Simplex Truncation in Japanese: A Dictionary-Based Survey of Simplex and Pseudo-Compound Words

Akiko S. TANAKA

Abstract: Simplex truncation is a highly productive word formation process in Japanese, where words with five or more moras are shortened to forms with four moras or fewer. Numerous studies have been conducted to predict a single output for a given input, but each faces its own set of challenges. To empirically evaluate the mechanisms proposed thus far, we conducted a dictionary-based survey and examined three key hypotheses: the accent-based hypothesis, the pseudo-compound hypothesis, and the tripartite structure hypothesis. The results of our survey indicated that simplex truncation is best explained by the tripartite structure hypothesis, which posits that the prosodic structure of a word comprises three components. Additionally, we identified the possibility that Japanese contains true pseudo-compounds—words whose underlying structure is bipartite.

Key words：simplex truncation, loanword, prosodic structure, pseudo-compound, Japanese

1. Introduction: General constraints on simplex loanword truncation in Japanese

Simplex truncation is a word formation process that typically shortens words of five moras or more into forms with four moras or fewer. Citing examples from Japanese and other languages, Kubozono (2002: 89–94, 2023: 147–9)[1),2)] highlighted that (i) longer words are more likely to be truncated, (ii) frequently used words tend to undergo truncation, and (iii) the initial part of the word is typically preserved. This process is highly productive in Japanese, where many truncated forms are derived from simplex loanwords, which are often longer than native words. Some examples are provided in (1), with dots /./ denoting syllable boundaries.*1

(1) a. su.to.rai.ki “labor strike” → su.to
in.te.ri.gen.cha “intelligent people” → in.te.ri
b. pu.ro.fii.ru “profile” → pu.ro.fu
sai.ke.de.rik.ku “psychedelic” → sai.ke
c. bi.ru.din.gu “building” → bi.ru
de.mon.su.to.ree.shon “demonstration” → de.mo

Several studies have examined simplex truncation, with Itô (1990)[3)], Kubozono (2003)[4)], and others demonstrating that the constraints outlined in (2) (Kubozono 2020: 26)[5)] accurately predict the resulting truncated forms.

(2) a. Five-mora or longer words are subject to truncation
b. Outputs are at most four moras long
c. Outputs are at least two moras long
d. Outputs are at least two syllables long
e. Outputs may not be Light-Heavy bisyllables

The input constraint in (2a) and the output constraint in (2b) are specific to Japanese. Kubozono (2004, 2005)[6),7)] and Kubozono and Ogawa (2005)[8)] affirmed that the distinction between words with four or less moras and those with five moras or more is evident in other Japanese phonological phenomena, including compound truncation, phone number accentuation, onomatopoeia pronunciation, and rendaku.*2

While the constraints in (2) accurately predict the observed truncated forms in (1), they do not eliminate all forms that do not actually occur. Some of them are correctly eliminated, but others are not, as illustrated in (3), where the former is denoted by /*/ and the latter by /ˀ/.

(3) a. su.to.rai.ki → su.to, *su, ˀsu.to.ra, *su.to.rai
in.te.ri.gen.cha → in.te.ri, *i, *in, ˀin.te, *in.te.ri.ge, *in.te.ri.gen
b. pu.ro.fii.ru →pu.ro.fu, ˀpu.ro, ˀpu.ro.fii
sa.i.ke.de.rik.ku → sai.ke, *sa, ˀsa.i,

（2024年10月１日受付，2024年12月12日受理）

ˀsa.i.ke.de, *sa.i.ke.de.ri

c. bi.ru.din.gu → bi.ru, *bi, ˀbi.ru.di, ˀbi.ru.din
de.mon.su.to.ree.shon
→ de.mo, *de, *de.mon, ˀde.mon.su,
*de.mon.su.to, *de.mon.su.to.ree

The fact that most simple words have only one truncated form suggests that the constraints in (2a–e), which predict multiple outputs for a single input, cannot fully explain the truncation process. To address this limitation, Labrune (2002)[11], Kubozono (2004)[6], Kubozono and Ogawa (2005)[8], and Kubozono (2020)[5] among others proposed various truncation mechanisms. However, as we will demonstrate, each mechanism produces exceptions and presents both advantages and disadvantages. We conducted a dictionary-based survey to empirically evaluate the three mechanisms, hypothesizing that output forms can be determined by the underlying prosodic structure of the words. In Section 2, we review how each mechanism determines output length. Section 3 provides a comparative analysis of the explanatory power of Labrune's (2002)[11] and Kubozono's (2020)[5] mechanisms based on the collected data. In Section 4, we discuss the predictability of the mechanism proposed by Kubozono (2004)[6] and Kubozono and Ogawa (2005)[8]. Finally, Section 5 offers a brief conclusion to this paper.

2. Previous studies

2.1 Labrune's (2002)[11] accent-based hypothesis

Labrune (2002)[11] addressed the issue raised in the previous section from the perspective of accentuation and proposed the hypothesis in (4) (Labrune 2002: 106)[11].

(4) [T]he base is truncated immediately before the accented mora.

The interaction between the output constraints in (2b–e) and the constraint in (4) predicts the truncations presented in (5), wherein the apostrophe /'/ indicates the placement of the accent, denoting where the pitch falls.

(5) a. su.to.ra'i.ki → su.to
in.te.ri.ge'n.cha → in.te.ri

b. pu.ro.fi'i.ru → *pu.ro
sai.ke.de.ri'k.ku → *sai.ke.de

c. bi'.ru.din.gu → *ø → (MR) bi.ru
de.mon.su.to.re'e.shon → *de.mon.su.to

While Labrune's accent-based hypothesis can predict the truncations in (5a), it fails to accurately predict the outputs from the base words in (5b) and *demonsutoreeshon* in (5c). If the base was truncated right before the accented mora *bi* of *birudingu* in (5c), the entire word would be eliminated; thus, minimal repair (MR) is implemented to ensure that the shortest candidate meets the constraint in (2c).

As we will discuss later, not only Labrune's mechanism but also the other two are not free from exceptions. The issue, as Kubozono and Ogawa (2005)[8] affirmed, is that Labrune's analysis does not explain why the accented nucleus—i.e., the most critical part of a word—would be deleted.*3 Furthermore, this hypothesis does not account for the input constraint in (2a) or the output constraint in (2b). For instance, the incorrect output *demonsuto* in (5c), which represents a non-existent four-syllable word, cannot serve as the input for a new truncation process because this hypothesis relies on actual word accents.

2.2 Kubozono (2004)[6] and Kubozono and Ogawa's (2005)[8] pseudo-compound hypothesis

Kubozono (2004)[6] and Kubozono and Ogawa (2005)[8] proposed a hypothesis that addresses two general constraints that the accent-based hypothesis cannot accommodate. They argued that five-mora words or longer are phonologically complex—essentially pseudo-compounds—despite being morphologically simplex. According to Kubozono (2002)[1], Japanese employs two truncation mechanisms for compound words. The examples in (6a) demonstrate that the elements following the compound boundary /#/ are deleted, with the remaining element serving as the truncated form. The examples in (6b) illustrate that the elements before and after the compound boundary are partially combined, resulting in a shortened form.

(6) a. kee.tai # den.wa "mobile phone" → kee.tai
suu.paa # maa.ket.to "supermarket" → suu.paa

b. po.ket.to # mon.su.taa "pocket monster"
→ po.ke.mon
a.ru.koo.ru # chuu.do.ku "alcoholic"
→ a.ru.chuu

Kubozono (2004)[6] argued that the simplex truncation process is nearly identical to the compound truncation

process depicted in (6a). According to Kubozono (2023: 166)[2], the principles involved in splitting words with five moras or longer are those listed in (7), which are observed in various phonological phenomena in Japanese.*4

(7) a. Pseudo-compound boundaries do not break up syllables and always align with syllable boundaries.
b. Both parts should be of equal length, or the first half should be longer than the second.

Of the two constraints mentioned, the constraint in (7a) dominates the one in (7b). These constraints work with the undominated constraints in (2a-e) to predict the truncated forms in (8), where /+/ denotes the pseudo-compound boundary.

(8) a. su.to + rai.ki → su.to
in.te.ri + gen.cha → in.te.ri.
b. pu.ro + fii.ru → *pu.ro
sai.ke.de + rik.ku → *sai.ke.de
c. bi.ru + din.gu → bi.ru
de.mon.su.to+ree.shon → *demonsuto
→ *de.mon→(MR) de.mo

In each case, the element following the pseudo-compound boundary is removed, while the word-initial part is retained as the output form.*5 Considering that the first half of *demonsutoreeshon* is *demonsuto*, which is five moras long, it undergoes further truncation; the output form becomes *demon*, with the correct form *demo* resulting by means of an MR.

Similar to the accent-based hypothesis, the pseudo-compound hypothesis makes incorrect predictions, such as those represented in (8b). The strength of this hypothesis is that it can explain why the input must be five moras or longer and why the output must be four moras or shorter. Moreover, it can generalize the truncation process for both morphologically simplex and complex words.*6

2.3 Kubozono's (2020)[5] tripartite structure hypothesis

Kubozono (2020)[5] proposed that the default prosodic structure of words in Tokyo Japanese is the tripartite structure illustrated in (9) (Kubozono 2020: 120)[5].

(9) { ..... } { H / LL / HL } { H / L } #
Pre-head Head Tail

This structure comprises three components and is calculated from the end of the word. "Tail" corresponds to the final syllable of the word regardless of whether it is heavy (H), i.e., bimoraic, or light (L), i.e., monomoraic. "Head," the most crucial part of the word, must be filled by a heavy syllable, a LL or a HL bisyllable. The remaining syllables are inserted in "Pre-head," which may be empty. This structure is based on Tanaka's (2008)[14] observation of players' names in baseball chants, as illustrated in (10), where the name of a batter is inserted in the three X slots.

(10) *Kattobase, X-X-X! Pitchaa taose-yo!* "Hit a home run, X-X-X. Beat the pitcher."
cf. *Kattobase, Ro-baa.to-son!* "Hit a homerun, Robertson."

Kubozono (2020)[5] elucidated that the default structure in (9) can also explain the accentuation of various types of words and the loanword truncation process. He also claimed that an accent falls on the initial syllable in the Head. This generalization aligns with the default accent rule in Tokyo Japanese, as outlined in (11), which essentially mirrors the same rule found in Latin (Kubozono 2008, 2023: 173).[15),2)]

(11) Accent on the rightmost, non-final foot.
cf. kyo'to "Kyoto", na'goya "Nagoya",
ro.san.ze'.ru.su "Los Angeles"

With the default structure in (9), the truncation process can be explained by deleting the Head and the Tail. This is illustrated in (12), where any part predicted to be eliminated is crossed out.

(12) a. {su.to}{rai}{ki} → {suto}~~{rai}{ki}~~
{in.te.ri}{gen}{cha} → {in.te.ri}~~{gen}{cha}~~
b. {pu.ro}{fii}{ru} → *{pu.ro}~~{fii}{ru}~~
{sai.ke.de}{rik}{ku} → *{sai.ke.de}~~{rik}{ku}~~
c. {bi.ru.}{din}{gu} → {bi.ru}~~{din}{gu}~~
{de.mon.su.to}{ree}{shon}
→ *{de.mon.su.to}~~{ree}{shon}~~
→ *{de}~~{monsu}{to}~~ → (MR) {de.mo}

In the case of *demonsutoreeshon* in (12c), the monomoraic

de is produced through a two-step truncation process. An MR is implemented to ensure that the shortest candidate fulfilling the undominated constraints is selected as the optimal output. Similar to the pseudo-compound hypothesis, the tripartite structure hypothesis offers principled accounts for the output condition in (2b). However, it shares a common weakness with the accent-based hypothesis: the inability to explain the input condition for simplex truncation in (2a).

As outlined, each hypothesis possesses both strengths and weaknesses. In the following sections, we will evaluate the explanatory power of these three hypotheses.

3. Data and analysis

3.1 Grouping for hypothesis testing

To examine the predictability of the three hypotheses outlined in the previous section, a dictionary-based survey was conducted. The *Sanseido Pocket Katakanago Jiten*, Premium Second Edition[16] was used as the data source, and only truncations with simplex loanwords as their base were collected. From a total of 161 five-mora or longer words, we excluded two, *entaateiment* "entertainment" → *entame* and *bookaroido* "Vocaloid" → *bokaro*, because in these cases, it is not only the first part of the words that is retained as truncations.*7 Table 1 presents the overall results.*8

Table 1. Number of predictable and unpredictable words using each truncation mechanism

Number of the simple words for which ...		Total
The accent-based hypothesis predicts correct truncations	100	159
The accent-based hypothesis predicts wrong truncations	59	
The pseudo-compound hypothesis predicts correct truncations	79	159
The pseudo-compound hypothesis predicts wrong truncations	80	
The tripartite structure hypothesis predicts correct truncations	107	159
The tripartite structure hypothesis predicts wrong truncations	52	

Given the size of the dictionary (approximately 11,000 headwords) and its policy of omitting "*katakana* words that are too obvious and common" (as noted in the explanatory notes), evaluating the explanatory power of these hypotheses based solely on the number of words each hypothesis can explain is challenging.*9 To address this, we created the following word grouping to better capture the characteristics of each hypothesis.

Table 2. Grouping to discuss the explanatory power of the three hypotheses

Number of the simplex loanwords for which ...	
All three hypotheses predict correct truncations	58
All three hypotheses predict wrong truncations	39
Only the accent-based hypothesis predicts correct truncations	2
Only the accent-based hypothesis predicts wrong truncations	10
Only the pseudo-compound hypothesis predicts correct truncations	9
Only the pseudo-compound hypothesis predicts wrong truncations	38
Only the tripartite structure hypothesis predicts correct truncations	1
Only the tripartite structure hypothesis predicts wrong truncations	2
Total	159

The 97 words in the first two groups of Table 2 share 30 syllable structures; however, specifics are not included here due to space constraints. What is crucial for our study is the syllable structures of the remaining 62 words, as each hypothesis is characterized by either the truncations it can predict or those it cannot.

3.2 The disadvantage unique to the accent-based hypothesis

The two truncations that only the accent-based hypothesis can predict are listed in (13) with their syllable structures, which are not shared by other words in the present study.

(13) LLLHHL: koresupo'ndensu "correspondence"
→ koruresu
LLHLLL: purofe'sshonaru "professional"
→ puro

Of these, *puro* may have been borrowed directly from the English abbreviation *pro*. It is also possible that the base word of *koruresu* is *koruresupondensu*, which is the spelling pronunciation of *correspondence*. In this case, the pseudo-compound hypothesis may also make a correct prediction: *koruresu+pondensu* → *koruresu*. Thus, it cannot be said that the ability to explain these two truncations is an advantage of the accent-based theory alone.

Conversely, the existence of truncations that only this

hypothesis cannot explain is a possible disadvantage of the hypothesis. The ten base words in question have one of the six syllable structures in (14), where examples are listed with their attested truncations in brackets.

(14) a. HLLHL: konsa'rutanto "consultant" → ʔkon [konsaru]
LLLHL: sa'purimento → *ø → (MR) ʔsapu [sapuri]
b. LHLHL: depa'atomento "department store" → *de → (MR) ʔdepaa [depaato]
LHLLHH: demonsutore'eshon "demonstration" → ʔdemonsuto [demo]
c. LLHHL: aguri'imento → ʔagu [agurii]
LLLLLL: toradi'shonaru → ʔtora [toraddo]

The syllable structures in (14a) are shared by nine words. As the examples in (15) illustrate, not only the accent-based hypothesis but also the other two hypotheses make incorrect predictions regarding truncations.

(15) HLLHL: anpurifa'ia / anpuri+faia / {anpuri}{fai}{a} "amplifier" → ʔanpuri [anpu]
LLLHL: ruporuta'aju / ruporu+taaju / {ruporu}{taa}{ju} → ʔruporu [rupo]

Moreover, *agurii* and *toraddo* in (14c) may be borrowed directly from the English verb *agree* and the English abbreviation *trad*. Even without considering these points, the inability to explain the truncations in (14a, b) is a serious flaw in the accent-based hypothesis.

One possible explanation for why only the accent-based theory fails to explain the truncations in (14) is that the accent of the base word shifts forward from the default position predicted by the accent rule in (11) discussed in the previous section. This is illustrated in (16), where the predicted foot is indicated by parentheses /( )/, and the syllable with the actual word accent is highlighted in bold.

(16) kon**sa**ru(tan)to, **sa**puri(men)to, de**paa**to(men)to, agu**rii**(men)to, tora**di**(shona)ru

It is now clear that the base word is truncated immediately before the predicted foot, not the actual accent foot. The accent-based hypothesis cannot predict the truncations in (14) because it discusses the morphological process of loanword truncation from the perspective of accentuation. Before proceeding to the next subsection, it is important to note that the tripartite structure hypothesis accurately predicts the outcomes for the words in (14), as illustrated by the examples in (17).

(17) {konsaru}{~~tan~~}{~~to~~}, {sapuri}{~~men~~}{~~to~~}, {depaato}{~~men~~}{~~to~~}, {agurii}{~~men~~}{~~to~~}, {toradi}{~~shona~~}{~~ru~~}

3.3 The advantage unique to the tripartite structure hypothesis

The discussion in the previous subsection leads us to conclude that the tripartite structure hypothesis can predict a wider range of truncations than the accent-based hypothesis. The truncations in (18) are the only cases for which the tripartite structure hypothesis cannot provide an explanation.

(18) a. LLLLH: {sere}{buri}{tii} "celebrity" → *sere [serebu]
b. LLLLLH: {ekisu}{poji}{shon} "exposition" → *ekisu [ekisupo]

Both truncations may have been borrowed directly from the English abbreviations *celeb* and *expo*. The inability to explain these two truncations is not a flaw in this hypothesis. Note that six words have the same syllable structures as the words in (18). These words exhibit the default accent placement, and both the accent-based hypothesis and the tripartite structure hypothesis accurately predict the truncated forms, as illustrated in the examples in (19).

(19) a. LLLLH: heriko'putaa / {heri}{kopu}{taa} "helicopter" → heri
b. LLLLLH: porunogu'rafii / {poruno}{gura}{fii} "pornography" → poruno
cf. HLLLH: kontinyu'itii / {konti}{nyui}{tii} "continuity" → konte

The truncation that can be explained only by the tripartite structure hypothesis is the one in (20) below. None of the words in this study shares the same syllable structure.

(20) LLLLLLLHL: {risutorakuchua}{rin}{gu} "restructuring" → {risutora} {kuchu}{a} → risutora

The four-mora truncation is accurately predicted by a

two-step truncation process from the underlying prosodic structure. The fact that there is only one truncation that only this hypothesis can explain may not in itself be a strength of this hypothesis. What is important is that the tripartite structure hypothesis can account for a wider range of truncations than the accent-based hypothesis because the former predicts truncated forms from the default underlying prosodic structure.

Now that the tripartite structure hypothesis has been demonstrated to be superior to the accent-based theory, the explanatory power of the remaining hypothesis, the pseudo-compound hypothesis, must be examined. In the following section, we will examine the truncations from the last two groups in Table 2: those that only the pseudo-compound hypothesis can predict and those that only this hypothesis cannot predict.

4. The bipartite structure hypothesis: real pseudo-compounds in Japanese vocabulary

Building on the conclusions of the previous section, this section evaluates the pseudo-compound hypothesis in greater detail. The claim that Japanese simplex words have an underlying bipartite structure is central to this hypothesis. By analyzing the truncations from the last two groups in Table 2—those that the pseudo-compound hypothesis uniquely predicts and those that it fails to predict—we demonstrate the existence of real pseudo-compounds in the Japanese lexicon.

The nine words listed below have truncated forms that can only be explained by the pseudo-compound hypothesis.

(21) a. LHLL: furan+neru "flannel" → furano, neru
b. LLHH: animee+shon "animation" → animee → anime
defuree+shon "deflation" → defuree → defure
rifuree+shon "reflation" → refuree → refure
c. LLLH: terebi+jon "television" → terebi
d. LLLLL: aperi+tifu "aperitif" → apero
fakushi+miri "facsimile" → fakkusu
makisi+mamu "maximum" → makishi
rokomo+tibu "locomotive" → rokomo

The base word in (21a) supports Kubozono (2004)[6] and Kubozono and Ogawa's (2005)[8] hypothesis that words of five moras or more are phonologically complex, as *furanneru* has two truncated forms: *furano* and *neru*, representing the elements before and after the boundary, respectively. Following Kubozono (2000)[17], we assume that the truncated forms in (21b) have a word-final short vowel because long vowels in the word-final position undergo weight neutralization in Japanese. However, the number of truncations that only this hypothesis can explain may be six out of nine above, as *apero*, *fakkusu*, and *makisi* in (21d) may have been borrowed directly from the truncated forms in French *apéro*, English *fax*, and English *maxi*, respectively.

It might be argued that these six truncations are exceptions to the loanword truncation process and that accounting for them does not necessarily strengthen this hypothesis. As many as 28 words, four of which may have borrowed their truncated forms from English, fall into one of the three syllable structures in (21b-d). All of their truncated forms can be correctly predicted by both the accent-based hypothesis and the tripartite structure hypothesis. Some examples are provided in (22).

(22) LLHH: anaunsaa "anouncer" → ana
opereeshon "operation" → ope
LLLH: ekorojii "ecology" → eko
konekushon "connection" → kone
LLLLL: gurotesuku "grotesque" → guro
puroguramu "program" → puro

In addition to these 28 (or 24) words, the six words with the same syllable structures as *cereburitii* or *ekisupojishon* in (18) are exceptions to the pseudo-compound hypothesis alone, although one of these words may have borrowed its truncated form from English. Furthermore, the hypothesis fails to account for the truncations in (23a, b) below.

(23) a. LLHLL: memoran+damu "memorandum" → *memoran [memo]
difaren+sharu (gia) "differential gear" → *difaren [defu]
b. LLHLHL: apoin+tomento "appointment" → *apoin [apo]
cf. LLLLH: heriko+putaa"helicopter"→*heriko [heri] (=19a)
HLLLH: kontinyu+itii "continuity" →

*kontinyu [konte] (cf.19b)

The discussion thus far has affirmed that, like the accent-based hypothesis, the pseudo-compound hypothesis has many shortcomings. The previous section's conclusion that the tripartite structure theory is superior to the other hypotheses still holds.

However, this does not rule out the existence of pseudo-compounds in Japanese. When we assume that truncation involves eliminating the head and the following element in the underlying default structure, it is intriguing that there are two types of loanwords: those with a tripartite structure and those with a bipartite structure. Given that the words in (21) are well established in Japanese, they may indeed be true pseudo-compounds for native Japanese speakers. It seems important to separate the fact that simplex words have a tripartite structure as the default prosodic structure from the possibility that some true pseudo-compounds exist in Japanese. There is also another strong reason to believe in the existence of pseudo-compounds: the fact that the tripartite structure hypothesis cannot explain these truncations is not a weakness of the hypothesis if those words are not simplex but complex. It just means that the pseudo-compound hypothesis and the tripartite structure hypothesis complement each other in explaining the process of loanword truncation.

5. Conclusion

In the present paper, we evaluated the explanatory power of the three hypotheses proposed to account for the simplex truncation process in Japanese and affirmed that Kubozono's (2020)[5] tripartite structure hypothesis has the greatest explanatory power among the three. Our analysis validated that truncation is dependent on the internal phonological structure of the word, rather than accent structure, and that the internal structure is tripartite, not bipartite. Additionally, we highlighted the possibility that some words, such as those in (21), may be true pseudo-compounds for native Japanese speakers. The two words excluded from this analysis—*bookaroid* and *entaateimento* mentioned in Section 3—and the well-known exception to simplex truncation, *hankachiifu* "handkerchief," may also be explained as words with pseudo-compound word boundaries.*[10] This is illustrated in (24), where the retained parts before and after the boundary are underlined.

(24) boo.ka + roi.do → bokaro
en.taa.tei + men.to→ entame
han.ka + chii.fu → hankachi

Further research is crucial from both phonological and other viewpoints to discover which words constitute pseudo-compounds for native Japanese speakers.

Another issue that must be addressed is that the tripartite structure hypothesis cannot explain the input condition in (2a). As mentioned in subsection 2.3, the tripartite word structure is based on Tanaka's (2008)[14] observation of baseball chants and can account for the segmentation patterns of three- and four-mora names, which are common among Japanese players. Although excluded from the present study, the *Katakana Dictionary* contains 19 truncated forms with bases that are four moras or shorter.*[11] The tripartite structure hypothesis, like the accent-based hypothesis, can predict the right truncations for the 18 of these forms, while only one truncation cannot be accounted for by either of the two hypotheses. Some examples are provided in (25).

(25) a. LLLL: {a}{machu}{a} "amateur" → *{a} → (MR) ama
LLH: {ø}{gyaba}{jin} "gabardine" → *{ø} → (MR) gyaba
LHL: {pi}{ket}{to} "picket" → *{pi} → (MR) pike
LLL: {po}{ri}{su} "police" → *{po} → (MR) pori
b. LLLL: {ba}{rasu}{to} "ballast" → *{ba} → (MR) *bara (barasu)

Although it is crucial to separately address why so few foreign words with four moras or fewer have truncated forms, especially considering that not all loanwords with five moras or more are truncated, the ability to explain those 18 truncations may present an additional advantage of the tripartite structure hypothesis. This, however, is left for further study.

Acknowledgments

This is a slightly revised version of my paper, which was submitted in August 2023 to the Tokyo Institute for Advanced Studies of Language. I would like to express my gratitude to Professor Emeritus Haruo Kubozono of

the National Institute for Japanese Language and Linguistics for all that I learned in his course at the Tokyo Institute for Advanced Studies of Language. Needless to say, any inadequacies are my own. The author declares no conflicts of interest associated with this manuscript.

Notes

*1 Our transcription is largely phonemic and follows the Hepburn style of Romanization with some modifications: vowel length is indicated by doubling the vowel; *ji*, *hi*, and *fu* stand for [dʒi], [ɕi], and [ɸu] respectively. The symbol *n* denotes the moraic nasal in Japanese, realized as a nasal glide that assimilates in place to following stop consonants. Additionally, the two vowels in a falling diphthong are treated as part of the same syllable.

*2 Furthermore, Kubozono and Ogawa (2005: 170)[8] and Kubozono (2023: 166)[2] present Dauer's (1983)[9] research, which identified that a four-mora word typically spans 0.5 seconds. They also referenced Allen's (1975)[10] study, which affirmed that this 0.5-second duration serves as a fundamental unit of time for various human motor activities. This duration is further hypothesized to represent the fundamental time unit of the brain clock responsible for regulating such activities.

*3 Kubozono and Ogawa (2005: 158-9)[8] and Kubozono (2023: 155)[2] also challenged Labrune's (2002)[11]—who employed the framework of optimality theory (Prince and Smolensky 1993/2004)[12] rather than lexical phonology (Kiparsky 1985)[13]—assumption that stress assignment as a phonological process precedes truncation as a morphological process.

*4 Kubozono (2023: 168-70)[2] demonstrated that the principle of syllable boundary is applicable to various phenomena such as the accent placement of several types of words and *zuja-go* formation, a form of back slang. He further affirmed that the 2 moras + 1 mora structure, where the first half is longer than the second half, is preferred in baby talk words (motherese), *zuja-go* formation, and baseball chants.

*5 *Birudingu* is segmented as biru+dingu rather than birudin+gu because if the first half is four-mora long and the second half is one-mora long, it will depart greatly from the principle in (7b), which requires both parts to be of equal length.

*6 Kubozono (2023: 139)[2] highlighted that simple words with five or more moras are phonologically complex and that their accent patterns align with those predicted by compound accent rules.

*7 Among the surveyed words, *konpu* and *konpuri* are listed in the dictionary as truncated forms of *konpuriito* "complete." In this study, the former is counted as the only truncated form based on our native speaker intuition.

*8 If two or more words with the same syllable structure share the same truncated form, only one of them is counted as the base word. For instance, the two base words *konsarutingu* "consulting" and *konsarutanto* "consultant" are counted as one case because they share the same syllable structure, HLLHL, and both of them are shortened to *konsaru*.

*9 For instance, the headword *choko* is neither listed in the dictionary, nor is it recognized as an abbreviation under the headword *chokoreeto* "chocolate."

*10 The headword listed in *Super Daijirin*[18] is *kaachiifu*. The OED[19] also mentions "HAND + KERCHIEF," although it remains unclear whether speakers are aware of the etymology.

*11 Among the shortened words listed in the dictionary, there are 19 words with four moras or fewer, including the three-mora word *porisu* "police" → *pori*, and the Japanese word *gokiburi* "cockroach" → *goki*.

References

1) Kubozono H, Shingo ha Kōshite Tsukurareru [How New Words Are Created], Iwanami, Tokyo (2002).

2) Kubozono H, Ippan Gengogaku kara Mita Nihongo no Gokeisei to On'in-kōzō [Japanese Word Formation and Phonological Structure from the Perspective of General Linguistics], Kurosio, Tokyo (2023).

3) Ito J, Prosodic Minimality in Japanese, CLS **26-II**: Papers from the Parasession on the Syllable in Phonetics and Phonology, 213-239 (1990).

4) Kubozono H, The Syllable as a Unit of Prosodic Organization in Japanese, In: The Syllable in

Optimality Theory, Fery C, van de Vijver R (eds.), Cambridge University Press, Cambridge, 99–122 (2003).

5) Kubozono H, Default Word Prosody and Its Effects on Morphology, Japanese/Korean Linguistics **26**, 19–31 (2020).

6) Kubozono H, On'in-kōzō kara Mita Tanjungo to Gōseigo no Kyōkai [The Phonological Boundary Between Simplex and Complex Words in Japanese], In: Bumpo to Onsei [Speech and Grammar] **4**, Spoken Language Working Group, Kurosio, Tokyo, 123–143 (2004).

7) Kubozono H, Rendaku, Its Domain and Linguistic Conditions, In: Voicing in Japanese, van Weijer J et al. (eds.), Mouton de Gruyter, Berlin, 5–24 (2005).

8) Kubozono H, Ogawa S, 'Sutoraiki' ha Naze 'Suto' ka? Tanshuku to Tango-bunsetsu no Mekanizumu [Why is Sutoraiki 'Strike' Shortened to Suto? The Mechanism Underlying Word Truncation and Segmentation], In: Gendai Keitairon no Chōryū [Trends in Modern Morphology], Oishi T, Nishihara T, Toyoshima Y (eds.), Kurosio, Tokyo, 155–174 (2005).

9) Dauer RM, Stress-timing and Syllable-timing Reanalyzed, Journal of Phonetics **11**, 51–62 (1983).

10) Allen GD, Speech Rhythm: Its Relation to Performance Universals and Articulatory Timing, Journal of Phonetics **3**, 75–86 (1975).

11) Labrune L, The Prosodic Structure of Simple Abbreviated Loanwords in Japanese, A Constraint-based Account, Journal of the Phonetic Society of Japan, **6-1**, 98–120 (2002).

12) Prince A, Smolensky P, Optimality Theory: Constraint Interaction in Generative Grammar (1993, distributed as manuscript), Blackwell, Malden (2004).

13) Kiparsky P, Some Consequences of Lexical Phonology, Phonology **2**, 85–138 (1985).

14) Tanaka S, Rizumu Akusento no Yure to On'in-Keitai-kōzō [Rhythmic and Accentual Variations and Phonological/Morphological Structure], Kurosio, Tokyo (2008).

15) Kubozono H, Japanese Accent, In: The Handbook of Japanese Linguistics, Miyagawa S, Saito M (eds.), Oxford University Press, Oxford, 165–191 (2008).

16) Sanseido Henshusho (ed.), Sanseido Poketto Katakanago Jiten [Sanseido Pocket Dictionary of Katakana Words] Second Ed., Sanseido, Tokyo (2019).

17) Kubozono H, Nihongo no Goi to Onsetsu-kōzō [Vocabulary and Syllable Structure in Japanese], Nihongo Kenkyu **20**, 1–18 (2000).

18) Matsumura A (ed.), Super Daijirin [Daijirin Dictionary] Fourth Ed., Sanseido, Tokyo (2019).

19) Simpson JA and Weiner ESC, The Oxford English Dictionary, Second Edition on CD-ROM Version 4.0, Clarendon Press, Oxford (2009).