A Morphological Account of Japanese Western Loan Word Truncation

Johnny Geor ge

Abstract

This paper examines the truncation or shortening of Japanese Western loan words and shows how such a process mirrors the general truncation pattern for all Japanese words. This work emphasizes the ability of loan word borrowers to decompose source words at the morphological level in order to apply a top-down truncation strategy. This study utilizes a corpus of over 200 loan truncata to predict salient truncation boundaries based on the collocations of segments in a database of approximately 30,000 Japanese words. Prior analyses of Japanese loan truncation, such as Ito (1990), Ito & Mester (1992), Buena (1995), and Kurisu (2001), represent Japanese loan truncation as a bottom-up process driven by a bimoraic template so fully motivate truncation based on syllabic weight. The current analysis calls attention to the decomposition of the source word as a strategy which applies to even non-Western Japanese loans. The top-down approach has several advantages over prior analyses: it expands the application of pre-existing constraints on Japanese truncation without the creation of additional mechanisms for a separate class of loan words; it provides an account for variation in truncation outputs; and it recognizes the source of loan words as bilingual speakers who can apply their knowledge of the source language to generate novel outputs.

Introduction

This study investigates the truncation process for *gairaigo* 'Western loan words' in Japanese, and finds evidence to support the conclusion that Japanese speakers utilize the morphology of a given source word to produce truncated outputs or truncata. Speakers demonstrate sensitivity to the morphemic composition of input lexemes so they can adapt strategies such as deleting inflectional and derivational components, or identifying licit 'word' strings in order to isolate monomorphemic outputs. Japanese loan word clipping has been a topic of great interest in the literature due to its prosodic similarity to various bimoraic derivational processes in Japanese, especially as outlined by Poser (1984 & 1990), and Ito, Kitagawa, and Mester (1996). The current study focuses on parallels between the truncation processes of *gairaigo* and words from the Sino-Japanese lexicon.

The morpheme-sensitive truncation framework involves a top-down approach to the truncation process in which the speaker first tries to decompose a source word morphemically based on its indexation to the speaker's mental lexicon. The bulk of previous work applies a bottom-up approach in which the speaker targets moraic units—measures of syllabic weight, to serve as the building blocks for licensed outputs. Typical analyses claim that Japanese truncation is motivated principally by phonotactics, or word structure (Inoue 1981, Ito 1990, Ito & Mester 1992, Buena 1995, Kurisu 2001). However, such accounts

fail to determine how speakers choose among several licit outputs of differing length. For instance, the word *animesyoN* 'animation' could potentially truncate as *ani*, *anime*, or *animesyo* but only a single output exists in the lexicon. The morpheme-sensitive framework presented here, in contrast to the ubiquitous approaches based upon a set number of moraic units such as the work of Ito (1990), provides an account for truncata that accommodates variation in output length.

Morpheme-sensitive truncation has support from various sources. The fact that Sino-Japanese word truncation and compound word truncation are morpheme sensitive processes suggests strongly that the same processes may extend to single word loans. Consultants surveyed as a part of this study show clear evidence of targeting morphemic strings in a given token, whether the input word consists of a Western loan compound or single word. The examination of bigram troughs (Schiller et al. 1997), or places in a word which have a low transition frequency between particular phonemes, offers insight into how speakers may identify morpheme chunks. Relation of source word length to truncatum output also serves as evidence. Sociolinguistic considerations driven by grammars applied by individual speakers ultimately determine the dominant forms or sets of forms utilized by the speaking community.

This paper provides a profile of Japanese loan word truncation, compares the loan word truncation process with Sino-Japanese word truncation, covers a bimoraic centered analysis, and finally offers a method to produce a morpheme driven truncation account.

1. A Profile of Japanese Truncation

Truncation serves as an extremely productive process in Japanese. Frequently cited examples include outputs from compound words (1a) and single words (1b) (Ito 1990, Shibatani 1990, Ito & Mester 1992, Buena 1995, Nishihara et al. 2001, Labrune 2002). Compound truncata make up the class of shortenings that have source words consisting of two free morphemes. Single truncata represent the shortened outputs of source lexemes made up of single words. Two common patterns of loanword shortening appear below.

1) Common Western Loanword Truncation Patterns

a.	Compound truncatum	*waa(do) puro(seQsaa)*	*waapuro*	'word processor'
b.	Single truncatum	*ope(reesyoN)*	*ope*	'operation'

Usually an output consists of the leftmost bimoraic segment(s). A mora (μ) represents a unit of syllabic weight. For Japanese, a syllabic nucleus, which consists of a vowel (V), counts as a single mora. A Japanese coda, or syllable final consonant (C), also adds moraic weight. The chart below shows a few examples. Each segment representing moraic weight appears in bold.

2) Moraic and Syllabic Structure

a.	*ope*	***O.PE***	=	2 mora	'operation'
		V.CV			
b.	*waapuro*	***WAA.PURO***	=	4 mora	'word processor'
		CVV.CV.CV			
c.	*toraNsu*	***TO.RAN.SU***	=	4 mora	'transformer'
		CV.CVC.CV			

A number of processes in the Japanese language, such as hypocoristic or nickname formation (Poser 1984) rely upon moraic weight. As noted by Ito (1990), a significant number of truncatum outputs consist of bimoraic feet; *bimoraic* refers to pairs of morae such as *ope* or *tora*, and *foot* refers to a bimoraic string, which reflects the underlying composition of the word. A detailed compositional analysis appears in Ito (1990).

The syllable (σ) acts as another significant aspect of loan truncation. In the chart above each syllable boundary is delimited by a period or stop. Note that the moraic weight does not necessarily correspond to the syllable boundaries. Words consist of light and heavy syllables. A light syllable (σ_{μ}) consists of one moraic weight; therefore, V and CV syllables such as *o* or *pe* class as light syllables. A heavy syllable ($\sigma_{\mu\mu}$) consists of two morae, so CVV and CVC syllables such as *waa* or *raN* act as heavy syllables. Typically in the literature on Japanese, an output of the form CV_1V_2, such as *roi* with different vowels, represents two light syllables CV.V rather than a single heavy syllable. Truncatum outputs disfavor certain weight asymmetries, discussed in section two below.

Compound truncata are derived from source lexemes consisting of at least two free morphemes. They appear much more commonly than non-compound shortenings (Motwani 1993). Four strategies for reducing the length of compounds exist (Kanno 1985, Nishihara et al. 2001). For instance, for the shortening of the name of the hit movie *batoru roiyaru* 'Battle Royale' the following outputs are potentially felicitous:

3) *Compound Truncation Patterns*

	Represent both words and…		
a.	truncate both strings	*bato(ru) roi(yaru)*	*batoroi*
b.	truncate one string	*batoru roi(yaru)*	*batoruro*
	Represent one word and…		
c.	truncate it	*bato(ru roiyaru)*	*bato*
d.	don't truncate	*batoru (roiyaru)*	*batoru*

The clipped portion can include either the beginning or the end of the word, although preservation of the leftmost portion of the word occurs most commonly. Pattern (3a) serves as the most common pattern (Shibatani 1990, Nishihara et al. 2001) and pattern (3d) appears ubiquitously as well (Kanno 1985).

Single or non-compound truncata may clip several morae from either the left or right end of the word, with right end clipping being more prevalent. The following examples below illustrate the range of patterns attestable for Japanese truncation. Example (1b) above, *ope(reesyoN)*, is a string that preserves the left foot, and the examples labeled (4a) below show the preservation of the rightmost segments. Non-contiguous segments may also become deleted as illustrated in list (4b) below. Additional patterns include the insertion of segments shown in set (4c), and even semantically motivated constraints as seen in (4d). The elicited example, 'lease Poland' potentially results in a taboo output, which the speaker avoids by not including the coda *N* of the initial syllable. Some forms have more than one documented output in circulation. The presence of such variation indicates that a particular form may emerge due to sociolinguistic considerations rather than purely structural processes or constraints.

4) Singles or Free Morpheme Truncation Patterns

a. **Non-initial strings**			
(aru)beito	*beito*	'part-time job'	
(sya)neraa	*neraa*	'Chanel maniac'	
b. **Non-contiguous outputs**			
mo(ru)hi(ne)	*mohi*	'morphine'	
ore(nzi) (su)ka(Qsyu)	*oreka*	'orange squash' soft drink	
waNpi(isu) + ota(ku)	*waNpiota*	'One Piece maniac'	nonce combination
se(buN) (su)ta(a)	*seQta*	'Seven Stars'	
(se)buN (su)ta(a)	*buNta*	'Seven Stars'	
c. **Insertion**			
gu(Qdo)	*gu**u***	'good'	V lengthening
sut(yuwaadesu)	*suQ**tii***	'stewardess'	"hypocoristic"
d. **Taboo Avoidance**			
ti(Ntai) po(o)ra(Ndo)	*tipora*	'lease Poland'	Taboo avoidance (*elicited*)
c.f. daN(su) pa(atii)	*daNpa*	'*dance party*'	

2. The Bimoraic Account for Truncation

The most influential works on the process of Japanese *gairaigo* truncation are Ito (1990) and Ito & Mester (1992). They characterize Japanese loan truncation as a process driven by two primary factors, a bimoraic template and a restriction on particular weight asymmetries.

Ito & Mester (1992) posit that Japanese Western loan truncation is one of many processes driven by a bimoraic template, so they construct an underlying lexical structure driven by the creation of bimoraic feet. This implies that a two moraic truncatum better meets the felicity condition of maintaining the shortest recoverable output than a longer truncatum, which consists of three or more morae. For instance, ani conforming to a simple bimoraic template represents the optimal output for a word such as *animesyoN* 'animation' independent of any other constraints; other phonological factors, such as syllabic weight asymmetry discussed in the following paragraph, motivate the creation of truncata of varied length. As a result of an underlying structure made up of bimoraic feet, Ito & Mester (1992) conclude that well-formed outputs consist of two, three or four morae. Compound truncata naturally fit their schema, and the balance of compound outputs conform to a bimoraic template consisting of two feet, such as in the output for 'word processor' *waa(do) puro(seQsaa)* [waa][pu.ro] $[\sigma_{\mu\mu}][\sigma_{\mu}\sigma_{\mu}]$. In contrast, it will be shown in section three that the Ito & Mester (1990) account underspecifies the criteria needed to motivate outputs of single word truncata consisting of two, three or four morae.

Ito (1990) introduces a syllable weight asymmetry restriction or constraint, which entails that words with heavy word-initial syllables receive preference in contrast to words of equal moraic length containing medial heavy syllables. For example, heavy-light syllabic $[\sigma_{\mu\mu}\sigma_{\mu}]$ pairs and heavy-light-light $[\sigma_{\mu\mu}\sigma_{\mu}\sigma_{\mu}]$ strings are favored while light-heavy $*[\sigma_{\mu}\sigma_{\mu\mu}]$ and light-light-heavy $*[\sigma_{\mu}\sigma_{\mu}\sigma_{\mu\mu}]$ strings are disfavored. The avoidance of these syllable weight combinations in Japanese clippings is evidenced in the independent

corpora of Ito (1990), Labrune (2001), Yoneyama (2001), as well as in the current study. While detailed discussion of the motivation for the preference for heavy word initial syllables goes beyond the scope of this paper, such a phenomena may reflect the influence of Positional Markedness in which phonological peaks are limited to prominent positions (Zoll 1998). Speakers may tend to shorten words and preserve heavy syllables in the psychologically strong initial position, which carries a significant lexical processing load (Beckman 1998). Ultimately the Japanese lexicon favors the alignment of prominent features, with longer durational segments appearing in the most prominent location in a given word (Yoneyama 2001). Ito & Mester (1992) introduce a larger analysis involving a family of constraints to produce truncatum outputs motivated by bimoraicity. Their account centers on the phonological motivations for truncatum outputs, and has influenced the work of a number of other authors (Buena 1995, Kurisu 2001, Labrune 2002).

3. The Truncata Corpus

The corpus used for this study comprises a collection from four dictionaries, Ishino (1992), Motwani (1993), Yonekawa (1997), and Yoshizawa et al. (1979) unless otherwise noted. There are 454 tokens including 206 compound and 248 single truncata. For this study every truncatum single found was included in the corpus; however, due to the regularity of output formation, only a random sample of compound forms were collected. A clipping such as koNti(*nyuitii*) 'continuity' represents a single truncatum since its input form consists of one free morpheme. The output has three morae: the first two morae form a bimoraic foot containing a heavy syllable ($\sigma_{\mu\mu}$) and the third mora is a light syllable (σ_{μ}), therefore it represents a heavy-light output ($\sigma_{\mu\mu}\sigma_{\mu}$).

The table below gives a breakdown of the corpus classified by length and truncatum type. The compound truncata collected do not necessarily make up a representative distribution as only a subset of available tokens entered the collection; however, four morae outputs dominate the compound truncata lexicon to a significant degree as shown below. The singles collection reflects the distribution of moraic length for all single truncata found.

5) Corpus Word Count by Moraic Length

	206 compounds		248 singles	
2 μ	7 (3%)	mo(daN) ga(aru) 'modern girl'	74 (30%)	seme(sutaa) 'semester'
3 μ	64 (31%)	pote(to) ti(Qpusu) 'potato chips'	93 (38%)	koNti(nyuitii) 'continuity' zyairo(sukoopu) 'gyroscope'
4 μ	114 (55%)	puro(heQsyonaru) resu(riNgu) 'professional wrestling'	78 (31%)	asupara(gus) 'asparagus'

As discussed in section two, truncation commonly is referred to as a bimoraic foot driven process, so single truncata should consist of only two morae; however, three-morae outputs make up the largest proportion of the singles. Chart (6) below shows the distribution of single truncata length in four independent corpora.

6) Distribution of trucatum singles by length in 4 corpora

	224 (Kanno)	199 (Ito)	325 (Labrune)	248 (George)
2 μ	35%	43%	36%	30%
3 μ	**39%**	27%	**38%**	**38%**
4 μ	21%	28%	26%	31%

While Ito (1990) shows bimoraic outputs to be dominant in her sample, the distribution of Labrune's (2002) data reflects the dominance of three-morae outputs for singles. Kanno (1985), who presents the most extensive study of all types of loan word shortening for Japanese, also shows a preponderance of three-morae outputs for singles. Part of the reason for the difference of distribution among the four independent corpora is that in some cases differing criteria are utilized for identifying a particular shortening as a truncatum or not. Ito (1990) includes target words derived from source words with derivational endings such as *–mento* '-ment', and words suffixed with inflectional endings such as *paakuiNgu* 'parking'. Kanno (1985) includes all types of shortened words and puts them into various classes. Labrune (2002) does not include some derivational words such as those ending in –mento '*-ment*'; however, he has words with other types of derivational endings.

The current study includes all words with truncatum forms whether consisting of derivational or inflectional endings. All tokens were included as this study makes a case for the creation of truncata via the isolation of morpheme or morpheme-like segments. The isolation of such segments represents the central process driving all truncation in the Japanese lexicon, whether for *gairaigo* or Sino-Japanese forms.

4. Sino-Japanese and Western Compound Truncation

This section will present Sino-Japanese (S-J) word truncation and illustrate the similarity between this morphemic top-down word shortening process and Western loan truncation. With regards to words of Sino-Japanese origin it is clear that forming a truncatum output requires sensitivity to the morphemic composition of a given word. In the case of S-J truncation, determining the most felicitous break requires access to the internal morphemes, or orthographic representation, in which each character represents a single morpheme. Commonly the left morpheme of each given word remains in the output. In each case below, the clipping occurs independently of prosodic considerations. Sometimes the semantics play a role in which segments are preserved as in the shortened form of 'driving school', which preserves the characters for 'car' and 'teach'.

7) Sino-Japanese Truncata

東(京)大(学)	*too(kyoo) dai(gaku)*	→	東大 *toodai*	'Tokyo University'
自(由)民(主)	*zi(yuu)miN(syu)*	→	自民 *zimiN*	'Liberal Democrat'
(自動)車教(育習所)	*(zido)syakyoo(ikusyuuzyo)*	→	車教 *syakyoo*	'driving school'
現(代)社(会)	*gen(dai)sya(kai)*	→	現社 *geNsya*	'modern society'

The samples above demonstrate the variety in output lengths and syllabic weights of S-J truncata. Epiphenomenally, truncated S-J outputs primarily consist of two bimoraic feet. This outcome is likely due to the overwhelming number of S-J loans formed as compound words made up of bimoraic morphemes.

8) The Moraic Composition of a Sino-Japanese Truncatum

	too	*(kyoo)*	*dai*	*(gaku)*	→	*too*	*dai*
Tier 1	東	(京)	大	(学)	→	東	大
Tier 2	μ μ	(μμ)	μμ	(μμ)	→	μμ	μμ

Western compound truncation has two ways in which it can potentially mimic S-J truncation (also discussed in Ishiwata 1983; Ito 1990; Nishihara et al. 2001; and Shibatani 1990). This output could be generated via a bottom-up approach and utilize the prosodic tier consisting of morae (tier 2), or it can be the result of a top-down process utilizing the morphemic tier made up of meaning units (tier 1).

The bimoraic foot preservation pattern for Western compound truncation mimics S-J shortening prosodically since the length of Western loan clippings reflects the typical length of S-J truncata. Similarly to S-J shortening separate morphemes must be identified in a compound to target boundaries for clipping. Compound word truncation is distinct in that clipping targets a bimoraic foot from two free lexemes rather than a single word as in the case of the truncation of loan singles. Identification of word breaks must either be through the discovery of potential free morphemes that make up the compound or via an orthography that shows the break between the pair of words forming the compound.

Compound truncation may also involve the preservation of one of the two free morphemes that make up the compound word.

9) Free Morpheme Preserving Compound Truncata

a. ereki (gitaa) ereki ‘electric guitar’
b. daNpuu (kaa) daNpuu ‘dump truck’

This process conceptually parallels morpheme targeting in the S-J truncation pattern since one or more input morphemes from the compound drop and result in a shortened output. This method works independently of the prosodic consideration of the input morphemes. Oftentimes the left morpheme is kept, but sometimes other considerations such as semantics will result in preservation of the right morpheme.

As for Western loan singles, some truncata such as *mo(ru)hi(ne)* listed with the examples from list (4b) above, involve the inclusion of skipped segments, thus mirroring the S-J and compound truncation patterns. The skipping of segments reflects the propensity of the truncation process to target morpheme-like strings for shortening. The dominant pattern which preserves the leftmost segment of a source word, such as in the token *ope(reesyoN),* still can reflect the strategy of identifying morpheme like segments for preservation as in the compound cases listed in (9). Consider a token such as *(puraQto)hoomu* which truncates to *hoomu* ‘platform’, the source word could have ambiguity between identification as a compound or single free morpheme. Such words may support the tendency to process all source inputs, single or compound, similarly applying the top-down morphemic truncation process.

5. A Top-Down Morpheme Truncation Process

The previous sections described prosodic and morphemic processes for the truncation of Western compound loans, and this section will show how truncation of single word inputs may also be motivated by the same S-J influenced top-down process. The study covered in this section shows that speakers can target perceived morphemes within a given Western loan single in order to produce licit truncatum outputs. Consultants demonstrate that different speakers can uniformly identify the same super-syllabic boundaries so share intuitions about where breaks can occur within a given word.

Section five provides three supports for a morphemic boundary driven truncation process. First, with

respect to the source word, truncata often break along derivational or inflectional boundaries so support the premise that source words receive a S-J type morphological decomposition. English/Japanese bilingual Japanese consultants evidence use of morpheme targeting strategies for forming nonce truncation forms. Finally, an examination of source/output word pairs reveals a correspondence between the length of source words and truncatum outputs, which can more readily be explained by a morpheme targeting account rather than one based on a bimoraic template analysis.

5.1 Loss of Derivational and Inflectional Segments

The first evidence supporting a top down analysis consists of truncata that appear to preserve the non-derivational portion of an input. A number of authors make a similar observation (Kanno 1985, Ito 1990, Ito & Mester 1992). Any truncatum output follows as the result of two polarized forces, one that maximizes shortening and another that preserves lexical recoverability. It follows that redundant features such as inflectional or nominalizing suffixes are readily eliminated in the course of producing a shortened output. The loss of inflectional suffixation is expected cross-linguistically in loans (Japanese examples include: *noQkiNgu* → *noQku* 'knocking', *paakiNgu* → *paaku* 'parking'), so the loss of inflection in truncation comes as no surprise. Additionally, common derivational suffixes may serve as targets for clipping. Some quintessential examples follow.

10) *Derivational Suffix Dropping Truncata*

a. anime(syoN)	'animation'	c. roke(esyoN)	'location
b. teribi(syoN)	'television'	d. apaato(meNto)	'apartment'

With the exception of the loss of the vowel length for *roke*, each word simply loses its derivative suffix and leaves a shortened output of two to four morae. This pattern also provides a paradigm for the preservation of a potential monomorphemic string, which mimics the compound truncation process discussed at the end of section four. Taking the preservation of a single morpheme into consideration, the following truncata can receive an account.

11) *Morpheme-like Segment Preserving Truncation*

a. inhura(sutorakutyaa)	infrastructure
b. heli(koputaa)	helicopter
c. saamo(sutaQto)	thermostat
d. saike(deriQku)	psychedelic

Each word could be interpreted by a speaker as a pair of single morphemes. It does not matter whether a potential word has interpretation as a compound or single form since the patterns of truncation illustrated earlier through the tokens under table (9) can apply to strings consisting of free or bound morphemes. Many words like *heli(koputaa)* do not necessarily break along actual morphemic lines (helico-pter). Instead, the speaker having familiarity with English morphology will recognize 'copter' as a *potential* morpheme and drop it; '-pter' does not seem like a viable morpheme considering English phonotactic constraints. The break along assumed morpheme boundaries mirrors the type of truncation evidenced for Sino-Japanese loans. More compelling evidence for morpheme targeting in singles comes from the variation in the truncata produced by consultants as covered in the next section.

5.2 Consultant Production of Truncata

For this study five native Japanese speakers from four parts of Japan were asked to generate truncata

from a list of approximately 100 Western loan words. There were three types of words in the list: one set consisted of loan words with listed forms, another set consisted of loan words without standardized truncata, and another group consisted of words adapted from English which are not loan words currently found in Japanese dictionaries. Each individual was given a list of the words given in the *katakana* script along with brief directions and an example. All of the participants were university students relatively fluent in English. There was no time limit placed on the task.

There exist varied consultant outputs that receive a uniform account under a morpheme-centered analysis, which cannot be provided for as readily with only a moraic evaluation. The following data from the consultant corpus appears below.

12) Consultant Truncata Outputs

			Tokyo 1	Aomori	Shiga	Okinawa	Tokyo 2
abandonment	abaNdoNmeNto	C		adomeN	abaNdoN	abameN	baNdoN
franchise	huraN**tyaizu**	C	tyaizu	huraNtya	tyaizu	huratya	huratyai
appointment	apoiNto**meNto**	L	apo	apo	apoiNto	apoiNto	apome
Transformer	toraNsu**hoomaa**	L	hoomaa	torahoma	torahoomaa	torahoma	torahaa
Frankhurt	huraNku**huruto**	L	huraNku	huraNku	hurahuru	huraNku	hurahuto
dynamite	dainamaito	L	maito	daimai	maito		dainama
plankton	puraNkutoN	N	puraNku	puratoN	puratoN		puraNku
apendicitis	aQpeNzisaitisu	L		apezi	apeNda	apesai	apesai
enterprise	eNtaa**puraisu**	L		eNpura	eNpura	eNpura	eNpura
transporter	toraNsu**pootaa**	N	pootaa	pootaa	torapootaa	torapota	torapoo

L= Listed-a loan w/clipping C= Coined-a loan w/o clipping N= Non-standard loan word made up

Even though many of the outputs vary, one could say that there is a level of uniformity in the placement of larger than syllabic boundaries within the words. These breaks can represent 'perceived morpheme boundaries' since the consultants share similar intuitions about where breaks should appear independent of some objective notion of what constitutes a morpheme per se. Many of these truncata patterned similarly to the compound truncata via the monomorpheme preservation strategy under (9) in section four. Consultants also produced morpheme motivated outputs similar to the compound truncation pattern as shown in (1a) and the S-J truncation pattern under (7). The words *abaNdoNmeNto* 'abandonment' and *apoiNtomeNto* 'appointment', although not compound forms, have derivations of *abameN* and *apome* with complementary outputs *abandon* and *apoiNto* that respectively target the same cut-off point, just before the string *–meNto*. These outputs provide evidence that the consultants can potentially target some type of boundary that they objectively intuit. The same pattern applies throughout the set; consultants vary in the outputs they produce while maintaining the same intuitions about some super-syllabic boundary in each given token. 'Transporter' is represented by four unique outputs, but the break points are all the same, between *toraNsu* and *pootaa*, indicated by the fact that *po* appears as the initial segment after the source word breakpoint in each truncatum. This uniform account of some of the consultant productions requires reference to some perceived morpheme break. In contrast, a primarily structural analysis based on the moraic structure of the truncata offers no ready explanation for the selection of non-contiguous outputs.

5.3 Correspondence Between the Length of Source Words and Truncata

Consultants' encounters with source words of substantial length increased the possibility of one of the 'compound truncation' patterns emerging—either the single morpheme preserving pattern as the tokens under (9) or the S-J pattern under (1a) and (7). Such an outcome reflects the fact that consultants initially

relied upon a top-down processing of a given input before resulting in a bottom-up strategy of composing a suitable output at the purely moraic or foot level. The consultants' performance calls into question the need to classify truncated forms into single and compound groups since the length of the input string acts as the focal consideration. The preference for higher order processing of a string via morphemic composition may relate to recoverability. Certain well established outputs based on the clipping of simple bimoraic feet from an older productive pattern (Inoue 1981) exist in the lexicon and recoverability is not an issue for such clippings, even when they represent longer source inputs; however, when modern speakers with greater exposure to English (this refers to any speakers subject to post-war Japanese education reforms in English education) are confronted with opportunities to produce novel truncata, length becomes a meaningful consideration. Taking a look at relative length of input strings to truncatum outputs in the dictionary corpus shows a correspondence in that longer inputs tend to result in longer outputs (c.f. Kanno 1985).

13) Relative length of source inputs to truncatum outputs

	Input length						
	3 μ	**4 μ**	**5 μ**	**6 μ**	**7 μ**	**8 μ**	**9 μ**
Output length							
1 μ	1						
2 μ	6	19	29	13	3	3	1
3 μ		3	23	33	22	9	2
4 μ			3	19	23	19	10
5 μ				1	0	1	0
Total	7	22	55	66	48	32	13

If the distribution of output truncata lengths were independent of source word length considerations, one would expect the distribution of outputs to remain consistent even as the input strings become longer; however, the number of smaller outputs decreases relative to the increase in the length of input strings. A purely moraic-based structural framework provides no direct motivation for the relationship between input and output length. Reference to morphemic units provides an adequate account since the perceived morphemic targets lead to tokens being split into even portions; therefore proportional increases of output truncatum lengths relative to source word input lengths are anticipated.

Although many Japanese speakers who learn English do not become balanced bilinguals, those who introduce loans to the language most likely belong to groups with the greatest exposure to English as a whole, including: members of the media, scientists, academics, and businesspeople; these groups use much of the terminology associated with Japanese loan word truncata. By extension, the truncation of loans is most likely done by those who utilize the relevant terminology most, and these groups happen to include those who likely have source word familiarity. A comprehensive study of French borrowings across five different languages by Pardee and Lacharité (1997) provides evidence that bilinguals in a society may serve as the primary initiators of loan word borrowing. Japan has universal English education for all children from middle school onward, and its strong international economic presence insures the existence of many bilinguals, especially in the metropolitan areas. As shown by consultants on this project, subjects potentially interpret the source word morphology in truncatum production.

6. A Structural Account for Identifying Source Word Boundaries (Bigrams)

The final section develops a theoretical account for the systematic identification of potential

truncation boundaries for a given source word. Given that consultants are able to intuitively access some uniform division due to identifying potential morpheme strings, there may be a means of objectively measuring the criteria applied by consultants. One such way to investigate the resultant outputs is through the investigation of bigrams troughs (Schiller et al. 1997) or places in a word which have a low transition frequency between particular phonemes. The basic idea is that consultants will tend to preserve high frequency strings over lower frequency ones. For instance, when a derivational string gets clipped, it is not due to its identity as a derivational string per se, but due to its regular appearance in variety of contexts. In a word such as *anime(syoN)* 'animation' one may find that there is a relatively low occurrence of the syllable pair *mesyo* which provides evidence of its eligibility as a suitable break point.

One way to measure the frequency of any given bisyllabic string is to make a count based on its appearance within a dictionary. For the purposes of this examination, a corpus was created by extracting approximately 30,000 words from the *Excite* online Japanese/English dictionary. An additional script was written to calculate the probability that a given syllabic character will follow another in order to create a matrix.

14) Bigram Matrix

*	カ **ka**	ガ **ga**	キ **ki**	ギ **gi**	ク **ku**	グ **gu**	ケ **ke**	ゲ **ge**	コ **ko**
サ **sa**	0.0025	0.0034	0.0042	0	0.0119	0.0017	0	0	0.0051
ザ **za**	0	0	0.0066	0	0.0099	0.0198	0	0	0.0033
ス **su**	0.0367	0.0011	0.0361	0.0003	0.0499	0.0003	0.0129	0.0013	0.0294
ズ **zu**	0.0022	0	0.0022	0	0.0022	0.0045	0.0045	0	0.0022

In the matrix extract above, observe that in the intersection of *su* on the vertical axis and *ku* on the horizontal is the value 0.0499. This value means that in the database of 30,000 words for all appearances of *su* (not including *su* in word final position), *ku* follows *su* 4.9% of the time. In this way values are produced for all possible syllabic pairs in a fixed order. The expectation is that derivational morphemes, perceived morphemes and the like are broken up by points of relatively low probability syllabic pairs that serve as borders between morpheme-like strings.

The extraction of such strings were tested for correspondence with consultant intuitions about where to divide a given word. Bigram values for words from chart (12) appear below.

15) Bigram Measures for a Set of Consultant Source Lexemes

	Bigram troughs	Values
abandonment	a-baNdoN--meNto	a (0.00690) ba (0.10330) N (0.08300) do (0.04650) N (0.00400) me (0.22890) N (0.09500)to
franchise	huraN-tyai--zu	hu (0.09160) ra (0.18040) N (0.02960) ti (0.18130) ya (0.03640) i (0.01390) zu
appointment	a--poiNto-meNto	a (0.00900) po (0.03720) i (0.18550) N (0.09500) to (0.01080) me (0.22890) N (0.09500) to
transformer	toraNsu--hoomaa	to (0.12200) ra (0.18040) N (0.05520) su (0.00220) ho (0.30150) o (0.01920) ma (0.14090) a
Frankhurt	huraNku--huruto	hu (0.09160) ra (0.18040) N (0.02970) ku (0.00710) hu (0.03500) ru (0.06670) to
dynamite	daina--maito	da (0.11680) i (0.01500) na (0.00910) ma (0.08040) i (0.05260) to
plankton	puraN-kutoN	pu (0.16770) ra (0.18040) N (0.02970) ku (0.07340) to (0.09250) N
appendicitis	aQ--peNzi—sai--tisu	a (0.02390) Q (0.00510) pe (0.17820) N (0.02800) zi (0.00000) sa (0.15620) i (0.00500) ti (0.01510) su
enterprise	eNtaa-purai--zu	e (0.09520) N (0.04220) ta (0.38450) a (0.02260) pu (0.16770) ra (0.11830) i (0.01390) zu
transporter	toraNsu-pootaa	to (0.12200) ra (0.18040) N (0.05520) su (0.0167) po (0.2233) o (0.033) ta (0.3845) a

The bigram troughs have a level of correspondence with the clipping patterns produced by consultants. The bigram pairs with the lowest probabilities fall along the same boundaries as those produced by the

consultants for: *toraNsu—hoomaa*, *huraNku—huruto*, *daina—maito*, *toraNsu—pootaa* and *a-baNdoN—meNto*. In these cases the lowest probability pairs form the basis for the ultimate clipping representation whether contiguous or non-contiguous. The lowest troughs of the set, *huraN-tyai—zu*, *eNtaa-purai—zu*, *a—poiNto-meNto* do not correspond with consultants' breaks, however these points do not produce truncatum outputs of two to four morae length and are ineligible for consideration as outputs. The next best bigram troughs do correspond with consultants' breaks. In conjunction with appropriate constraints, the correct boundaries can be extracted.

This initial investigation demonstrates a possible way to utilize a bigram trough analysis to show one way in which speakers can uniformly identify divisible elements in a given input string. Unfortunately, examination of the full corpus of singles truncata does not yield a statistically significant result with regards to using a bigram analysis based on dictionary counts to predict outputs. Rather than utilizing type frequencies, it may be fruitful to input token frequencies from a text corpus in order to see if bigrams can serve as a significant predictor. A fuller analysis would assume application of a number of sub-phonological morphological well-formedness constraints to assure that truncatum outputs met minimal criteria such as observing the moraic asymmetry constraint covered in section two.

Conclusion

Ultimately there is a preponderance of evidence which shows that speakers in the production of truncata delete derivational morphemes, provide alternative clippings based on shared intuitions about 'morpheme-like' chunks, and provide consistency in the relative length of input to output strings. Without an account that captures such elements, a complete analysis of Japanese truncation will not be possible. Since the morpheme preservation account has its basis in Sino-Japanese truncation, such a treatment provides a related account for both compound and single truncatum production.

Japanese Western loan truncation is a process that involves a "bottom-up" approach whereby speakers use bimoraic units as building blocks to produced licit truncatum outputs and a "top-down" approach in which speakers eliminate derivational affixes or preserve single morphemes. Speakers have multiple strategies for truncating words and show differing judgments when creating novel truncatum outputs. Ultimately, the form that wins out will be determined primarily by who decides to use a given truncatum, the context in which is it utilized and how is the form disseminated to other speakers. This paper also recognizes the importance of the relationship between loan word alteration and bilinguals who introduce such words into Japanese. Experiential familiarity with Japanese morphemic structure along with some level of understanding of the source word provides loan word adaptors an avenue for truncating words at a higher-order level.

References

Beckman, Jill (1998). Positional Faithfulness. Univ of Mass Dissertation. ROA 234-1297-Beckman-6-0

Buena, Laura (1995). Identity effects in morphological truncation. In Jill Beckman, Suzanne Urbanczyk and Laura Walsh (eds.) Papers in Optimality Theory. *University of Massachusetts Occasional Papers in Linguistics 18*. Amherst: Graduate Linguistic Student Association. 77-136.

Inoue, Hisashi (1981). *Shikaban Nihongo Bumpo <An Informal Japanese grammar>*. Shinchoubunko.

Ishino, Hiroshi (1983). *Gendai gairai go*. *<Modern Western Loan Words>*. Tokyo. Taishukan.

Ishiwata, Toshio (1983). Waseieigo to kokusaituuyoogo. <Japanese English vs An International Language> *Gairaigo*. 89-104. Kawade Shoboo Shinsha.

Ito, Junko (1990). Prosodic minimality in Japanese. *Proceedings of the Chicago Linguistic Society* 26. 213-39. Chicago, IL.

Ito, Junko, and R. Armin Mester (1992). *Weak Layering and Word Binarity*. Manuscript.

Ito, Junko, Yoshihisa Kitagawa, and Armin Mester (1996). Prosodic Faithfulness and correspondence: Evidence from a Japanese Argot. *Journal of East Asian Linguistics* 5.217-294.

Kanno, Ken (1985). Yogo no ryakugokei <Western shortened word forms>. *Nihongogaku* 4-9. 54-64.

Kurisu, Kazutaka (2001). The Phonology of Morpheme Realization. *ROA* 490-0102.

Labrune, Laurence (2002). The prosodic structure of simple abbreviated loanwords in Japanese: a constraint-based account. *ROA* 532-0802.

Nishihara , Tetsuo, Jeroen Van De Weijer and Kensuke Nanjo (2001). Against Headedness in Compound Truncation: English Compounds in Japanese. Jeroen van de Weijer and Tetsuo Nishihara eds. *Issues in Japanese phonology and Morphology*. Mouton de Gruyter. NY. 299-323.

Paradis, Carole and Darlene Lacharité (1997). Preservation and Minimality in Loanwords. *Journal of Linguistics* 33(2).379-430.

Poser, William J. (1984). Hypocoristic Formation in Japanese. *WCCFL* 9. 218-29.

Poser, William J. (1990). Evidence for foot structure in Japanese. *Language* 66 (1).78-105.

Schiller, Niels O., Antje S. Meyer And Willem J. M. Levelt (1997). The Syllabic Structure of Spoken Words: Evidence from the Syllabification of Intervocalic Consonants. *Language and Speech*. 40(2). 103-40.

Shibatani, Masayoshi (1990). *The Languages of Japan*. Cambridge University Press.

Vance, Timothy (1987). *An Introduction to Japanese Phonology*. NY, State University of NY Press.

Yoneyama, Akiko. Structural Aspects of the Japanese Lexicon (2001). University of Ohio Manuscript.

Zoll, Cheryl (1998). Positional asymmetries and licensing. Ms., MIT. ROA #282.

Sources for corpus

Ishino, Hiroshi (1992). *Masukomi ni yoku deru tanshukugo ryakugo kaidoku jiten <Dictionary of Common Shortenings and Abbreviations in Mass Communication>* Sotakusha, Tokyo.

Motwani, Prem (1993). *A Dictionary of Abbreviations and Truncations*. Maruzen, Tokyo.

Teacher's Network in Kitakyushu (2003). *Wakamono yoogo tukatteiru? jiten*. <Do you use youth slang? A dictionary> www.tnk.gr.jp

Yonekawa, Akihito (1997). *Wakamono Kotoba Jiten <Young People's Slang Dictionary>* Tokyo Doshuppan. Tokyo.

Yoshizawa, Norio and Toshio Ishiwata (1979). *Gairaigo no gogen, Yoshizawa jiten, Kadokawa Shoujiten 26*, *<Yoshizawa's Dictionary of Gairaigo Etymology, Kadokawa's Concise Dictionary 26>* Kadokawa Shoten, Tokyo.

Excite Japanese-English dictionary. http://www.excite.co.jp/dictionary/japanese_english/