sound distinct to the human ears. Also note that Kyouro's spelling rules are designed to be tolerant to minor analytic inconsistencies. (See Chapter Seven.)

A potential minimal pair of the offglide *e* and the nucleus *e* arises when the verbs *kaeru* 'to go back' and *kaeru* 'to change' (transitive), which differ in the lexical (nF vs. F) tones in their lemmas (the forms of dictionary headings, not necessarily the lexical forms), take suffixes to suppress the tonal contrast: *kaeraremashita* ({'go back'/'change'}-[+ honorific, + past]). Supposedly, if the *e* in the stem is a nucleus, it is the 'change' verb, and the 'go back' otherwise (See also the note below).

A major—and probably the most noticeable to many including myself—phonetic difference between the two lies in the realization of the IP-initial rising tone. The contrast suppressed by the said suffixes is of the falling tone.

kaeru	'to go back'	R-F-F
kaeru	'to change'	R-R-Lv
kaeraremashita	'went back (honorific)	R-Lv-Lv-Lv-Lv-F-F
kaeraremashita	'changed (honorific)	R-R-Lv-Lv-Lv-F-F

This is a relatively loose contrast. According to my introspection, in careful speech the *e* nucleus appears in the "change" verb but not in the "go back" verb, and the *e* offglide vice-versa, but in sloppy speech the *e* offglide can appear in both.

Kaeru as well as phonologically similar *kaesu* are quite interesting both phonologically and lexically. »exorcist. »glucose-harmonica.

The offglide *u* [u̯] can occur between *a* and a coronal consonant (*t, ts, d, s, z, sh, j, ch, r*) (e.g., *hausudasuto* 'house dust'), provided that no syntactic boundaries intervene. Cases in which *u* in the same environment is realized as a nucleus have been found but they are rare (e.g. *naute* 'well-known,' *naushika* 'Nausicaä' (a Studio Ghibli animation)). The corresponding tonal difference between the two parallels to the contexts of the other offglides—whether it is *au, ae,* or *ai*, VG cannot be R-R. The difference in the vowel quality or articulation is not difficult to notice when compared to the case where the vowel *u* is followed by a noncoronal segment, in which case the vowel is further back or more rounded (e.g., *augusutusu* 'Augustus').

hausudasuto	'house dust'	R-Lv-Lv-Lv-F-F
augusutusu	'Augustus'	R-R-Lv-Lv-F-F
naute	'well-known'	R-R-Lv

Here and elsewhere, tone marks given to voiceless moras are conveniently assumed. In reality, voiceless segments do not exhibit phonetic tones.

An onset can be a consonant cluster. In writing, an onset cluster is interpreted to have a voiceless high vowel in-between. The supposed vowel in a cluster, in the following examples indicated by a superscript vowel letter, is not very clearly pronounced if at all.

k^{u}sai	'smells bad'
s^{u}koshi	'a small amount'
k^{i}taku	'arriving home'
ts^{u}kau	'to use'

The sequence 'ts' in the Hepburn system denotes a single consonant: IPA [t͡s].

Syllabification

Offglides are not allowed after a long vowel. When a segment that would otherwise be an offglide *i* follows a long vowel, it turns into a nucleus *i*. This is observed when the adjective suffix *i* is attached to a stem that ends with a long vowel, such as *tō* 'is far.'

chika	+	i	—>	chikai	CCVG	'is near'
tō	+	i	—>	tō.i	CV.V	'is far'

If a nasal glide follows an oral offglide, typically, the rhyme breaks and the original oral offglide turns into a nucleus. This is observed when *n* ('because') follows the present declarative form of adjective:

usu + i + n + da	—>	u.su.in.da	nF-nF-nF-F-F

But the pre-nasal rhyme breaking seems to be, at least under certain condition, optional as the following pronunciation is also allowed:

usu + i + n + da	—>	u.sui.n.da	nF-nF-F-F-F

A similar example is *osanai n da* ("oh, you won't push it"), which can be nF-nF-nF-nF-F-F or nF-nF-nF-F-F-F (my introspection). A further comparable case of the pre-nasal rhyme breaking is seen in *shinai n da* as appears in:

- *konoyoni reinōryoku nante sonzai shinai n da tte*

(TRICK, Ep. 7, at around 0:33:20.)

A nasal mora before a following long stop consonant (cf. *nihon tte*) feels like its own syllable. The articulatory phenomena that seem to contribute to the syllabic impression include the tongue position and the duration of oral closure. Questions as to how this phenomenon should be characterized are tentatively put aside. The syllabicity of a nasal mora before a long consonant matters in cases like: *kan tte* "speaking of intuition," *san tte* "speaking of three," and *bon tte* "speaking of the Bon festival."

In loan words and Sino-Japanese words, the sequence *ai* is usually a diphthong unless a morpheme boundary intervenes. However, when followed by the moraic nasal, the same letter sequence is realized disylabically:

Although the term morpheme is used here to talk about the phenomenon in terms of the common sense, whether a sequence of vocoids is a diphthong or a hiatus can be specified in the lexicon without referring to the traditional notion of morpheme. In general, the notion of morpheme has no use in the kind of a grammar we are dealing with here. For an illustration, »sky-snapshot.

ai.do.ru	VG.CV.CV
a.in.shuta.in	V.VG.CCV.VG

When a moraic nasal follows a long vowel, instead of breaking the long vowel into two, the nasal becomes syllabic. This is observed when the aforementioned *n* follows *yasashii, oishii,* and other adjective forms that have a long *i* (spelled as *ii* in Hepburn, pronounced as *ī*) at the end:

yasashii	(ya.sa.shī)	CV.CV.CV
oishii	(oi.shī)	VG.V
yasashii n da	(ya.sa.shī.n.da)	CV.CV.CV.V.CV
oishii n da	(oi.shī.n.da)	VG.CV.V.CV

This applies to orthographic long vowels before a moraic nasal. *Gurīn* ('green') is pronounced as *gu.rī.n* or CV.CV.N unless it resorts to a vowel sequence as in *gu.ri.in* or CV.CV.VG. (I suggest understanding this as a case of doublets, not of optional phonological rules.) NHK's pronunciation dictionary lists both possibilities of *kōncha* ('corn tea'):

kōncha	(kō.n.cha)	CV.V.CV
kooncha	(ko.on.cha)	CV.VG.CV

For the latter, the literal notation of the dictionary is as if it is *kōncha*, but the recorded audio is as noted here.

THE TRIANGLE CONSTRAINT

Distribution of R

Since a sentence contains one or more IPs, it must have at least one R (or a sequence of consecutive R's of an AP, henceforth treated without discrimination in this section) within itself. If it has only one R, it comes at the top.

When multiple R's are in a sentence, their positions are restricted in certain ways. In the following case ("Burnable garbage is collected on Mondays."), *dashimasu* can bear R only if *getsuyōbini* does so. R is marked by capitalizing the first letter.

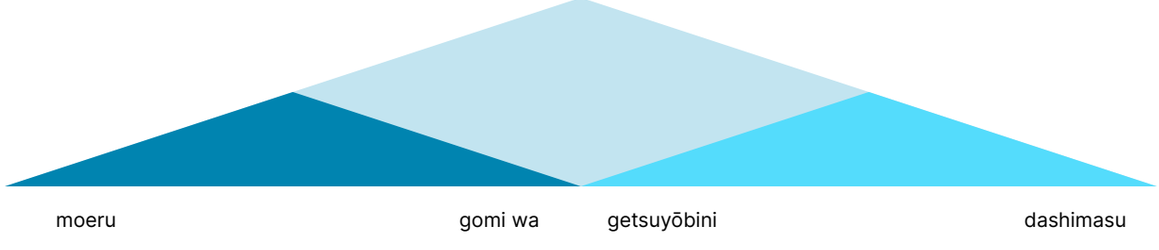
	Moeru	gomi wa	getsuyōbini	dashimasu
*	Moeru	gomi wa	getsuyōbini	Dashimasu
	Moeru	gomi wa	Getsuyōbini	dashimasu
	Moeru	gomi wa	Getsuyōbini	Dashimasu
	Moeru	Gomi wa	getsuyōbini	dashimasu
*	Moeru	Gomi wa	getsuyōbini	Dashimasu
	Moeru	Gomi wa	Getsuyōbini	dashimasu
	Moeru	gomi wa	Getsuyōbini	Dashimasu

In *yamamoto wa katōno hohoni kisu shita* ("Yamamoto kissed Katō on his cheeks"), *hoho* can be capitalized only if *katōno* is, *kisu* only if *katōno* does so, and *shita* only if *kisu*.

	Yamamoto wa	katōno	hohoni	kisu	shita
*	Yamamoto wa	katōno	hohoni	kisu	Shita
*	Yamamoto wa	katōno	hohoni	Kisu	shita
*	Yamamoto wa	katōno	hohoni	Kisu	Shita
*	Yamamoto wa	katōno	Hohoni	kisu	shita
*	Yamamoto wa	katōno	Hohoni	kisu	Shita
*	Yamamoto wa	katōno	Hohoni	Kisu	shita
*	Yamamoto wa	katōno	Hohoni	Kisu	Shita
	Yamamoto wa	Katōno	hohoni	kisu	shita
*	Yamamoto wa	Katōno	hohoni	kisu	Shita
	Yamamoto wa	Katōno	hohoni	Kisu	shita
	Yamamoto wa	Katōno	hohoni	Kisu	Shita
	Yamamoto wa	Katōno	Hohoni	kisu	shita
*	Yamamoto wa	Katōno	Hohoni	kisu	Shita
	Yamamoto wa	Katōno	Hohoni	Kisu	shita
	Yamamoto wa	Katōno	Hohoni	Kisu	Shita

Triangle diagram

The fact that an AP can only carry R if a particular other AP carries R means that these APs are in a dependency relationship. Consider the following diagram.

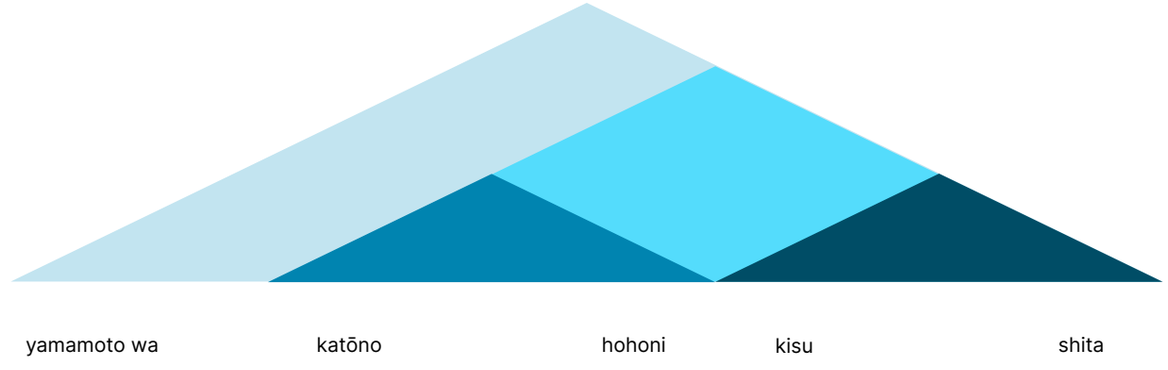


The APs at the bottom of a triangle belong to the nearest corner. *Moeru* belongs to the left corner of the dark blue triangle. Similarly, *getsuyōbini* belongs to the left corner of the bright blue triangle. These two triangles in turn belong to the larger triangle behind them.

For the left corner of any one of the triangles, its AP must bear R if the triangle contains an AP with R. If *dashimasu* bears R, then *getsuyōbini* and *moeru* must bear R because *dashimasu* is at the right corners of the bright blue triangle and the large triangle.

> Certain lexical items such as *tabun, chotto, sore,* etc., when used as fillers, can evade these constraints. More investigations are necessary to correctly capture the phonological nature of these items.

Similarly, the following triangle layers represent the text *yamamoto wa katōno hohoni kisu shita* and correctly capture the interdependence of its APs.



The triangle diagram satisfies the following conditions:

- Each bottom corner of a triangle corresponds to exactly one AP. (If two or more triangles share a corner, the shared corner corresponds to one AP.)
- There are no orphan APs. (Each AP corresponds to one of the bottom corners.)
- At most two triangles can be directly layered on top of a larger triangle. If two triangles are directly on a larger triangle, the larger triangle's two bottom corners are both shared with one of the two smaller triangles.

A triangle diagram can be replaced with parentheses. The aforementioned structures can be represented as:

```
| ((moeru      gomi wa)  (getsuyōbini     dashimasu))      |
| (yamamoto wa     ((katōno    hohoni)   (kisu shita))) |
```

These are not syntactic representations. An AP is a phonological unit and is not a terminal node of a syntactic tree. Rules to convert a syntactic representation to a phonological representation will be discussed in Chapter Six.

Chapter Three

LEXICAL FORMS

THE NATURE OF THE LEXICON

Phonology derives the phonetic form from the output of syntax

In the previous chapter, we briefly mentioned how onset clusters are represented in standard orthography. Here are some additional examples of words that contain consonant clusters.

its^{u}kushimajinja	kak^{u}shin	yamash^{i}ta
ik^{u}sa	hakush^{i}katei	tokush^{i}tsu
tok^{u}shima	mok^{u}teki	kakuek^{i}teisha
asak^{u}sa	k^{i}shakaiken	kak^{u}san
ish^{i}kariheiya	f^{u}kushima	s^{u}toraiku
kakush^{i}ki	kok^{u}seki	kakk^{i}teki

It is immediately apparent that the superscript vowels are uniformly bounded by orthographic voiceless consonants. This is part of the syllable template's requirements. When the onset is a consonant cluster, both C's must be voiceless.

A plosive as the second consonant of an onset tends to have a shorter (though positive) VOT than a plosive after a voiced segment. Orthographically, both of the consonants in a cluster are treated as voiceless. The discussion that follows in this section suggests that the consonants in the said environment are underlyingly voiceless.

The fact that the superscript letters are limitedly distributed raises a question as to whether the CC structures are derived or lexical. At least in some instances, CC structures are conditioned by the following word's initial segment, in which case it is reasonable to think that the surface CCs are derived from underlying CVCs:

aruku	+	kara	—>	aruk^{u} kara	"because you walk"
aruku	+	nara	—>	aruku nara	"if you walk"

although the reverse analysis (insertion) is conceivable:

ar^{u}k^{u}	+	kara	—>	aruk^{u} kara	"because you walk"
ar^{u}k^{u}	+	nara	—>	aruku nara	"if you walk"

Individuating the single most economical analysis is beyond the current scope of Kyouro. The point to be made here is that the output of syntax is subject to phonological rules that derive the phonetic form. Studying phonetic forms is part of pursuing the nature of the output of syntax. Syntax builds up a structure containing enough information for phonology to derive the phonetic form of the sentence. Unpredictable phonetic specifications are supplied from the lexicon.

Some phonological processes are applied to a local syntactic structure. Accent switching occurs recursively as suffixes merge recursively with the base (»globalist). Accessing local structures is also necessary to determine the positions of R (Chapter Six). These facts suggest that the precedence of syntax over phonology only holds for each subprocess and it is not the case that everything in syntax completes before phonology begins.

Information supplied by the lexicon

Unpredictable information is provided by the lexicon. This includes boundary labels and phonological symbols as well as semantic information although semantics is not discussed here. This information comes in a lexical item, whose components are linked to each other so that if any partial information of a lexical item is available for the grammar all else of the same item is also available at some point of derivation. However, modification processes, including deletion and alternation, are irreversible, and the same information cannot be retrieved twice to repair what has been altered.

Phonological symbols supplied by the lexicon include segments, accents, and AP boundaries. An AP boundary can be a boundary fall. Syntactic specifications will be discussed in Chapters Four and Five.

Segments are organized into a sequence, which is interpreted by phonology to derive syllable structures. Certain modification processes take place before syllabification. Syllables are associated with moras. This syllable-mora structure constitutes part of the prosodic hierarchy.

AP boundaries, including boundary falls, are segmental in the sense that their positions are defined with respect to the segmental sequence. They are, however, not pronounced as a phonetic segment and not considered as a constituent of a syllable, and they always coincide with a syllable boundary as implied by the strict layer hypothesis. The prosodic hierarchy completes after R allocation (Chapter Six) is done.

The segmental standing of accents is disputable. In the Kyouro spelling, accents, indicated by the symbols ˈ and ˌ, are written in the same sequence as segments, and a basic algorithm deriving the correct phonetic form from the spelling form is provided (»forest). The Kyouro spelling form of an item is therefore taken to be a fairly good approximation of the lexical form, as far as accent is concerned. Potential problems of treating accents as segmental include the fact that some lexical items require an accent to be placed in the middle of a bimoraic segment (»glucose). This problem is concealed by the typeability principle: bimoraic segments are spelled with digraphs.

Spelling form:	Phonetic form:	Lexical form:
kıno ˈu	*kinō*	kinow, kinō, or ...?

Another potential problem is that an accent can be placed apart from the segmental sequence. Some suffixes have an accent one mora behind its true segmental sequence. It seems that this is the leftmost position an accent can be placed in in the lexical form. It is not clear whether or how this should be characterized in the segmental sequence in the lexicon. In Kyouro, items that have an accent apart from the sequence of true segments, when spelled in isolation, are noted with the symbol '**(_)**,' representing a placeholder of one mora.

Spelling form:	Phonetic form:	Lexical form:
ˈ**(_)r·u**	nF-{F/nF}	ˈ(_)ru, ru (2), or ...?

A lexical item can have at most one upper accent and one lower accent per AP. Each accent present in a lexical item is specified with the position. If two accents are present in one AP, as a corollary of what is assumed about the two types of accents, the upper accent precedes the lower accent (»generativist).

Some word boundaries are AP boundaries

Some words do not bear an AP. What this means is exemplified by the following. APs are separated by tab spaces.

- tarō ga ikebana o kanshō shita.
- hanako ga fuchūi o kōkai shita.
- basu ga mokutekichi ni tōchaku shita.
- makku nara iroirona koto ga kantanni dekiru.

Items such as *ga, ni,* and *o* do not bear an AP. They belong to the preceding word's AP. In contrast, items such as *tarō, ikebana, shita,* and *basu* bear an AP. Their left edge always coincides with an AP boundary. Whether a given item initializes an AP in the sentence partly depends on the syntactic standing. Items with certain syntactic properties obligatorily initialize an AP under certain phonological condition. Other items usually belong to the preceding word's AP, but some of them begin with a new AP.

If the item's left boundary is a *word boundary* (see Chapters Four and Five for details) and the item's first AP has an accent, its left edge is an AP boundary. This AP boundary can be accounted for by a syntax-phonology mapping rule, and therefore it is not considered as a lexical property. Lexical specification of AP boundaries is necessary if the item accompanies an AP boundary under other conditions.

As explained in Chapter Five, certain syntactic boundaries are classified as word boundaries. This classification is part of the spelling design of Kyouro. Word boundaries in the spelling form of a sentence satisfy the following conditions:

- Word boundaries are marked by a space.
- If there are no other spaces between a space and the first accent that follows, upper or lower, the space corresponds to an AP boundary.

The following items, for example, accompany an AP boundary at their left edge. We know this from the spelling because they have a word boundary on the left side and have an accent. On the left is the phonological specifications of the item, written in Kyouro, and on the right is the syntactic specifications. The notation of syntactic properties is explained in Chapter Four, but for the following items, the labels placed on the left side of the bar indicate that the item has a word boundary on the left.

- ˈ**wakega** R1|A4
- **ɪˈnu** R2|m-ii
- **ɪˈkebana** R2|m-ii
- **kannsyouˈ** .|G1
- **sɪˈta** G2|R1

In contrast, the following items, despite having a word boundary on the left, belong to the preceding items's AP because they do not have an accent.

- **ga** A1|A4
- **wo** A1|A4
- **nɪ** A1|A2
- **nara** A2|A4
- **hȧ** A3|A4

Items that do not have a word boundary on the left, unless otherwise specified in the dictionary, belong to the preceding word's AP regardless of whether they have an accent. These items are called *inflectional suffixes* and have an *inflectional symbol*, **-** or **·**, somewhere in the spelling. (Deletion rules may be applied. See Chapter Seven.) Some of the inflectional suffixes are shown below:

- **-nı** km-iii|A2
- **-ˌno** m-i|R2
- **·ˈmasu** d·i|P2
- **ˈ(_)r·u** d·i|R1
- **ˈ(_)·ı** k·i|R1

Knowing these facts is enough to extract all unmarked AP boundaries from sentences written in Kyouro.

- **ˈtarou ga ıˈkebana wo kannsyouˈˎ sıˈta.**
- **ˈhanako ga huˈtyuui wo ˈkoukaıˎ sıˈta.**
- **ˈbasu ga mokuteˈkıtı nı toutyakuˈˎ sıˈta.**
- **ˈmakku nara ıroiroˈna koˈto ga kanntannˈ-nı ˈdeˈkir·u.**

An AP boundary can be a boundary fall

An AP boundary can be a boundary fall. Some boundary falls that occur between words are predictable. Unpredictable boundary falls must be specified in the lexicon. Kyouro's spelling rule discriminates among different types of predictable and lexical boundary falls. »stibine.

Unlike AP boundaries, a boundary fall is always directly shown in the spelling form of a sentence with a *boundary fall symbol*, ` or ˎ, regardless of whether the boundary fall in question is predictable. The former, the upper boundary fall symbol, is used when the corresponding boundary fall is phonetically realized. The latter, the lower boundary fall symbol, is used when the potential boundary fall is canceled due to the phonological environment. When a lexical item is spelled in isolation, only lexical boundary falls are shown.

The lower boundary fall symbol does double duty. Taking advantage of the fact that the presence of a boundary fall symbol implies an AP boundary even when the boundary fall is canceled, the same symbol is used to indicate a lexical AP boundary that does not accompany a boundary fall—it can be interpreted as a potential boundary fall that just happens to never be realized due to the fixed phonological condition.

In isolation, if the boundary fall is never or only conditionally phonetically realized, the lower boundary fall symbol is used, and if it is always realized phonetically, the upper boundary fall symbol is used. A space is never inserted within a lexical item. The following are some items that are spelled with a boundary fall symbol in isolation.

ˈhı-ˌnoˎkuruma˝	moˈno`mous	ˎdaˈkega	ˎdaˈroukara	
ˈgogonoˎkoutya˝	otoˈko`ıxtupıˈkı	ˈkannˎmuryou˝	ˎdaˈke	
ˎdaˈroukedo	ˈımıˎhumeı˝	ˈdaıˎgozyuunıˈkaı	ˎdeˈsyoukedo	
ˈzennˎzyuuˈgyouinn	ˈkousyoˎkyouhusyou˝	tuˈgı`		

tuˈgı` is assumed to be |A4 because per our spelling convention, a boundary fall symbol must correspond to an AP boundary and therefore the symbol is the upper one even in isolation.

SEGMENTS

Verbs

Consider the following:

taberu	tabeta	tabenai
tsukiru	tsukita	tsukinai

It seems that the forms *ru, ta,* and *nai* are inflectional suffixes.

Now consider the following:

tsukuru	tsukutta	tsukuranai
yobu	yonda	yomanai

The suffix *ta* shows an allomorph *tta* in *tsukutta,* and in *yonda* it takes the form of *da*. Also, the suffix *ru* in *yobu* takes the form of *u*. As we will see below, it is also reasonable to consider the phonetic *ru* in *tsukuru* as a sequence of the last consonant *r* of the stem and the *u* of the suffix *ru*.

> The term "allomorph" is conveniently used to refer to any alternative phonetic forms of a given syntactic unit.

The allomorph *u* appears after a consonant as observed in the following list of verb forms.

kaku	kaita	kakanai
nugu	nuida	nuganai
hanasu	hanashita	hanasanai
katsu	katta	katanai
shinu	shinda	shinanai
kamu	kanda	kamanai
tsukuru	tsukutta	tsukuranai
warau	waratta	warawanai

Since in these forms the consonants preceding *u* differ, they must be part of the stem, not the suffix.

Hence, the following rule can be thought of:

- r —> [o] /C_.

which reads "r *is deleted after a consonant.*"

This chapter lays out phonological rules to describe several allomorphs of stems and suffixes. The rules are not necessarily system-wide. Some rules are better restricted to the context of inflection. The distinction however is not made explicit.

Similarly, the alternation of the suffix *ta* observed in these forms can be described by a set of segmental rules.

First, it is obvious that the form *da* appears after a voiced stop (*g, n, b, m*) from the forms *nuida, shinda, kanda,* and *yonda*.

- t —> d /{g, n, b, m}_.

"*t* becomes *d* after *g, n, b,* or *m*."

Second, *tta* appears after *t, r,* and *w* as in *katta, tsukutta,* and *waratta*.

- t —> tt /{t, r, w}_.

Now, there needs to be subsequent rules to alter or delete the stem's final consonant.

k and *g* become *i* before *t* or *d,* but the syllabicity of *i* depends on the preceding segment. If it follows *e*, it is a nucleus, and otherwise, it is an offglide.

- {g, k} —> i (offglide)/_{t, d},
- i —> i (nucleus)/e_.

Generally, the order of application matters.

Consonants are deleted before *tt*.

- C —> [o] /_tt.

m, b, and *n* become *n* before *t* or *d*.

- {m, b, n} —> n /_{t, d}.

s becomes *shi* before *t*. (*hanasu–hanashita*)

- s —> shi /_t.

Now we turn to *nai*, which a close examination of the aforementioned list reveals is underlyingly *anai* (e.g., *tsukuru–tsukuranai*). The allomorph *nai* appears after a vowel ending stem.

- a —> [o] /V_.

This may seem trivial, but this rule provides a simple way to reveal stem-final consonants. In fact, before *a* is the only environment where the stem-final *w* is clearly phonetically and graphically revealed. In (most) other vocalic contexts, the stem-final *w* is deleted.

- w —> [o] /_{i, e, o}.

The case of *u* admits multiple interpretations. As the original *u* corresponds to a syllabic *u* on the surface, and since *w* is by definition a nonsyllabic counterpart of *u,* it is possible to see the *w* as the onset before *u*. In Kyouro, when a stem-final *w* is followed by *u,* it is written as if the stem-final *w* remains intact.

waraw'	+	'(_)r·u	—>	wa'ra'w·u

The vowel letters *e* and *o* of verb suffixes are underlyingly nuclei and, when *w*-deletion is applied, their syllabicity remains intact. The syllabicity of *i* in the same environment depends on the suffix. (These suffixes will be discussed in the following pages.)

waraw'	+	'(_)r·ėba	—>	wa'ra'·ėba
waraw'	+	y·'ȯu	—>	wara'·'ȯu

The following table summarizes the obtained lexical forms of the stems and the suffixes as well as the corresponding surface forms of when they are coupled.

	ru	ta	anai
kak	kaku	kaita	kakanai
kag	kagu	kaida	kaganai
hanas	hanasu	hanashita	hanasanai
kat	katsu	katta	katanai
shin	shinu	shinda	shinanai
yob	yobu	yonda	yobanai
yom	yomu	yonda	yomanai
tsukur	tsukuru	tsukutta	tsukuranai
waraw	warau	waratta	warawanai
tsuki	tsukiru	tsukita	tsukinai
tabe	taberu	tabeta	tabenai

Using our preceding discussion as a basis, we will continue to examine the remaining suffixes. Let us begin by examining the following list of inflectional forms of verbs.

kaku	kakare	kakeba
kagu	kagare	kageba
hanasu	hanasare	hanaseba
katsu	katare	kateba
shinu	shinare	shineba
yobu	yobare	yobeba
tsukuru	tsukurare	tukureba
warau	waraware	waraeba
tsukiru	tsukirare	tsukireba
taberu	taberare	tabereba

The appearance and disappearance of *r* are consistent for each stem. It appears after a vowel and disappears after a consonant. Hence, it seems safe to say that here we are dealing with the suffixes that are underlyingly *rare* and *reba,* for which no additional rules are required.

Now observe the following:

kaku	kakō
kagu	kagō
hanasu	hanasō
shinu	shinō
yobu	yobō
tsukuru	tsukurō
warau	waraō
tsukiru	tsukiyō
taberu	tabeyō

We can see that the segment *y* appears and disappears simultaneously with *r*. Hence, the suffix is *yō* and we need to modify our rule accordingly:

- {r, y} —> [o] /C_.

Finally, observe the following:

kakanai	kakimasu	kaki
kaganai	kagimasu	kagi
hanasanai	hanashimasu	hanashi
katanai	kachimasu	kachi
shinanai	shinimasu	shini
yobanai	yobimasu	yobi
tsukuranai	tsukurimasu	tsukuri
warawanai	waraimasu	warai
tsukinai	tsukimasu	tsuki
tabenai	tabemasu	tabe

It is clear that the vowel *i* appears and disappears in the same way as *a*. Hence, we need to modify our rule of *a*-deletion to include the deletion of *i*.

- {a, i} —> [o] /V_.

Consequently, the suffixes are *imasu* and *i*.

Implicit in the list is that when *i* follows *w* as in *waraimasu* and *warai*, the phonetic *i* is a nucleus for the suffix *i* and an offglide for *imasu*. This implies that we are observing two different *i*'s in these suffixes. We therefore suppose that the *i* in the suffix *imasu* is underlyingly an offglide while the *i* in the suffix *i* is underlyingly a nucleus. This assumption necessitates a further modification of the vowel deletion rule to capture both types of *i*.

- {a, i (nucleus), i (offglide)} —> [o] /V_.

And the *w*-deletion rule must include the case of the offglide *i*.

- w —> [o] /_{i (nucleus), i (offglide), u, e, o}.

And to capture the fact that the offglide *i* turns into a nucleus when preceded by a consonant, we need the following rule.

- i (offglide) —> i (nucleus) /C_.

The table on the next page summarizes our observations so far. As a spelling system, Kyouro requires any form to conform to this table regarding segmental alternations in order for it to be classified as a verb stem or a verb suffix. If the form in question necessitates additional phonological rules, it will not be considered a verb. Simply put, irregular inflections are not inflections.

Where the distinction appears to be critical, the terms *token* and *tokenized form* are used to refer to Kyouro's unit items and their represented forms, respectively, as opposed to *lexical item* and *lexical form*, which refer to scientific notions of basic units of linguistic expressions and their representations. The tenet that forms not conforming to the phonological rules listed here are not verbs is an artificial praxis, not a scientific argument. The item *iku* 'to go' is treated as a token as per our convention, but this does not entail the claim that it is an indivisible lexical item.

	ru	rare	reba	anai	yō	ta	imasu (i is offglide)	i (nucleus)
kak	kaku	kakare	kakeba	kakanai	kakō	kaita	kakimasu	kaki
kag	kagu	kagare	kageba	kaganai	kagō	kaida	kagimasu	kagi
hanas	hanasu	hanasare	hanaseba	hanasanai	hanasō	hanashita	hanashimasu	hanashi
kat	katsu	katare	kateba	katanai	katō	katta	kachimasu	kachi
shin	shinu	shinare	shineba	shinanai	shinō	shinda	shinimasu	shini
yob	yobu	yobare	yobeba	yobanai	yobō	yonda	yobimasu	yobi
yom	yomu	yomare	yomeba	yomanai	yomō	yonda	yomimasu	yomi
tsukur	tsukuru	tsukurare	tsukureba	tsukuranai	tsukurō	tsukutta	tsukurimasu	tsukuri
waraw	warau	waraware	waraeba	warawanai	waraō	waratta	waraimasu (i is offglide)	warai (i is nucleus)
tsuki	tsukiru	tsukirare	tsukireba	tsukinai	tsukiyō	tsukita	tsukimasu	tsuki
tabe	taberu	taberare	tabereba	tabenai	tabeyō	tabeta	tabemasu	tabe

Some miscellaneous facts about the segmental shapes of verb stems:

- *kaer* and *kaes* are the only known "simplex" with an offglide *e* in the entire lexicon of the reference dialect. (I try to avoid discussing cases "compounds" like *furikaeru* ('to look back'/'to change (the date)'), where the suspected diphthong cannot be distinguished from a hiatus by the distribution of R tone, although for this particular case it seems reasonable to use the distribution of F tone as evidence.) The peculiarity of these two stems can be highlighted by comparing them with *aeg,* a verb stem whose second last mora is a nucleus *e*.

- The second last mora being an offglide *i* is not rare in the entire lexicon but is extremely rare for a verb. Some of known instances are *hair* ('to enter') and *hais* ('to abolish'). (*aisu* 'to love' is probably not a verb form. Author's note to self: a discussion on this should be given.)

- The second last mora being the second half of a long vowel (obviously not rare in the entire lexicon) is found in a small number of verb stems including *tōs, tōr, shōs,* and *kōr*. Compare them with *moyoos,* where the second last mora is a nucleus *o* which happen to follow another *o*.

- The potential verbs are mostly predictable but less so than the imperative form. A vowel verb on one hand typically has its potential counterpart ('potential verb') that theoretically can be derived by adding **rarė** to the base. A consonant verb on the other hand typically has its potential counterpart derived by adding **ė** to it. Besides their imperfect predictability, since this type of segmental alternation (**rarė** to **ė**) is not allowed in Kyouro, the potential verbs' *rare* and *e* must be written as part of the stem.

- The imperative form of a verb is predictable. It is obtained by adding **ro** if it is a vowel-ending verb ("vowel verb") and **e** if it is a consonant-ending verb ("consonant verb"). Since Kyouro does not allow as a rule the alternation between **ro** and **e** (it allows certain arbitrarily-chosen phonological operations but not the **ro**-to-**e** alternation; it is a whitelist system), the imperative form must be written as a unit form

Adjectives

Every adjective stem ends with a vowel, *a, i, u, o,* or marginally *ō*.

Some examples of adjectives are: *samu* ('cold'), *atsu* ('hot'), *sukuna* ('small in number'), *urusa* ('loud'), *hiro* ('wide'), *sema* ('narrow'), *kitana* ('dirty'), *mabushi* ('bright').

The noun group

Nouns, temporal nouns, and adjectival nouns are phonological similar and called *the noun group*.

The following are some members of the noun group: *yama* ('mountain'), *umi* ('ocean'), *te* ('hand'), *kinō* ('yesterday'), *kyō* ('today'), *ashita* ('tomorrow'), *nihongo* ('Japanese language').

ACCENT

The upper accent

Recall that the AP's template is nF1F0 (Chapter Two). This means that for any AP it is sufficient to specify its last nF mora to phonologically characterize it in terms of nF and F.

Consider the following nouns followed by *nara*. These nouns do not include an AP boundary in their underlying form and with *nara* constitute a single AP.

ke nara	nF-nF-F
te nara	nF-F-F
kaki nara	nF-nF-nF-F
yama nara	nF-nF-F-F
haru nara	nF-F-F-F
karada nara	nF-nF-nF-nF-F
onna nara	nF-nF-nF-F-F
tamago nara	nF-nF-F-F-F
remon nara	nF-F-F-F-F

Notice that for all cases the last mora is F. This is because *nara* is a two mora item that does not inherently bear nF and for the noun group any item can bear nF only up to its right segmental edge's next mora. For any item in the noun group, if it is N mora long, the longest nF sequence possible when followed by *nara* is N + 1 mora long.

Since an AP must begin with an uninterrupted sequence of nF moras, the fact that the longest possible nF sequence for the initial N mora item followed by *nara* is N + 1 readily means that the number of possible nF distribution patterns under the same condition is also N + 1, all of which can be uniquely characterized by specifying the right edge of the potential nF sequence.

In Kyouro, the said specification is indicated by the special symbol ˈ (modifier letter vertical line) inserted immediately before the final mora of the potential nF sequence. (This is the first approximation and will be modified later.) This symbol as well as the specification—the phonological feature that is supposed in the given position or that specifies the given position—is called the *accent* of the form, simplex or otherwise.

Accordingly, the list is written in Kyouro as the following:

ke' nara	nF-nF-F	0
'te nara	nF-F-F	1
kakı' nara	nF-nF-nF-F	0
ya'ma nara	nF-nF-F-F	1
'haru nara	nF-F-F-F	2
karada' nara	nF-nF-nF-nF-F	0
onn'na nara	nF-nF-nF-F-F	1
ta'mago nara	nF-nF-F-F-F	2
'remon nara	nF-F-F-F-F	3

The *accent number* is a number assigned to an AP or relevant form that is used to refer to the position of the accent. For any given form, if the accent is at the right segmental edge (i.e., immediately after the ultimate mora), the accent number is 0; if the accent is immediately behind the rightmost mora, the accent number is 1; if the accent is right behind the penultimate mora, the accent number is 2; and so on and so forth. When the given item's accent number is N, we also say that the item's accent is N. In some older documents, accent numbers are written in parentheses. As of version 5, parenthesizing accent numbers is optional. In the list above, the accent number indicated in each line is the one given to the item on the left. For example, the accent number of **ya'ma** is 1 because the accent is immediately before the last mora of the word.

For some groups of lexical items, the position of the accent is more limited.

The accent number of a verb stem is revealed when it is followed by, among others, *ta*. When a verb stem is followed by *ta*, the following tonal forms are possible.

kita	nF-nF	0
nusumimita	nF-nF-nF-nF-F	1
tsukita	nF-F-F	2

The accent number of a verb stem can only be either 0, 1, or 2. 1 is marginal.

For adjectives, the accent numbers 0, 1, and 2 are possible, but 2 is available almost exclusively for a two-mora stem, and 1 is mostly found in three-mora or longer stems. Hence the accent numbers 1 and 2 for adjective stems are in a near complementary distribution.

The accent number of an adjective stem is revealed when it is followed by, among others, *ku*.

koku	nF-F	1
akaku	nF-nF-nF	0
aoku	nF-F-F	2
yasashiku	nF-nF-nF-nF	0
kawaiku	nF-nF-nF-F	1

It seems that the stem of *koku* is the only one-mora adjective stem. (As we will see later, *yoku* and *naku* are not considered adjective forms due to otherwise irregular behaviors in inflection.)

Some suffixes have an accent, and therefore one AP can have multiple accents. This is observed when, among other times, a verb stem takes the suffix *ru*, whose accent number is 2 (The nF sequences are underlined):

uita	uku
kaita	kaku
wareta	wareru
tabeta	taberu
koroshita	korosu
hanashita	hanasu
kurabeta	kuraberu
osameta	osameru
hamideta	hamideru

Take *tabeta-taberu*. On one hand, the nF sequence of *tabeta* extends only through the first mora. This suggests that the stem *tabe*'s accent number is 2 and nF is inherently given to the first mora only. On the other hand, the nF sequence of *taberu* extends through the second mora. The nF of *be* must have been given by the suffix *ru*, meaning that the accent number of *ru* is indeed 2.

As the list exemplifies, for each accent present in the AP, the nF sequence extends through the accent's adjacent moras. Between the stem's accent and the accent of *ru,* whichever is placed on the right is phonetically effective.

In the following example, it is shown that the accent of the suffix *masu*, whose accent is 2, is effective with all stems.

uita	ukimasu
kaita	kakimasu
wareta	waremasu
tabeta	tabemasu
tabeta	tabemasu
koroshita	koroshimasu
hanashita	hanashimasu
kurabeta	kurabemasu
osameta	osamemasu
hamideta	hamidemasu

In both of these cases, whichever is on the rightmost position is phonetically effective. Recalling that the accent is supposed immediately before the final nF mora, however, to generalize our previous observation, the only necessary rule is:

- *For each accent within the limits of an AP, the mora immediately after the accent and all moras behind it receive nF; the remaining moras receive F.*

  I suppose this rule feels intuitive for many people under the assumption that all moras are by default F and nF's are given by accents.

This rule ensures only the rightmost accent in the given AP is phonetically effective—given the rightmost accent, all other accents giving nF are redundant.

When inflection is recursive, an AP can have more than two accents. Still, the rule is the same. For ˈ**tabė·ha**ˈ**zımė·**ˈ**masu**, the rightmost accent is on *masu*, and that is the only accent that phonetically matters.

Correlation on version 5.2.3: the *rightmost* accent is on *masu*, not *leftmost*. I thank @atri_nen for pointing out the error.

The lower accent

Examine the following, in which nF sequences are underlined:

ke nara	ke-kara
te nara	te-kara
kaki nara	kaki-kara
yama nara	yama-kara
haru nara	haru-kara
karada nara	karada-kara
tamago nara	tamago-kara
remon nara	remon-kara

The ablative suffix *-kara* extends the nF sequence when the preceding item's accent number is 0.

Now, consider the following:

ke nara	ke-no
te nara	te-no
kaki nara	kaki-no
yama nara	yama-no
haru nara	haru-no
karada nara	karada-no
tamago nara	tamago-no
remon nara	remon-no

The possessive suffix *-no* extends the nF sequence when the preceding item's accent number is 1 and the preceding item's accent is not on the left edge.

It seems that certain items can move the position of the preceding item's accent on a selective condition. To determine exactly and generally when and to where a relevant item moves the preceding item's accent, now consider the case of adjectives:

ko-ku	ko-katta
aka-ku	aka-katta
ao-ku	ao-katta
yasashi-ku	yasashi-katta
sukuna-ku	sukuna-katta

Now it is observed that the adjective suffix *-katta* can move the preceding item's accent to the left when the accent number is 0.

What is common among the cases of *-no, -kara,* and *-katta*? It seems that a suffix can move the preceding item's accent only if the target accent is not placed on the AP's left edge. In the case of *-no* and *-kara,* the AP's accent moves from the position 2 to the position 1, and *-katta,* 3 to 4. It seems that the "goal" of the affected accent is specified as a property of the suffix causing the move. Observations such as these allow us to suppose the following generalizations as the first approximation:

- There exists some position specified for each of these suffixes to which the preceding form's accent is moved; and
- For the move to occur, the affected accent's original position must be within one mora from the specified position and not the left edge of the AP.

Our first approximation, however, fails to capture the following fact. Certain verb suffixes, when attaching to a verb stem whose accent number is 0, move the stem's accent to a specific position. This can be observed below:

tsuketa	tsukerareta
tometa	tomerareta

To capture this, we need to modify the second statement as the following:

- For the move to occur, the affected accent's original position must be **at the right edge of the base or** within one mora from the specified position and not the left edge of the AP.

Now, our generalization can be restated as the following:

- *Certain items move the immediately preceding form 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 Kyouro, in the tokenized form, P is marked by the special symbol (˵) (modifier letter low vertical line) and is called the *lower accent*. The regular accent (ˈ) may be called *the upper accent* if necessary, in which case *accent* may be used as a hypernym.

When A is moved to P, P is then marked by ˈ to show that it is phonetically effective (unless superseded by another accent) and A's original position is then marked by ˵ to show the trace. (P and A can be seen as a pair; when ˵ shifts up, ˈ shifts down.) Otherwise, accents are shown in the tokenized form.

Inflection can be recursive. If the resulting form of an inflectional process undergoes a further inflectional process, the said form acts as a holistic form. If in inflection the form on the left has multiple upper accents as a result of the preceding processes, only the rightmost one is subject to paring. After all processes are done, the placement of the rightmost upper (not the lower) accent determines the surface positions of nF and F.

An example of recursive inflection within the *verb group* (the verbs and the adjectives) is the following:

- *kappuru ga tagai o mitsumeatta*. "A couple contemplated each other."
- *futakumi no kappuru ga tagaini tagai o mitsumeawaseta*. "A pair of couples made each other contemplate each other."
- *futakumi no futakumi no kappuru ga tagaini tagaini tagai o mitsumeawasesaseta*. "A pair of pairs of couples made each other make each other contemplate each other."
- *futakumi no futakumi no futakumi no kappuru ga tagaini tagaini tagaini tagai o mitsumeawasesasesaseta*. "A pair of pairs of pairs of couples made each other make each other make each other contemplate each other."

The form *mitsumeawasesasesaseta* is written in Kyouro as **mıtume ˈȧw·ˈasė˵s·ˈasė˵s·ˈasė˵t·a** and this is derived through the following process.

which can be represented as the following tree:

mıtumė'·'aw·'asė,s·'asė,s·'asė,t·a

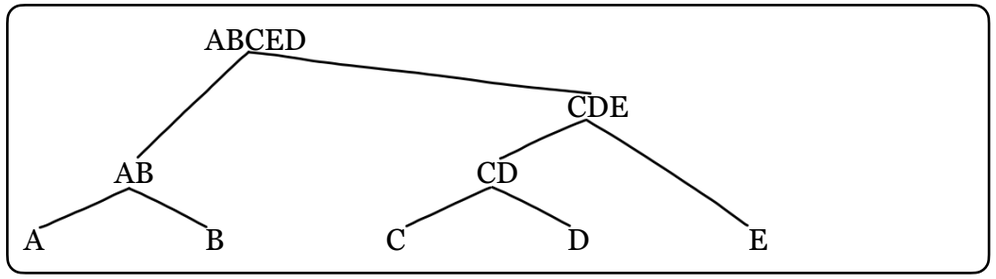
mıtumė'·'aw·'asė,s·'asė,s·'asė,

mıtumė'·'aw·'asė,s·'asė,

mıtumė'·'aw·'asė,

mıtumeȧw s·'asė, s·'asė, s·'asė, t·a

The tree diagrams used in this document do not show branches graphically. Branches between nodes are indirectly yet unambiguously shown by the positions of the nodes. The following serves as an example of how branches can be complemented:



In the derivational process, forms are concatenated working from left to right. Without accent switching, it may be less obvious why the order matters. The following example involves accent switching.

The form *warawaserareru* 'to be made laugh' is related to *warau* 'to laugh' and *warawaseru* 'to make (someone) laugh.' The form *warawaserareru* is derived in the following way.

1.	waraw'	+ s·'asėr,	+ r·'arėr, + '(_)r·u
2.	waraw,	+ s·'asėr'	+ r·'arėr, + '(_)r·u
3.	wara,w·'asėr'	+ r·'arėr,	+ '(_)r·u
4.	wara,w·'asėr,	+ r·'arėr'	+ '(_)r·u
5.	wara,w·'asė,r·'arėr'		+ '(_)r·u
6.	wara,w·'asė,r·'a'rė'r·u		

which can be represented as the following tree:

wara,w·'asė,r·'a'rė'r·u

wara,w·'asėr,·'arėr'

wara,w·'asėr'

waraw' s·'asėr, r·'arėr, '(_)r·u

Notice how, when accent switching is involved, working from left to right is necessary to derive the surface form correctly.

The bimoraic rhyme rule

Consider the following:

hachiji nara	hachiji-no	hachiji-kara
kinō nara	kinō-no	kinō-kara

kinō indicates that the nF region extends exactly through the first mora of nō.

According to our accent theory, *hachiji* can be appropriately characterized and spelled as **haˈtızı** in Kyouro. How do we spell *kinō*? Since in the *nara* context it is nF-nF-F, **kıˈnou** can be thought of, but that would not explain the form *kinō-no*, nF-nF-nF-nF. In order for *no* to be nF, the accent of the preceding item, in this case, *kinō,* needs to be 1 or 0, that is, **kınoˈu** or **kınouˈ**. At the same time, the *kara* context suggests that it is not 0, since if it was 0, *kara* would be nF-nF. This would rule out **kınouˈ**. Now, between **kınoˈu** and **kıˈnou**, which is the correct analysis?

To solve the question, observe the following:

kōban nara	0
jōdan nara	2
nēsan nara	4

All of these words are two syllables and four moras. All syllables are with a bimoraic rhyme, a rhyme that is two-mora long. Now, notice that the list does not include words with the accent number 1 or 3. In a perfect world, the list should look like the following:

	kōban nara	0
*	rōnan nara	1
	jōdan nara	2
*	mūkin nara	3
	nēsan nara	4

**rōnan* and **mūkin* are intended as nonwords.

As implied by the asterisks, a word with the accent number 3 or 1 in this environment is not realized as-is. In fact, in the entire lexicon, no known items clearly have a bimoraic rhyme whose second mora is nF when immediately followed by an F mora.

Let us now turn our attention back to the original question. If *kinō* is **kınoˈu**, it will not be realized in the *nara* context as *nF-nF-nF. Such a form is never possible in any comparable contexts. We can suppose the following phonological rule while conforming to our observations of the phonetic forms.

- *An nF mora on the second half of a two-mora rhyme becomes F if it is immediately followed by an F mora.*

Now we can correctly characterize and write *kinō* as **kınoˈu** and this explains the behaviors of *kinō* and a number of other items, including *ototoi* **ototoˈı** ('the day before yesterday'), *nihon* **nıhoˈnn** ('Japan'), *nippon* **nıxtupoˈnn** ('Japan'), *mukō* **mukoˈu** ('over there'), *sō* **soˈu** ('like that; that way'), *kō* **koˈu** ('like this; this way'), *takusan* **takusaˈnn**('a lot'), *daitai* **daıtaˈı** ('approximately'), *kurai* **kaˈraˈ·ı** ('is dark'), *usui* **uˈsuˈ·ı** ('is thin'), and *omoi* **oˈmoˈ·ı** ('is heavy').

In isolation, for *kinō,* both nF-nF-F and nF-nF-nF are accepted, but for *nihon,* only nF-nF-F is accepted. Perhaps this discrepancy may be attributed to the segmental makeup, maybe the syntactic standings, or a combination of them. It is also possible that *nihon* is not a noun. With regard to this issue, for now, isolation forms are not used as evidence for deciding the correct spelling form.

In one other context, the bimoraic rhyme is relevant to tone allocations. If a boundary fall follows, an nF-nF bimoraic rhyme becomes nF-F. This is illustrated in the following example:

- *taimumashinde kinō itta*. "said yesterday in the time machine."
- *taimumashinde kinō itta*. "went to yesterday by the time machine."

Only in the second example *kinō* is pronounced as nF-nF-F. This is because in the given environment a boundary fall is otherwise inserted (a break in the underline aligned to a space indicates a boundary fall):

- *kurumade yama itta*. "went to the mountain by car."
- *shinkansende otōsanno tokoro itta*. "went to dad's place by bullet train."

Observe that these two sentences have a syntactically identical structure as the last example: [means]-de [destination] itta. "went to [destination] by [means]." In this structure a boundary fall is found on the left edge of *itta*, as long as the preceding mora is nF.

Let us suppose that at some stage of derivation *kinō* is nF-nF-nF in the time travel sentence as illustrated by the following notation (notice the asterisk).

- **taimumashinde kinō itta. "went to yesterday by the time machine."*

A boundary fall is indicated between *kinō* and *itta*. Comparing this form with the actual surface form, we can naturally suppose that another environment where an nF-nF rhyme turns into nF-F is immediately before a boundary fall.

Hence, we can revise our rule as the following:

- *An nF mora on the second half of a two-mora rhyme turns into F if it is immediately followed by an F mora or a boundary fall.*

After the application of this rule, since a boundary fall never appears after an F mora, the phonetic boundary fall between *kinō* and *itta* needs to be deleted, although, in the Kyouro spelling, the boundary fall symbols remains there to show the tone of the preceding mora.

A form that has a bimoraic rhyme at the end whose accent number is 1 is said to belong to *kino group*.

> The term *kino group* (originally "*kinōrui*") was coined by @xhioe in their tweet. My 2022 blog post on a relevant topic is available here. I thank @xhioe for coining the convenient term and promoting the concept.

Accent shift

As we have seen, where the stem and the suffix abut, the latter one of two adjacent vowels is deleted. An example of this is when the verb ˈ**tabė** ('to eat') takes the suffix **i·hả** ('as for doing...') to form ˈ**tabė·hả** ('as for eating'). There the vowel **i** of the suffix is deleted.

When the suffix's vowel that is subject to deletion immediately follows an accent, lower or upper, the accent moves to one mora left from its original position: **hurėˈ + i·hả —> huˈrė·hả**

The accent shift takes place after switching paired accents. Hence, **hurėˈ + ˈa·na˛1 —> hurė˛ + ˈa·naˈ1 —> hu˛ˈrė·naˈ1**

Similarly, when a suffix's initial consonant undergoes deletion, the suffix's accent, upper or lower, is placed according to the suffix's accent number and its segmental shape after the consonant deletion. Take *ru,* for example. The accent number of *ru* is 2, which means the accent intervenes in the stem and is placed one mora left of the segmental edge of the stem (ˈ**tabė + ˈ(_)r·u —> ˈtaˈbėr·u**). When it follows a consonant ending stem, its accent is placed one mora left of the segmental edge of the stem not counting the stem-final consonant: **waravˈ + ˈ(_)r·u —> waravˈ + ˈ(_)·u —> waˈrav ˈ·u** (a further spelling rule readjusts the position of accents placed in the middle of a mora: **—> waˈraˈw·u**)

Chapter Four

WORD-LEVEL SYNTAX

BASIC OBSERVATIONS

Syntactic compositionality

A sentence consists of smaller forms. The compositionally of a sentence is observed in sentences like *toshokan ni itta* ('went to the library'), whose former parts *toshokan ni* can be replaced with *kōen ni* ('to the park') to make *kōen ni itta* ('went to the park').

Syntactic compositionality is distinguished from phonological compositionality or that a piece of speech sound is comprised of smaller units such as segments. Syntactic compositionality is when smaller forms in a sentence can be identified by the speaker as coherent linguistic units instead of arbitrary sequences of segments, syllables, morphemes, words, etc.

Coherency minimally implies that both phonological and semantic representations are constant. Phonological compositionality and semantic compositionality must be observed simultaneously. Many "compounds" fail here. The latter part of *tībaggu* ('tea bag') may be analyzed to have the same phonological form as *baggu* ('handbag'), and a teacher may recklessly attribute them to a single syntactic form, but adding the form *tī* to what in isolation means 'handbag' to reasonably derive 'teabag' is impossible. They must be treated as unrelated syntactic entities.

This does not exclude the possibility that they are historically related.

Syntactic boundary

Not all forms are mutually replaceable. The former part *ōkina* of the sentence *ōkina kōen* ('the big park') can be replaced with *kireina* ('beautiful') to form *kireina kōen* ('the beautiful park') while the same *kireina* cannot be used in place of *toshokan ni* as in *toshokan ni itta*: **kireina itta*.

The fact that only certain forms can stand in a certain spot of a sentence necessitates some means to identify the spot in describing the properties of syntactic forms. In Kyouro, this is done by labeling the syntactic boundaries. In this case, the boundary at the left edge of *itta* is called A4, and the right edge of *kireina* and *ōkina* R2. *Itta* cannot follow *kireina* and *ōkina* because they do not share the boundary where they abut. The forms *toshokan ni* and *kōen ni* can stand before *itta* because their right edges are A4.

Some of these subparts of a sentence can be further divided into smaller forms. Both *toshokan ni* and *kōen ni* have *ni* in common, and *toshokan* and *kōen* can replace each other with itself. The left edge of *ni* and the right edges of *toshokan* and *kōen* are the A1 boundary.

Composition of forms

The boundary labels are indicated alongside the phonetic form in the following notation:

- *kōen* R2|A1

where the label on the right side of the symbol '|' ("bar") indicates the right boundary, and on the left side the left boundary of the syntactic form whose phonetic form is indicated on the left. In this example, the syntactic form's phonetic representation is *kōen,* its left boundary is R2, and its right boundary is A1.

Sometimes the vertical bar is used to represent only one of the two boundaries of a form. If a form is said to be |X ("bar-X"), it means that the form's right boundary is X. Likewise, a form whose left boundary is X is said to be X| ("X-bar").

The other items we have considered so far are the following.

- *toshokan* R2|A1
- *kōen ni* ?|A4
- *toshokan ni* ?|A4
- *ni* A1|?
- *toshokan ni itta* ?|?
- *kōen ni itta* ?|?
- *ōkina* ?|R2
- *kirei na* ?|R2

Some components are phonetically null and accordingly left implicit.

An unknown boundary is indicated by the symbol '?'.

Forms such as *kōen ni* are derived from its subparts. In this case, we need to characterize *kōen* and *ni* and a derivational rule that derives *kōen ni* from them.

The forms *kōen ni* and *toshokan ni* share the right boundary. Since they share the form *ni* in common, the right boundary A4 might have been given by the form *ni*. Additionally, observe the following:

- *ōkina kōen* ?|A1
- *kōen* R2|A1

Despite the difference on the left part (*ōkina* is absent in the second form), they share the right boundary.

These observations suggest that the right boundary of a complex form is identical to the right boundary of its right subpart. Hence:

- *ni* A1|A4

The left boundary of a complex form in some cases seems to be identical to the left boundary of its right subpart:

- *toshokan* R2|A1
- *ōkina toshokan* R2|A1
- *kirei na toshokan* R2|A1

But this is not always the case:

- *ni* A1|A4
- *toshokan ni* .|A4

The boundary label "." (called *period*) indicates that nothing can be attached to it. Nothing can be the left sibling of *toshokan ni*. More on this will be discussed later.

Zero

Consider the following:

- *toshokan ni itta*
- *toshokan ni nara itta* ('went to the library in particular')

The forms *toshokan ni* and *toshokan ni nara* seem to share the right boundary. Does this mean that the form *nara* has the same A4 boundary on its left and right edges? If so, the grammar should allow *nara* to be added recursively, but that is not the case:

- **toshokan ni nara nara nara nara itta*.

The problem may be solved if we assume a *zero* in the first example:

- *toshokan ni [0] itta*
- *toshokan ni nara itta*

The phonetic form of a zero is indicated by the symbol '[0]'. (Brackets are given lest it be confused with the accusative *o*.) Zero is a title given to any syntactic form corresponding to an empty phonological form. Simply put, it plays some role in syntax but is not pronounced.

In the new notation, what corresponds to the phonetic form *ni,* which we previously described to be *ni* A1|A4, is described as:

- *ni [0]* A1|A4

which can be further divided into:

- *ni* A1|A2
- [0] A2|A4

Now, since *ni*'s right boundary is A2, *nara*'s left boundary should be A2 as well. Hence:

- *nara* A2|A4

This correctly captures the fact that *nara* cannot be recursively added after its duplicate.

We will later see that [0] A2|A4 can be further divided into [0] A2|A3 and [0] A3|A4.

Zeros are implicit in the Kyouro spelling.

Naming scheme

We have seen the following boundaries so far: R2, A1, A2, A3, A4, period. Other boundary labels used in Kyouro include *R1, C1, C2, C3, C4 (R2c), R2b, P1 (R2d), P2, P3,* and *P4 (C3b)*. Inside parentheses are aliases. The same type of boundary is called P4 or C3b depending on the context.

Boundaries are named so that zeros can be left implicit. For two syntactic forms that share the prefix (R, A, C, P, or G) where they touch each other, if the number (1-5) is greater on the right form, the two forms can be in effect siblings. For example, *toshokan ni* R2|A2 can be followed by *wa* A3|A4 because they share the prefix (A) where they abut and the number is greater on the right (3 versus 2). Theoretically, this is explained by assuming [0] A2|A3 in-between, but the naming scheme eliminates the necessity to explicitly mention zeros every time.

Branching

Repeatedly dividing parts into subparts readily means that a form is represented as a binary tree. The sentence *toshokan ni nara itta* is represented as the following:

```
                toshokan ni nara itta A4|R1
          toshokan ni nara .|A4
    toshokan ni .|A2
toshokan R2|A1     ni A1|A2    nara A2|A4      itta A4|R1
```

In this example, branching consistently takes place on the left node. If a sentence structure is always left-branching, it would be practically a flat structure. (The branching direction would not matter.) That branching directions can be mixed can be seen in the following example:

```
                kawaii ōkina toshokan ni nara itta A4|R1
              kawaii ōkina toshokan ni nara .|A4
          kawaii ōkina toshokan ni .|A2
    kawaii ōkina toshokan R2|A1
              ōkina toshokan R2|A1
A4|R1   A4|R1     R2|A1     A1|A2     A2|A4     A4|R1
kawaii  ōkina     toshokan  ni        nara      itta
```

Boundaries of the terminals are given on top of the phonetic forms to save space.

Correction in version 5.1.2: The third line is *kawaii ōkina toshokan* ***ni***. I thank @ho_hasios for pointing this out.

The branching direction can be contrastive in the sense that different branching directions correspond to different meanings. The following examples show that the phonetic form *kinō katta dōnatsu o tabeta* can mean "(I) ate a donut that (I) yesterday bought" or "yesterday (I) ate a donut that (I) bought" depending on the branching direction.

- *(I) ate a donut that (I) yesterday bought.*

```
                kinō katta dōnatsu o tabeta A4|R1
              kinō katta dōnatsu o .|A4
          kinō katta dōnatsu R2|A1
    kinō katta A4|R1
R2|A4     A4|R1     R2|A1     A1|A4     A4|R1
kinō      katta     dōnatsu   o         tabeta
```

- *yesterday (I) ate a donut that (I) bought.*

```
    kinō katta dōnatsu o tabeta A4|R1
                    katta dōnatsu o tabeta A4|R1
                  katta dōnatsu o .|A4
              katta dōnatsu R2|A1
R2|A4     A4|R1     R2|A1     A1|A4     A4|R1
kinō      katta     dōnatsu   o         tabeta
```

Different branching directions can be reflected differently in the phonetic form. Syntax-phonology mapping will be discussed in Chapter Six.

Merging

Merging is branching seen from below. When a phrase AB branches into A and B, it is said that A and B merge, A merges with B, or A and B are merged. Formation of a sentence is expressed in terms of what subforms merge with what in which order. The phrase ABC can be built by merging A and BC, which has been formed from B and C merging, or by merging AB, which has formed from A and B merging, and C, and can itself merge with another phrase to form a greater form which then can merge again; the process repeats until the sentence completes.

Syntactic forms are expressed in the following format:

- *to* P1\A2
- *hayaku* A4/A2
- *sugoku* ./A4

On the left is the phonetic form. On the right is the form's syntactic boundaries: the left boundary on the left and the right boundary on the right. The slash in the middle of the boundary labels indicates the merging direction: '/' indicates that the form merges rightward and '\' leftward. The symbol '|' is used to indicate boundaries without specifying the merging direction. As with the case of the vertical bar, slashes can be used to indicate a boundary for only one side of the slash. Vertical bar is useful for noting superficial observations and was used earlier in this section. The notation will be extended as we proceed with our analysis.

The relationship between the resulting form and its components can be complex. Here we lay out some conventions and principles to ease our discussions. They are general laws or rules which will be modified for specific contexts.

The right boundary principle: We assume that on a leftward merge, the right boundary is inherited to the resulting form. Taken literally, this is a blank statement (unless a major boundary is involved; see below) because, by definition, the right boundary does not matter when it merges leftward. We say that a leftward-merging form has a boundary X on its right to indicate that the resulting phrase will have X on its right.

The left boundary principle: If a form has R2 or A4 (called the *major* boundaries) on its left and merges leftward, the resulting form will have the same major boundary on its left. This is contrasted with the case of the minor boundaries (any boundary that is not major is *minor*), where the resulting form will have period on its left. (Later we will see that this can be overwritten in several instances.)

The direction principle: If the form has a major boundary on the left side of a rightward slash, the form is bidirectional; it can merge rightward or leftward. (By convention, we do not place period on the right side of a rightward slash as the symbols' meanings would contradict.) Forms that do not meet this criteria are unidirectional. We use a leftward slash when the item's left boundary is major and the item is leftward unidirectional.

- *hayaku* A4/A2

  If neither side of a slash is period, the form is usually leftward-merging and the right boundary is indicated to satisfy the right boundary principle. The direction principle can be understood as a statement of an exception of this generality—that if the left side of a rightward slash is major the form is allowed to merge in the opposite direction.

Boundary generation

Observe the following:

- *samukereba* A4/A4
- *samusa* R2/m-ii

These forms seem to share the stem:

- *samu* ./k·ii

Therefore the suffixes must be:

- *kereba* k·i\A4
- *sa* k·i\m-ii

But according to the left boundary principle, merging these suffixes with the stem should generate:

- **samukereba* ./A4
- **samusa* ./m-ii

which are wrong as demonstrated by:

- *sonnani samukereba*
- *huyuno samusa*

How should we solve this? I initially attempted to capture these facts by stating syntactic rules that generate these left boundaries conditioned by the right boundary of each case such as:

- X —> A4|A4
- X —> R2|m-ii

But the collection of these syntactic rules quickly grew painfully complex and it was evident that alternative approaches were necessary.

Instead of setting up several general rules, the left boundary of the resulting form is ascribed to the relevant lexical items as their properties:

- *kereba* k·i\A4/A4
- *sa* k·i\R2/m-ii

Where two slashes are used for one form, the leftmost boundary indicates the form's left boundary, and the two boundaries indicated on the right are for the left and right boundaries of the resulting form. Here it is stated that the suffix *kereba* generates A4/A4 and *sa* generates R2/m-ii when merging. These item-specific properties supersede the left boundary principle.

Ascribing these specific properties to each relevant item is potentially a backward development of the theory since setting up general rules can be, if successful, more economical, at least lexically. The intention here is to keep the system easy to handle for the human mind and to obtain accurate descriptions of lexical items, and to hope that elegant generalizations will be discovered later based on these descriptions and as a result a greater economy will be achieved.

Branch notation

Recall that the sentence *kawaii ōkina toshokan ni nara itta* has the following syntactic structure:

- ((((kawaii (ōkina toshokan)) ni) nara) itta)

If the same structure is visually expressed, it will look like the following:

```
                    kawaii ōkina toshokan ni nara itta
              kawaii ōkina toshokan ni nara
        kawaii ōkina toshokan ni
  kawaii ōkina toshokan
                    ōkina toshokan
kawaii        ōkina        toshokan   ni     nara        itta
```

A two-dimensional diagram like this is intuitive but space-consuming. To aid visual parsing while keeping everything in one line, what is called branch notation is employed:

- kawaii/ ōkina/ toshokan\\/ ni\/ nara\/ itta\

This equals to the following form in the reverse Polish notation:

- kawaii ōkina toshokan\\ ni\ nara\ itta\

Branch notation is simply reverse Polish notation with redundant slashes added to represent left branches for easier eye parsing. (It is called branch notation because you can see the slashes as branches squeezed into one line).

This notation was originally described in a blog post of mine. The original version allowed multiple distribution patterns of slashes for the same structure. The idea was to utilize the redundancy to reflect additional analyses or intuitions about the sentence, such as headedness. In this version, the notation is more restricted, so that the positions of slashes are uniquely determined. I thank @hedalu244 for discussing formalism of this notation back in 2019.

WORD BOUNDARIES

Quotation (P1, P4, P5)

We begin our investigation with quotation because it concerns approximately the outermost layers of a sentence. Almost any seemingly autonomous sentence can be quoted by some special function words such as *tte* or *to* and be incorporated into a larger sentence.

At least three types of quotation must be recognized. The first type (type-I quotation) introduces a fact or idea into a larger sentence. This can be called an indirect quotation. It semantically resembles the English *that* clause that introduces a fact or idea into a larger sentence. As an example of the resemblance and supporting evidence that it is a type of indirect quotation, in the following, *ashita* refers to the day after the time of the utterance.

- *ashitamadeni sūtsu o kurīninguni daseba, kekkonshikini wa maniau to omotta*. "thought that if I brought my suit to the cleaners tomorrow, it would be ready in time for the wedding."

The second type of quotation (type-II quotation) embeds the entirety of the target sentence (or "near sentence"—a discussion on this follows) nearly word-to-word into the larger sentence.

- *"demo sonna koto shitara maichan kaette korenai jan" tte katō ga iu kara, ōrora o mini iku no wa kekkyoku akirameta*. "Because Kato said, 'but if we do such a thing Mai won't be able to come back,' ultimately we gave up going to see auroras."
- *"isshoni momijigari ni ikimasenka" nante iwarete, aitsu ga kotowaru wake nai desho*. "If you ask him, 'Would you go to see the autumn leaves with me?' there's no way he will turn you down."

The third type (type-III) deals with potentially paralinguistic expressions. The quoted portion can violate the usual phonology of Japanese. This is illustrated by the following example.

- *jimu wa watashini "I'll take you to New York" to itta*.

(Fujita 1994, p. 75.)

The quoted part "I'll take you to New York" can be pronounced like English. Phonetic katakanization is not obligatory.

Inconveniently the phonetic forms of the quoting particles used in these different types of quotations partially overlap: *tte, to, nante,* and so on. But they are syntactically distinct.

The items used to make a type-II quotation are labeled as P4|.

Since these items merge leftward, the backslash symbol is used in place of the bar.

- *to* P4\A2
- *tte* P4\A4
- *nante* P4\A4

Similarly, type-III quoting particles can be characterized as the following:

- *to* P5\A2
- *tte* P5\A4
- *nante* P5\A4

And type-I quoting particles:

- *to* P1\A2
- *tte* P1\A4
- *nante* P1\A4

The type-III quotation is where we find the outermost layer of the structure of an utterance. The type-II deals with a unit clearly smaller than the utterance and nearly equal to but seemingly slightly smaller than the sentence.

Iori (2012), citing Minami (1974, 1993), treats the target of a direct quotation, called the D-rui or "type D" (per Iori 2017), as virtually identical to a sentence (p. 176, pp. 199-204). A motivation to make a distinction between the sentence and type D and a distinction between type-III and type-II can be illustrated by the following examples:

- *"kōtarōsan kin'yōbino yoru aitemasuka?" tte miyazawasan ga ittemashita yo.*
- *"a, kōtarōsan kin'yōbino yoru aitemasuka?" tte miyazawasan ga ittemashita yo.*

On one hand, the former can be a case of a type-II or type-III quotation. In a real-life speech, if it is a type-III quotation, the quoted portion can be pronounced with a different tone of voice from the rest of the utterance for the effect of imitating the supposed original speaker. If it is a type-II quotation, ignoring uncontrolled performance factors—stutter, hesitation, etc—the tone of speech will be consistent throughout the utterance. On the other hand, the latter is clearly a case of a type-III quotation due to the presence of the filler *a*. In a real-life speech it is likely, though not obligatory, that a different tone of voice is used to show that it is a type-III quotation. Certain "less meaningful" items seem to be unavailable for type-II quotation, and *a* used here seems to be one of them.

Another item incompatible with type-II is *tteba*.

- *"kin'yōbi wa itsumo isogashii n da tteba" to miyazawa ga itta.*

This is inevitably a case of a type-III quotation due to the presence of *tteba*. For this nature, the boundary between *tteba* and *to* can be seen as an approximation of the right boundary of a sentence. The sentence can then be approximated as any syntactic unit that is .|P5.

> Some more work needs to be done to pinpoint the layperson's notion of a sentence. Being .|P5 alone is not enough to exclude an utterance of multiple sentences and a type-III-quoted portion of a larger sentence.

In novels and other forms of literary works, a type-III quoting particle is often phonetically null (zero). Observe that in the following example the quotation cannot be a type-II quotation.

- *"kōtarōsan kin'yōbino yoru aitemasuka?" miyazaki ga itta.*

Predicate (P1, P2, P3)

The predicate sits approximately on the right edge of a syntactic form that is subject to a quotation.

Around the main clause predicate are P1, P2, P3, and P4, the last of which we have seen in the previous subsection. To see the difference between P1, P2, and P3, observe the following:

- *mochiron kekkon wa shitai yo*
- *mochiron nishūkande kekkon nante muri darō yo*
- *mochiron kekkon wa shitai kedo*
- **mochiron nishūkande kekkon nante muri darō kedo*

darō kedo is not allowed. Compare the phonetic form of *muri darō* in the second example with the intended correct form corresponding to the fourth example. Unlike the second, in the fourth one the first two moras of *darō* are pronounced in nF. The two instances of phonetic *darō* are not identical, and it is therefore necessary to see the latter as not the same item but as part of another lexical entry whose phonetic form is *darōkedo*.

	muri darō yo	nF-F-F-F-F-F
*	muri darō kedo	nF-F-F-F-F-F-F
	muri darōkedo	nF-F-nF-nF-F-F-F

The fact that *kedo* is not allowed after *darō* demonstrates the distinction between P3 and P2. *Kedo* cannot appear after *darō* because:

- *kedo* P2\A4
- *darō* C4\A4/P3

Darō needs to generate A4| because it can follow |C3 to constitute a predicate that receives A4 modifications as shown in the second example.

Unlike *kedo,* as illustrated in the examples, *yo* can follow *darō*:

- *yo* P3\P5

In the same syntactic category as *yo* are:

- *tteba* P3\P5
- *ne* P3\P5

(Correction in version 5.0.4): **sa* P3\P5 —> *sa* C2\P5. (Cf. **desu sa*). Also *darōsa* C4\A4/P5 is necessary. I thank @wyoreorm for pointing this out.

To account for *darōkedo* we need:

- *darōkedo* C4\A4/A4

Some other items distributed similarly are:

- *darōkara* C4\A4/A4
- *deshōkedo* C4\A4/A4
- *deshōkara* C4\A4/A4

The difference between P2 and P1 can be found in a similar fashion. Consider the following:

- *daigakusei ni natta karaniwa*
- **daigakusei ni narimashita karaniwa*
- *daigakusei ni natta kara desu*
- **daigakusei ni narimashita kara desu*
- *daigakusei ni natta kara, arubaito o shitemiru no mo ii kato omotte*
- *daigakusei ni narimashita kara, arubaito o shitemiru no mo ii kato omoimashite*

The first thing to point out is that *karaniwa* cannot follow *narimashita* while it can follow *natta*. (Those who feel otherwise may be thinking of some special registers that a tour attendant might use.) This is captured by stating:

- *natta* |P1
- *narimashita* |P2
- *karaniwa* P1\A4

A bit more tricky is the case of *kara* and *kara desu*. Taking *kara* alone, it appears that both *natta* and *narimashita* can precede it. But only when *kara* follows *natta, desu* can appear after *kara*. To see what conditions the distribution of *desu,* notice that in the example "*daigakusei ni narimashita ...*" of this page *narimashita kara* is followed by another clause. The following examples may make the point easier to see:

- *akarichan wa isogashii kara asobenai.*
- *jōjikun wa ashi ga hayai kara moteru.*

The point is that when followed by another clause *kara* is |A4 and when followed by *desu* it is |C2. They are different items:

- *kara* P1\C2
- *kara* P2\A4

Kara P1\C2 cannot follow *narimashita* A4|P2 and *kara* P2\A4 can follow it as indicated by the boundary labels.

Some similar items are:

- *ga* P2\A4
- *kedo* P2\A4
- *shi* P2\A4
- *keredomo* P2\A4
- *kedomo* P2\A4

Around relative clauses (R1, R2, R2b, R2c, R2d)

»nix

Recall first that R2d is an alias of P1. All statements applicable to R2b apply to P1 and vice versa. »humanity

Study the distribution of *noni* and *kara* in the following examples:

- *daigakusei ni natta noni arubaito mo dekinai*
- *daigakusei ni natta kara arubaito ga dekiru*
- *daigakusei na noni arubaito mo dekinai*
- **daigakusei na kara arubaito ga dekiru*
- **daigakusei da noni arubaito mo dekinai*
- *daigakusei da kara arubaito ga dekiru*

Both *noni* and *kara* can appear after *natta* but only *noni* can appear after *na* and only *kara* can appear after *da*. This is because their left boundaries differ:

- *noni* R2b\A4
- *kara* R2d\A4

and

- *da* |R2d
- *na* |R2b
- *natta* A4/R1

As per our naming scheme, for the right boundary of a form to be compatible with both R2b| and R2c|, it has to be named R1. Accordingly, *natta* is |R1. In the previous subsection we have seen that *natta* is |P1, but that is because P1 is another name of R2d, and (again, as per our naming scheme) a |R1 form can behave as |R2d without changing its corresponding phonetic form.

The following are some other examples of items that are R2b|.

- *node* R2b\A4
- *noka* R2b\A4
- *n* R2b\C3
- *no* R2b\C3

The other boundaries that can follow R1 are, besides R1 itself, R2 and R2c. R2c is observed in sentences such as:

- *asatte tenjikai iku nara, watashi mo tsuretette yo*
- *datte, hontōni jugyō ga kyūkō ni natta nara, isogashiku nai desho*

where

- *iku* A4/R1
- *natta* A4\R1
- *nara* R2c\A4

Since R2c is an alias of C4, *nara* R2c\A4 can follow C1, C2, and C3:

- *misakichan no koto nara watashi ga yattoku kara, mō agatte daijōbu da yo*
- *hitori de wa muri demo, sannin nara dekiru*
- *moshi yamadasan ga maketa no ga aite ga miyamotosan datta kara nara, miyamotosan wa yamadasan yori tsuyoi koto ni naru*
- *shutchō ni iku n nara sono maeni katazuke owarasete*

where

- *koto* |C2
- *sannin* |C1
- *kara* |C2
- *n* |C3

R2 is, informally speaking, the left edge of a typical noun. Also, the genitive *no*'s right edge is R2, which is why the genitive *no* can be followed by a noun. The following examples demonstrate these facts:

- *satōkunno apāto*
- *misakichanno obentō*

where

- *satōkunno* ./R2
- *misakichanno* ./R2
- *apāto* R2/C2
- *obentō* R2/C2

The difference between R2 and R1 can be observed with:

- *ka* R1\A1 'whether'

(Cf. **kireina ka* 'whether it is beautiful'; *kawaii ka* 'whether it is cute'; *kireina* is |R2.)

> *ka* is one of the few items that is |A1 that is not an inflectional suffix. The existence of this item, among others, justifies the A1 boundary as a word-level, as opposed to inflectional level, boundary. »journalist

On a related note, the past form of the copula *datta* 'was' is C3\R1 while the present *da* 'is' is C3\P1. Hence:

- *hito* R2\C2
- *suki* A4\C2
- *suki datta hito* 'the person (I) liked'
- **suki da hito* 'the person (I) like'

Around copulas (C1, C2, C2b, C3, C3b, C4)

»pun

Using a copula is one of the strategies to constitute a predicate from a noun and other items that are |C2 or |C3. Copulas include:

- *da* C3\A4/P1 'is'
- *datta* C3\A4/R1 'was'
- *desu* C3\A4/P2 'is'
- *deshita* C3\A4/P2 'was'
- *desuka* C3\A4/P4 'is ... ?'
- *deshitaka* C3\A4/P4 'was ... ?'
- *na* C2\A4/R2b
- *ka* C2\A4/P5 'oh'

Some of them are C3|. These forms can follow *n* ('because'), which is R2b\C3. Items that are C2| cannot follow *n*. This justifies the distinction between C2 and C3.

Correction in version 5.1.2: "Some of them are C3|" not |C3. I thank @ho_hasios for pointing this out.

- *doko ni iku n desuka?* 'Where are (you) going? Lit. Because going where?'
- *aitai hito ga iru n desu.* 'I want to see someone. Lit. Because there is a person (I) want to see.'
- **aitai hito ga iru n na n na n* ... 'because it's because it's because ... I want to see someone'
- **kaimono ni iku n ka* 'oh, because (you're) going shopping?'

Items that have C2 or C1 on the right, such as *kara* P1\C2 'because' and *futari* R2/C1 'two people', can be followed by C3.

- *futari datta kara desuka?* 'Because it was two people?'

C1 is characterized by *no* C1\R2 '(amount/quantity) of' (as in "two of the apples"). It is important not to confuse this with the possessive *no*.

- *futatsu no ringo* 'two pieces of the apples'
- *hitori no gakusei* 'one person of the students'

Some other notable words that have C1 on the left are:

- *wa* C1\A4 'at least'
- *mo* C1\A4 'as much as'
- *nisen'en wa moratta* 'received at least two thousand yen'
- *nisen'en mo tsukatta* 'spent as much as twenty thousand yen'

Although we provisionally treat these *wa* and *mo* as C1|, these items are incompatible with some C1| items. »membrane

C2b, C3b, and C4 are the aliases of A4, P4, and R2c, respectively. This means C1 can freely precede an A4 boundary:

- *gakusei ga sannin aruiteiru* 'three students are walking. lit. students are walking in a group of three'

And C1 and C2 can be followed by P4 (relevant places unitalicized):

- *watashino kaono shiwa ga* suki tte *kare wa itta*. "He said he liked the wrinkles of my face."
- *tsugini kuru no wa* nihonkabu tte *senpai ga itteta yo*. "Our senior said Japanese stocks are booming next."

And R1 can be followed by C4:

- *darō* C4\A4/P3 'I suppose'
- *janaika* C4\A4/P5 'could have been known'
- *nara* C4\A4/A4 'if'
- *ame wa ashitaniwa yamu darō*. "I suppose the rain will stop by tomorrow."
- *kin'yōbi wa isogashii to itta janaika*. "I told you I'm busy on Fridays."
- *zarigani ga shima o tsukutta nante yappari uso janaika*. "You could have known that the story of crawfish creating an island was a lie."
- *sonnani nihon ni ikitai nara hitoride itte*. "If you want to go to Japan so bad go by yourself."

Arguments and adverbial phrases (A1, A2, A3, A4)

Recalling the version 2 terminology, A1 is roughly "nominal" and A4 "adverbial." These two boundaries are characterized by, among others, the following items:

- *ga* A1\A4 (nominative)
- *o* A1\A4 (accusative)

A4 is typically found as the left boundary of the predicate:

- *basu ga kita*. "The bus arrived."
- *fujisan o mita*. "saw Mt. Fuji."

Other items that are |A4 include *wa* and *mo*, but their left boundaries are not A1 as demonstrated by the following examples.

- *Satōsanni* wa *maketaku nai n desu*. "Because I don't want to be defeated by Sato specifically."
- *Kyōto dakede naku, tōkyōde* mo *sakura o miru koto ga dekiru*. "Not only in Kyoto, but in Tokyo too cherry blossoms can be seen."
- *Hikkishiken ga dame de* mo *mada mensetsude bankai suru chansu ga aru*. "Even if you messed up the paper test, you still have a chance to recover on the interview."

- *Hikkishiken ga kore de* wa *ikura mensetsu ga yokutemo goukaku* wa *dekinai*. "If the paper test is this (bad), there's no chance of passing no matter how you do well on the interview."

In the examples above, *satōsanni* ('by Sato') and *tōkyōde* ('in Tokyo') are |A2, and *de* ('given that ... is') is |A3. *Wa* and *mo* can connect to these boundaries (|A2 and |A3) because they are A3|. Another example of an item that has A3 on its left is *koso*:

- *Satōsanni* koso *maketaku nai n desu*. "Because I don't want to be defeated, precisely by Sato."
- *Kyōto dakede naku, tōkyōde* koso *sakura o miru koto ga dekiru*. "Not only in Kyoto, but precisely in Tokyo cherry blossoms can be seen." (This is semantically awkward, albeit not ungrammatical.)
- *Hikkishiken ga dame de* koso *mada mensetsude bankai suru chansu ga aru*. "Precisely because you've messed up the paper test, you still have a chance to recover on the interview." (Again, this is grammatical.)

The difference between |A2 and |A3 is manifested by the distribution of *nara* A2\A4 'if it is'.

- *Satōsanni* nara *maketaku nai n desu*. "Because I don't want to be defeated if it is by Sato."
- *Kyōto dakede naku, tōkyōde* nara *sakura o miru koto ga dekiru*. "Not only in Kyoto, but if it is in Tokyo cherry blossoms can be seen." (This is semantically awkward, albeit not ungrammatical.)
- **Hikkishiken ga dame de nara mada mensetsude bankai suru chansu ga aru*. Intended: *"Given that if it is you've messed up the paper test, you still have a chance to recover on the interview."

Nara A2\A4 can follow *satōsanni* and *tōkyōde* because they are |A2, but not *de* because it is |A3.

To summarize:

- *ga* A1\A4 (nominative)
- *o* A1\A4 (accusative)
- *wa* A3\A4 'specifically'
- *mo* A3\A4 'even'
- *koso* A3\A4 'precisely'
- *nara* A2\A4 'if it is'
- *de* C2\A4/A3 'given that ... is' (literary)

Some other items that have A1, A2, A3, or A4 are:

- *shika* A2\A4 'except' (used in a negative clause)
- *nante* A2\A4 'of all (places, people, things, etc.)'
- *kosowa* A2\A4 'precisely'
- *made* A2\A4 'even ... of all (places, people, things, etc.)'
- *ni* A1\A2 'to'
- *de* C2\A4/A4 'is ... and' (plain)

The ontology of G1, G2, and C1

Finishing Chapter Five is recommended before reading this subsection.

G1 and G2 are closely related to both what has been informally called the verbal nouns and the verbs (d·i, d·ii, d·iii, d·iv, d·v, and d·vi). They are newly recognized boundaries and they or their equivalents used to be misidentified as C1. After some close examinations of phenomena around C1 and *suru*, it was evident that some separate boundaries, now named G1 and G2, were necessary.

To repeat, C1 is a boundary found at, among other places, the right edge of some quantifiers such as *minna* and *futari* (»jeans). |C1 can be followed by *de* C1\A4, among others, as in:

- *kono mafurā wa futari de amimashita.* "The two people (=we) together knitted this scarf."
- *minna de gojira o yattsukemashita.* "Everyone (=we) together beat Godzilla."

Since quantifiers are phonologically similar to nouns, and since the function of the phonetic *no* that follows them is at least superficially similar to the noun suffix *no*, it is easy to mistakingly mix them up. A clear case that demonstrates that |C1 is not a noun boundary is the following. It differs from the noun possessive suffix in the phonetic form:

otoko	otokono
minna	minna no

In another usage, *no* placed at |C1 is placed after a fait accompli, followed immediately by a statement made based on it.

- *sonna fukuzatsuna kin'yūshōhin ga dekita no mo, sore ga toripuruēni kakuzuke sareta no mo, chantoshita gakujutsutekina rironno konkyo ga atte no koto yo.*

(マチネの終わりに.)

which is parsed as:

```
                                R2|
                          ./R2
                    A4/C1
          ./A4
./A1        A1\A4     A4/C1     C1\R2     R2|
... konkyo  ga        atte      no        koto      ...
```

For both quantitative and fait accompli cases, |C1 freely connects to A4|. For this reason, C2b is an alias of A4.

- *pokemon o nihiki tsukamaeta.*
- *huryō ga sannin aruitekita.*
- *sakamotosan ga ōenni haittekurete, yatto dōnika shigoto ga mawaru yōni natta.*

We have seen that |C1 can be made with a verb form. It can also be made with a copula or an adjective.

- *katte mo ii to wa itta kedo, sore wa jūbun* {yasukattara/anka dattara} *no hanashi da yo.*

where

- *yasukattara* A4/C1
- *anka* ./km-ii
- *dattara* C3\A4/C1

Caution must be exercised not to confuse the fait accompli/quantifier *de* C1\A4 with the copular *de* C2\A4/A3 as in:

- *jōrenkyaku ga tatta futari de wa, totemo eigyō o tsuzukerarenai.* "If the regulars are only two people, the business cannot possibly be carried on." (*de* C2\A4/A3)
- *shojikin ga sen'en de mo nomeru mise wa aru.* "Even if what you have is a thousand yen, you can drink at some of the places." (*de* C2\A4/A3)

When the subject of the clause of the copular *de* C2\A4/A3 is phonetically absent, the phonetic distinction can be lost:

- *(shojikin ga) sen'en de mo nomeru mise wa aru.* (*de* C2\A4/A3)
- *sen'en de mo nomeru mise wa aru.* "even with a thousand yen, you can drink at some of the places." (*de* C1\A4)

It is not intended to claim that the "omission" is a phonetic-level process.

It seems reasonable to recognize at least two semantic usages of *de* C1\A4, the fait accompli usage and the quantitative, conditioned by the syntactic or semantic environment.

Although they are called "usages" here, they are likely separate lexical items.

- *sourukara tōkyōe wa sanjikan de ikeru.* (quantitative)
- *sono hanashi suru no, basude de ii?* (fait accompli)

The quantitative usage must be distinguished from the instrumental *de* form of a noun. Assuming that the instrumental *de* in question is C1| requires a zero suffix in the noun paradigm that makes |C1, which, since C2b is an alias of A4, results in a phonetically bare noun being followed by A4|. The problem of this consequence is that A4| when following a phonetic bare noun accompanies a boundary fall but when following |C1 it does not.

- *atama butsuketa.*

In the example above, a boundary fall is inserted between *atama* and *butsuketa*. Compare this with the following quantitative example:

- *hayaku shinaito minna itchau yo.*

where no boundary falls are found.

Hence, by contradiction, the said instrumental *de* is not C1|.

In addition, the fact that *no* following a noun conforms with the phonological rule of **-¦no** suggests this conclusion. »journey

Other notable usages of C1 are noted in the noun and the temporal noun sections. »procrastinator, »reality

Let us now turn to the issues of G1 and G2. G2 is a boundary found, among other places, between *suru* (and its related items) and an immediately preceding certain verb suffixes including *iwa, imo, isae,* and so on. In version 4, today's G2 was equated with C1 without much of a good reason. If anything, G2 at least feels adverb-ish. A native speaker of Japanese can probably see this in the following example:

- *kodomo futario daigaku ni ikaseru tameni jitaku o urisae shita.* (»snapshot for why *futario* not *futari o*.)

But this *shita* does not appear immediately after a |C1 quantifier.

- *shakkin o minna de hensai shita.*
- **shakkin o hensai minna shita.*

The same *shita* appears after what can be—and has been—called verbal nouns, those that are phonologically noun-like and semantically verb-like. *Hensai* in the examples above is one of them.

- *suwabeno chōshu o tantō shita hikara tōkahodo tatta aruhi, okino wa futatabi, mogamino shitsumushitsuni yobareta.*

(検察側の罪人.)

- *karera wa minzokufunsōno heiwatekikaiketsusaku o miidasōto kenmeini doryoku shita.*

(小学館日韓韓日辞典.)

In the examples above, *tantō* and *doryoku* are what we call verbal nouns. Usual adverbs do not intervene between these and *shita*, but *sae, mo,* and *wa* (and some others) can intervene.

- *yoku tantō shita.*
- **tantō yoku shita.*
- *tantō sae shita.*
- *chōkikan doryoku shita.*
- **doryoku chōkikan shita.*
- *doryoku wa shita.*

This is analogous to the cases where zeros can be replaced with non-zero items without changing the right boundary of the phrase. (»humanity.) Hence, the same strategy—set multiple boundaries whose gap can be filled with zeros—can be employed and G1 and G2 are separately postulated. Accordingly,

- *wa* G1\G2
- *mo* G1\G2
- *sae* G1\G2
- *shita* G2|R1

Some may find it necessary to admit an additional suffix *o* to account for when an instance of the accusative particle *o* appears in one place for the semantic object of the verbal noun and another appears between the verbal noun itself and the following *suru* (**su**ˈ**ru**) or a related item. Such a construction is usually considered ungrammatical (Masuoka & Takubo 1992, p. 20), but it has been claimed that in real-life speeches they are repeatedly found (Kurosawa 2008). The situation will be properly described if we assume in those who accept the said construction's lexicon some forms that semantically are verbal noun-like, phonetically have *o* at the end, and syntactically are G2 on its right. That is:

- *onegaio* |G2
- *kenkyūo* |G2

The exact syntactic nature of *shita* still needs to be investigated in some more detail to be determined. Before doing so, some investigation on G1 would be worthwhile.

First, recognize that *deki* and *shita* shows some distributional similarity:

- *konankun wa kuruma o unten dekita.*
- *konankun wa kuruma o unten shita.*

Their difference however becomes apparent when a G1|G2 item intervenes:

- **konankun wa kuruma o unten mo dekita.*
- *konankun wa kuruma o unten mo shita.*

This suggests that *deki* is G1|.

Now, notice that replacing *o* with *no* makes the first example well-formed:

- *konankun wa kurumano unten mo dekita.*
- *konankun wa kurumano unten mo shita.*

The fact that *unten* is preceded by *kurumano* suggests that *unten* there is somehow R2|.

Simply stating that *unten* can be R2| does not explain the unacceptability of the following:

- **konankun wa kuruma o jōzuna unten mo shita.*

as opposed to:

- *konankun wa kuruma o jōzuni unten mo shita.*

Compare:

- *konankun wa kurumano jōzuna unten mo dekita.*
- *konankun wa kurumano jōzuna unten mo shita.*

Suppose now that *jōzuna unten mo shita* (this is wellformed if uttered in isolation) is somehow R2| so that *kuruma o* .|A4 cannot precede it. This presupposition is immediately falsified by:

- *nandomo jōzuna unten mo shita.*
- *konankun wa jōzuna unten mo shita.*

It seems that our boundary-labeling approach cannot efficiently handle the situation. This seems to be one of the cases where we need to approach acceptability from the point of view of semantics. More on this will be discussed in p. surname.

Recall that phonetically identical *mo, sae, wa* are available along the A series boundaries, as in:

- *konankun wa kuruma mo motte iru.*
- *konankun wa kuruma sae motte iru.*
- *konankun wa kuruma wa motte iru.*

where

- *mo* A3\A4
- *sae* A2\A4
- *wa* A3\A4

We can then suppose that the phonetic *mo dekita* is allowed after |A3 but not after |G1. In fact, it is not difficult to observe that *unten* when preceded by |R2 behaves as a noun (|m-ii), which is shown partly by:

- *konankun wa kurumano untenkara sesunano sōjūmade dekita.*
- *kurumano untenno jōzuna konankun.*

where *untenkara* is pronounced exactly like **untenn˛-ka˙ra**. Of course, *kara* is not allowed when *unten* is preceded by *o*:

- **konankun wa kuruma o untenkara sesuna o sōjūmade dekita.*

Putting our observations so far together, it seems that the lexicon has a zero that turns a |G1 item into a noun (R2/m-ii, to be more specific):

- [0] G1\R2/m-ii

and that typical |G1 items such as *unten* and *sōjū* are lexically ./G1 since they do not allow a preceding |R2 unless they are turned into a noun:

- *unten* ./G1
- *sōjū* ./G1

That they are not A4/G1 is suggested by:

- **konankun wa kurumano jōzuni unten shita.*
- **konankun wa sesunano sukoshi sōjū dekita.*

The remaining questions are as to how *deki* and *shita* are lexically characterized. We have already seen that *deki* is G1| and *shita* is G2|. One of the questions is how they can follow |A4. Since *unten* and *sōjū* are ./G1, *deki* and *shita* must generate A4| as a result of merging. That is:

- *shita* G2\A4/R1

The necessity of *deki* being G1\A4/d·i is not obvious since as long as it is a verb (|d·i) it has a chance to obtain A4| through a merge with an inflectional suffix. For now, I suggest the following:

- *deki* G1\d·i

Note that it cannot be rightward merging since merging breaks the G1 boundary.

Finally, let us confirm that a verbal noun merges rightward before it merges with an *o*-phrase in sentences such as *konankun wa kuruma o unten shinakatta*.

	Konankun wa	kuruma o	unten	shinakatta.
*	Konankun wa	kuruma o	unten	Shinakatta.
*	Konankun wa	kuruma o	Unten	shinakatta.
*	Konankun wa	kuruma o	Unten	Shinakatta.
	Konankun wa	Kuruma o	unten	shinakatta.
*	Konankun wa	Kuruma o	unten	Shinakatta.
	Konankun wa	Kuruma o	Unten	shinakatta.
	Konankun wa	Kuruma o	Unten	Shinakatta.

This confirms our assumption as from this observation the triangle constraint suggests the following structure:

konankun wa/	kuruma o/	unten/	shinakatta\\\	

Our analysis now allows us to visually express how verbal nouns work in a sentence:

A4/R1			
	A4/R1		
		A4/R1	
./A4	./A4	./G1	G2\A4/R1
konankun wa	kuruma o	unten	shinakatta

More tree diagrams of some of the sentences and nonsentences we have seen in this subsection follows:

Terminals are not necessarily tokens.

- *konankun wa kuruma o unten dekita.*

```
A4/R1
                A4/R1
                          A4/R1
                      ./d·i
./A4        ./A4      ./G1      G1\d·i    d·i\A4/R1
konankun wa kuruma o  unten     deki      ta
```

- *konankun wa kuruma o unten shita.*

```
A4/R1
                A4/R1
                          A4/R1
./A4        ./A4      ./G1      G2\A4/R1
konankun wa kuruma o  unten     shita
```

- **konankun wa kuruma o unten mo dekita.*

```
                          *
                      ./G2
./A4        ./A4      ./G1      G1\G2     G1\d·i
konankun wa kuruma o  unten     mo        deki  ...
```

Illformed since ./G2 and G1\A4/R1 cannot merge.

- *konankun wa kuruma o unten mo shita.*

```
A4/R1
                A4/R1
                          A4/R1
                      ./G2
./A4        ./A4      ./G1      G1\G2     G2\A4/R1
konankun wa kuruma o  unten     mo        shita
```

- *kurumano untenno jōzuna konankun.*

```
                R2/m-ii
            A4/R2
        ./A4
    R2/m-ii
./R2        R2/m-ii   \A4       A4/R2     R2/m-ii
kurumano    unten     no        jōzuna    konankun
```

- **konankun wa kurumano jōzuni unten shita*

```
                *
                          A4/R1
                                  A4/R1
./A4        ./R2      A4/A4     ./G1      G2\A4/R1
konankun wa kurumano  jōzuni    unten     shita
```

Illformed since ./R2 and A4/R1 cannot merge.

- *konankun wa kurumano unten mo dekita*

A4/R1							
			A4/R1				
		./A4					
	./A1						
	R2/m-ii						
		R2/m-ii				A4/R1	
./A4	./R2	./G1	G1\R2/m-ii	m-i\A1	A3\A4	./d·i	d·i\A4/R1
konankun wa	kurumano	unten	[0]	[0]	mo	deki	ta

Observe how *unten* needs a zero to acquire R2/m-ii.

- *konankun wa kurumano untenkara sesunano sōjūmade dekita*

A4/R1								
		A4/R1						
					A4/R1			
	./A4				./A4			
	R2/m-ii				R2/m-ii		A4/R1	
./A4	./R2	R2/m-ii	m-i\A4	./R2	R2/m-ii	m-i\A4	./d·i	d·i\A4/R1
konankun wa	kurumano	unten	kara	sesunano	sōjū	made	deki	ta

A map of the boundary system

```
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                
```

A summary of the boundary system. [sıka>, with [aligned to A2 and > aligned to A4, indicates **sıka** A2|A4. <sı], with < aligned to A4 and] aligned to P2, indicates **sı** P2|A4. Not all items or boundaries are listed. Errors are expected. Note: On 5.2.0, **sıka** C1|A4 was removed from the map, and copulas misplaced in C2| were fixed and moved to C3|.

Chapter Five

INFLECTION

PRELIMINARY REMARKS ON INFLECTION

On the (non)nature of the inflectional boundaries

By dividing a sentence in the way we have seen in the previous chapter, the analysis eventually reaches a lexical form, a syntactic form that is indivisible without breaking the syntactic coherency.

Some of these lexical forms fall into the classes of the traditional parts of speech, such as adjectives, verbs, and, nouns. As part of the effort of organizing and presenting the grammatical description in a way that is comfortable for myself and for those who are familiar with the traditional grammatical terminology, select boundaries are classed into *inflection* groups and are given familiar names: the *verb* boundaries, the *adjective* boundaries, the *noun* boundaries, the *adjectival noun* (roughly corresponds to *keiyōdōshi*) boundaries, etc.

Deciding which boundaries are word-level and which are inflectional was necessary also for the sake of spelling aesthetics. Nobody would want to put a space at every syntactic boundary. This pertains to the parseability principle; with dozens of grammatical boundary types, having some means to distinguish certain groups of them from the rest seemed useful.

Hence, the distinction between the word-level boundaries and the inflectional ones is arbitrary and has no real significance in the grammar. It is a matter of presentation. Select hand-picked boundaries are called *inflectional* so that the terminology feels friendly and is convenient for designing spelling rules. All else are *word*-level.

Which boundaries are inflectional?

As a native speaker of Japanese, I could not help but feel that certain boundaries were more “inflection-ish” than the others. The boundary found on the left side of the adjective ending *i*, found in such forms as *samui* and *yasashii*, is one of the inflection-ish boundaries. I believe most Japanese speakers feel the same.

The following are some items whose left side is one of the more “inflection-ish” boundaries:

- *ku* as in *yasashiku* and *samuku*
- *te* as in *tabete* and *aruite*
- *ru* as in *taberu* and *tsumeru*
- *sō* as in *kireisō* and *taihensō*

What are common among (some of) the inflection-ish boundaries? What contributes to the naive impression that they are inflection-ish?

Syntactically, these boundaries tend to be more "internal" than some of the less inflection-ish boundaries. We will later see that there is reason to believe the middle item *yasashi* as in *sugoku yasashii* merges rightward (*yasashi* + *i*) first and then the resulting phrase merges leftward (*sugoku* + *yasashii*). The left boundary of *i* is more internal than the left boundary of *yasashi*.

Phonologically, the inflectional boundaries are seldom (or, perhaps, never) aligned to an AP boundary. None of the items listed above as "inflection-ish" are disallowed to follow another item within the same AP as demonstrated by *yasashiku wa*, *tabete no*, *taberu kara,* and *taihensō da*. This amounts to saying that they tend to be phonologically more internal than some other boundaries.

Another factor is that they tend to allow otherwise rare phonological operations, namely, segmental alternations and deletions and accent shifts.

Additionally, the less inflection-ish boundaries feel less "contained" in one paradigm. *Wa* and *mo,* both of which feel very word-ish, can follow nouns, verbs, adjectives, adjective nouns, and others. They can finish almost any kind of item's AP. This is in contrast to more inflection-ish *i,* which, if appears in an AP, the AP begins mostly with an adjective. (Some inflectional suffixes "change" the inflection group of the stem.)

Based on these observations, it was decided that preferably 1. If a boundary allows segmental alternations or deletions or accent shifts or paring, it is an inflectional boundary; 2. If a boundary is inflectional, all boundaries that are more "internal" are also inflectional; 3. If a boundary can follow forms of multiple inflectional paradigms, it is a word boundary; 4. The stem of an inflectional paradigm must have at least one inherently corresponding nF mora.

Although these criteria are by no means strict rules, what it means by "internal" in criterion 2 is worth some clarification.

- The boundary at which a merge occurs first is more internal.

For the structure A/ B\/ C\, the boundary between A and B is more internal.

For the structure A/ B/ C\\, the boundary between A and BC is less internal.

- Two boundaries are not compared in internality if a word boundary intervenes.

For the structure A/ B\/ C\, if the boundary between AB and C is a word boundary, the boundaries between A and B and the right or left of ABC are not compared in internality.

- In an AP, the boundary closer to the left edge but not the left edge itself is more internal.

~~The major boundaries seem to always coincide with an AP boundary~~. (Note added in version 5.1.1: this is likely wrong. We still need to account for the non-copular *desu* which seem to be A4|, as in *darega desuka? — watashi ga desu*. Ver. 5.2.0: »oak-obeche)

Nevertheless, factors considered when deciding certain boundaries to be inflectional or word-level do not bind future decisions. Yet they are likely to be referenced when making decisions.

Kyouro's spelling rules are designed so that spaces and accents in combination show AP boundaries in the sentence unambiguously. »echo.

Spelling design and boundary notation

Inflectional boundaries are graphically marked with special symbols, the middle dot (·) or the hyphen-minus (-), inserted into tokens with an inflectional boundary at its left edge.

Inflectional boundaries are further classed into the *noun boundaries (m-ii* and *m-i),* the *temporal noun boundaries (tm-ii* and *tm-i),* the *adjectival noun boundaries (km-iii, km-ii,* and *km-i),* the *verb boundaries (d·vi, d·v, d·iv, d·iii, d·ii,* and *d·i),* and the *adjective boundaries (k·ii* and *k·i).*

The Roman numeral given to the name of a boundary assumes the same kind of relationship found in the word boundaries but in the opposite direction: an inflectional boundary with a greater number can freely precede an inflectional boundary with a smaller number as long as the prefixes are the same.

Inflectional boundaries do not have aliases for indicating zeros. We will later see that a zero [0] tm-ii|m-i is reasonably supposed, but m-i is not called by other names.

Semantic features

In this chapter, some suffixes are introduced with what can be regarded as semantic features with binary values. If a form is indicated as [+ formal], the form carries certain grammatical formality. Likewise, [+ past] indicates what can be loosely called the past tense. These features are used to characterize competing suffixes within the same paradigm. The semantic substances that they should express are yet to be defined. They are used as convenient organizing devices. As the semantic component of the theory advances these tentative features will hopefully be replaced or substantiated.

- [formal] or frm
- [habitual] or hab
- [past] or pst
- [irrealis] or irr
- [negative] or neg
- [literary] or lit
- [interrogative] or int
- [imperative] or imp
- [complement] or cmp
- [collective] or col
- [possessive] or pos
- [ablative] or abl

In each context, only those features that are useful to organize the paradigm will be shown.

ADJECTIVES

Initial observations

Two syntactic boundaries are classed as adjective boundaries: k·i and k·ii. Forms that are |k·i or |k·ii are called adjectives.

Although a form's real-world meaning does not determine the boundary type, as a tendency, the meaning of an adjective relates to states and feelings.

The lemma of an adjective contains the *i* suffix. This suffix makes a predicate and the resulting predicate takes an argument (usually the subject) and optional modifiers.

- *soto ga sugoku samui*. "It is so cold outside."

The past tense is made by adding the suffix *katta*.

- *soto ga sugoku samukatta*. "It was so cold outside."

Some adjective suffixes have phonetic *desu*, but this should not be confused with the copular *desu*, which is C3\A4/P2.

- *soto ga sugoku samuidesu*. "It is so cold outside." (formal)

The adverbial form takes the *ku* suffix. *Kute* also appears. They slightly differ semantically.

- *samukute soto ni derarenai*. "It is cold and I cannot go out."
- *michihaba ga semaku, ōkina kuruma wa tōrenai*. "The road is narrow and a big car cannot pass through."

In the "basic" usage the adjective functions as a predicate, but when followed by the suffix *sa*, it functions as a noun.

- *tōkyōno natsu wa atsusa ga kibishii*.

When it functions like a noun, it can receive a nominal modifier.

- *kibishii atsusa ni taerarenakatta*.

The *garu* form functions as a verb.

- *kodomo ga amarini itagaru kara kusuri o nomaseta*.

The *sō* form functions as an adjectival noun.

- *oishisōna aisukurīmu*.

Main clause

Some of the suffixes that can follow k·i are the following:

		int	frm	pst	neg
ˈ(_)·ı	k·i\A4/R1	–	–	–	–
ˈ(_)·ıdesu	k·i\A4/P2	–	+	–	–
ˈ(_)·ıdesuka	k·i\A4/P4	+	+	–	–
ˌ(_)·katta	k·i\A4/R1	–	–	+	–
ˌ(_)·kattadesu	k·i\A4/P2	–	+	+	–
ˌ(_)·kattadesuka	k·i\A4/P4	+	+	+	–

These items can constitute a main clause predicate. The semantic features [interrogative], [formal], and [past] represent the clause types. For example, ˈ**(_)·ıdesu** makes a predicate that is declarative, formal, and nonpast.

In the list, items that are k·i\A4/R1 are generally given [– interrogative]. Since a main clause predicate that these items make can be turned interrogative just by adding a corresponding predicative intonation (R%), it is reasonable to assume they are by default [– interrogative] (or underspecified) and the value [+ interrogative] is given by the intonation.

The [+ negative] forms are constituted with the "bridging" *ku* k·i\A3:

		int	frm	pst	neg	lit
·ˈku	k·i\A3	–	–	–	–	–

followed by:

		int	frm	pst	neg	lit
ˈnaı	A4\R1	–	–	–	+	–
ˈnaıdesu	A4\P2	–	+	–	+	–
arımaˈsenn	A4\P2	–	+	–	+	+
ˈnaıdesuka	A4\P4	+	+	–	+	–
arımaˈsennka	A4\P4	+	+	–	+	+
ˈnakattadesu	A4\P2	–	+	+	+	–
arımaˈsenndesıta	A4\P2	–	+	+	+	+
ˈnakattadesuka	A4\P4	+	+	+	+	–
arımaˈsenndesıtaka	A4\P4	+	+	+	+	+

Those that are [+ formal, + negative] have two different phonetic forms for each. The *-masen* series feels slightly more literary than the *-desu* series, so the *-masen* series are marked as [+ literary] and the *-desu* series [– literary].

Unlike these negators, the bridging *ku* k·i\A3 does not generate A4| when merging. Hence, any |A4 forms that are semantically "connected" to the main adjective that *ku* is attached to must merge after the negator merges:

- *kyō wa* atsuku *mo naku,* samuku *mo nai, chōdoii kion da.*

```
A4/R1
                    A4/R1
          A4/A4                A4/R1
./A4    ./A4      A4\A4     ./A4            A4\R1
kyō wa  atsuku mo  naku     samuku mo       nai   ...
```

- *keishikishugi ga suki na hito wa* sukunaku *wa nai ga, ōi wake demo nai.*

```
A4/R1
                    A4/R1
./A4            ./A4            A4\R1
... hito wa     sukunaku wa     nai
```

That *kyō wa* does not merge with *atsuku (mo)* is shown by:

```
| *  Kyō wa    atsuku mo  Naku     Samuku mo     Nai     |
```

Similarly, that *... hito wa* does not merge with *sukunaku (wa)* is shown by:

```
| *  ... hito wa   sukunaku wa   Nai                    |
```

Subordinate clause

The same k·i boundary can constitute a subordinate clause. The suffixes that constitute a subordinate clause predicate include:

		hab	pst	irr	lit
ˌ(_)·kereba	k·i\A4/C1	–	–	+	–
ˌ(_)·kute	k·i\A4/A4	–	+	–	–
·ku	k·i\A4/A4	–	+	–	+
ˌ(_)·kattara	k·i\A4/C1	–	+	+	–
ˈ(_)·ɪto	k·i\A4/A4	+	–	+	–
ˈ(_)·ɪˌto	k·i\A4/A4	+	–	+	–

In p. jeans, I listed *wa* and *mo* as C1|, but these two forms do not follow *kereba* k·i\A4/C1 and *kattara* k·i\A4/C1. This is one of the "unpatched defects" (»surname) of the current system. This applies to some verb suffixes listed in p. northwest. Same may apply to many other items that are currently labeled to have C1 on either side.

What is described with the [past] feature may need some attention here. Consider the following pairs of sentences.

- *oishikattara nandemo ii tte wake ja nai.* "It's not like anything will be okay {once it turns out to be/if it has turned out to be} delicious."
- *oishikereba nandemo ii tte wake ja nai.* "It's not like anything delicious is okay."
- *moshi watashino hōga umakereba, watashi ga rīdā ni natte ii?* "Can I be the leader if I do it better (than you)?"
- *moshi watashino hōga umakattara, watashi ga rīdā ni natte ii?* "Can I be the leader if it has been shown that I do it better (than you)?"

In these examples, *-kattara* is used with the assumption that there is some particular event or time that the fact in question is revealed, based on which what is stated in the main clause then holds, and *-kereba* is used to refer to some law present in the context without implying a specific event or time.

Some may question the validity of the use of the feature [past] here. As I stated earlier, the semantic features are used as organizing devices and the exact semantic content corresponding to each value of each feature is not given.

Regardless, it is interesting that the temporal relations between the *-katta* clause and the main clause are relative past and present, respectively, since the same relative tense phenomenon can be observed in relative clauses such as:

- *daigakuni nyūgaku suru toki wa kurō shita.*
- *daigakuni nyūgaku shita toki wa kurō shita.*

When referring to a particular event expected before what is stated in the main clause can happen, *kereba* sounds awkward:

- *tsurakattara itte.* "let me know if you are uncomfortable."
- *tsurakereba itte.* "let me know if you ever happen to be uncomfortable"

The former expects that the agent may be uncomfortable. The latter is stated as if it is a conditional rule that the agent is asked to comply with, without referring to any particular event where the condition holds true.

For [– irrealis], a somewhat similar contrast between [– past] and [+ past] seems to be available between *kute* and *i kara*:

- *sono atarashii gēmuki wa takakute zenzen urenakatta.* "The new game console didn't sell at all because it had turned out to be expensive."
- *sono atarashii gēmuki wa takai kara zenzen urenakatta.* "The new game console didn't sell at all because it was expensive."

As is obvious, *katta kara* is available as the "past" counterpart of *i kara,* and it is hardly synonymous with *kute:*

- *sono atarashii gēmuki wa takakatta kara zenzen urenakatta.* "The new game console didn't sell at all because it had been expensive."

The feature [habitual] is used here to capture the difference between *ito* and *kereba*. The value [+ habitual] assigned to *ito* but not to *kereba* reflects the difference in generality of the law in question. *Ito* is typically used to state a constant law:

- *suiminjikan ga mijikaito byōki ni naru.* "Short sleeps make you sick."
- *kuraito michini mayoiyasui.* "A dark path is hard to follow."
- *heya ga hiroito nanikato benri.* "A spacious room is convenient."

Kereba [– habitual] is often used to refer to a conditional rule provisionally claimed by the speaker.

- *omoshirokereba shichōritsu wa toreru.*
- *toshi o tottemo kimochi ga wakakereba nandemo dekiru.*
- *okane ga hoshikereba baito demo sureba ii daro.*

[+ negative] subordinate predicates are constituted with the same *ku* k·i\A3 used in the main clause, followed by:

		hab	pst	irr	lit	neg
ˈnakereba	A4\C1	–	–	+	–	+
ˈnakute	A4\A4	–	+	–	–	+
ˈnaku	A4\A4	–	+	–	+	+
ˈnakattara	A4\C1	–	+	+	–	+
ˈnaıto	A4\A4	+	–	+	–	+

The bridging *ku* k·i\A3 has also a negative counterpart:

		int	frm	pst	neg	lit
ˈnaku	A4\./A3	–	–	–	+	–

A4\./A3 instead of A4/A3 indicates that it is obligatorily leftward-merging and when it merges it generates ./A3.

This ensures negation is recursively applicable.

- *kyō wa samuku nai.*
- *kyō wa samuku naku nai.*
- *kyō wa samuku naku naku nai.*
- *kyō wa samuku naku naku naku …*

The “resultative” usage

The “resultative” suffix *ku* k·i\A4/A2 functions like:

- *kare yori umaku wa narenakatta.*

```
                    A4/R1
              ./A4
        A4/A2
   ./A3                A4/A2
./A1    A1\A3    ./k·ii    k·i\A4/A2  A3\A4    A4\R1
kare    yori     uma       -ku        wa      narenakatta
```

That *narenakatta* merges last (that it is <u>kareyori/ umakuwa\/ narenakatta\</u> and not *<u>kareyori/ umakuwa/ narenakatta\\</u>) is shown by:

| Kare yori umaku wa Narenakatta |

It has been called “resultative” because, typically, it appears in a sentence in which the adjective describes a result of a change.

- ˈ**samu·ku** ˈ**natt·a.** “became cold.”
- **oo**ˈ**kı·ku sı**ˈ**ta.** “made bigger.”
- **oo**ˈ**kı·ku** ˈ**mıėt·a.** “appeared big.”

Forms that can sit under the phrase of the resultative *ku* seem to be limited. From my quick introspection, it seems that *yori* phrases are one of the few types.

- *asa ga samuku natta.*

| * Asa ga samuku Natta |

The R positions suggest:

```
  A4/R1
              A4/R1
./A4     A4/A2     A4\R1
asa ga   samuku    natta
```

Interestingly, replacing *asa* with body parts seemingly changes the structure:

- *ashi ga samuku natta.*

```
|  Ashi ga     samuku     Natta  |
```

```
        A4/R1
  A4/A2
./A4     A4/A2     A4\R1
ashi ga  samuku    natta
```

These two usages and their differing structures can be captured by listing two *samu*'s in the lexicon, one for holistic sensations and the other for sensations of specific body parts. More on these will be discussed later in this section. »nationalist

It appears therefore that some adjectives have an extra "spot" that accommodates certain types of |A4. *Hoshii* is one of these:

- *akemi wa furūto ga hoshiku wa naranakatta.*

```
|  Akemi wa   Furūto ga   hoshiku wa      Naranakatta  |
```

```
  A4/R1
                          A4/R1
                     ./A4
                A4/A2
./A4          ./A4        A4/A2     A3\A4     A4\R1
akemi wa      furūto ga   hoshiku   wa        naranakatta
```

and, among others, *itai*:

- *onaka ga itaku naranai yōni, yoku kande tabeta.*

```
|  Onaka ga   itaku Naranai    yōni  Yoku Kande     tabeta  |
```

```
                ...
            ./A4
        A4/R1
  A4/A2
./A4          A4/A2     A4\R1     R2b\A4
onaka ga      itaku     naranai   yōni      ...
```

The negative form of the resultative usage accompanies *naku* A4\./A2 as in:

- *kareno taido ga itsumohodo yasashiku naku omoeta.*

When the semantic condition is such that *naku natta* would appear, instead, *nakunatta* tends to be used:

- *suki datta rāmen ga rinyūarude oishiku nakunatta.*

This is likely best analyzed as *nakunar* ./d·ii and *ta* d·i\A4/R1.

The ku aru form

Sometimes [– negative] items are used in place of the aforementioned [+ negative] series.

This happens often when *ku* is followed by *wa, mo, koso,* etc.

- *saishinkishuhodo de wa nai ga, soredemo takaku wa aru.*
- *chīsaku mo aru shi ōkiku mo aru.*

It is usually awkward to use *ku aru* without inserting *wa, mo,* or such in between, but it is grammatical:

- *chīsaku aru shi ōkiku mo aru.*

In the main clause the following items are used:

		imp	int	frm	pst	neg	lit
ˈaru	A4\R1	–	–	–	–	+	–
arıˈmasu	A4\P2	–	–	+	–	+	–
arıˈmasuka	A4\P4	–	+	+	–	+	–
arıˈmasıta	A4\P2	–	–	+	+	+	–
arıˈmasıtaka	A4\P4	–	+	+	+	+	–
ˈare	A4\P4	+	–	–	–	–	–

In a subordinate clause the following items are used:

		hab	pst	irr	lit	neg
ˈareba	A4\C1	–	–	+	–	+
ˈatte	A4\A4	–	+	–	–	+
ˈarı	A4\A4	–	+	–	+	+
ˈattara	A4\C1	–	+	+	–	+
ˈaruto	A4\A4	+	–	+	–	+

The “adjunct” ku is not an adjective suffix

It seems as if some adjectives could take *ku* k·iii\A4/A2, where k·iii is a temporarily supposed boundary for discussion, as in:

- *te o hageshiku furimawashita.*
- *yasashiku hohoenda.*

This ostensible suffix *ku* appears only on limited adjectives. (Which is why a new label would be necessary.)

- **te o ureshiku furimawashita.*
- **modokashiku hohoenda.*

These forms can be further modified by an adverb:

- *te o sugoi hageshiku furimawashita.*

	Te o	Sugoi	hageshiku	Furimawashita.
*	Te o	sugoi	Hageshiku	Furimawashita.

A4/R1			
		A4/R1	
	A4/A2		
./A4	./A4	A4/A2	A4/R1
te o	sugoi	hageshiku	furimawashita

Until version 4, this ostensible suffix was intentionally undistinguished from another *ku* |A2 due to the identical syntactic right boundaries. For version 5, the decision was made to recognize them separately to reflect the semantic and distributional differences.

The rationale for not distinguishing the two forms is documented briefly in 『形容詞について 4.1』, originally prepared for version 3.0, which however stayed local until December 2023, referenced only internally.

If we were to recognize *ku* k·iii|, however, the problem is that some would-be |k·iii adjectives then would be incompatible with k·ii|, despite having a greater Roman numeral, as shown by:

- *ureshigaru*
- **hageshigaru*

Switching the statuses (the present k·ii to k·iii and the supposed k·iii to k·ii) will not solve the problem:

- **ureshiku furimawashita.*
- *hageshiku furimawashita.*

The choice was then between setting up a neutral, non-k boundary and regarding these adjunct forms as unit items. As of version 5, we choose the latter, considering:

- The adjectives are a closed class and notoriously scarce; the lexical burden inflicted by treating the adjunct forms as unit items is, although not necessarily very small, limited.
- While preserving the spelling with no added cost would be the best choice if attainable, simplifying the spelling at the said cost on the lexicon is not the worst either and is reasonable.

Hence, the first two examples are now written as:

- **ˈte wo haˈgesıku hurımaˈwasıt·a.**
- **yasasıˈku hohoˈènnd·a.**

The quantifier *ōku* modifying the “existential” *aru* is a case of an adjective-like adjunct:

- *kyonenno harowinde wa yotta wakamono ga jidōshato sesshoku suru jiko ga ōku arimashita.*

A4/P2		
	A4/P2	
./A4	|A3	A4/P2
jiko ga	ōku	arimashita

Suffixes that "derive" forms of other paradigms

The suffix *sō* k·i\km-i turns an adjective into an adjectival noun:

- *terashimakun ga tsukutta kēki wa totemo mazusō datta.*
- *sore na noni, takasō na osaifu o misebirakasarete, mukatsuite, yatchatta.*

kawaii does not take *sō*. *Kawaisō* is "pitiable" not "seems to be cute." Accordingly, **kawaˈii, kawaˈikute, kawaˈikatta,** etc., are written as unit items.

The negative counterpart of *sō* is given by:

- *nasasō* A4|km-i

which follows *ku*.

- *zenzen tanoshiku nasasō.*
- *konaida okutte kureta sanpuru waruku nasasō datta yo.*

The suffix *sōni* k·i\A4/A2 is necessary because the adjectival noun making *sō* is |k-i and the adjectival noun suffix *ni* is k-ii|.

- *anmari oishisōni taberu kara kotchi made onaka ga suitekita.*

The negative counterpart of *sōni* is *nasasōni* A4|A2.

- *miyazakikun wa jugyōchū wa itsumo tanoshiku nasasōni shite iru.*

The suffix *sa* k·i\R2/m-ii turns an adjective into a noun.

- *munashisade karada ga ugokanaku natta.*
- *yukino omosade yane ga tsubureteshimatta.*

As *ōkii* does not take *sa,* **ooˈkıı, ˈookıku, ooˈkıkatta, ookısaˈ,** etc., are written as unit items.

When the host of *sa* is a derived adjective of a verb, such as *tabeta,* the resulting phrase such as *tabetasa* makes an interesting (though it may be colloquial) structure:

- *amaimono o tabetasa ga gaman dekinakute daietto ga obotsukanai.*

```
                ./A4
          R2/m-ii
   A4/k·ii
./A4          A4/k·ii     k·i\R2/m-ii      A1\A4
amaimono o    tabeta      sa               ga    ...
```

A premise of the labels in the tree above is that those derived *ta* forms are A4|k·ii while the typical lexical adjectives are ./k·ii or ./k·i. Some forms, those that are conventionally treated as "irregular," are, since in Kyouro they must be treated as tokens, lexically A4|k·ii. They include:

- *shita* A4/k·ii "want to do"
- *kita* A4/k·ii "want to come"
- *ikita* A4/k·ii "want to go"
- *toita* A4/k·ii "want to ask"

The [– negative] counterpart of *sa* is *nasa:*

- **·ku˵ˈnasa** k·i\R2/m-ii

as in:

- *yumiko wa jibunno yasashikunasa ga yurusenakatta.*
- *renshū o yaritakunasani taerarenaku natte, gasshuku o nigedashita.*

This is a colloquial term and some speakers/teachers may simply dismiss it as a nonword.

It would be worth drawing a tree for the latter as it has the same kind of interesting structure as the aforementioned case:

```
                ./A2
          R2/m-ii
   A4/k·ii
./A4            A4/k·ii     k·i\R2/m-ii      m·i\A2
renshū o        yarita      kunasa           ni    ...
```

The suffix *sugi* k·i\d·i turns an adjective into a verb.

- *hoshii hon ga atta kedo, takasugite kaenakatta.*
- *naraken wa aruite ikuniwa tōsugiru.*

The suffix *gar* k·ii\d·v turns an adjective into a verb.

- *tsumano shi o itsumademo kanashigatte bakariwa irarenai.*

(A Handbook of Japanese Grammar Patterns for Teacher's and Learner's.)

- *kodomo ga omocha o hoshigatte jibetani suwarikonde naite ita.*

(A Handbook of Japanese Grammar Patterns for Teacher's and Learner's.)

Author's note to self: the *garu* construction seems to be |d·v as opposed to |d·vi, but unlike typical |d·v forms, this accepts an o-phrase.

As shown, the resulting verb accepts an o-phrase.

```
   A4/A4
                    A4/A4
./A4            ./A4        A4/A4
kodomo ga       omocha o    hoshigatte ...
```

This property of the *gar* suffix reveals an interesting nature of *atsui*. As is commonly known, this adjective can refer to a specific object that is hot or a spatial or atmospheric condition (room temperature, climate, weather, etc) that is hot:

- *kōhī ga atsukute nomenai.*
- *shinshitsu ga atsukute nemurenai.*

These two usages may seem to be realizations of an ambiguous meaning of the word, but that they are distinct syntactically is revealed when *gar* is attached:

- *kōhī o atsugatte nomanai.*
- **shinshitsu o atsugatte nemuranai.* Intended: "She won't sleep as she complains about how she feels hot in the bedroom"

A phonetic equivalent of the latter is an acceptable sentence when the sentence refers to the bedroom as an object as opposed to a condition.

- *#shinshitsu o atsugatte nemuranai.* “She won’t sleep as she complains about how the bedroom is too hot (to touch/to go in/ etc).”

From these observations, it appears that *atsui* is actually at least two homophonic adjectives:

- *atsu* ./k·ii “is hot (of an object)”
- *atsu* ./k·i “is hot (as a holistic sensation about the room temperature, weather, etc)”

The latter is judged here as |k·i to reflect the fact that the *garu* form behaves differently from the |k·ii adjectives. As a result, *atsugar* must be admitted as a simple verb:

- *atsugar* ./d·v

The aforementioned two usages of *samu* can receive a similar treatment. *samu* ./k·ii, *samu* ./k·i and *samugar*./d·v can be listed in the lexicon.

- *kodomo ga te o samugaru kara tebukuro o katte ageta.*

A4/R1					
		A4/R1			
			A4/R1		
		./d·v			
./A4	./A4	./k·ii	k·ii\d·v	d·i\A4/R1	
kodomo ga	te o	samu	gar	ru	...

- *otto ga samugatte ita kara danbō o ireta.*

A4/R1				
		A4/R1		
	./A3			
./A4	./d·v	d·i\A3	A4/R1	
otto ga	samugar	te	ita	...

A quick Twitter search (December 2023) revealed that some speakers have *samugar* that takes a place, time, or garment as an object, as exemplified by:

- *shachō ga kondo wa sukāto o samugatte iru.*

  https://x.com/mamama_s/status/1558739075284103168?s=61&t=P5zngLY-0SHK2OgjuSYsug

- *saikin shashin agete naidesu kedo, inu wa genkini ningennamini fuyu o samugatte imasu.*

  https://x.com/maxi_3960/status/1604466052842553344?s=61&t=P5zngLY-0SHK2OgjuSYsug

I have not consulted with any of the users in the search result on their native language, but if this usage of *samugar* should be incorporated into the present description of the language, there must be a syntactic form that is *samugar* ./d·i. Whether this is of an (almost) unit form such as *samugar* ./d·v or a complex form such as *samu* ./k·ii followed by *gar* k·ii\d·v, I am currently unable to identify the lexical item, either because of some unknown factors that prevent me from accessing the knowledge or because I simply do not have this item in my vocabulary. (December 2023)

VERBS

Initial observations

Six boundaries are classed as verb boundaries: d·i, d·ii, d·iii, d·iv, d·v, and d·vi.

The lemma of a verb contains the suffix *ru*. This form makes a predicate and can take arguments and adjuncts.

- *kuizuni seikai suruto hāgendattsu ga moraeru.*

The past tense is constructed with the suffix *ta*.

- *kuizuni seikai shite hāgendattsu o moratta.*

The adverbial form is formed with the suffixes *te, reba,* and *ruto* among others.

- *gyūnyū o mainichi nomeba kenkō ni nareru.*
- *sensei ga watashino nigaoe o kaite kontesutode shō o totta.*

In the "basic" usage the verb functions as a predicate, but when followed by the suffix *ikata* ('the way of') it functions like a noun.

- *tabekata ga kitanai.*

When it functions like a noun, it can receive a nominal modifier and not an adverbial modifier.

- *kareno tabekata ga amarini kitanakute watashi wa zekku shita.*

The *itai* form functions as an adjective (|k·ii)

- *worumāto ni ittara zettaini hottodoggu o tabetai.*

The *isō* form functions as an adjectival noun.

- *imanimo shinisō datta kara tsui tasukete shimatta.*

Main clause

Some of the main clause verb suffixes are the following:

		imp	frm	pst	neg	int	lit
ˈ(_)r·u	d·i\A4/R1	–	–	–	–	–	–
ˈa·naˌı	d·i\A4/R1	–	–	–	+	–	–
t·a	d·i\A4/R1	–	–	+	–	–	–
ˈa·ˌnakatta	d·i\A4/R1	–	–	+	+	–	–

ɪ·ˈmasu	d·i\A4/P2	–	+	–	–	–	–
ɪ·ˈmasuka	d·i\A4/P4	–	+	–	–	+	–
ɪ·maˈsenn	d·i\A4/P2	–	+	–	+	–	–
ɪ·maˈsennka	d·i\A4/P4	–	+	–	+	+	–
ɪ·ˈmasɪta	d·i\A4/P2	–	+	+	–	–	–
ˈa·ˌnakattadesu	d·i\A4/P2	–	+	+	+	–	–
ɪ·maˈsenndesɪta	d·i\A4/P2	–	+	+	+	–	+
ˈa·ˌnakattadesuka	d·i\A4/P4	–	+	+	+	+	–
ɪ·maˈsenndesɪtaka	d·i\A4/P4	–	+	+	+	+	+
t·e	d·i\A4/P4	+	–	–	–	+	–
ˈa·ˌnaɪde	d·i\A4/P4	+	–	–	+	+	–
t·eˎkudaˈsaɪ	d·i\A4/P4	+	+	–	–	–	–
ˈa·ˌnaɪdeˎkudaˈsaɪ	d·i\A4/P4	+	+	–	+	–	–

The plain series can be the predicate of a relative clause while the formal series cannot. This fact is accounted for by assuming |R1 for the plain series and |P2 for the formal series.

- *tarō ga hon o kakimashita.*

- *tarō ga* {kaita/*kakimashita} *hon wa omoshiroidesu.*

(Both from Iori 2012, p. 50.)

The plain series of the verb are considered |R1 reflecting the fact that they can be followed by *ka* and *kadōka* 'whether' unlike the adjectival noun's *na* form.

- **kouhannˈ ga ˈɪzɪˎˈdeˈkir·u kaˎˈdoukaˎ daˈkewo kannˈgaėt·eˌru nokamo.**

(検察側の罪人)

- **atarashii ie ga kireina kadōka wa satokonitotte jūyō da.*

The interrogative forms have partially regular correspondence with the declarative forms. For the formal series, adding *ka* to the end of the declarative form gives the corresponding interrogative form. For the plain form, in contrast, the interrogative form has the same phonetic form as the declarative form, except the interrogative form comes with R%.

The interrogative forms are here assumed to be |P4, which means quoting particles such as *tte* can follow them without breaking a sentence. The rising intonation (R%) stays on the predicate when followed by *tte*.

The imperatives such as *tabero* and *nome* are not considered as part of the paradigm. The imperative forms we will discuss in this section are separate from those.

A typical intonation used with the plain imperative form is the rising intonation.

- **¿asonnˈd·e** “play (with me)!”

The imperative forms with and without the rising intonation are pragmatically distinct and the native speaker chooses one over the other at each time.

- **asonnˈd·e** “play (with me)!” (assertive)

Other imperative suffixes include *ina* and *inasai*.

- *ina*
- *inasai*

-yō and *-imashō* are used for making invitations.

- **y·ˈou**
- **ɪ·maˈsyou**

The interrogative counterparts of them are used for making suggestions:

- *yōka*
- *imashōka*

Subordinate clause

See also the note in p. membrane about the C1 boundary.

In a subordinate predicate the following suffixes, among others, can appear:

		hab	pst	neg	irr	lit
ˈ(_)r·ėba	d·i\A4/C1	–	–	–	+	–
ˈa·ˌnakereba	d·i\A4/C1	–	–	+	+	–
ˈa·naɪˌto	d·i\A4/C1	–	–	+	+	–
t·e	d·i\A4/C1	–	+	–	–	–
i·	d·i\A4/A4	–	+	–	–	+
t·ara	d·i\A4/C1	–	+	–	+	–
ˈa·ˌnaɪde	d·i\A4/C1	–	+	+	–	–
ˈa·zuˌnɪ	d·i\A4/A4	–	+	+	–	+
ˈa·ˌzu	d·i\A4/A4	–	+	+	–	+
ˈa·ˌnakattara	d·i\A4/C1	–	+	+	+	–
ˈ(_)r·uˌto	d·i\A4/A4	+	–	–	+	–
ˈa·naɪˌto	d·i\A4/A4	+	–	+	+	–

Semantic features are used here mostly in the same way as in the adjectives section.

- *kaijō ni tsuitara, mazu uketsukede nēmutagu o uketottekudasai.* “When you’ve arrived at the site, first receive your name tag at the reception desk.”
- *#kaijō ni tsukeba, mazu uketsukede nēmutagu o uketottekudasai.* “If you happen to arrive at the site, first receive your name tag at the reception desk.”

The latter does not particularly expect that the agent is going to arrive at the site. Rather, it is stated as a conditional rule that the agent is asked to comply.

The use of a [+ irrealis] suffix when referring to an event that is usually taken for granted leads to unnaturalness.

- *aki ni nareba taiwan ni ikō.*

The |C1 series need some attention here. In the following example, the verb *hukumete* |C1 'including' is followed by *no* C1|:

- **nıgatu**ˈ˪ˈ**nızyuu**˪ˈ**kunıtı wo hu**ˈ**kumėt·e no hana**ˈ**sı wo sı**ˈ**te** ˈ**r·u nn desu yo.** "I'm talking about [it], including [the matter] on February 29th."

(検察側の罪人. Translation is mine)

In the following example, the verb *shitara* |C1 'if have done' is followed by the copula *da* C3\A4/P1, which we have seen in p. jeans.

- *keihin ga moraeru no wa kuizuni seikai shitara da.* "It is if [you] correctly answer the quiz that you can receive the prize."

It is important to recognize that the copula *da* generates A4|. As I briefly mentioned in p. london, there seems to be a series of copula-like, non-copular d-words that are used to make, semantically speaking, a sentence fragment that complements another sentence, as exemplified by the following:

- *annna koto ieru tte ningensei okashiidesu. yoriniyotte isha ga desu yo.* "There must be something wrong with [their] personality given that [they] can say such a thing. [They are] a doctor of all others."

(From https://x.com/mfa2s0w2fgir0o1/status/1813367364932767917?s=61&t=P5zngLY-0SHK2OgjuSYsug. Punctuations are modified. Translation is mine.)

- *rekishi ga subete hontou da to wa omoenai nn desu yo ne. watashi wa desu ga.* "I can't seem to think history is right about everything, though it is just me [thinking this]."

(From https://x.com/10000000aobara/status/1610167299184676864?s=61&t=P5zngLY-0SHK2OgjuSYsug. Assumed typos are corrected. Translation is mine.)

In both of these examples, *desu* seems to be placed on |A4, that is, it seems that it is *desu* A4|, and because the right boundary seems to be similar to the copular *desu* C3\A4/P2, it is probably A4|P2. Since the "fragment" supported by *desu* A4|P2 does not seem to follow another |A4, here I judge it as *desu* A4\./P2 (»minimum).

The following example (acceptable to me) might appear as a counter example of the hypothesis that the *desu* fragment cannot follow |A4.

- *a, kore wa akumademo "watashi wa" desu. hokano hitoni wa zenzen kankei nai.* "Oh, this is just 'I.' It is completely irrelevant to other people."

(from https://x.com/c2shock/status/1764618193820406254?s=61&t=P5zngLY-0SHK2OgjuSYsug. Translation is mine.)

I think the graphically quoted portion, *"watashi wa,"* is indeed syntactically quoted, and accordingly the following *desu* is neither C3\A4/P2 nor A4\./P2.

We can test if a given *desu* or *da* is a true copula or a pseudo one by checking whether the resulting phrase can follow |A4. In the aforementioned example "*keihin ga moraeru no wa kuizuni seikai shitara da*," the part "*kuizuni seikai shitara da*" follows |A4 "*keihin ga moraeru no wa*." In the tree notation, this can be represented as:

A4/P1		
	A4/P1	
.|A4	A4|C1	C3\A4/P1
keihin ga moraeru no wa	kuizuni seikai shitara	da

Now, notice that while *desu* A4\./P2 (the fragmentizer *desu*) can, by definition, follow both |C1 and |A4, *desu* C3\A4/P2 (the copular *desu*) on the other hand is allowed to follow |C1 but not |A4. This readily means that the same test can be used to test whether a given boundary is |A4 or |C1. Given that it is either |A4 or |C1, if putting *desu* immediately after it can make an A4| phrase, it is |C1, and if the same operation never makes an A4| phrase, it is |A4.

In the following example, the verb *nattara* |C1 'if have become' is followed by *nara* C1\A4.

> *Nara* C1\A4 coexists with *nara* C4\A4/A4. The fact that the latter constitutes an A4| phrase but the former does not justifies their seperate entries in the lexicon.

- *gakunen'ichii ni nattara nara tanjōbini aifon o kau to yakusoku shite ageru*. "I'll promise you that [I'll buy you an iPhone] if that is on the condition that you'll have gotten the highest score in your grade."

These |C1 forms incorporate a fait accompli or a precondition of what is stated in the main part of the sentence.

To verbal nouns

The following suffixes turn a verb into a verbal noun.

- *iwa* d·i\A4/G2
- *imo* d·i\A4/G2
- *ikoso* d·i\A4/G2
- *isae* d·i\A4/G2

These items differ from typical original verbal nouns syntactically:

- naokoni/ ayamariwa\/ shinakatta\
- naokoni/ shazai wa/ shinakatta\\

	Naokoni	ayamariwa	Shinakatta
*	Naokoni	shazai wa	Shinakatta
	Naokoni	Shazai wa	Shinakatta

Progressive and perfect

This subsection will be filled in later.

Other suffixes

> Many of the suffixes listed in this subsection are assumed to have **i** at the beginning. These are based on my introspection, and other native speakers of Tokyo Japanese may find **ɪ** descriptively more accurate for some or all of the suffixes in question. The speller is responsible for deciding what phonetic forms should be reflected in the spelling.

- **i·ˈwuru** d·ii\A4/R1 'it is possible that'
- **i·ˈė** d·ii\d·i 'it is possible that'
- **i·ˈta**ˌ d·ii\A4/k·ii 'want to'

> The following example involves multiple groups of inflectional boundaries through *gar* and *tai*.
>
> - *samui*. "I'm cold."
> - *samugaru*. "I will act like I'm cold."
> - *samugaritai*. "I want to act like I'm cold."
> - *samugaritagaru*. "I will act like I want to act like I'm cold."
> - *samugaritagaritai*. "I want to act like I want to act like I'm cold."

Generally, potential verbs are not followed by these items, and so are judged to be |d·i.

- **r·ˈarė**¸ d·iii\d·i ‘annoyingly’
- **r·ˈarė**¸ d·iii\d·ii (honorific)
- **i·ˈsugi** d·iii\d·ii ‘excessively’
- **i·ˈkata**¸ d·iii\R2/m-ii ‘the way of’
- **i·ˈkakė** d·iv\d·iii ‘partially’
- **i·ˈtutu** d·iv\A4/A4 ‘while’
- **i·ˈtutu** d·iv\A4/A4 ‘even though’
- **i·ˈtutuno** d·iv\A4/R2 ‘done simultaneously with’
- **i·ˈnaga¸ra** d·iv\A4/A2 ‘while’
- **s·ˈsasė**¸ d·iv\d·vi ‘cause’
- **i·hazımė** d·v\d·iv ‘to start ’
- **i·ȯˈwar** d·v\d·iv ‘to finish’
- **i·ˈȯė** d·v\d·iv ‘to finish’
- **r·ˈarė**¸ d·vi\d·v ‘to be done’

The true passive rare and the annoyance rare

To understand the difference between the boundaries d·vi and d·v, observe that there are two homophonic “passive” suffixes.

- *otōtoni purin o taberareta*. “my brother ate my pudding”
- *otōtoni nakareta*. “annoyingly, my brother cried”

In the first example, the suffix *rare* refers to the fact that the speaker’s pudding was eaten. The pudding receives the action of eating. In the second example, in contrast, nothing receives the action of crying. If anything, the speaker expresses her annoyance about her brother crying.

It is possible that in the first example the suffix *rare* is used to express annoyance, in which case the pudding does not have to belong to the speaker. If she has been annoyed about the fact that her brother was eating a pudding, the suffix *rare* can be used.

At a glance these two readings of *rare,* received action and annoyance, may seem to be contextual variations or ambiguity of the same syntactic form. However, that they are not can be observed in the following pair of sentences:

- *otōtoni purin o taberarehajimeta*. “my brother started eating my pudding.”
- *otōtoni purin o tabehajimerareta*. “annoyingly, my brother started eating a pudding.”

In the first example, the suffix *rare* is used for a received action. In the second example, annoyance. It is observed that the received action *rare* cannot follow *hajime,* but the annoyance *rare* can. Hence, the annoyance *rare* must have a “shallower” left boundary than the received action *rare*.

This and other observations are summarized as the following:

- *rare* d·vi|d·v ‘to be done’ (passive)
- *rare* d·iii|d·i ‘annoyingly’
- *hajime* d·v|d·iv ‘to start doing’

ADJECTIVAL NOUNS

The ontology of adjectival nouns

What we call the adjectival nouns or their close equivalents have received various treatments in grammatical theories. This subsection lays out some ontological bases of adjectival nouns (km-i, km-ii, k-iii) as opposed to adjectives (k·i, k·ii) and nouns (m-i, m-ii).

See Masuoka and Takubo (1992), Takahashi et al. (2005), and Kubo (1994) for existing treatments.

First, observe that not all items with |C2 can be followed by *na mono* 'the thing that.'

kuruma deshita	*	kurumana mono
sakura deshita	*	sakurana mono
kirei deshita		kireina mono
taihen deshita		taihenna mono

Similarly, some |C2 items can be followed by *sa* 'the degree of' and *sō* 'seemingly' while the others cannot.

*	kurumasa	*	kurumasō
*	sakurasa	*	sakurasō
	kireisa		kireisō
	taihensa		taihensō

These facts seem to justify the distinction between nouns and adjectival nouns.

Next, observe the difference between adjectives (k·i, k·ii) and adjectival nouns (km-i, km-ii, km-iii). Only the former of the two can be followed by *idesu* and only the latter can be followed by *desu*:

*	kowa desu		kowaidesu
*	sukuna desu		sukunaidesu
	kirei desu	*	kireiidesu
	taihen desu	*	taihen'idesu

Both of these classes have a phonetic *sa* that turns them into a noun (m-ii), but they are two separate items as becomes apparent in the phonological representations:

utukuˈsı	+	·saˌ	—>	utukuˈsı·saˌ
sukuˈna	+	·saˌ	—>	sukuˈna·saˌ
ˈhaya	+	·saˌ	—>	ˈhaya·saˌ
ˈkıreı	+	ˈ(_)-saˌ	—>	ˈkıreˈı-saˌ
taıhennˈ	+	ˈ(_)-saˌ	—>	taıheˈnnˌ-saˈ

Sō that follows adjectives and the one that follows adjectival nouns appear to be identical in their phonological forms:

utukuˈsı	+	·ˈsouˌ	—>	utukuˈsı·ˈsouˌ
sukuˈna	+	·ˈsouˌ	—>	sukuˈna·ˈsouˌ
ˈkıreı	+	-ˈsouˌ	—>	ˈkıreı-ˈsouˌ
taıhennˈ	+	-ˈsouˌ	—>	taıhennˌ-ˈsouˈ

For the sake of the simpler syntax, however, these *sō*'s are separately admitted to the lexicon.

- *sō* km-ii\km-i
- *sō* k·i\km-i

General observations

Adjectival nouns are phonologically noun-like and semantically adjective-like.

The phonetically bare form of them can be followed by an item with C2|, some of which are copulas, to describe the state of the subject.

Some of the phonetically non-null adjectival noun suffixes are *na, ni, sō,* and *sa*.

- **yuuguredokı-ˈno namımaˈ-nı uˈkaˈb·u ˈkıreı-na ˈho hả, ˈtyotto sutekıˈ desu.** "It is so lovely in the evening a beautiful sail is floating on the waves." (丘の家のミッキー 1) (Translation is mine.)
- **konnˈna ıkagennˈ-na ˈhonn wo ˈˈdas·u no wo ˈˈmir·uˌto ko ˈno sakkaˈ mo ˈmou osımaıˈ da ˈna.** "Judging from the fact that he has published a book of such a sloppy work, his career as an author is finished." (小学館韓日日韓事典) (Translation is mine.)

The suffix *sō* can be attached to an adjectival noun and in turn forms an adjectival noun.

- **ˈRınndoˌsyounennˈ hả Husıgıˌ-ˈsouˈ-na kaȯˈ wo sıˈte Watasıˈ-nı Gaˈgannbo-ˌno koˈto wo Kııˈt·a.** "The boy Rindo asked me about gagambos with a curious face."

But the form with *sō* cannot be further followed by *sō*. This necessitates setting up multiple adjectival noun boundaries.

Some adjectival nouns take *ni* to describe an event but the others do not. (e.g., *kyokutanni* 'extremely' :: **yūshūni* intended: 'excellently'). This too can be described with multiple boundary types.

- **ˈYumı hả Toukyouˌ-ˈno Naˌtu-ˈno ˈatu·sa wo Kyokuˈtann-nı kıratˈt·e, Natuˈyasumı hả Kareiˈzawa nı ıkıˈtaˈ·ı to ˈTıtı-nı Seˈgannd·a.** "Yumi extremely hated the summer heat and begged her father to take her to Karuizawa for the summer vacation."

km-i

The suffixes that follow |km-i include:

- [0] km-i\C2
- *na* km-i\A4/R2

> *daiji* cannot be followed by *na* and therefore it is not an adjectival noun. (Compare *daijina*'s tones with *sukina*, *hetana*, *jōzuna*, etc. This fact as well as the decision that *daijina* is not part of the adjectival noun paradigm was announced in this 2019 blogpost of mine.)

The former is necessary to explain the fact that adjectival nouns can be followed by copulas.

- *kumamotoken wa nihonno kyūshūchihōni ichisuru ken de, sekaide mo kokode shika saibai sareteinai kankitsu, pomeroman ga yūmei desu.*

(The sentence inspired by: https://akumamoto.jp/archives/217974)

- *yūmei* ./k-ii

Na makes a relative clause.

- *taino yūfukuna hitobito wa hishochitoshite hoahin wo yoku riyō shimasu.*
- *yūfuku* ./k-ii
- *hitobito* R2\m-ii (»sine. Probably *hitobito* R2/m-i coexists.)

Do not confuse **-na** with the copular **na** C2\A4/R2b that connects to **no, nn, noka,** and **nnda**.

> Many adjectival nouns have *teki* at the end and many of them are |km-i.

km-ii

km-ii can be followed by:

- ˈ**(_)-sa**˛ km-ii\R2/m-ii 'the degree of'
- **-**ˈ**sou**˛ km-ii\km-i 'seemingly'
- **-**ˈ**sou**˛**nı** km-ii\A4/A2 'appearing'
- **-**ˈ**sugi** km-ii\d-i

> *Yō* as in *kare wa koko ni sundeiru yō da* is not |k-ii since it cannot be followed by *sō*. As to whether it is |k-i we need some more investigation.

Some examples:

- *hibino kurashino taikutsusa wa kaette yumi o shigotoni muchū ni saseta.*
- *watashito rindo wa, hima sae areba tshokande ryokōgaido o asari, kyarantinno kireina machinami o utsushita shasinnado o mite, toshide no kurashi wa kaitekisou de urayamashii nadoto katariatta mono datta.*
- *sono seijitsuna otokonohito wa ikanimo seijitsusouni hanashita.*
- *kireina doresu o kitai ga, kabisugitemo ikenai to yumi ga iu node, watashi wa apparerushoppu o nankenmo hashigo shinakerebanaranakatta.*

where

- *taikutsu* ./km-iii
- *taikutsusa* R2/m-ii
- *kaiteki* ./km-iii
- *de* C2\A4/A3
- *seijitsu* ./km-ii
- *hanashita* A4|
- *kabi* ./km-ii
- *itemo* d·i\A4

Some may question why *sōni* independently of *sō* is necessary. We will later see *ni* km-iii\A4/A2. An obvious fact is that we have to distinguish this *ni* from the phonetic *ni* part of *sōni* since, in order for *ni* km-iii| to connect to supposed *sō* km-ii\km-i, it has to be km-i|. Compare:

yūmeisōni	*	yūmeini
yūbōsōni	*	yūbōni
tekisetsusōni		tekisetsuni
seimitsusōni		seimitsuni

Do not confuse *ni* here with the copular *ni* C2\A4/A2:

- *yūmei ni natta*
- *yūbō ni natta*
- *tekisetsu ni natta*
- *seimitsu ni natta*
- *yūmeina hito ni natta*
- *yūbōna hōhō ni natta*
- *tekisetsuna setsumei ni natta*
- *seimitsu na chizu ni natta*

To discern whether a phonetic *ni* is *ni* km-iii\A4/A2 or a sequence of [0] km-ii|C2 and *ni* C2|A2, replacing the adjectival noun part with a noun phrase can be useful because a noun phrase can be connected to C2|. For example, *gijutsu ga kōdo ni hattatsu shita* 'the technology has been highly developed' can be compared to *gijusu ga kōdona gijutsu ni hattatsu shita* 'the technology has become a high technology'. Here *kōdo* and *kōdona gijutsu* seem to be interchangeable and therefore it seems that this *ni* is C2|A2. The suffix *ni* km-iii|A2 can be used to describe the manner in which the event in question happens. *Kōdo ni hattatsu shita* refers to the high state of the resulting technology and not to the way in which it has been developed.

Another class of items that can be confusing is those such as *igaini* 'unexpectedly.' This is a unit item and the phonetic *ni* in it is neither a copula nor an adjectival noun suffix. One way to tell *igaini* is not an adjectival noun is to observe if in the typical setting the modified predicate receives the R tone, e.g., *Dame dato omotta ga, Igaini Daijōbu datta*, where the R tone is indicated by a capital letter. If *igaini* semantically modifies *daijōbu*, as in "it was okay in the way that it was unexpected," it is unlikely that the predicate receives the R tone by default. That is, if *daijōbu* is modified by some adverbial phrase and still receives the R tone, it is less likely that it is perceived as "neutral." The correct interpretation of *igaini daijōbu datta* is "unexpectedly, it was okay = it was unexpected that it was okay." Compare it with "*Dame dato omotta ga, Girigiri daijōbu datta*." where *girigiri* modifies the predicate.

Lastly, it may be worth mentioning that *sō* km-ii\km-i is not km-i| or |km-ii because otherwise the grammar would allow the following unacceptable forms:

- **taihensōsō*

km-iii

km-iii can be followed by:

- **-ni** km-iii\A4/A2 '-ly'

as in:

- *kore ga saigono chansu na node, kakujitsuni seikō sasetekudasai.*

where

- *kakujitsu* ./km-iii
- *seikō* ./G1
- *sasetekudasai* G1|

Ni km-iii\A2 constitutes an "adverbial" phrase that semantically modifies the predicate of the same clause. The boundary km-iii as opposed to km-ii is postulated in order to describe the distribution of this suffix. Some |km-ii items can take this suffix and others cannot. To see this, study the following list:

kappatsuni	*	idaini
junsuini	*	kichōni
seijitsuni	*	gōkaini
kaitekini	*	yūbōni
genjūni	*	yūryokuni
seidaini	*	kimyōni
tekisetsuni	*	igaini
yūgani		
tekidoni		

(For making the list I referred to Lianping 2009.)

kireini 'completely' is not an adjectival noun. This is contrasted with *kirei* 'beautiful' and *kireini* 'beautifully'.

Although semantics needs a separate investigation, we can observe the general tendency that those adjectival nouns that cannot take *ni* semantically modify *things* specifically, as opposed to *events,* and those that can take *ni* semantically modify both things and events.

Therefore, at least tentatively, it seems reasonable to postulate two subcategories of adjectival nouns, one of which is the *thing modifying adjectival nouns* and the other of which is the *event modifying adjectival nouns*. It follows then that an event modifying adjectival noun is usually ./km-iii lexically and a thing modifying adjectival noun is usually ./km-ii lexically.

NOUNS

Some typical behaviors of a noun stem

Typically, a noun stem (|m-ii, |m-i) can be modified from the left.

- *tairyōno mikanjūsu*
- *kokusaitekina jizendantai*

This is captured by assuming:

- *mikanjūsu* R2/m-ii
- *jizendantai* R2/m-ii

As evidence that R2| belongs to the stem instead of some larger structure, the R2| "slot" remains open even with suffixes present:

- *tairyōno mikanjūsude onaka o kowashita.*
- *kokusaitekina jizendantaikara tegami ga todoita.*

The suffix *-ppo* turns a noun into an adjective. When this suffix is present, both A4| and R2| modifications can be available, but the former must come before the latter.

- *ikanimo tōkyōno usagippoi.*
- **tōkyōno ikanimo usagippoi.*

This behavior of a noun stem is contrasted with those of adjective, verb, and adjectival noun stems. For these types of stems, typically, whether A4 or R2 modifications are allowed depends on the suffix.

Adjectives:

	odoroku hodo shiroi	*	odoroku hodono shiroi
*	odoroku hodo shirosa		odoroku hodono shirosa

Adjectival nouns:

	odoroku hodo kirei	*	odoroku hodono kirei
*	odoroku hodo kireisa		odoroku hodono kireisa

Verbs:

	nintendō o taoshita	*	nintendōno taoshita
*	nintendō o taoshikata		nintendōno taoshikata

That is,

- *mikanjūsu* R2/m-ii
- *shiro* ./k·ii
- *taos* ./d·vi

Some noun suffixes

m-i can be followed by:

- **-ˌno** m-i\R2 ‘of’
- **-ˌno** m-i\R2 ‘which is’
- **-ˌno** m-i\C2 predicative
- **-kaˌra** m-i\C1 ‘from’
- **-kaˌrano** m-i\C2 ‘from’ predicative
- **-to** m-i\R2 ‘and’
- **-to** m-i\A1 ‘and…’
- **-to** m-i\A2 ‘with’
- **-made** m-i\C1 ‘up to’
- **-hė** m-i\A2 ‘toward’
- **-hėno** m-i\R2 ‘toward’
- **-de** m-i\A2 ‘using’
- **-nı** m-i\A2 ‘to’
- **-nı** m-i\A4 ‘about’
- [0] m-i\C2 predicative
- [0] m-i\A1 nominative

m-ii can be followed by:

- **-nado** m-ii\m-i ‘including’
- **-toka** m-ii\m-i ‘including’
- **-yo** m-ii\P5 vocative
- [0] m-ii\P5 vocative

Nouns, noun-like forms, and copulas

This subsection pertains to a broader category than just nouns as defined in terms of m-ii and m-i. For explanatory purposes, we take nouns as examples.

Nouns typically have R2 on the left.

- *sutekina kaban*
- *kinō totta shashin*

Despite this, a noun can merge with a copula to make a predicate that is A4|.

- *kore wa sutekina kaban desu*
- *kore wa kinō totta shashin desu*

The fact that a predicate "headed" by a copula is A4| is explained if the copula is C3\A4/P2.

A copula can be phonetically null (the zero copula).

- *kore wa sutekina kaban.*
- *kore wa kinō totta shashin.*

Unlike *desu* C3\A4/P2, the zero copula cannot be followed by items that are P2\ or P3\ such as *yo* (not the feminine *yo*) and *tte* (type-I quoter):

- **kore wa sutekina kaban yo*
- **kore wa kinō totta shashin tte iwareta*

but it allows type-II and type-III quotations:

- *kore wa sutekina kaban tte itteta*
- *kore wa kinō totta shashin tte iwareta*

and it cannot follow *n* R2b\C3:

- **ashita komike ni iku yōji ga aru n.*
- *ashita komike ni iku yōji ga aru n desu.*

which means that the zero copula is:

- [0] C2\A4/P4

part of which is reflected in our boundary inventory as C3b, the alias of P4.

TEMPORAL NOUNS

The ontology of temporal nouns

As temporal nouns are superficially similar to nouns, this subsection is dedicated to establishing the case for the temporal noun boundaries tm-ii and tm-i separate from the noun boundaries.

The nouns and the temporal nouns are phonetically similar. Also, they both can be |C2 or |A1 in the phonetically bare form. But their typical meanings are substantially different. Typically, a noun refers to things and events while a temporal noun refers to a point in time.

Some of the clear cases where they differ are the suffixes *niwa, nimo,* and *madeni*. *Niwa* is used to refer to a point in time *no later than which* an event is located. *Nimo* is used to give an early (or the earliest) possible time for the event in question, as in "he *can* be here *as soon as* Monday." Both of them seem to be difficult to relate with noun forms while maintaining semantic coherency. The last one, *madeni*, which is used to give the latest possible time of an event, is inconceivable with an AP beginning with a noun.

- *yuki wa yoruniwa yamu hazu da.*
- *gojira wa kyōnimo ōsaka ni tōtatsu suru to mirarete iru.*
- *kōji wa kurisumasumadeni kanryō suru.*

Some phonetically identical suffixes differ in meaning between nouns and temporal nouns.

De as a noun suffix signifies the means of an action while the same-sounding temporal noun suffix refers to the point in time at which a state completes.

- *hōchōde ninjin o kirōto shitara yubi o kitteshimatta.*
- *natsuyasumi wa kyōde owari.*

De, as in the following examples, that signifies a cause of an event is treated as a copula: *de* C2\A4/A4.

- *musume ga kaze de watashi wa shigoto o yasunda.*
- *kanojono yokogao ga amarini kirei de, kokoro ga suikomareteshimatta.*
- *karui kimochide hajimeta arubaito ga omotta yori taihen de, suguni yameteshimatta.*
- *dōsōkai ga saibanto onaji hi de, sanka dekinakatta.*

Kara as a noun suffix signifies a source of a thing or event, but as a temporal noun suffix, it signifies the point in time at which an event or state begins.

- *aitsukara moratta nuigurumi o itsumademo daijini totteoite iru rashii.*

- *kyōkara natsuyasumi da kara, kankokugono benkyō o hajimeta.*

The following are some of the suffixes whose phonetic forms are shared between the noun and temporal noun paradigms.

-kaˌra	m-i\C1	tıànnˈ ga ˈwaˈru·ı ˈerıà hà kokoˌ-kaˈra desu.
		aıtuˌ-kaˈra hà ˈzıkenn-ˌno koˈto hà nanıˈmo kııˈt·e ˈnaı.
		kımıˌ-kaˈra no tegamıˌ-ˈno okageˈ de, tansınnˈhunınn wo gannˈˈbar·u koˈto ga ˈdekit·a.
	tm-ii\C1	huyuˈyasumı hà asıˈta-kaˌra desu.
		haˈzımete da kara ˈkonnkaı hà syougaˈnaı kedo, kouiˈwu koˈto hà koreˌ-kaˈra hà sıˈnaıde ˈne.
		nıˈgakkı-kaˌra no ˈzyugyou hà tanoˈsımı da kedo, huànnˈ mo ˈaru.
-made	m-i\C1	koˈno ˈretu-nı narannˈd·e ıˈru hıˌto-ˈno ˈnakade, ˈkyou obakeˈyasıkı-nı haıˈˈrèr·u no hà kımıˈ-made da.
		ˈkyouto-kaˌra toukyouˈ-made no ˈkyorı hà oyoˈso yonnhyakkıroˈme—toru da.
	tm-ii\C1	ˈse—ru hà ˈmou owatˈt·a kara, ısogaˈˈsı·ı no hà ˈkyou-made da.
		syūmatuˈ-made no yoteıˈ wo osıèˈt·e kureˈt·ara, ˈato hà kotˈtı-de tyouseıˈ sıte˪ˈoˈk·u kara.
-ˌno	m-i\R2	soreˈ˪ watasıˌ-ˈno ˈhonn da kara kaˈesıt·e.
	tm-i\R2	asıtaˌ-ˈno nomıkaıˈ hà zannˈnenn da kedo kotoˈwatt·a.
	m-i\C2	soˈno ˈhonn hà watasıˌ-ˈno desu.
	tm-i\C2	koˈno gyuunyuuˈ˪ ˈkyonenn-ˌno zyanaı?
	m-i\A1	osaıhouˈsetto hà zettaıˈnı purıkyuàˌ-ˈno ga ˈıı tte ˈmama-nı ıtˈt·e˪ˈoıt·a.
	tm-i\A1	yuuhannˈ hà kınoˌu-ˈno ga noˈkott·eˌru kara tekıtouˈnı ˈtabèt·e.
-to	m-i\R2	ˈpenn-to ˈno—to wo tsukuèˌ-ˈno uˈè-nı ˈdasıt·e˪kudaˈsaı.
	tm-i\R2	ˈkyou-to asıˈta hà nanıˈmo sıˈnaıde hıˈtorı de ıˈè-de suˈˈgos·u tumorıˈ.

Notes on temporal nouns (tm-ii, tm-i)

A temporal noun is any syntactic form with *tm-ii* or *tm-i* on its right edge.

Since some items that refer to a point in time such as *kinō* and *ashita* function like adverbs (|A4), it is easy to erroneously infer that this "adverbial" behavior is a common property of temporal nouns. That it is not the case can be seen from the following:

kinōmadeni		kurisumasumadeni
kinō tabeta kurowassan	*	kurisumasu tabeta kurowassan

These items that seemingly function like adverbs and like temporal nouns are better described as homophones across two separate categories, one in the adverb class (|A2, |A3, or |A4) and the other in the temporal noun class (|tm-ii or |tm-i).

Also, it is easy to misapprehend the phonetic *ni* part found in some items that refer to a point in time as a temporal noun suffix. That it is not a temporal noun suffix is shown by the following examples:

- **kinōni tabeta kurowassan*
- *kurisumasuni tabeta kurowassan*

Again, it is better to recognize that some of them are phonetically similar adverbs, in this case *kurisumasuni* R2/A2. (They differ from typical adverbs in that they are R2|.) Some other adverbs of this type are:

- *aidani*
- *madeni*
- *atoni*
- *getsuyōbini*
- *gogonijinijūnifunni*

Time-referring adverbs that refer to a point in time relative to the speech time tend not to have *ni* and those that refer to a point in time not relative to the speech time tend to have *ni*.

kinō	getsuyōbini
ashita	suiyōbini
asatte	mokuyōbini
ototoi	nichiyōbini
ototsui	
raishū	daisanshūni
senshū	daiisshūni
konshū	dainishūni
kotoshi	nisennijūsannenni
rainen	nisennijūyonenni
kyonen	nisennijūninenni

Asa, yoru, hiru, and *toki* seem to be exceptions. Another way to think about it is that those that are numerals (the names of the days of the week are essentially numerals considering how they work) have *ni* and others do not. Still, *tokini, kurisumasuni,* and several others are exceptions. Either way, this is not considered as a syntactic or phonological phenomenon. These items are treated as unit forms.

The bare form of a temporal noun typically has R2 on its left.

Some temporal noun suffixes are:

- **-ˌno** tm-i|R2 ‘at’
- **-to** tm-i|R2 ‘and’
- **-de** tm-i|A2 ‘having [a point in time]’
- **-nıhȧ** tm-i|A4 ‘no later than’
- **-nımo** tm-i|A4 ‘as soon as’
- [0] tm-i|C2 predicative
- [0] tm-i|A1 nominative
- **-nado** tm-ii|m-i ‘including’
- **-toka** tm-ii|m-i ‘including’
- **-made** tm-ii|C1 ‘through’
- **-madenı** tm-ii|A4 ‘by’
- **-kaˌra** tm-ii|C1 ‘from’
- [0] tm-ii|m-i noun (a temporal noun used as a noun behaves like a place noun)

river

Chapter Six

MISCELLANEOUS TOPICS

SYNTAX-PHONOLOGY MAPPING

Syntactic and phonological trees

Recall that the sentence *kawaii ōkina toshokan ni nara itta* has the following syntactic structure:

- kawaii/ ōkina/ toshokan\\/ ni\/ nara\/ itta\

```
                    kawaii ōkina toshokan ni nara itta
              kawaii ōkina toshokan ni nara
        kawaii ōkina toshokan ni
  kawaii ōkina toshokan
              ōkina toshokan
kawaii  ōkina      toshokan  ni   nara      itta
```

All sentences are subject to the triangle constraint, but the said constraint does not need the syntactic tree's all terminal nodes. Certain simplification is necessary for a sentence to be applied to the triangle constraint. In the structure simplified for the triangle constraint, the terminal nodes are APs. In this case, the simplified structure looks like the following:

- kawaii/ ōkina/ toshokanninara\\/ itta\

```
          kawaii ōkina toshokanninara itta
  kawaii ōkina toshokanninara
              ōkina toshokanninara
kawaii  ōkina          toshokanninara      itta
```

We call this kind of simplified structure the *phonological tree* of a sentence as opposed to the *syntactic tree*.

Let us take another example:

root

- Syntactic tree: fuyu/ no\/ samu/ sa\\/ ni\/ taerare/ nai\\

```
                    fuyuno samusani taerarenai
              fuyuno samusani
        fuyuno samusa
  fuyuno          samusa                taerarenai
fuyu    no    samu      sa    ni    taerare    nai
```

- Phonological tree: fuyuno/ samusani\/ taerarenai\

```
        fuyuno samusani taerarenai
  fuyuno samusani
fuyuno        samusani        taerarenai
```

The basic procedure to obtain the phonological tree (Trimming)

Let us suppose we have a sentence that is syntactically represented as:

- A B\

where a letter or a sequence of letters represents a syntactic form and slashes are used as operators of the reverse Polish notation.

If the forms A and B are both AP-initial, the phonological tree of this sentence will be identical (conversion is not necessary):

- A B\

Structure-changing operations are necessary if one or more forms in the syntactic tree are non-AP-initial. Let us now represent non-AP-initial forms in lowercase letters:

- A b\

The structure a B\ has not been found as far as I am aware. If it ever appears, how it is mapped onto phonology must be investigated separately.

This syntactic form will be converted to the phonological form:

- Ab

which stands for the phonological sequence corresponding to the syntactic sequence A b. Since this is a single AP form, slashes are not necessary.

Now, observe that when converting A b\ to Ab, as the form b joined A, the slash was removed.

- A b\ —> Ab

When A and b join together, if A has a slash—that is, if A is immediately followed by a slash—unlike b's slash, A's slash must be placed on the right side of the new sequence Ab. That is,

- A\ b\ —> Ab\

This implies,

- C A\ b\ —> C Ab\

and if A has multiple slashes,

- C D A\\ b\ —> C D Ab\\

If b has multiple slashes and joins A, exactly one slash of b will be removed.

• C A b\\ —> C Ab\

We call this operation *trimming* as it involves cutting down unwanted branches of a tree. Trimming converts a syntactic tree to a phonological tree.

Let us take *kawaii ōkina toshokan ni nara itta*. To restate, the syntactic tree of this sentence looks like the following in the reverse Polish notation.

• KAWAII ŌKINA TOSHOKAN\\ ni\ nara\ ITTA\

As per our new convention, the AP initial forms are in capital letters and the other forms are in lowercase letters.

Our task is to find convertible local structures. Working from left to right, (a) check whether trimming is applicable and apply if it is, and (b) after trimming a local structure, apply the same procedure until all constituents of the sentence are AP-initial.

The first step is to convert TOSHOKAN\\ ni\ to TOSHOKANni\\.

• KAWAII ŌKINA TOSHOKANni\\ nara\ ITTA\

Next, since TOSHOKANni\\ nara\ can be converted:

• KAWAII ŌKINA TOSHOKANninara\\ ITTA\

Now, as the constituents of the structure are all AP-initial, we say that the phonological tree has been obtained.

Let us now see how *fuyuno samusani taerarenai* is trimmed down into the phonological tree:

1. FUYU no\ SAMU sa\\ ni\ TAERARE nai\\
2. FUYUno SAMUsa\ ni\ TAERAREnai\
3. FUYUno SAMUsani\ TAERAREnai\

The phonological tree has been obtained.

We ignored zeros for ease of exposition. Zeros are treated in the same way as the non-AP-initial forms.

$-THEORY

Restating the triangle constraint in terms of the abstract feature '$'

Consider the following statement:

- In the phonological tree of a sentence, a phrase has R at its beginning iff its inner phrase has R.

The statement holds everywhere but terminal nodes.

Now, consider the following:

- In the phonological tree of a sentence, a phrase has R at its beginning iff it contains $, where $ is a feature that can be given to an AP.

The revised statement holds everywhere including terminal nodes.

Distribution of $

It is assumed here that $ is a feature at most one instance of which can be given to an AP. An AP in a sentence either does or does not have $ ("dollar"). Also, as implied by the triangle constraint and the fact that a sentence must begin with R, it is assumed that at least one AP of a sentence must have $.

Now consider the following sentence, where R is indicated by the capital letter.

• *Ame ga futta*. "It rained."

Since a sentence must have at least one $ and the only AP that has R is *Ame ga*, a $ must be given to *Ame ga*, which can be represented by inserting the symbol $ at the top of the corresponding AP:

- *$Ame ga futta*. "It rained."

Compare this with the following, in which the second AP has R.

- *Ame Futta?* "Did it rain?"

To account for this distribution of R, two different patterns of $ can be thought of. One is to place a $ on *Futta* only, and the other is to place a $ on both *Ame* and *Futta*.

- *Ame $Futta?* "Did it rain?"

- *$Ame $Futta?* "Did it rain?"

As we will see later, in some cases one of the two analyses may be preferred over the other. From the phonetic evidence alone, however, nothing can be said to rule out one in favor of the other.

$ and information structure

In a question, $ tends to be given to the AP that semantically corresponds to the information being asked for. In a yes-no question, $ tends to be given to the predicate of the main clause. In a wh-question, $ tends to be given to the wh-word.

• *Iida o koroshita no wa $Dare?* "Who is it that killed Iida?"

When a question starts with the word *nande* ('why'), the sentence typically has only one $, which is given to the initial AP. This type of sentence yields a globally downward pitch contour.

• *$Nande iida wa korosareta no?* "Why was Iida killed?"

Answering these questions would yield declarative sentences. In these sentences, the AP that corresponds to the information that was originally asked for tends to have $.

- *Iida o koroshita no wa $Dare?* "Who is it that killed Iida?"
  - *Iida o koroshita no wa $Yamamoto da.* "Yamamoto is the one who killed Iida."
  - *$Ashikawa ga iida o koroshita n da.* "It is Ashikawa who killed Iida."

In the examples above, the question asks for *who* and the corresponding sentences have $ at the AP where the *who* is provided.

When the question is a why question, the information that is asked for can be more complex, but the correspondences are not difficult to see.

- *$Nande yamamoto wa iida o koroshita no?* "Why did Yamamoto kill Iida?
  - *Yamamoto wa Iida ga $Kirai datta n da.* "Because Yamamoto *hated* Iida."
- *$Nande ashikawa wa iida o koroshita no?* "Why did Ashikawa kill Iida?"
  - *Ashikawa wa Iida ga $Suki datta n da.* "Because Ashikawa *liked* Iida."

When the why-word is in the middle of a question, typically, $ sticks to the why-word. The why-word "carries" $. (Observe that the positions of R are in accordance with the triangle constraint.)

- *Yamamoto ga iida o koroshita no wa $Naze?* "For what reason did Yamamoto kill Iida?"
- *Yamamoto wa Iida o $Naze koroshita no?* "Why did Yamamoto kill Iida?"

A statement does not always have to be an answer to an explicit question. Yet, the positions of $ are easier to predict under the assumption that the statement corresponds to a question. The sentence *iida ga korosareta* ("Iida was killed") has a few possible versions depending on what implicit question it corresponds to:

- *$Iida ga korosareta* (when it corresponds to "who was killed?")
- *Iida ga $Korosareta* (when it corresponds to "What happened to Iida?")

A third version is imaginable. It corresponds to something like "Why are you crying?" but the phonetic form is identical to the second version, just as predicted by the triangle constraint.

- *$Iida ga $Korosareta* (when it corresponds to "Why are you crying?")

When the first AP of a multiple-AP word has R, whether the remaining AP's have R is not completely predictable from the positions of $. When the word *nintokutennō* (Emperor Nintoku) receives R, it is given only to the first AP *nintoku* while for the word *tōkyōdizunīsī* (Tokyo Disney Sea) R is given to each of the two AP's. It seems that the different distribution patterns of R for these words need to be characterized lexically. Some common expressions that seem to come in both forms interchangeable (at least for some speakers), such as *arigatōgozaimasu*, may receive the same treatment: the different forms are the results of different lexical properties, not of different information structures.

The basic idea of $-theory and the triangle constraint is explained in this 2019 blogpost of mine.

TOPICS ON LEXICAL ITEMS

Items that require a modifier

Certain items, including **hɪˈto**, ˈ**tokɪ** (noun-like), ˈ**tokɪ** (adverb-like), **kuˈnɪ** and some others, cannot merge rightward, but a phrase having them as its right node can merge rightward. That is, they can only merge leftward first and the resulting phrase can merge rightward.

- **watasɪˈ ga ˈkɪta ˈtokɪ hả dareˈmo ɪˈnakatta.** “When I arrived, nobody was there.”
- *ˈ**tokɪ hả dareˈmo ɪˈnakatta.**

It is reasonable to postulate the following:

- ˈ**tokɪ** R2\tm-ii. (noun-like)
- ˈ**tokɪ** R2\A4. (adverb-like)
- **kuˈnɪ** R2\m-ii

Another interesting item **hɪˈto** cannot follow *ano, kono,* or *sono*.

- ***aˈno hɪˈto hả ˈdare?** “Who is that person?”
- ***koˈno hɪˈto hả ˈdare?** “Who is this person?”
- ***soˈno hɪˈto hả ˈdare?** “Who is that person?”

When this construction is otherwise expected, the unit forms **a ˈnohɪto**, **koˈnohɪto**, and **soˈnohɪto** are used.

- **aˈnohɪto hả ˈdare?** “Who is that person?”
- **koˈnohɪto hả ˈdare?** “Who is this person?”
- **soˈnohɪto hả ˈdare?** “Who is that person?”

These facts can be properly described if we split R2 into two: R2’ and R3, and assume *ano, kono,* and *sono* are ./R3, so that only items that are R3| can follow them. Under this assumption, most items we now consider R2| will be R3|, most items we now consider |R2 will be |R2’, and **hɪˈto**, which cannot follow *ano, kono,* and *sono*, will be R2’|.

<u>The proposed split of R2 remain unofficial for now. We for the time being pretend that **hɪˈto** just happens to not follow *ano, kono,* or *sono* in all cases. The R2 split will likely be reflected in a future version.</u>

Dake and its/their curious phonetic forms

This is an updated version of this blog post of mine, originally written for version 1.7.

The morpheme *dake* ('only') concisely demonstrates (for native speakers of Japanese) that a naive parsing can fail to find a syntactic boundary and the notion of morpheme has no place in syntax. The following sentences could not be properly characterized if *dake* (unitalicized) was assumed as a common component.

i. *Chichioya ga daijin na* dake *de honnin wa nanno torie mo nai.* "His only takeaway is his father being a minister and he himself is a good-for-nothing."

ii. *Yaru* dake *yatte mimashō.* "Let's do it as far as we can."

iii. *Kyō ga saishūbi de aru* daken*i, kiai ga haitte iru.* "As today is the last day, they are full of spirit."

iv. *Watashi* dake *itte mo shōganai.* "If I come alone, I'll be no use."

v. *Watashi* daken*i osiete kureta n desu.* "They told this only to me."

vi. *Watashi* dake*ga sankasha ja nai.* "I am not the only one participating."

vii. *Watashi* dake*o suki de ite kureru.* "(She) likes only me and will stay that way."

viii. *Watashini* dake *oshiete kureta n desu.* "They told this only to me."

In i, vi, and vii, *dake* may seem like behaving noun-like. The phonological shape as it is suggested by the same sentences, in addition to iii and v, seems to be something like **da**ˈ**ke**. If it is a noun and the accent number is one, it should yield a boundary fall in ii, iv, and viii. In reality, no boundary falls are found in ii or iv.

In viii, a boundary fall is found between *watashini* and *dake*, which may seem to suggest that the boundary fall is part of the lexical form of *dake* instead of a result of the A1 boundary. (It is not expected that *watashini* has |A1. Cf. **watashini ga*.) But ii and iii contradict this hypothesis as no boundary falls are found in *yaru dake* or *aru dake*.

These "problems" are not real. They are pseudo-problems. No evidence suggests that all of these instances of the phonetic *dake* are of the same lexical item. The fact is that these phonetic *dake*'s behave differently both in syntax and in phonology. If the same phonetic form across multiple sentences behaves differently syntactically or phonologically, they are different lexical items.

In i, *dake* is R2b|C2 and phonologically **da**ˈ**ke**. In ii, it is R2b|A4 and phonologically **da**ˈ**ke**. In iii, it is part of *dakeni* which is R2b|A4 and **da**ˈ**kenɪ**. In iv, it is A1|A4 and **da**ˈ**ke**. (Cf. *otoko dake* 'only men'. Also notice that no boundary fall arises between *dake* and *itte*.) In v, it is A1|A4 and **da**ˈ**kenɪ**. In vi, A1|A4, **da**ˈ**kega**. In vii, A1|A4, **da**ˈ**kewo**. And in viii, A2|A4, ˎ**da**ˈ**ke**.

i.	daˈke	R2b\C2
ii.	daˈke	R2b\A4
iii.	daˈkenı	R2b\A4
iv.	daˈke	A1\A4
v.	daˈkenı	A1\A4
vi.	daˈkega	A1\A4
vii.	daˈkewo	A1\A4
viii.	ˏdaˈke	A2\A4

Below is another set of sentences corresponding to each of the items above.

i. *Mojisū ga onajina dake de naiyō wa zenzen chigau.* "They just have the same number of letters and the contents are totally different."

ii. *Okane o morattara moratta dake tsukatte shimau.* "Whenever you receive money, you'll spend as much as you've received."

iii. *Wakai dakeni kaifuku ga hayai.* "It's your youth that brings you such a swift recovery." (Kenkyusha's New Japanese-English Dictionary 5th Edition)

iv. *Pasokon dake attemo sofuto ga naito yakuni tatanai.* "If all you have is a computer, it's useless without software."

v. *Tāgettono nomimono dakeni suimin'yaku o ireta.* "I put sleeping pills only to the target's drink."

vi. *Subete no kadai o teishutu shita hito dakega sotsugyō dekiru.* "Only those who submitted all assignments can graduate."

vii. *Risaikuru dekiru gomi dakeo aoi fukuroni ireta.* "I put only recyclable garbage into the blue bag."

viii. *Jugyō o kesseki shita hitoni dake hoshūkadai o dashita.* "I gave make-up assignments only to those who were absent from class."

Despite the fact that some of their grammatical behaviors are dissimilar, the native speaker of Japanese tends to think they stem from the same entity. The naive assumption may be reinforced by the notion of morpheme, which is commonly explained as the smallest linguistic unit that retains a pairwise relation of form and meaning. However, the notion of morpheme has no place in grammar. Any attempt to incorporate it into grammar will be deceptive at best.

Quantifiers

This section is planned to be filled in later.

- *gakusei hutariga aruitekita*
- **gakusei hutari aruitekita*
- *furyō ga gakusei o sannin nagutta.*
- **furyō ga gakusei o gakusei sannin nagutta.*
- *furyō ga gakusei sannin'o nagutta.*

Chapter Seven

SPELLING GUIDELINES

ABOUT KYOURO VERSIONS AND OLDER MATERIALS

Version management

New versions of Kyouro are released by modifying existing documents or releasing new documents. When modifications or new documents become public, only the parts of the theory or the rules that are revised in the new texts are considered updated. Old analyses and rules remain valid unless they have been invalidated by newly given statements.

To keep Kyouro documents organized and keep clear which document supersedes which, all Kyouro documents are given dot-separated three-digit version numbers.

The rightmost digit is incremented when minor modifications, including fixing typos, have been made to the document. This digit is only used for indicating the version of the document per se to avoid confusion when referencing multiple versions of the same document.

The remaining two digits indicate the present Kyouro version as of the release of the document. When multiple documents provide specific rules or analyses on the same topic, only those of the greater version number are effective. Provided that, when an old document receives correction, if the corrected version still does not represent the latest version of Kyouro, only the rightmost digit is incremented lest contents once overwritten become effective again.

When incrementing a digit, all digits on the right are set back to zero.

When incrementing either of the left two digits, usually, only the middle digit is incremented. The leftmost digit is incremented at arbitrary times based on the significance of the change or for any reason that I feel reasonable. The leftmost digit is used somewhat symbolically. Since the first version (released in 2018) through version 5, the leftmost digit has been incremented once per year. It is treated as the “major version” indicator and *Kyouro 5* or *version 5* refers to a version of Kyouro whose version number’s leftmost digit is 5.

None of the official documents, including this one, are meant to be a complete statement of how the spelling system works. Kyouro materials are intended to be referenced in conjunction with the other existing materials.

Obsolete materials

- Starting with version 5, all files with a version number lower than 2.0.0 are obsolete.

THE DOT ABOVE

The general rules of the dot above

- In Kyouro, the vowel letters, including **i**, are dotless by default: **a, ı, u, e, o**. Where certain conditions are met, a vowel letter receives a dot above: **ȧ, i, ė, ȯ**.

- The letter **u** never graphically receives a dot above. When the letter **u** immediately following another vowel letter is under such circumstance that **a, ı, e,** or **o** would receive a dot above, the letter **u** is replaced with the sequence **wu.** We say "**u** is dotted," "**u** receives a dot" and so on to refer to the process where **u** immediately after another vowel is replaced with **wu**.

- As of version 5, all rules given in earlier versions to add or remove dots to or from a vowel letter are now deprecated.

The kanji rule

- A vowel letter immediately following another vowel letter receives a dot above if it corresponds to the left edge of a kanji.

はくあい 博愛	ha'kuȧı
はいしゃ 歯医者	'haisya
しかいしかい 歯科医師会	sıkai'sıkaı
きいはんとう 紀伊半島	kıi'hanntou
おうえんだん 応援団	ou'ėnndann

The katakana rule

- Where it is spelled in katakana, all vowel letters immediately following another vowel letter are dotted: **ȧ, i, wu, ė, ȯ**. This does not include a vowel letter that follows a long vowel symbol.

ネイルアート	neiru'ȧ—to
エーアイ AI 絵	e—'aiė
プレイステーション	pureisu'te—syonn
メインストリーム	meinnsuto'rı—mu

The vowel verb rule

- A vowel letter that corresponds to the final vowel of a vowel verb, either *e* or *i*, is dotted: **ė, i.**

食べた	ˈtabėt·a
閉めた	ˈsımėt·a
伸びた	ˈnobit·a
落ちた	ˈotit·a

The syllabicity rule

- A vowel letter immediately after another vowel letter receives a dot if it consistently (i.e., regardless of the phonological environment of the lexical item it belongs to) corresponds to a syllable nucleus's first mora.

A vowel letter corresponding to an underlying offglide that gains syllabicity only under certain phonological contexts of the lexical item it belongs to is written as nonsyllabic.

寒い	ˈsaˈmu·ɪ
遠い	toˈoˈ·ɪ
多い	ˈoˈo·ɪ

In the examples above, the adjective suffix *i* is written dotless in *tōi* and *ōi*, in both of which the said suffix is realized as a nucleus, as it is theorized that its syllabicity is a product of the phonological context. (Offglides gain syllabicity after a bimoraic rhyme. »disinfection)

Some may question whether the second vocoids of *kaetta* and *haitta* are syllabic. Regardless, since these vowel letters correspond to offglides in other contexts than before a long consonant, they are consistently written dotless:

入った	haˈɪtt·a
帰った	kaˈett·a

- A vowel letter corresponding to a syllable nucleus that follows another vowel letter in the context where the original preceding consonant undergoes deletion receives a dot.

The inflectional suffix *reba* is written as ˈ**(_)r·ėba** because the *r* is subject to deletion when the stem-final segment is a consonant.

笑えば	waˈraˈ·ėba
話せば	haˈˈnas·ėba
着れば	ˈkiˈr·ėba

The rule of velar verbs

- A vowel letter *i* derived from an underlying *k* or *g* in a verb stem is written dotted iff it follows *e*.

ゆらめいた	yuraˈmeit·a
はなやいだ	hanaˈyaɪd·a
羽ばたいた	habaˈtaɪt·a

The rule of w

- The letter *i* at the beginning of a verb suffix is consistently dotted if it gains syllabicity when following an underlying *w*.

行い	okonaˈˈi
行います	okonaˈɪ·ˈmasu

The rule of h

- は is written as **hȧ** if it is pronounced as *wa*.
- へ is written as **hė** if it is pronounced as *e*.

戦地へ赴く	ˈsenntɪ-hė omoˈˈmuk·u
今日は暖かい	ˈkyou hȧ atataˈˈka·ɪ

THE BOUNDARY FALL (`, ̀)

The basic rules of the boundary fall symbols

- In Kyouro, spaces are dedicated to marking the word-level syntactic boundaries. Where an AP boundary is present without a word-level syntactic boundary coinciding, a boundary fall symbol, ` or ̀, is inserted.

男一匹	otoˈko`ɪxtupɪˈkɪ
日本国語大辞典	nɪˈhonǹkokugoˈ̀daɪˈzɪtenn
現首相	ˈgenǹsyusyouˈ

- A phonetically effective boundary fall is marked with the upper boundary fall symbol (`). Whether and where a space is placed in conjunction with a boundary fall symbol depends on the syntactic context.

咳してる	seˈkɪ` sɪteˈru
行くだろうけど	ɪˈku `daˈroukedo
ダメこれじゃ	daˈme ` koreˈ zya

In this section, spaces adjacent to a boundary fall symbol are shown in grey to avoid potential confusion.

- If a would-be boundary fall is canceled, the AP boundary is marked by the lower boundary fall symbol, regardless of whether and where space is adjacently present. This happens when the form on the left side of the AP boundary ends with an F mora.

本もらった	ˈhonǹ moratˈt·a
帰るだろうから	kaˈˈer·u ̀daˈroukara
帰るから明日	kaˈˈer·u kara ̀ asɪˈta

- If a would-be boundary fall is “converted” to the F tone of the preceding mora (»harmonica), the AP boundary is marked by the upper boundary fall symbol, regardless of whether a space is adjacently present. This happens when the form on the left side of the AP boundary ends with a bimoraic rhyme and the accent is placed at position 1 of the AP.

日本帰った	nɪhoˈnn` kaˈˈett·a
辛いだろうし	tuˈraˈ·ɪ `daˈrousɪ
辞めたいこんなの	yamė̩·ˈtaˈ·ɪ ` konˈnano

The A1 boundary fall

- The A1 boundary fall arises at an AP boundary where the left AP is |A1, not counting zeros between A1 and A4.
- The A1 boundary fall is marked by a boundary fall symbol with a space only on the right side.

山登った	yaˈma` nobotˈt·a
頭ぶつけた	ataˈma` butukėˈt·a
家買った	ɪˈė` katˈt·a

For this type of boundary fall to be phonetically effective, the left AP's accent number must be 1.

富士山登った	ˈhuzɪsann` nobotˈt·a
肩ぶつけた	ˈkata` butukėˈt·a
車買った	kurumaˈ` katˈt·a

When the left AP |A1 is not a one-word AP, which happens when *ka* (**ka** C2|A1 or **ka** R1|A1) is involved, both an active and inactive boundary falls seem to be accepted (my introspection into the following examples, December 2023):

- *nande yatta no ga jibun ka wakaranai no?*
- *nande jūsho ga amerika ka wakatta no?*
- *nande rainenno torendo ga endaka ka shirabeteru no?*

In general, if multiple pronunciations are possible, any of the possible forms that can be expressed in Kyouro can be chosen. For this specific case, if the student needs to understand the precise rules of the boundary fall, the following explanations may be worth considering:

- The A1 boundary fall can be phonetically effective only if the left AP is a one-word AP.
- When the left AP is of more than one word and the AP's accent number is 1, both the active and inactive A1 boundary falls can be accepted.
- The A1 boundary fall can be phonetically effective if the left AP's accent number is 1 regardless of the number of words the AP consists of.

unless better evidence for or against any of these options is available.

The G1 boundary fall

- The G1 boundary fall arises at an AP boundary where the left AP is |G1, not counting zeros between G1 and G2.
- The G1 boundary fall is marked by a boundary fall symbol with a space only on the right side.

日本を旅した	nɪhoˈnn wo taˈbɪ` sɪˈta
楽団を指揮した	gakudannˈ wo sɪˈkɪ` sɪˈta
和解を意図した	wakaɪˈ wo ˈɪto` sɪˈta

Lexical boundary falls

- A boundary fall in a tokenized form is written as part of the token. This includes previously excluded instances where a boundary fall is found at the right edge of a token.

次私の番ね	*tɯˈgɪ` watasɪ₊-ˈno ˈbann ˈne*
私にだけ教えてくれた	*watasɪˈ-nɪ `daˈke osɪėˈt·e kurėˈt·a*
もう終わるでしょうから	*ˈmou oˈwaˈr·u `deˈsyoukara*

THE INVERTED QUESTION MARK (¿)

The usage of the inverted question mark

- The inverted question mark ¿ is used to indicate the R% intonation given to a predicate.
- ¿ is attached to the beginning of the corresponding word. The absence of ¿ indicates the absence of R%.
- In version 4 and earlier, the R% intonation was indicated by the regular question mark (?). As of version 5, the symbol '?' is dedicated to transcribing the question mark of the source text, regardless of whether R% is present.

透くん見なかった？	toorukunnˈ、¿ˈˈmi·ˌnakatta?
どうすればいいでしょうか？	ˈdousureba ˈıı desyouka?
帰るよ	kaˈˈer·u ¿ˈyo

GENERAL SPELLING GUIDELINES

Source text

- The source text must be written in a variety of Japanese that is compatible with the grammatical description of Japanese as part of the theory of Kyouro. The speller is responsible for judging whether the sentence to be spelled is compatible with Kyouro.

- The source text may be pre-modified to simplify the Kyouro spelling to the extent that the pronunciation and the literal meaning are preserved. Whether the source text should be modified to a simpler style or the speller should be letter-to-letter faithful to the source text depends on the purpose and the nature of the transcription and the source text. However, since only specified symbols are available in Kyouro, in many instances it is impossible to be completely faithful to the source text. Examples:

- ねえ (ˈ**nexe**) and ねー (ˈ**ne—**) can be replaced with ねえ (ˈ**nee**).

- Different styles of quotation marks (“ ”, ‘ ’, 『 』, 【 】,etc.) can be replaced with the regular Japanese quotation marks: 「 」(**[ ]**).

- Different styles of parentheses (“( )”, “ 《 》 ”, “ 〈 〉 ”, etc.) can be replaced with the regular parentheses: ( ).

- The question exclamation mark (?!) and other versions of the question mark or the exclamation mark can be replaced with the regular question mark (?), the regular exclamation mark (!), or a combination of them (?!, !!, ??, etc.).

- A nonstandard spelling, such as writing a native word in katakana for the effect of a foreigner’s speech or writing a kanji word in hiragana for the effect of a child’s speech, can be replaced with the standard spelling.

- When the source text allows multiple pronunciations, any one of them that can be transcribed in Kyouro can be chosen.

- Numerals are spelled out.

  - 3歳 and 3丁目 are written as ˈ**sannsaɪ** and **sanntyou**ˈ**me**

Limitations of the present analyses of boundaries

The boundary system we have laid out captures some aspects of the syntax of the language, but it is by no means a complete description. Several factors are left out of the analysis and the theory needs to be modified to accommodate them.

One of the yet-to-be-described factors is polarity. It is well-known that some items require a negator. Such items are called negative polarity items (NPI). One of the NPIs is *shika*:

- *shinigami wa ringo shika tabenai.*

(Death Note, Season 1, Episode 10, at about 9 min 10 sec.)

as opposed to:

- **shinigami wa ringo shika taberu.*

A negator is required for each *shika* that appears in the sentence:

- *miyamotosan wa suiyōbi shika shussha shinai.*
- **miyamotosan shika suiyōbi shika shussha shinai.*
- *miyamotosan shika suiyōbi shika shussha shinaku nai.*

Other syntactic factors we have ignored include valence and case.

- *ashita wa dōbutsuen ni iku.*
- **ashita wa dōbutsuen o iku.*
- *keikaku o jikkō shita.*
- **keikaku o jikkō o shita.*

Our syntactic description fails in several other places. Some defects that I was aware of during my editing of this document have received treatment, but many are left unpatched. This is partly due to time constraints, but also partly because I expect that a significant portion of those factors is better handled with semantics.

> Kyouro's semantic component has been in preparation and the initial release is coming hopefully in late 2024.

When a teacher or textbook writer puts a sentence in Kyouro, it is advised that a minor defect should not interfere with a boundary identification. *Taberu* cannot complete a clause that includes *shika,* but *taberu* should still be regarded as A4|R1. The list of the boundaries is given with general characterizations and examples, and the most suitable label for a given syntactic boundary should be chosen from them unless not recognizing the boundary appears to be more appropriate. Analyses provided in this document and other Kyouro documents are meant to serve as references.

Also, note that Kyouro's spelling rules are designed to be tolerant of minor analytic inconsistencies. Whether a certain boundary is A2 or A3 does not change the spelling as both are marked by a space.

If a boundary seems to be not working at a specific place in a sentence, ask if some semantic explanation could rescue it. If a simple and consistent semantic explanation can be provided (e.g., "This verb does not accept a patient," "This is an NPI," etc.), that is, if you can "shift the blame" from syntax to semantics, consider regarding the boundary identification as a pass. If the boundary system appears to be inescapably broken, consider not recognizing some would-be boundaries of the sentence as a potential workaround. (Ultimately, Kyouro allows, on the speller's responsibility, spelling a sentence—no matter how complex it may be—as a unit item, although it is not recommended as the first choice.)

Romanization table

- The following romanization table is recommended.
- The table can be modified based on students' needs. Minimally, the spelling must produce the correct orthographic form on the input system the student is expected to use. (When a specific spelling is given in this section, it is based on the recommended table here. Should the table be modified, change the given spelling accordingly.)
- If yomigana and the real pronunciation (if given) contradict, yomigana is reflected in the spelling.
- *fuinki* or *fun'iki* ('atmosphere') is written **hunnˈıkı** as long as that is the yomigana that is adopted.
- Adaptation of a nonstandard yomigana, such as **huˈinnkı**, is within the speller's discretion.

Tokenized form

- The spelling of a tokenized form represents the unit form of the item according to the syntactic and phonological analyses as illustrated in this document and other official materials of Kyouro.
- Phonological or syntactic environments do not alter the spelling of a tokenized form unless the typeability principle demands so.

	a	ı	u	e	o	ya	yı	yu	ye	yo	n
	あ	い	う	え	お	や		ゆ	いぇ	よ	
k	か	き	く	け	こ	きゃ	きぃ	きゅ	きぇ	きょ	
g	が	ぎ	ぐ	げ	ご	ぎゃ	ぎぃ	ぎゅ	ぎぇ	ぎょ	
s	さ	し	す	せ	そ	しゃ	しぃ	しゅ	しぇ	しょ	
z	ざ	じ	ず	ぜ	ぞ	じゃ	じぃ	じゅ	じぇ	じょ	
t	た	ち	つ	て	と	ちゃ	ちぃ	ちゅ	ちぇ	ちょ	
d	だ	ぢ	づ	で	ど	ぢゃ		ぢゅ	ぢぇ	ぢょ	
n	な	に	ぬ	ね	の	にゃ	にぃ	にゅ	にぇ	にょ	ん
h	は	ひ	ふ	へ	ほ	ひゃ	ひぃ	ひゅ	ひぇ	ひょ	
b	ば	び	ぶ	べ	ぼ	びゃ	びぃ	びゅ	びぇ	びょ	
p	ぱ	ぴ	ぷ	ぺ	ぽ	ぴゃ	ぴぃ	ぴゅ	ぴぇ	ぴょ	
m	ま	み	む	め	も	みゃ	みぃ	みゅ		みょ	
r	ら	り	る	れ	ろ	りゃ	りぃ	りゅ	りぇ	りょ	
w	わ				を		ゐ		ゑ		
wh		うぃ		うぇ	うぉ						
f	ふぁ				ふぉ		ふぃ	ふゅ	ふぇ	ふょ	
v	ゔぁ	ゔぃ	ゔ	ゔぇ	ゔぉ	ゔゃ		ゔゅ		ゔょ	
th	てゃ	てぃ	てゅ	てぇ	てょ						

Accent

- The accent symbols are inserted according to the unit forms and the surface forms.
- An AP has at least one upper accent. If evidence shows the AP lacks lexical accents, an accent is assumed according to the surface form.
- Although not recommended as a primary choice, if necessary, the accent symbols can be omitted altogether. It is also allowed to omit only the lower accent. Mixed notations are not allowed.
- If evidence is agnostic about two possible accent numbers of a form:
  - If it belongs to the verb group, the one that fits best in the lexical system is preferred.

喋った	syaˈbett·a
喘いだ	aˈėid·a
帰った	kaˈett·a

  - If the AP requires a boundary fall symbol at its end, 1 is preferred if the boundary fall is phonetically present, and 0 is preferred otherwise.
  - If it neither belongs to the verb group nor is adjacent to a boundary fall symbol, the greater number is preferred.
- If an accent, lower or upper, coincides with a boundary fall, lower or upper, the accent takes the more internal (further from the space, for example) position.
- If an accent is otherwise placed in the middle of a mora, it is shifted to immediately behind the mora. However, an accent does not shift over an inflection symbol.

Long vowels

- The kana long vowel symbol is written as em dash (—).
- If a long vowel is spelled as two vowels in the original spelling, the latter part of the vowel is written as an offglide (a dotless vowel).
- If a vowel of a token is spelled as a short vowel and is consistently pronounced as a long vowel regardless of the environment, the bar is added to the vowel. E.g., *nīten san* (‘2.3’) is written ˈ**nītenn̩sann**ˈ.
- If a phonetic long vowel is spelled as a sequence of a regular letter and a small letter (e.g., ねえ pronounced *nē*), **x** is used for the small letter.
- いう is written as **ɪwu** (with or without a dot on the **ɪ**) where it is pronounced as *yū*. Non-alphabetic symbols should be placed as per usual.
- A one-mora AP is commonly pronounced longer in the A1 boundary fall (lower or upper) environment. Long vowels that arise for this or comparable reasons are unmarked.

Inflection symbols

- The hyphen-minus (-) is used for the noun group and inserted immediately before the leftmost segment of the suffix.
- The middle dot (·) is used for the verb group.
  - If the suffix's first letter undergoes alternation or deletion, the middle dot is inserted immediately after the segment.
  - Otherwise, the middle dot is inserted to the left of the leftmost segment of the suffix.
- Inflection symbols are deleted if as a result of inflection the symbol is placed at the right edge.

ˈtabė	+	i·	—>	ˈtabė
yoroˈkob	+	i·	—>	yoroˈkobi

- If evidence suggests that an unaccented item is inflectional, it is written as part of a larger token. Inflection symbols cannot be attached to an unaccented form.

Punctuations

- The Japanese quotation marks (「」) are written as square brackets (**[ ]**). The part bounded by the quotation marks can be incorporated into a sentence with various syntactic properties. If it is treated as a noun, the inflectional symbol is directly attached to **]**. »texas.
- The Japanese comma (U+3001 、 IDEOGRAPHIC COMMA) is written as the regular comma (**,**).
- When a comma comes at a syntactic boundary, exactly one space is inserted between the comma and the next word.
- If the comma is part of a token, such as *ni, san* ('two to three'), a space is not inserted around the comma (ˈ**nı,sann**).
- If a comma comes right after a boundary fall symbol, the boundary fall symbol is deleted.
- The Japanese period (U+3002 。 IDEOGRAPHIC FULL STOP) is written as the regular period (**.**).
- Exactly two spaces are inserted between two adjacent orthographic sentences.
- An original period is written according to the pronunciation.
  - If it is part of a number notation (i.e., a decimal point) that is pronounced as *ten*, it is spelled out as *ten* (**tenn**).
- Ellipses (either the Japanese ellipses (U+2026 … HORIZONTAL ELLIPSIS) or consecutive periods (...)) are written as a single period surrounded by spaces ("**.**"). If ellipses are followed by a regular period, it will be SPACE-PERIOD-SPACE-PERIOD ("**..**").
- The Japanese middle dot (・; do not confuse this with the regular middle dot used in suffixes) used in foreign names is written as a slash (/).
  - If an AP boundary coincides with a slash, the lower or upper boundary fall symbol can be on either side of the slash.

Small tsu (doubled letters, xtu, and k)

- If the small tsu in question corresponds to a katakana or hiragana:
  - double the following letter where it produces a small tsu (**ai ˈpaddo**, **kamuˈtyakka**, **hıyokkoˈ**).
  - use **xtu** where necessary. (**ˈıxtunu**, **hagaxtunya**)
- (The rule of K): If the small tsu in question is at the end of an on'yomi,
  - if the kanji's corresponding on'yomi in isolation ends with *ku* and the next letter is **k**, then use **k**
  - otherwise, use **xtu**.

The original version of the rule of K was inspired by the following tweet of @sosoBOTpi: https://x.com/sosobotpi/status/1055734601031704576?s=61&t=P5zngLY-0SHK2OgjuSYsug

Proper names

When the title of a literary work, such as a film, a novel, or a manga, is used as a noun to refer to the work in a sentence, it is treated as a unit item as it is obviously a proper name. Hence, referring to *Boku Dakega Inai Machi,* aka *Erased*, by Kei Sanbe, "I read *Erased*." can be translated as **ˈboku˵daˈkega˵ınaımaˈtı wo ˈyonnd·a.**

When it is treated as an autonomous sentence, however, it can be written as a sentence itself. *Boku Dakega Inai Machi* can be called in Kyouro **ˈboku ˵daˈkega ınaˈı maˈtı**. This includes when the title is bounded by the Japanese quotation marks (「」)—like how an utterance can be incorporated into a sentence with quotation marks. The aforementioned sentence can be written as **[ˈboku ˵daˈkega ınaˈı maˈtı] wo ˈyonnd·a.**

When the part bounded by **[ ]** is incorporated as an inflectional, phonological processes, including accent switching, may be evaded. If this happens, the spelling form should reflect the typed and phonetic forms:

- "*boku dakega inai machi*"*no syujinkō*.
- **[ˈboku ˵daˈkega ınaˈı maˈtı]-˵no syuˈzınnkou.**

Capital letters

- IP boundaries are indicated by capitalizing the first letter of each IP. If necessary, IP boundaries can be left implicit, in which case no capital letters are used. (Mixed notations are not allowed.) Most of the sentences provided as examples in this document are given without capital letters because the IP boundary positions depend on the placement of $. When transcribing a sentence with a reference audio and sufficient contexts, the full notation, where all relevant symbols including capital letters are appropriately used, is recommended as the default option.

NOTES AND REFERENCES

REFERENCES

References

- Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. Harper & Row.
- Fujita, Y. (1994). 引用されたコトバの記号論的位置づけと文法的性格. 『詞林』 16, pp. 73-85. 大阪大学古代中世文学研究会.
- Hasegawa, Y. (1995). Against Marking Accent Locations in Japanese. *Japanese-language Education Around the Globe*, Vol. 5, 95-103
- Hata, Y. (2012). 日本語の文末音調の機能 : ノダ文における発話音声の音響分析を中心に. 創価大学大学院紀要. Volume 34, pp. 431-455.
- Hattori, S. (1961). Prosodeme, Syllable Structure and Laryngeal Phonemes. [日本の言語学 第二巻 音韻]
- Iida, J. (2021). Prosody of Tokyo Japanese Revisited: “Boundary Fall” in the Sentence-Level Pitch Contour. *2021 UWM VIRTUAL Undergraduate Research Symposium*.
- Ikeuchi, M. & Goro, T. (2013). 『生成文法研究の現在』ひつじ書房.
- Iori, I. (2017). 5. The layered structure of the sentence. [Shibatani, Miyagawa, & Noda (2017), pp. 157-186].
- Iori, I. (2012). 『新しい日本語学入門（第2版）』. スリーエーネットワーク
- Kawakami, S. (1953).「花高し」と「鼻高し」.『首声学会会報』82.
- Kawakami, S. (1963). ピッチグラムで見た日本語のアクセント. 『音声の研究』10.
- Kawakami, S. (1985). 書評・紹介　Y. Higurashi著　The Accents of Extended Word Structures in Tokyo Standard. 『言語研究』87 号 pp. 156-171.
- Kuroki, K. (p.c.). A conversation on Twitter in 2022. https://x.com/croquis_kuni/status/1631342017224802305

- Kurosawa, A. (2008). 二重ヲ格は会議でどのように使われてきたか. 『日本語教育連絡会議論文集 v.20』
- 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
- Kori, S. (1997). 「当時の村山首相」の2つの意味と2つの読み　名詞句の意味構造とアクセント弱化について. 『文法と音声』くろしお出版, pp.123-146.
- Kori, S. (2004). 東京アクセントの特徴再考. 『国語学』第55巻2号
- Kori, S. (2008a). 東京方言におけるアクセントの実現度と意味限定. 『音声研究』第12巻1号
- Kori, S. (2008b). 東京方言における平叙文末の下降増大現象　平叙文末は平調か下降調か. 『音声言語VI』
- Kori, S. (2011). 東京方言の平叙文文末に見られる下降増大現象について(再考).『音声研究』 15 巻, 3 号, pp. 76-77.
- Kori, S. (2017). Accent Reduction and Restrictive Modification in Tokyo Japanese. *Journal of the Phonetic Society of Japan*. Volume 21, Issue 1, pp. 105-126.
- Kubo, M. (1994). *Japanese Syntactic Structures and Their Constructional Meanings*. ひつじ書房
- Lianping, L. (羅蓮萍). 「ナ形容詞」と「ノ形容詞」のイメージ：日本語母語話者の使用意識. 『山口国文』 Volume 32, pp. 120-110. 山口大学人文学部国語国文学会.
- Maekawa, K. (2018). アクセント句頭のFo上昇は条件異音ではない. 日本音声学会第32回全国大会予稿集, pp.78-83.
- Masuoka, T. & Takubo, Y. (1992). 『基礎日本語文法—改訂版—』くろしお出版.
- McCawley, J. (1968). *The Phonological Component of a Grammar of Japanese*. Mouton.
- Minami, F. (1974). 『現代日本語の構造』大修館書店.
- Minami, F. (1993). 『現代日本語文法の輪郭』大修館書店.
- Nespor, M. & Vogel, I. (1986) *Prosodic Phonology*. [Nespor & Vogel (2007)]
- Nespor, M. & Vogel, I. (2007). *Prosodic Phonology: With a New Foreword*. Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110977790
- Okada, H. (1999). Japanese. [Handbook of the International Phonetic Association.]

- Saito, M. (2013). 日本語埋め込み文の意味的・談話的性質–比較統語論への招待. [Ikeuchi and Goro 2013], pp. 221-251.
- Saito, S. (2006). 『日本語音声学入門』 三省堂.
- Selkirk, E., Shinya, T. & Sugahara, M. (2003). Degree of Initial Lowering in Japanese as a Reflex of Prosodic Structure Organization.
- Shibata, T., Kitamura, H. & Kindaichi, H. (Eds.). (1980)『日本の言語学 第二巻 音韻』大修館書店.
- Shibatani, M., Miyagawa, S., & Noda H. (Eds.). (2017). *Handbook of Japanese Syntax Volume 4*. De Gruyter, Mouton.
- Masayoshi Shibatani Shigeru Miyagawa Hisashi Noda
- Sugito, M. (1969). 音程動態測定による大阪・東京アクセントの一考察 “A Study on the Accentual System of the Tokyo and Osaka Dialects with Regard to Pitch Movement.” 『国語学』第79集, pp. 71-89.
- Takahashi, T., Kaneko, N., Kaneda, A., Sai, M., Suzuki, T., Suda, J., & Matsumoto, H. (2005). 『日本語の文法』ひつじ書房
- Tronnier, M., Dantsuji, M. (1993). Some Acoustic Characteristics of Glottal and Palatal Fricatives in Japanese and German. *Lund University, Dept. of Linguistics 213 Working Papers* 40, pp. 213-228
- Tserdanelis, G., Beckman, M., Kong, E., Li, F., & Syrika, A. (2005). Comparing Greek, English and Japanese velar and palatal(iz)ed stops. *7th International Conference in Greek Linguistics*. The University of York.
- *Handbook of the International Phonetic Association : a guide to the use of the International Phonetic Alphabet*. (1999). International Phonetic Association.

Dictionaries and literary sources

- 雫井脩介 (2017)『合本　検察側の罪人【文春e-Books】』文藝春秋
- 久美沙織 (2001)『丘の家のミッキー1』Apple Books
- 平野啓一郎 (2019)『マチネの終わりに』 Apple Books
- 『小学館 韓日・日韓辞典』(2012) 小学館 [Dictionaries]
- *A Handbook of Japanese Grammar Patterns for Teacher's and Learner's*. (2023). [Dictionaries]
- *Kenkyusha's New Japanese-English Dictionary 5th Edition*. (2021) [Dictionaries]
- *Kenkyusha's New English-Japanese Dictionary 6th Edition*. (2020) [Dictionaries]
- Monokakido Co. Ltd. (2019). *Dictionaries*. AppStore
- 『NHK 日本語発音アクセント新辞典』 (2016) [Dictionaries]
- PUNPEE. (2020) "Gizmo (Future Foundation)." Apple Music (Lyrics only).
- 日本テレビ放送網株式会社, 有限責任事業組合D.N.,ドリームパートナーズ, バップ, 集英社, pマッドハウス. (2006-2007) *Death Note* Amazon Prime Video.
- テレビ朝日, 東宝. (2000)『TRICK』Abema TV
- テレビ朝日, 東宝. (2003)『TRICK 3』Abema TV

APPENDIX

- **noka** ‘oh’ • **nnda** ‘id.’
- **ka** ‘oh’
- **darou** ‘I suppose’ • **desyou** ‘id.’
- **zyanaıka** ‘after all, it’s just …’ • **zyanaıno** ‘id.’ • **zyann** ‘id.’ • **kayo** ‘id.’
- **nara** ‘if it is’ • **nannte** ‘no way’ • **sıka** ‘only’ • **kosohả** ‘at least’ • **made** ‘through’
- **desıta** ‘was’ • **desu** ‘is’
- **hả** ‘especially’ • **mo** ‘also’ • **koso** ‘particularly’
- **ka** ‘whether’
- **da** ‘is’
- **tteba** ‘I’m telling you’ • **yo** ‘I tell you’
- **kke** ‘did you say that … ?’ • **monn** ‘because (I insist that) … !’
- **nı** ‘to’
- **na** ‘because it is (state)’
- **desıtaka** ‘was it?’ • **desuka** ‘is it?’
- **wo** ‘(object), but’ • **ga** ‘(subject),

R1	R2	A1	A2	A3	A4 (C2b)	C1	C2	C3	C4 (R2c)	R2b	P1 (R2d)	P2	P3	P4 (C3b)	P5

- **no** ‘the fact that’
- **de** ‘given that … is’ (literary)
- **dato** ‘if’ • **zya** ‘id.’ • **dattaraba** • desıtara • **desıtaraba** • **dattara**
- **datta** ‘was’
- **no** ‘given that; (the amount) of; (shared possession)
- **nn** ‘because’ • **no** ‘id.’
- **demo** • **dehả** • **zya** • **nara** • **sıka** • **nannte** • **hả**
- **kara** ‘because’
- **nı** ‘(becomes) into’
- **nara** ‘if’ • **naraba** ‘id.’
- **ka**
- **nonı** ‘despite’ • **node** ‘because’ • **noka**
- **karanıhả** • **tte**
- **to** ‘(think) that …’
- **keredomo** • **ga** • **kedomo** • **kedo** • **keredo** • **sı** • **kara**
- **tte** • **to** ‘(say) …’

Note:
This chart was created as a draft of the “map” on p. kangaroo based on 『詞と語について』(ver. 3.0). This page is kept for the record and the aforementioned map should be primarily referenced. Some changes since ver. 5.0 may be not reflected here.
Kyouro is in boldfaced letters.
Above the grey line, the line below the the list of forms indicates the left and right boundaries of the forms. The line is aligned to the grey line so that the line’s left edge indicates the left boundary, and the right edge the right boundary.
Below the grey line, the line above the list of forms indicates the boundaries. The line’s right edge indicates the left boundary, and the left edge indicates the right boundary.
Errors are expected.

GLOSSARIES

Notes

asdlkfj dk nf;asl d;lkasfj ad;lskdjf

Tests

- Kara test
  - *ashi ga samuku naranakatta kara*
  - **asa ga samuku naranakatta kara*
- Nande test

Sentences and nonsentences

This subsection lists examples. Some of them are not used in the body. Romanizations and translations are mine unless otherwise noted.

- *amarino utsukushisani iki o nonda.*
- *daisukina karē o tabesae shinakatta.*

Daisukina	karē o		tabesae	Shinakatta
		A4/R1		
		A4/G2		
	./A4			
	R2/m-ii			
./R2	R2/m-ii	A1\A4	A4/G2	G2\A4/R1
daisukina	karē	o	tabesae	shinakatta

- *kare wa mukashikara shōgi ga tsuyoku, bokunado wa, nimaiochi de yatto ii shōbu toiu tokoro datta.*
- *kodomo futario daigaku ni ikaseru tameni tochi o urisae shita.* (»snapshot)
- **konankun wa kuruma o sukoshi no unten shinai.*
- *moshi samukattara kore kite*. In case you're cold, wear this.
- *moshi samukereba kore kite*. If you feel cold, wear this.
- *onegai shite kureru wake nai janaidesuka*. (TRICK 3, ep. 2, 00:25:33). "There's no way she'll give it to you just because you ask her."
- *shimarisukun wa asobini muchūni naranakatta kara.*
- *sōda, takara sagasanakya*. (TRICK, ep. 10, 00:38:37)
- *tomodachini ayamarisae shinakatta.*

Tomodachini	ayamarisae	Shinakatta
	A4/R1	
	A4/G2	
./A2	A4/G2	G2\A4/R1
tomodachini	ayamarisae	shinakatta

- *tomodachini shazai sae shinakatta.*

*	Tomodachini	shazai sae	Shinakatta
	Tomodachini	Shazai sae	Shinakatta

	A4/R1		
		A4/R1	
	./G2		
./A2	./G1	G1\G2	G2\A4/R1
tomodachini	shazai	sae	shinakatta