

The prosodic structure of simple abbreviated loanwords in Japanese: a constraint-based account

Laurence LABRUNE

Bordeaux 3 University & CNRS (UMR 5610/Erss)

1. Introduction

The aim of this work is to provide a principled account of the morpho-phonology of Japanese simple (=non-compound) abbreviated loanwords (hereafter SALs) such as those in (1), in a constraint-based framework, focusing on the constraints that govern the prosodic organization and the length of the abbreviated form.

(1)

ana	< anaunsaa	“announcer”
kosume	< kosumetikku	“cosmetic”
rihabiri	< rihabiriteesyon	“rehabilitation”
paama	< paamanento (weebu)	“permanent (wave)”
baito	< arubaito	“Arbeit”

The study is based on a thorough analysis of an extensive body of data, which constitutes the empirical and statistical evidence for the present research. The main argument developed in this paper is that the length of an abbreviated word can be predicted from the prosodic structure of the base (i.e. the complete Japanized loanword acting as a source word) : we will show that apocope and aphaeresis generally occur just before the accented mora of a base. Another crucial aspect of abbreviated loanword morpho-phonology, already uncovered by previous research (Itô & Mester, 1992, Itô, Kitagawa & Mester, 1996) is that they must also constitute well-formed prosodic words. We will argue that a small number of general well-formedness constraints, as well as the more specific alignment constraint demanding that SALs be cut just before the head foot of the base (containing the accented mora) are sufficient to account for the prosodic diversity and the length of SALs. Thus the data at hand receive a straightforward explanation. This could not be achieved by previous treatments which were unable to account for length variation in SALs. The proposed analysis provides a consistent account of more than 70% of SALs in the corpus.

The paper is organized as follows : section 2 reviews previous scholarship, presents the theoretical framework, and provides some explanation on the corpus. It also establishes the distinction between Simple Abbreviated Words (SALs) and Complex Abbreviated Words (CALs), and between apocopes, aphaeresis and discontinuous abbreviations. Section 3 contains the analysis of truncated SALs in an OT framework, and ends with a discussion on the exceptions to the

proposed treatment, which happen to share some common properties. Sections 4 and 5 are concerned with aphaeresis and discontinuous abbreviation. Finally, a conclusion is offered in section 6.

2. Methodological and theoretical preliminaries

2.1. Previous studies

The first comprehensive study on the subject of Japanese foreign loan truncations from a morpho-phonological perspective is due to Itô (1990). Itô's pioneering study, based on a corpus of 199 words, uncovered the existence of gaps in the possible patterns, namely, the non-productivity of *L (= light syllable), *H (= heavy syllable), *LH, and *LHL truncations:

(2) Patterns of abbreviation (Itô 1990)

1-mora	*L				
2-mora	LL demo	*H			
3-mora	HL konpa	LLL arumi	*LH		
4-mora	HH baaten	HLL waapuro	LLH sukeboo	LLLL anakuro	*LHL

Itô (1990) considers that truncation of foreign loans constitutes another addition to the already extensive evidence for the existence of a bimoraic foot template in Japanese, and views restrictions on truncation patterns as a result of minimal word requirements which involve two different types of requirements: one based on the mora count for members of loan compounds, and another based on the syllable count for single loanwords. Itô supposes that “the morphological categories STEM and WORD have separate associated prosodic minimality constraints”; the minimal prosodic stem is bimoraic, while the word must be minimally disyllabic. This differentiation is supposed to provide an explanation for the fact that *kon* (from *konpanii* “company, party”) is a well-formed abbreviation if it belongs to a compound (for example *kura-kon* “class party”), but is impossible as an independent one (*konpa* is then the correct reduced form).

The impossibility of having *LH truncations results, according to Itô, from the interaction of the two prosodic categories of STEM and WORD. A truncation like **demon* from *demonsutoreesyon* “demonstration” is structurally blocked, because the STEM category is required to be equivalent to a bimoraic foot ($\mu\mu$), and thus cannot dominate *de* nor *demon*. If STEM dominates *mon*, then it violates the “left edge requirement” stipulating that the stem must be aligned with the left edge of the word. The four-mora pattern **demonsu* is ruled out because, putatively, it cannot be divided into two bimoraic halves.

In their 1992 paper, Itô and Mester propose an analysis of Japanese loanword abbreviation which does not rely on a templatic approach. They argue that no single species of prosodic words (minimal or otherwise) could be invoked to cover e.g. both bimoraic and fourmoraic forms. Itô & Mester (1992) consider instead that word clippings must be well-formed prosodic words. As such,

abbreviations must minimally contain one foot, by virtue of Proper Headedness. Proper Headedness, a minimal well-formedness condition, stipulates that every prosodic constituent must have a head. Thus, every Word must contain at least one Foot, every Foot at least one Syllable, every Syllable at least one Mora. But abbreviations are also free to contain more material, as long as Maximal Parsing, a well-formedness condition which demands that prosodic structure is maximally parsed (with all syllables grouped into feet), is obeyed.

Monosyllabic abbreviations such as **paa* are ruled out by the Word Binariness requirement, according to which prosodically derived words must be strictly binary in feet or syllables. The Word Binariness requirement does not apply to syllable internal branching, i.e. bimoraicity. Thus a bimoraic but monosyllabic abbreviation such as **paa* counts as structurally unary. The impossibility of LH forms such as **demon* is explained by the Left Edge Matching constraint, which requires that left word edges preferentially coincide with foot edges. Thus, *demon* being parsed as *de(mon)*, rather than *(demo)n* (which infringes syllable integrity), or *(de)(mon)* (where *de* constitutes a degenerate foot), it is not a well-formed abbreviation.

Itô & Mester (1992) constitutes an important contribution to the study of loanword abbreviation. The uselessness of templates and the binariness requirement are the two major points that we should retain of their analysis. However, while Itô & Mester (1992), Itô (1990), as well as other subsequent works, attempt to provide an explanation for the ill-formedness of *H, *LH, and *LHL patterns, none of them directly address the issue of the variation in the allowed patterns, that is, the factors which determine the exact length (between 2 to 4 moras) of the truncation for a given base. Another shortcoming shared by previous treatments is that they cannot account for the gaps observed in the HH and LLH patterns, which, as we will show, are not possible SALs. This is true, not only, of Itô (1990)'s left edge requirement, and of Itô & Mester (1992)'s Left Edge Matching constraint, but it is also true of Kubozono, Itô & Mester (1997) and Itô & Mester, (1995) (cited by Kubozono, in press), which invoke a nonfinality constraint of the form Nonfinality(σ).

In an attempt to propose a theoretical solution to the variability in the allowed patterns, Itô & Mester (1995, cited by Kubozono, in press) formulate the hypothesis that two-mora, three-mora and four-mora loanwords are subject to the same set of constraints, but according to different rankings. However, this just postpones the problem as to why two-mora, three-mora and four-mora abbreviations should display a different constraint ranking. Some of the problems encountered by previously proposed analyses stem from the fact that they do not consistently adopt a strict distinction between compound and non-compound abbreviation.

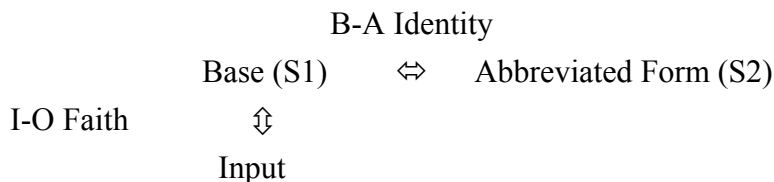
2.2. Correspondence Theory

Correspondence Theory is a framework originally developed within Optimality Theory by McCarthy & Prince (1995) to account for morpho-phonological phenomena such as reduplication and truncation. Correspondence is a relation between two strings (the "correspondents"), S1 and S2, which, in the case of derivational morpho-phonology such as reduplication or abbreviation, are two actual free standing forms of the language, and are therefore related to one another as output-to-output, rather than input-to-output (underlying form / surface form). This relation is

evaluated by identity constraints.

Benua (1995) proposes the following correspondence model for morphological abbreviation¹, inspired by the one proposed by McCarthy & Prince (1995) for reduplication.

(3) Abbreviation (adapted from Benua 1995)



The basic idea is that the base of the abbreviation (S1) is already an output form, and the abbreviated form (S2) is related to the surface form of its source, not to the original input of the base. The general identity constraints are dominated by other general prosodic constraints, which trigger abbreviation as such, and are responsible for the loss of segmental material between S1 and S2. One of the claims made by Benua (1995) is that “the phonological [properties] of truncated forms are identity effects forced by constraints demanding identity between truncated forms and their source word. These constraints [...] regulate the correspondence relation between the source word base and the truncated form [...]” So the theory predicts that phonological characteristics directly inherited from the output form of the base, themselves resulting from the application of constraints, will already be available at the stage at which truncation applies. We will see in this paper that the morpho-phonology of SALs provides additional evidence that this assumption is correct.

2. 3. The corpus

The first step of this research has been to collect as many items as possible to elaborate a corpus that would be representative and could serve as a basis for quantitative evaluations. We assume that significant and representative results can only be obtained from the exploitation of a large corpus, collected in a systematic fashion, and not just on the examination of a small number of words gathered randomly. This implies that definite criteria have to be explicitly adopted, and strictly respected, for the constitution of the corpus.

As has been frequently observed, most abbreviated words in Japanese (and in other languages as well) are often familiar or jargon-like derivations (Morioka, 1988), sometimes very local and primarily belonging to some specialized or professional vocabulary, created and used among definite and restrained socio-professional groups, before they eventually come to be used by a larger number of people, and their existence noticed by lexicographers. For this reason, the corpus also includes a few words which are not necessarily widely used, or words which might even have been ephemeral creations by a single speaker, but they have all actually been attested and used by native speakers. Given the specific jargon-like nature of some abbreviated forms, a narrow reference to some “average” or “ideal” speaker would jeopardize the significance of the results. What is of significance for the present study is that *sutetosukoopu* “stethoscope” yields *suteto* rather

than **sute* or **sutetosu*, and *misutaa doonatu* “Mister Donuts”, *misudo* rather than **misudoo*, **misudona*, or **mido*. Such forms were included because we assume that the morpho-phonology of abbreviation obeys very general principles, which apply regardless of word frequency or diffusion. The postulate is that because such principles exist, abbreviation is a productive process for creating new words in Japanese, and thus deserves theoretical attention².

However, this does not mean that just any abbreviated loanword has been randomly included in the corpus. The following criteria have been strictly observed in collecting the data:

i) the abbreviation functions as an autonomous word. Application of this criteria excludes items such as *supa* for “spaghetti”, *paa* for “percent”, etc. which only appear as bound forms (in *sarasupa* “salad spaghetti”, *hyakupaa* “100%”).

ii) the abbreviation is not based on the written form of the source word. Acronymic formations such as *ooeeru* (OL) (*ofisu redii* “office lady”) are thus excluded.

iii) the abbreviation has been created in Japanese, from a complete form attested in Japanese and originally borrowed recently from a foreign language, generally European. This excludes words like *tereson* “telethon”, *supekku* “spec (specification)”. Note that cases like *baio* “bio” or *anpu* “amp”, which exist as abbreviations in the source language (English *bio* < *biological*, *amp* < *amplifier*), but could equally well have been created after being borrowed in Japanese, have been included in the corpus.

iv) the abbreviation is perceived by Japanese speakers as a reduced form. Abbreviations such as *on'ea* “on the air”, *depaato* “department (store)”, *on za rokku* “on the rocks”, *sumooku saamon* “smoked salmon” (the element deleted in Japanese has been underlined) have therefore not been included. It has been frequently noticed in the literature (Ishino 1983 :121 *sq.*, and others) that numerous foreign loans used in Japanese are incorrect from the point of view of the original language, as they frequently omit some element (preposition, article, suffixes such as *-s*, *-ed*, *-ing*, *-ment*, etc.) which is grammatically necessary in the source language, but which is seen as superfluous in the Japanized form³. Note that words abbreviated through truncation of the suffix *-mento*⁴ (*-ment*) have been excluded from the corpus as a principle.

v) The truncated part in the source word does not correspond to a complete lexical word. This excludes abbreviations which are syntactic or semantic in nature (ellipsis), like *amerikan* for *amerikan koohii* “American coffee”, *waiaresu* for *waiaresu maikurohon* “wireless microphone”, *desin* for *kureepu de sin* “crêpe de Chine”.

vi) The abbreviated form has been derived from an attested Japanized complete form: words such as *ryuumati* and *roimati*, from English “rheumatism” and German “Rheumatismus” are not included since the *katakana* form of the adapted words is unknown.

When the Japanized base is clearly identified, but not used any longer, or too rare to be included in accent dictionaries, its accentuation has been verified with native speakers of standard Japanese. However, such words represent only a small proportion of the total corpus.

In sum, the words collected for this study are words which can be analyzed as resulting from a derivation process which can be termed phonological in nature. They all correspond to autonomous abbreviated forms satisfying the above criteria, and their use has been verified with native speakers

belonging to the socio-professional groups who use them. A few pre-Meiji (1868) items, mentioned by dictionaries, have also been included in the corpus for the sake of representativeness and diversity.

The corpus has been gathered through systematic exploitation of dictionary sources⁵. These have been completed with a large number of oral and written forms collected mainly in Tokyo during 2001. A single occurrence of an abbreviated form (oral or graphic) was considered sufficient to justify its inclusion in the corpus, provided that it met the criteria set above.

Following the application of the above principles, a corpus of 675 abbreviated Japanese loanwords has been collected. Among those, 325 words (48.1%) correspond to SALs (“simple” (= non-compound) abbreviated loans). This paper will be solely concerned with the analysis of this latter type. The reason for this choice will be exposed below.

2. 4. Different types of abbreviated loanwords

Previous studies on the morpho-phonology of abbreviated loanwords in Japanese have brought significant results concerning the general properties of this part of the lexicon, but have failed to account for the actual length of derived truncated words, which are generally 2, 3 or 4 moras long, as we will see below. One of the reasons for this flaw lies in the fact that these works did not operate a sufficiently strict distinction between simple abbreviated loanwords (SALs) and compound abbreviated loanwords (CALs). These two types clearly display different lexical and prosodic properties, as table 1 shows.

SALs are made by combining moras extracted from a single word⁶ as in (4). The reduction process consists in the deletion of one or several moras of the source word, generally from the right end, but in some cases from the left end (aphaeresis) or medially (discontinuous abbreviations). Note that the base can be a compound, but the phonological material appearing in the abbreviation comes from only one word of the base, as in *paama* or *konbi*.

(4)

terebi	< terebizyon	“television”
konbini	< konbiniensu (sutoa)	“convenience store”
neru	< huranneru	“flannel”
mentamu	< mensoreetamu	“mentholetum”

CALs are made by combining one or two moras (exceptionally three) taken from more than one word.

(5)

mo-ga	< modan gaaru	“modern girl”
ni-kado	< nikkeru kadomiumu	“nickel cadmium”
aru-saro	< arubaito saron	“arbeit salon”
eba-miruku	< ebaporetiddo miruku	“evaporated milk”

The distinction between these two types is crucial, since the principles governing the derivation of abbreviations produce a different result whether the form involved is a SAL or a CAL. The main difference lies in the observation that the prosodic patterns differ remarkably for 2, 3 and 4 mora long abbreviations, as shown in table 1.

Table 1 : Prosodic structure of simple abbreviations (SAL) and compound abbreviations (CAL)⁷

Prosodic Structure	SAL	CAL	Total
1 μ L	0	0	0
2 μ LL	113	5	118
H	3	5	3
	116		121
3 μ LLL	59	35	94
LH	1	4	5
HL	64	12	76
	124	51	175
4 μ LLLL	35	102	137
LLH	8	82	90
HLL	31	58	89
LHL	5	3	8
HH	4	36	40
	83	281	364
5 μ	0	10	10
6 μ	0	3	3
7 μ	2	0	2
Total	325	350	675

The crucial observation is that CALs are overwhelmingly four moras long (281 instances, i.e. 80.3% of the total number of CALs), whereas SALs display a more balanced repartition between two, three and four moras. The trimoraic pattern is the favored one for SALs (124, i.e. 38.2%), just ahead of the bimoraic one (116, i.e. 35.7%); the quadrimoraic pattern (83, i.e. 25.5%) is slightly less represented, but it is still quite common. This shows that SALs and CALs should be strictly differentiated.

Depending on the edge of the base where phonological material is subtracted, and on whether the deleted elements form a contiguous string or not, SALs can be further divided into three different types: apocopes (or truncations), aphaereses, and discontinuous abbreviations. In apocopes, the final part of the source word has been deleted, as in *biru* < *biru{dingu}* “building”, or *interi* < *interi{gentiya}*, “intelligentsia”. Such forms amount to 277 words. In aphaereses, the initial part of the source word has been deleted, as in *neru* < *{huran}neru* “flannel”, or *baito* < *{aru}baito* “Arbeit”. Such forms amount to 27. In discontinuous abbreviations a medial part of the source word, or several discontinuous parts of it, have been subtracted, or some material has been added, as in *torapen* < *tora{nsu}pe{are}n{sii}* “transparency”, or *koruresu* < *koresu{pondento}* “correspondent”. Such forms amount to 21.

Having shown how the different types of abbreviations should be distinguished, we will now turn to an examination of truncated SALs (apocopes).

3. A constraint-based analysis of truncated SALs

The analysis will now focus on apocopes, which represent the most numerous type of SALs.

3.1. Two well-formedness constraints : BIN and *H#

Let us look more closely at the prosodic organization of truncated SALs :

Table 2: Prosodic structure of truncated SALs.

Prosodic Structure		Total	
2μ	LL	103	105 (37.9%)
	H	2	
3μ	LLL	56	106 (38.3%)
	LH	1	
	HL	49	
4μ	LLLL	33	65 (23.5%)
	LLH	2	
	LHL	5	
	HLL	25	
	HH	0	
7μ	LLLHH	1	1
Total		277 (100%)	

Observe that quadrimoraic words, though less numerous than bimoraic and trimoraic ones, still represent a very common pattern (23.5%) for SALs. There is only one word longer than four moras : *asutorinzen* < *asutorinzento* “astringent”, which might have been borrowed from the oral form in English. There is no monomoraic form. This confirms that the optimal length for truncated non-compound abbreviations is two, three or four moras. Table 2 also reveals that LLH and HH patterns are marginal and non-productive for SALs. This is particularly significant since LLH and HH patterns are perfectly fine for CALs, as table 1 indicates. Therefore the dominant prosodic patterns for truncated SALs are much more restricted than what is generally assumed (see for instance (1) in § 2.1). It is clear that *LLH and *HH, although frequent in CALs, are unproductive in truncated SALs, just like the *H and *LH patterns⁸.

(6) Productive patterns for truncated SALs

1-mora *L

2-mora LL demo *H

3-mora HL konpa LLL arumi *LH

4-mora HLL bangura LLLL anakuro *LHL *LLH *HH

It turns out that LL, LLL, LLLL, HL and HLL constitute the only productive patterns for SALs. Next, we observe that two properties of the licit patterns in (6) stand out :

- i) They are prosodically binary: (LL), (LL)(L), (H)(L), (LL)(LL), (H)(LL)⁹
 (no *(L), *(L)(H)(L), *(H)(H)(H), *(H)(LL)(L), etc.).
- ii) They end with a light syllable : LL, LLL, HL, HLL, LLLL
 (no *H, *LH, *LLH, *HH, *HHH, etc.)

This generalization can be stated in the form of two well-formedness constraints :

- (7) BIN
 The output is prosodically binary (under foot, syllable, or mora analysis)
- (8) *H#
 The output must not end in a heavy syllable

BIN is roughly equivalent to the Word Binarity requirement (« P-derived words must be prosodically binary ») in Itô & Mester (1992), except that in our approach, BIN also applies to moraic binarity. Recall that in Itô & Mester (1992), syllable branching (bimoraicity) does not count as binarity because syllable internal structure is supposed to be opaque. Such a restriction was rendered necessary in order to rule out monosyllabic bimoraic outputs such as **paa*. However, in our analysis, the ill-formedness of **paa* results from its ending in a heavy syllable, not from its being unary, so an additional - and somewhat *ad hoc* - specific principle such as syllable opacity is no longer needed. The interaction of BIN and *H#¹⁰ provides a straightforward account for the fact that the productive patterns for truncated SALs are LL, LLL, HL, HLL and LLLL, as shown in line iii), table 3. These are precisely the patterns which respect both *H# and BIN :

Table 3: Interaction of BIN and *H#

Possible patterns of truncated loanwords	BIN	*H#
i) L, LHL, HLLL, LLHL, HHL, etc.	no	yes
ii) H, HH, LH, LLH etc.	yes	no
iii) LL, LLL, HL, HLL, LLLL	yes	yes
iv) HHH, LLLH, HLLH, etc.	no	no

The unproductivity of truncated forms whose prosodic shape is *L, *LHL, or which are over five moras (*LLLLL, *LHHL, *HHL, etc.) arises from the fact that none of these structures is prosodically binary. Contrary to other analyses, the ill-formedness of *H, *LH, *LLH, *HH now receives a unified treatment. This represents a significant advantage.

Moreover, building on the data provided in table 3, Itô's assumption that "a form like *irasuto*, although derived from a single loanword *irasutoreesyon* "illustration", is prosodically identical to the four-mora compounds", for example *sara-dore* or *rimo-kon*, cannot be supported. If this were so, four-mora simple truncated words ending with a heavy syllable would be tolerated, but as table 2

shows, this is not the case.

3.2. Determining the length of the truncated form : the correspondence constraint ALIGN(A, Right, HeadFoot of Base, Left)

We have just seen that constraints BIN and *H# account for the absence or quasi-absence of the patterns i, ii and iv in table 3. However, these two constraints cannot account for the actual length of the derived abbreviated word. BIN allows it to be two, three or four moras long, but it fails to predict the exact number of moras of the abbreviation. That is to say, considering that there exist five productive patterns, one could, in theory, expect that given a single base, two or even three different truncated forms, all satisfying BIN and *H#, would be possible. For example, the base *aruminiumu* “aluminium” could yield *aru*, *arumi* or *arumini*, and *banguradesyu* “Bangladesh”, *bangu* or *bangura*. Notice that the truncated forms just mentioned all perfectly conform to BIN and *H#. However, only the truncated forms *arumi* and *bangura* respectively are attested¹¹.

The crucial issue we now have to consider is : How do we predict the length of the truncated word ? In other words, where do we cut the base ? To answer this question, we must consider something that has not been pointed out in the literature : the connection of the truncation site with the accent location of the base. We will formulate the following hypothesis:

(9) HYPOTHESIS: the base is truncated immediately before the accented mora.

Indeed, a close examination of the corpus reveals that in a significant number of cases (134 out of 277), the derived form has been truncated just before the accented mora of the base¹², as in (10): (In the rest of the paper, the accent of the base will be indicated when relevant, and the accent of the SAL is ignored).

(10)			
LL	tero	< terorízumu	“terrorism”
LLL	kosume	< kosumetíkku	“cosmetic”
HL	sando	< sandoítti	“sandwich”
LLLL	irasuto	< irasutoréesyon	“illustration”
HLL	sinkuro	< sinkuronáizu	“synchronize”
LHL	huranku	< hurankuhúruto	“Frankfurt”

The statement that “the base is truncated immediately before the accented mora” is nothing more than the translation of an empirical observation. We assume that what is crucially involved here is prosodic structure : the split occurs just before the head foot of the base. Japanese feet being trochaic, the accented mora is situated at the left edge of the head foot, so what is really relevant is the foot structure of the base, rather than the location of the accented mora *per se*. Following some comments by two reviewers, we acknowledge that one should ask ‘why’ the base should be truncated at the left edge of the accented mora, and in no other possible location (right edge of the

accented mora, right edge of the head foot, etc.). We have no answer for the moment, and shall leave this issue to future research.

Another piece of evidence demonstrating the crucial role of the prosodic organization of the base lies in the observation that nearly all the bases which are unaccented, or accented on their first or second syllable, are regular in our treatment. In other words, irregular apocopes present the common property of being accented on a syllable which is over the third syllable of the base. We will provide further evidence that the output form of an abbreviation is directly related to the prosodic organization of the base when we turn to the examination of aphaereses in section 4.

To capture the fact that the split occurs immediately before the accented mora of a base, we introduce the following alignment constraint:

(11) ALIGN (A, Right, HeadFoot of B, Left)

Align the right edge of the SAL (A) with the left edge of the head foot (accented foot) of the base (B).

ALIGN (A, Right, HeadFoot of B, Left) is an output-to-output correspondence constraint demanding alignment of the right edge of the abbreviated word (A) with the left edge of the head (accented) foot of the base (B). This ensures that the site of the truncation coincides with the left hedge of the head foot of the base, i.e. the accented mora¹³. If the right edge of A and the left edge of the head foot of B coincide, like the two halves of a broken plate, as in *{kosume}+{tikku} < kosumetikku* “cosmetic”, the alignment constraint is satisfied; otherwise, it is violated, and the candidate which has fewer violations of ALIGN(A, Right, HeadFoot of B, Left) (thereafter ALIGNRL) is selected. Such violations to ALIGNRL are evaluated by counting the number of moras standing between the designated edge and the end of the truncated candidate.

The implication is that the correct output forms depend on the prosodic shape of the base, and not only on prosodic constraints on the final shape of the word. Moreover, as the default accent in Japanese foreign loans is assigned from the end of the word (McCawley 1968, Akinaga, 1981, Sibata, 1994, Kubozono, 1996), the tendency we observe for a long base to yield a relatively long truncated form, or, to put it differently, the fact that the length of the derived word is generally proportional to the length of the base, receives a straightforward explanation.

These findings confirm Benua’s (1995) proposal that abbreviations are faithful to the output form of the base, because foot organization or accent location is typically the kind of information which results from the ranking of constraints, and is therefore only available at the surface level.

In the case of bases bearing an accent on their first or second mora, ALIGNRL is violated in order to satisfy BIN and *H#. These two constraints then determine the length of the abbreviated form : two moras for bases starting with a light syllable (12a-c) ; three moras for bases starting with a heavy syllable (12d-e). This shows that BIN and *H# are ranked higher than ALIGNRL, because otherwise, the outputs for *gyarántii* or *konpóonento* would be **gya* and **kon* or **ko*. There are 56 instances of the types illustrated in (12), on a total of 277.

(12)

a. #L-	biru	μμ	*bi	< bírudingu	“building”
b. #L-	gyara	μμ	*gya, *gyaran	< gyarántii	“guarantee”
c. #L-	kyapa	μμ	*kya	< kyapásitii	“capacity”
d. #H-	konpa	μμμ	*ko, *kon	< kónpanii	“party”
e. #H-	konpo	μμμ	*ko, *kon, *konpoo	< konpóonento	“component”

Let us now turn to longer bases, accented on a mora which is located further than the fifth mora from the beginning. There are only ten words of this type in the corpus. Eight of them can be accounted for following the same line of analysis.

(13) Longer bases

		Form satisfying ALIGNRL		
LL	demo	*demonstoto	< demonsutoréesyon	“demonstration”
LLL	yuniba	*yunibaasi	< yunibaasiáado	“universiade”
HL	appe	*appendi	< appendisáitisu	“appendicitis”
LLLL	urazio	*uraziosu	< uraziosutókku	“Vladivostock”
LLLL	terekomi	*terekomyuni	< terekomyunikéesyon	“telecommunication”
LLLL	sutorobo	*sutorobosu	< sutorobosukóopu	“stroboscope”
HLL	insuto	*insutoru	< insutoruméntaru	“instrumental”
HLL	pankuro	*pankuroma	< pankuromatíkku	“panchromatic”

In (13), the candidates best satisfying ALIGNRL are over four moras long, and thus violate BIN. As we will see in the tableaux below, the correct output for longer words is the candidate which satisfies BIN and *H#, while minimally violating ALIGNRL¹⁴.

3.3. Additional constraints

Before proceeding with the detailed analysis, let us introduce three additional constraints. For the sake of brevity, only the decisive constraints are presented here. The difference between apocope and aphaeresis can be captured by positing an anchoring constraint of one edge of the SAL with one of the edges of the base :

(14) ANCHOR(*edge*)

The SAL must contain an element from the designated edge of the base

ANCHOR(*edge*) is crucial in explaining the difference between apocopes and aphaeresis. In apocopes, the designated edge is the Left edge (the beginning of the base), whereas in aphaeresis, it is the Right edge (the end of the base). As for discontinuous SALs, they appear to be Left-anchored, but for one exception. Abbreviations from longer words such as *demo* from *demonstotoréesyon* (23) tend to prove that ANCHOR(*edge*) is ranked higher than BIN, *H# and ALIGNRL. Otherwise,

**monsuto* would have been a better output for *demonstutoréesyón*, insofar as it respects the three constraints just mentioned. (See also the discussion on aphaeresis for further evidence that ANCHOR is ranked above *H#.)

(15) ALLFEET(*edge*)

Every foot stands at the designated edge of the SAL.

The relevant edge is Left for apocope, Right for aphaeresis. ALLFEET(*edge*) is necessary to account for the fact that apocopes and aphaereses from unaccented bases tend to be bimoraic. This is because ALIGN obviously cannot apply in unaccented bases, which renders the action of the lower-ranked constraint ALLFEET(*edge*) visible (see tableau 25 below). Violations to ALLFEET(L) are evaluated by counting the number of moras between the right edge of the initial foot and the right edge of the word.

As for the ranking, a form such as *denomi*, from *denominéesyón* (see 20 below), shows that ALLFEET(*edge*) is ranked lower than ALIGNRL, otherwise, **deno* would be the correct output for *denominéesyón*.

(16) CONTIG

Corresponding elements in the base and in the SAL form a contiguous string.

CONTIG is needed to account for the difference between apocopes and aphaereses, on one hand, and discontinuous abbreviations, on the other hand. Since CONTIG is always satisfied in apocope and aphaeresis, it is ranked higher than BIN and *H#. However, we find no definite evidence for ranking it with respect to ANCHOR, so we shall leave these two constraints mutually unranked at the top of the hierarchy.

The constraints posited so far are ranked according to the following hierarchy :

(17) Constraint hierarchy :

ANCHOR(*edge*) , CONTIG >> BIN >> *H# >> ALIGNRL >> ALLFT(*edge*)

We assume that BIN is ranked above *H#. However, we will not make a strong stand on this point, because the evidence for this assumption lies in the existence of only one SAL, namely *guu*, from *guddo* “good” (classified as a discontinuous SAL). The binary form *guu* appears as the optimal output for *guddo*, even though it violates *H#, rather than the light syllable ending form **gu*.

3.4. Tableaux

How the above set of constraints work to produce the correct outputs will be exemplified below. Due to space limitation, we will concentrate on a few representative cases. Candidates differing only by foot construction are not presented in the tableaux, and CONTIG is not taken into

account. Note that the presence of ANCHOR in the following tableaux might seem irrelevant at this point of the demonstration, because we are only dealing with apocopes. However, its full significance will become clear when we turn to the study of aphaeresis.

Tableaux (18) and (19) illustrate how monosyllabic L or H apocopes are ruled out by the ranking of BIN and *H# over ALIGNRL in #L- and #H- initial bases bearing an accent on their second syllable :

(18) gyarántii -> gyara LL “guarantee”

BASE: gyarántii	ANCHOR(L)	BIN	*H#	ALIGNRL	ALLFT(L)
(gya)		*!			
☞(gyara)				μ	
(gya)(ran)			*!	μμ	**
(gya)(ran)(ti)		*!		μμμ	***

(19) konpóonento -> konpo HL “component”

BASE: konpóonento	ANCHOR(L)	BIN	*H#	ALIGNRL	ALLFT(L)
(ko)		*!		μ	
(kon)			*!		
☞(kon)(po)				μ	*
(kon)(poo)			*!	μμ	**
(kon)(poo)(ne)		*!		μμμ	***

Tableaux (20) and (21) show the action of ALIGNRL :

(20) denominéesyon -> denomi LLL “denomination”

BASE: denominéesyon	ANCHOR(L)	BIN	*H#	ALIGNRL	ALLFT(L)
(de)		*!		μμ	
(deno)				μ!	
☞(deno)(mi)					*
(deno)(mine)				μ!	**
(deno)(mi)(nee)			*!	μμ	***
(deno)(mi)(nee)(syo)		*!		μμμ	****

(21) interigéntiya -> interi HLL “intelligentsia”

BASE: interigéntiya	ANCHOR(L)	BIN	*H#	ALIGNRL	ALLFT(L)
(i)		*!		μμμ	
(in)			*!	μμ	
(in)(te)				μ!	*
☞(in)(teri)					**
(in)(teri)(ge)		*!		μ	***
(in)(teri)(gen)		*	*!	μμ	****

The case of initially accented bases, yielding dissyllabic SALs, is illustrated in (22):

(22) bíruḡingu -> biru LL “building”

BASE: bíruḡingu	Anchor(L)	BIN	*H#	ALIGNRL	ALLFT(L)
(bi)		*!		μ	
☞(biru)				μμ	
(biru)(di)				μ!μμ	*
(biru)(din)			*!	μμμμ	**

Tableaux (23) and (24) illustrate the action of BIN in longer bases. In (23), satisfaction of ALIGNRL produces the derived word **demonḡuto*, which is not binary. The correct output candidate is *demo*, which satisfies the higher constraints BIN and *H#, although it has three violations of ALIGNRL.

(23) demonsutoréesyon -> demo LL “demonstration”

BASE: demonsutoréesyon	ANCHOR(L)	BIN	*H#	ALIGNRL	ALLFT(L)
(de)		*!		μμμμ	
☞(demo)				μμμ	
(de)(mon)			*!	μμ	**
(de)(mon)(su)		*!		μ	***
(de)(mon)(suto)		*!			****
(mon)(suto)	*!				**

In (24), observe that *uraziosu*, the candidate best satisfying ALIGNRL is ruled out because it infringes BIN. Consequently, quadrimoraic *urazio* is selected over trimoraic **urazi* and bimoraic **ura*, because it displays less violations of ALIGNRL among all the possible outputs which are both binary and end in a light syllable. This case illustrates that ALIGNRL is a gradient constraint.

(24) uraziosutókku -> urazio LLLL “Vladivostock”

BASE: uraziosutókku	ANCHOR(L)	BIN	*H#	ALIGNRL	ALLFT(L)
(u)		*!		μμμμ	
(ura)				μ!μμ	
(ura)(zi)				μ!μ	*
☞(ura)(zio)				μ	**
(ura)(zio)(su)		*!			***

Finally, we see in (25) that ALIGNRL is not applicable in unaccented bases. ALLFEET(L) is then responsible for the winning of bimoraic *ama* over trimoraic *amatyu*.

(25) amatyua -> ama (unaccented base) LL “amateur”

BASE: amatyua	ANCHOR(L)	BIN	*H#	ALIGNRL	ALLFT(L)
(a)		*!		n. a.	
☞(ama)				n. a.	
(ama)(tyu)				n. a.	*!
(matyu)	*!			n.a.	

3.5. Evaluation of the analysis

In order to establish the respective covering of each of the constraints involved in the

derivation of truncated SALs, we proceeded to a detailed breakdown of how many words could be accounted for if the set of the most relevant constraints was reduced to four, instead of five as in the tableaux above. Here are the result of this investigation :

(26) Number of truncated SALs accounted for given the following constraint subsets (ignored constraint crossed out) :

a. Without ANCHOR(L):

~~ANCHOR(L)~~ >> BIN >> *H# >> ALIGNRL >> ALLFT(L) = 0 (277 exceptions)

b. Without BIN:

ANCHOR(L) >> ~~BIN~~ >> *H# >> ALIGNRL >> ALLFT(L) = 133 (144 exceptions)

c. Without *H#:

ANCHOR(L) >> BIN >> ~~*H#~~ >> ALIGNRL >> ALLFT(L) = 181 (96 exceptions)

d. Without ALIGNRL:

ANCHOR(Left) >> BIN >> *H# >> ~~ALIGNRL~~ >> ALLFT(L) = 152 (125 exceptions)

e. Without ALLFEET(L):

ANCHOR(L) >> BIN >> *H# >> ALIGNRL >> ~~ALLFT(L)~~ = 193 (84 exceptions)

(27) Number of truncated SALs accounted for given the following complete constraint set :

ANCHOR(L) >> BIN >> *H# >> ALIGNRL >> ALLFT(L) = 196 (81 exceptions)

We therefore conclude that each of the four top constraints proposed within our analysis, including *H# and ALIGNRL, is necessary in order to account for the abbreviation patterns of truncated SALs. ALLFEET(L) might seem more accessory, but note that it will be more active with aphaeresis.

As a matter of fact, this approach captures the principles governing the truncation patterns of 70.8% of the total number of truncated words (196 from a total of 277), in a simple fashion, relying only on a few very general constraints.

3.6. Typology of the exceptions

The analysis developed thus far accounts for the abbreviation pattern of a large majority of truncated SALs. However, there remain 81 exceptions (29.2%). We will now turn to a close examination of these exceptions, and see that a majority of them actually share some puzzling properties. The first important point to be raised is that a large majority of irregular forms should have been ruled out by fatal infringement of the ALIGNRL constraint, as the following figures indicate:

(28) Number of exceptions incurred by a ‘should-have-been-fatal’ violation:

- Exceptions to BIN = 6
- Exceptions to *H# = 5
- Exception to ALLFT(L) = 2
- Exceptions to ALIGNRL = 68

These figures could cast doubt on the validity of ALIGNRL: they could be interpreted as meaning that ALIGNRL is actually not so relevant in the abbreviation process. However, there is good reason to dismiss such an interpretation.

The first argument lies in the fact that, as stated earlier, bases which are accented on the initial or second mora, and bases which are unaccented, are overwhelmingly regular (more than 95%). This constitutes undeniable evidence that the abbreviation pattern is correlated to the prosodic structure of the base, otherwise we would expect to find similar exception rates for all types of bases, regardless of their accentual structure. A second type of evidence is provided by aphaereses: we will see in section 4 that about 90% of aphaereses can be properly accounted for within the same line of analysis. The third argument comes from the observation that forms which are irregular with respect to the present analysis share some common and intriguing properties. A preliminary examination of the 81 irregular truncated SALs reveals that about 73% of the truncated forms under consideration apparently display one of the two following particularities (in some cases both of them):

i) The expected (but incorrect) output would have ended with an epenthetic vowel (*u* or *o*, underlined), such as in (29). This subgroup constitutes the main group of exceptions, with 35 examples (i.e. 43.2% of the exceptions).

(29)

inhure	*inh <u>u</u>	< inhuréesyon	“inflation”
kamuhura	*kamuh <u>u</u>	< kamuhuráazyu	“camouflage”
akuse	*ak <u>u</u>	< ákusesarii, akusésarii	“accessory”
misupuri	*misup <u>u</u>	< misupurínto	“misprint”
saamo	*saamos <u>u</u>	< saamosutátto	“thermostat”

ii) The expected (but incorrect) output would have ended with a *rV* mora (24 examples, i.e. 29.6%):

(30)

huraku	*hura	< hurákusyon	“fraction”
korokke	*koro	< korokkétto	“croquette”
deforume	*deforu	< deforuméesyon	“deformation”
huranku	*hura	< hurankuhúruto	“Frankfurt”
puremia	*pure	< puremiamu, purémiamu	“premium”

These examples suggest that the presence of an epenthetic vowel or of a rV mora occurring before the split might disrupt the abbreviation process, presumably by preventing satisfaction of ALIGNRL. To verify whether this assumption is correct, a more detailed investigation is necessary. As a first step, let us consider the abbreviation patterns involved in the exceptions, in comparison with those of regular words.

Table 4. Abbreviation patterns involved in regular and irregular outputs.

	Regular outputs	Irregular outputs	Total
LL	97	6	103
H	0	2	2
LLL	27	29	56
LH	0	1	1
HL	42	7	49
LLLL	15	18	33
LLH	0	2	2
LHL	0	5	5
HLL	15	10	25
LLLHH	0	1	1
Total	196	81	277

We notice first that the bulk of exceptions is concentrated in LLL, LLLL and, to lesser extent, in HLL outputs. Observe that, strikingly, LLL and LLLL exceptions even outnumber regular forms. Secondly, we shall verify whether an epenthetic vowel, or a rV mora occurring before the accented mora of a base¹⁵ are more likely to constitute a decisive factor for exception. In order to do so, we shall determine the number of epenthetic vowels immediately preceding the accented mora ($\underline{v}'\mu$) in the 196 regular forms, and compare it with the number of such vowels in the sub-group of 81 irregular forms, with regard to the prosodic structure of the output. For example, in the case of a base such as *saamosutátto* “thermostat” which yields the irregular HL abbreviation *saamo* (rather than the expected **saamosu*), we count one instance of $\underline{v}'\mu$ on the HL line in the irregular forms column.

We shall then consider whether the presence of a rV mora before the accented mora of the base ($rV'\mu$) is a factor for irregularity. We shall proceed exactly as we do for epenthetic vowels : for example, a word such as *intorodákusyon* “introduction” yielding the regular *intoro* is counted as one instance of $rV'\mu$ in the HLL line of the regular forms column.

The results of these surveys are presented in tables 5 and 6 below.

Table 5. Number of epenthetic vowels preceding the accented mora ($\underline{v}'\mu$). A comparison between regular and irregular forms, with regard to the abbreviation pattern.

	Number of $\underline{v}'\mu$ instances within bases		
	Regular forms	Irregular forms	Total
LL	18 90%	2 10%	20 100%
LLL	9 39.1%	14 60.9%	23 100%
HL	4 66.7%	2 33.3%	6 100%
LLLL	4 21.1%	15 78.9%	19 100%
HLL	5 71.4%	2 28.6%	7 100%
LHL	0 0%	3 100%	3 100%
LLLHH	0 0%	1 100%	1 100%
Total	40 (20.4% of 196)	39 (48.1% of 81)	79 (28.5% of 277)

Table 6. Number of rV moras preceding the accented mora ($rV'\mu$). A comparison between regular and irregular forms, with regard to the abbreviation pattern.

	Number of $rV'\mu$ instances within bases		
	Regular forms	Irregular forms	Total
LL	25 92.6%	2 7.4%	27 100%
LLL	4 30.8%	9 69.2%	13 100%
HL	0 0%	3 100%	3 100%
LLLL	6 75%	2 25%	8 100%
HLL	9 100%	0 0%	9 100%
LHL	0 0%	1 100%	1 100%
Total	44 (22.4% of 196)	17 (21% of 81)	61 (22% of 277)

Table 5 shows that the rate of epenthetic vowels is significantly higher (48.1%) in exceptions than in the general corpus (28.5%). Moreover, the results in table 5 confirm that LLL and LLLL truncated SALs derived from a base in which an epenthetic vowel precedes the accented mora display a stronger tendency to be irregular. They represent respectively 60.9% and 78.9% of all the LLL and LLLL bases with $\underline{v}'\mu$. Note also that three LHL forms, infringing BIN, have an epenthetic vowel before the accented mora in the base.

We see in table 6 that unexpectedly, the rate of irregular $rV'\mu$ forms is not higher in exceptions than in regular forms. It is even slightly lower. Observe, however, that irregular LLL

forms represent an abnormally high proportion of all $rV'\mu$ instances of the corpus, reaching a rate of 69.2%. This, in our view, constitutes significant evidence, although we do not know how it should be interpreted. Note also that contrary to forms with epenthetic vowels, LLLL forms involving a rV mora are not as often irregular, whereas all HL forms are. Another significant fact is that LLL irregular forms represent more than one half (9 out of 17) of the total number of exceptions having the $rV'\mu$ property.

The quantitative data above demonstrate that epenthetic vowels act as an undeniable cause of irregularity when occurring in bases which yield LLL, LLLL and LHL SALs, and that rV moras probably do, too. This means that most of the forms which first appear as irregular can probably receive a principled account. In other words, they are more than merely inexplicable exceptions. This issue clearly deserves further investigation, and we shall leave it to future research. Other possible causes to irregularity might include synchronic or diachronic accent variation, and the avoidance of homophones.

An additional piece of evidence for recognizing epenthetic vowels and /r/ as a major cause for irregularity comes from the data provided by discontinuous SALs in section 5 : moras including these elements are precisely among the ones which are first deleted word-internally, in violation to the CONTIG constraint. Moreover, we will see in section 4 that r is also involved in irregular aphaereses. Finally, it is worth pointing out that the two types of phonological objects involved here are independently known in their own right to have a special status in Japanese: epenthetic vowels often disrupt the accent pattern in foreign loans (Kubozono, 1996, 2001, Shinohara, 2000); the liquid r displays several properties which have led to interpret it as an underspecified or unmarked consonant (Itô & Mester, 1986, Labrune 1997).

While a full understanding of some of the facts uncovered here remains to be provided, we take the results above as constituting sufficient evidence that the analysis proposed in this paper, and especially the validity of the Alignment constraint, is correct.

4. Abbreviation through aphaeresis

We will now turn to the examination of aphaeresis. Aphaeresis involves deletion of a string standing at the beginning of the base, as in *baito*, from *arubáito* “Arbeit (German)”. This type is rare, with only 27 examples in the corpus. Examination of the prosodic structure of aphaeresis abbreviations reveals that HL is the major pattern.

Table 7: Prosodic structure of aphaeresis abbreviations

Prosodic structure	Total
2 μ LL	6
3 μ HL	13
4 μ LLH	4
HH	3
7 μ HLLLH	1
Total	27

Observe that, contrary to what occurs with apocopes, heavy syllables at the end of aphaereses are not avoided: there are eight cases of H ending abbreviations in this sub-group of the corpus. This is due to the action of the high-ranked constraint ANCHOR(Right), which demands that the right edge of the base and the right edge of the derived word coincide.

(31) ANCHOR(Right)

A SAL must contain an element from the Right edge of the base

As ANCHOR is ranked over *H#, no element can be deleted at the end of the base. It is precisely the specification of the relevant edge for ANCHOR which accounts for the difference between apocope and aphaeresis regarding final heavy syllables¹⁶.

A closer examination indicates that in 20 cases out of 27, the subtracted part corresponds exactly to the portion of the word preceding the accented mora in the base as in :

(32)

baito	< arubáito	“Arbeit”
dakusyon	< purodákusyon	“production”
ettaa	< daiéttaa	“dieter”
nyuumu	< aruminyúumu	“aluminium”
risurin	< gurísurin	“glycerin”

This confirms the claim that the prosodic structure of the base constitutes relevant information for the derivation, and that an output-to-output identity constraint plays a role in the derivation of SALs. It is clear, in apocope and in aphaeresis alike, that the left edge of the accented mora of the base constitutes the “key point” for the abbreviation process, the very place where splitting is most likely to occur, provided that the material obtained at the issue of this operation satisfies certain prosodic requirements.

Aphaeresis can be accounted for by the action of an alignment constraint identical to the one posited for apocopes, except for one edge specification. Instead of aligning the *right* edge of the derived word with the left edge of the headfoot of the base, we now have to align the *left* edge of the derived word with the left edge of the headfoot of the base. The fact that the same alignment constraint, with different parameters, is involved in apocope and aphaeresis constitutes a strong argument for its validity.

(33) ALIGN (A, Left, HeadFoot of B, Left)

Align the left edge of the SAL (A) with the left edge of the head foot of the base (B)

ALLFEET(*edge*) should now be specified as Right, i.e. ALLFEET(R). As for the constraint hierarchy, it is the same as in (17). The mechanism for aphaeresis is exemplified in (34):

(34) baito < arubáito “Arbeit”

BASE: arubáito	ANCHOR(R)	BIN	*H#	ALIGNLL	ALLFT(R)
(a)	*!	*		μμ	
(aru)	*!			μμ	
☞(bai)(to)					*
(ru)(bai)(to)	*!			μ	***
(ito)				μ!	

If the base is unaccented, or trisyllabic with an accent on the initial mora, only the final foot of the base is kept in the abbreviation. ALLFEET(R) guarantees that the outcome of unaccented bases will be bimoraic :

(35)

neru	< huranneru (n.a)	“flannel”
teki	< bihuteki (n.a)	“beefsteak”
nisu	< wánisu	“varnish”
nasu	< bóonasu	“bonus”

There are only three exceptions (11.1%) : *ketto*, **ranketto* < *buránketto* “blanket”, *pasu*, **ripasu* < *káripasu* “calipers” and *gootabiritii*, **biritii* < *sukeepugootabiritii* “scapegoatability”. Observe that two of them should have started with /r/, whereas the last one, *gootabiritii* is probably motivated by lexical reasons.

The question arises of why aphaeresis is sometimes preferred to apocope. We have no definite answer. According to Kanno (1985), loanwords derived by means of aphaeresis are generally older formations. But the phonetic nature of the initial segment might play a role in determining the anchoring edge. It seems that certain types of consonants constitute better word left margins than others : 12 out of 27 aphaereses start with a labial (p, b, m, w), whereas only four bases do, and 10 aphaereses start with a nasal (m or n) whereas no base does. As for fricatives, they seem highly disfavored at the beginning of the SAL¹⁷. However, given the somewhat limited number of examples involved, no definite conclusion can be drawn for the moment.

5. Discontinuous abbreviations

In the examples studied so far, the material taken from the base consists in a contiguous string of segments. But there exist 21 irregular cases where the abbreviation pattern is complex and unpredictable: a final – rarely initial - string of a base is deleted, just as in apocope or aphaeresis abbreviations, but a medial part of the base is also deleted (36a), or some segments absent in the base are inserted in the abbreviation, as in (36b). These processes constitute infractions to the constraint demanding contiguity between the segments of the base copied in the abbreviation. There are 21 examples of this sort in the corpus of SALs, for instance:

(36)

a.

	SAL	Base	Deleted moras
“unbelievable”	anbiba	anbiriiba(buru)	<i>ri</i> , R
“enterprise”	enpura	entaapura(izu)	<i>ta</i> , R
“conservative”	konsaba	konsaaba(tiibu)	R
“demonstrator”	demosuta	demonsutoreeta(a)	N, <i>to</i> , <i>re</i> , R
“transparency”	torapen	toransupearen(sii)	N, <i>su</i> , <i>a</i> , <i>re</i>
“instructor”	intora	insutora(kutaa)	<i>su</i>
“morphine”	mohi	moruhi(ne)	<i>ru</i>

b.

“correspondent”	koruresu	koresu(pondento)	<i>ru</i>
“pocket”	pokke	poket(to)	Q
“good”	guu	gud(do)	R

What is of interest here is that all the medially skipped or added elements – but for three exceptions– correspond to the second part of a heavy syllable (/N/, /Q/ or /R/) in 15 instances, or to the phonological elements already identified as special during the analysis of exceptions, namely a mora containing an epenthetic vowel (7 instances) and a mora starting with *r* (7 instances) :

6. Conclusion

We started this study of Japanese abbreviated loanwords by providing evidence that simple and compound abbreviations should be strictly differentiated. The data establish that CALs are overwhelmingly quadrimoraic, whereas SALs display a more balanced repartition between two, three and four moras forms. They also reveal that heavy syllable ending patterns are non-productive in SALs. The core of the abbreviation process has been captured by positing the action of two well-formedness constraints, BIN (imposing prosodic binarity to the output) and *H# (avoidance of final heavy syllable), as well as an output-to-output identity constraint, demanding alignment between the right edge of the SAL and the left edge of the accented foot in the base in the case of apocope, or between the left edge of the SAL and the left edge of the accented foot of the base in case of aphaeresis. This alignment constraint, which is ranked lower than BIN and *H#, accounts for the length of the derived form (two, three or four moras).

This analysis offers generalizations which are both theoretically pertinent and surface-true, taking into account a large amount of original empirical and statistical data. It represents an advance inasmuch as it captures the prosodic organization of SALs in a simple and unified fashion, and succeeds in accounting for the fact that productive prosodic patterns observed in SALs vary from two to four moras, an aspect of Japanese SALs morpho-phonology which had been left unaddressed in former studies. How the analysis presented here can be extended to the treatment of compound

abbreviated loanwords will be a subject for future research.

Acknowledgements:

This research was partly supported by a Japan Foundation Grant. Previous versions were presented in August and November 2001 at Meikai University, Japan, at the Erss seminar (CNRS) in Toulouse and at the Ecole des Hautes Etudes en Sciences Sociales, Paris. Thank you to Kondô Takako, Hiraide Naoya and Takahashi Nobuo for their help on the data. I am also indebted to Marc Plénat, Shinohara Shigeko, Carole Paradis, Philip Carr, Kubozono Haruo and two anonymous reviewers for helpful comments and discussions on some preliminary oral or written versions of this paper. Special thanks also to Susan Branz, Philip Carr and Natalie Smith. All remaining errors are mine.

NOTES :

¹ Benua and others use the term ‘truncation’, but as we use this term in a more specific way - reserving it for describing a morphological shortening from the *end* of a base - we will use the word ‘abbreviation’ (A) instead. In this paper, ‘abbreviation’ refers to any morphological shortening (initial, final or medial) operating on a base (B).

² This choice implies that there will necessarily be, among the data presented below, some examples which some Japanese native speakers might find dubious, since they themselves do not use them. This can only be so if they do not belong to the socio-professional group where the word is used.

Another point to be noted is that out of their context, a fair number of words seem rare or strange, but pose no interpretation problem when replaced in the original context.

³ Such forms belong to *wasei eigo* (or *wasei gairaigo*), “English (or foreign) word coined in Japan”, and are best described as back formations, rather than clippings. Note that the material which has been deleted has morphemic status.

⁴ See also Tanomura (1999) for arguments on the special prosodic and lexical status of *-mento* as a suffix in Japanese loanwords.

⁵ The following dictionaries were used for collecting the corpus:

Konsaisu katekanago jiten, Tôkyô, Sanseidô, second edition, 2000.

Masukomi ni yoku deru tanshukugo ryakugo kaidoku jiten, Ishino Hiroshi (ed), 1992, Tôkyô, Sôtakusha.

Nichiei ryakugo - ryakushô jiten, Prem Motvani, Tôkyô, Maruzen, 1993.

Gendai yôgo no kiso chishiki 2000, Jiyû Kokuminsha, 2000.

⁶ The usual definition problems have been encountered with the notion of ‘word’. We relied on the typographic criteria, i.e. whether or not two morphemes are separated by a blank in the orthography, and on the lexical/semantic criteria. There are a few cases where a more detailed examination was necessary, especially when the first morpheme of the base is exactly two moras long and could possibly be analysed as a prefix rather than an autonomous word on semantic and lexical criteria. This problem arose typically with elements such as *mini* “mini”, *noo* “no”, *non* “non”, *semi* “semi”, *puro* “pro” and *hai* “high”, as they are either separated from the following word by a blank or a hyphen, either just agglutinated to it, and with transparent and productive prefixes such as *suupaa* “super”, *intaa* “inter”, *appaa* “upper”, *syuuru* “sur”, *inhura* “infra”, *sabu* “sub”, *an* “un-”, etc, even if no blank is inserted in the orthography. A closer analysis of the morpho-phonological characteristics of loanwords starting with these morphemes finally revealed that they should rather

be classified as CALs rather than SALs. The reason for this decision is that abbreviations like *minisuka* “mini skirt”, *nookon* “no control” or *sabukon* “subcontractor”: i) are mostly quadrimoraic, like CALs ; ii) can end with a heavy syllable, like CALs, and unlike SALs.

⁷ When there is minor segmental variation in the abbreviated form, such as *meriken* / *merikan* for “American”, *handi* / *hande* for “handicap”, the word has been entered as a single item. The present paper will not deal with such segmental variation.

⁸ An anonymous reviewer has pointed out that the scarcity of LLH and HH patterns in abbreviations might also come from the absence of bases beginning with these syllable structures. Moreover, the same reviewer observes that LLH and HH abbreviated outputs seem to be legal if they end in /N/, the moraic nasal. In order to verify this point, we proceeded to a comparison of truncation patterns yielded by LLH- and HH- starting bases containing /N/, /R/ (a long vowel), /Q/ (the first part of a geminate) and /J/ (the palatal mora) at the end of the second foot:

Truncation pattern of bases beginning with #LLH- and #HH- :

base= abb.=	#LLH-				Total
	H- = CV/R/	H- = CV/N/	H- = CV/Q/	H- = CV/J/	
a. #LL#	23	13	6	0	42
b. #LLL#	10	0	3	0	13
c. #LLH#	0	2	0	0	2
Total	33	15	9	0	57

base= abb.=	#HH-				Total
	H- = CV/R/	H- = CV/N/	H- = CV/Q/	H- = CV/J/	
a. #H#	0	1	0	0	1
b. #HL#	2	2	2	0	6
c. #HH#	0	0	0	0	0
Total	2	3	2	0	7

These tables provide two types of evidence. First, bases beginning with #LLH- and #HH- are not rare in the sub-corpus of truncated SALs : 57 bases start with #LLH-, while 7 of them start with #HH-, that is to say a total of 64 forms out of 277 (23.1%). Clearly, this means that the scarcity of LLH and HH abbreviations is not due to the absence of bases beginning with such patterns.

Second, the data above show that #LLH- and #HH- starting bases containing /N/ at the end of the second foot are not likely to yield /N/ ending outputs. They therefore satisfy the *H# constraint. Observe that #LLH- and #HH- bases containing /N/ in their second foot amount to 18 (15+3) forms out of 64 (28.1%).

These results confirm that /N/ ending SALs are only regularly found among CALs.

Incidentally, the reviewer cites *baaten* < *baatendaa* “bartender” as a possible exception to *H#. However, we consider that *baaten* is a CAL (compound) and not a SAL, relying on lexical and semantic criteria (see note 6 above).

⁹ () denote foot boundaries. Throughout this paper, we assume that syllables are exhaustively parsed into feet, and that degenerate feet, (L), are licit. This latter assumption diverges from what seems to be assumed by many scholars, as one reviewer points out.

¹⁰ A question by one of the reviewers concerns the relationship between *H# and final vowel shortening. Final vowel shortening is widely attested in Japanese, and not only in loanwords, see for example *honto* < *hontoo* “really”. In this paper, we consider that two distinct and separate phenomena, expressed as two different constraints, are involved. The first phenomenon (resulting from the satisfaction of *H#) is specific to the SAL derivation process. The latter one, which can be translated as *V:#, is not particularly associated to prosodically derived forms in the way *H# is. (For a different assumption about the restricted nature of long vowel final shortening, see also

Nishihara et al., 2001).

One argument in favor of this approach is the fact that variation between a long and a short final vowel occurs frequently in abbreviations, especially CALs (see *ruuso* or *ruusoo* < *ruuzu sokkusu* “loose socks”), whereas there is no such example of a variation involving /N/, /Q/, or /J/ in our corpus. Another reason for maintaining the *H# constraint distinct from an eventual *V:# constraint lies in the fact that *V:# alone cannot account for the loss of /Q/ (which is regularly deleted at the end of SALs, with generally no phonological compensation, as in *basuke* < *basuketto* /*basukeQto*/), in the way *H# does.

One could eventually dispense with the need for a final V shortening constraint, and, instead, introduce a higher-ranked melodic faithfulness constraint. The fact that syllables ending with /R/ or /Q/ are, among heavy syllables, more often affected by final shortening than /N/ and /J/ ending syllables could be explained by the fact that CVR# -> CV# and CVQ# -> CV imply no segmental loss, whereas CVN# -> CV# or CVJ# -> CV# do. Thus, deletion of a final /N/ or /J/ in non-derived words infringes some sort of melody-faithfulness constraint, whereas the shortening of /R/ or /Q/ does not.

A full examination of the implications of the *H# constraint in Japanese morphology and phonology would be beyond the scope of this paper, so we leave this question for further study.

¹¹ In such cases, one could expect variation to occur. However, variation is rare in abbreviated loanwords. In this respect, they differ radically from hypocoristic formations, an instance of subtractive morphology where variation seems to be the common rule.

¹² If the base is made up of two or more words, the accent taken into account is the lexical accent of the one word undergoing truncation: for example *paamanénto* (*wéebu*) > *paama*.

¹³ Following a well-established tradition, we assume that Japanese feet are trochaic.

¹⁴ The exceptions are: *nitoroguriserin* / *nitoroguriserin* (n.a.)-> *nitoro* “nitroglycerin” and *hokkesyuteríngu* -> *hokke* “Hockestellung (ski figure)”. However, as an anonymous reviewer has observed, these apparent exceptions just use the initial member of the compound. They might therefore be seen as instances of lexical ellipsis.

¹⁵ We are now referring to the “presence of an epenthetic vowel” or “presence of a *rV* mora” before the accented mora of the base, because this allows a more factual approach. This is not strictly equivalent to saying that “the expected (but incorrect) output would have ended with an epenthetic vowel” or to saying that “the expected (but incorrect) output would have ended with a *rV* mora”, as formulated in i) and ii) above. The former formulation concerns the input level, the latter one the output level. Only a few examples exist which possess an epenthetic vowel or a *rV* mora before the accented mora, without producing an output ending with such elements. This happens with longer bases (i.e. *sutorobosukóopu* “stroboscope” yielding regular *sutorobo*) and with exceptions incurring a violation of BIN or *H# (i.e. *hurankuhúruto* “Frankfurt” yielding *huranku* rather than the expected **hura*).

¹⁶ There is only one case of a final heavy syllable being reduced to a light syllable: *puropéeraa* which yields *pera* (but *puropera* n.a. is also attested).

¹⁷ It is interesting to note that Plénat (1999), in his study of French reduplicated hypocoristics (most of which happen to be abbreviated forms), observes similar tendencies, i.e. a predisposition for certain classes of segments such as stops, and especially nasal stops, to be promoted as the initial consonant.

REFERENCES:

- Akinaga, Kazue (1981) *Meikai Nihongo Akusento Jiten*. Tokyo : Sanseidô.
- Benua, Laura (1995) "Identity Effects in Morphological Truncation, " in J.N. Beckman, L. Walsh Dickey, S. Urbanczyk (eds.) *Papers in Optimality Theory*. (pp. 77-136). Amherst : University of Massachussets.
- Ishino, Hiroshi (1983) *Gendai gairaigo kô*. Tokyo: Taishûkan.
- Itô, Junko (1990) "Prosodic minimality in Japanese," in M. Ziolkowski, M. Noske, K. Deaton (eds.) *The syllable in Phonetics and Phonology* CLS 26, vol 2. (pp. 213-239). Chicago :Chicago Linguistic Society.
- Itô, Junko and Armin Mester (1986) "The Phonology of Voicing in Japanese," LI 17, 49-73.
- Itô, Junko and Armin Mester (1992) "Weak Layering and Word Binarity," Report n°-92-09, Linguistic Research Center, University of California Santa Cruz.
- Itô, Junko and Armin Mester (1995) "Binarity," Paper presented at the GLOW phonology workshop, Tromso.
- Itô, Junko, Kitagawa Yoshihisa and Armin Mester (1996) "Prosodic Faithfulness and Correspondence: Evidence form a Japanese Argot," *Journal of East Asian Linguistics* 5, 217-294.
- Kanno Ken (1985) "Yôgo no ryakugokei," *Nihongogaku* 4-9, 54-64.
- Kubozono Haruo (1996) "Syllable and accent in Japanese : evidence from loanword accentuation," *Journal of the Phonetic Society of Japan* 211, 71-82.
- Kubozono, Haruo (2001) "Epenthetic vowels and accent in Japanese : facts and paradoxes , " in J. van de Weijer and T. Nishihara (eds.) *Issues in Japanese Phonology and Morphology*. (pp. 111-140). Berlin and New York : Mouton de Gruyter.
- Kubozono, Haruo (in press) "The syllable as a unit of prosodic organization in Japanese, " in C. Fery C. and R. van der Vijver (eds.) *The syllable in Optimality Theory*. Cambridge: Cambridge University Press.
- Kubozono, Haruo, Junko Itô & Armin Mester (1997) "Nonfinality in Japanese Phonology," in *Proceedings of the 16th International Congress of Linguists* (CDRom edition). Oxford: Pergamon Press.
- Labrune, Laurence (1997) "Inertie phonologique, absence de marque et sous-spécification : la consonne *r* en japonais" [phonological inertia, unmarkedness and underspecification : *r* in Japanese], in G. Deléchelle and M. Fryd (eds.) *Travaux Linguistiques du Cerlico* 13. (pp. 245-267). Rennes : PUL.
- McCarthy, John and Alan Prince (1995) "Faithfulness and Reduplicative Identity," in J.N. Beckman, L. Walsh Dickey, S. Urbanczyk (eds.) *Papers in Optimality Theory*. (pp. 249-384). Amherst : University of Massachussets.
- McCawley, James (1968) *The Phonological Component of a Grammar of Japanese*. The Hague / Paris: Mouton.
- Morioka, Kenji (1988) "Ryakugo no joken," *Nihongogaku* 7-10, 4-21.
- Nishihara, Tetsuo, Jeroen van de Weijer, Kensuke Nanjo (2001) "Against Headedness in

Compound Truncation: English Compounds in Japanese,” in J. van de Weijer and T. Nishihara (eds.) *Issues in Japanese Phonology and Morphology*. (pp. 299-324). Berlin and New York : Mouton de Gruyter.

Plénat, Marc (1999) “Prolégomènes à une étude variationniste des hypocoristiques à redoublement en français” [prolegomena to a variationnist study of reduplicated French hypocoristics], in J. Durand and C. Lyche (eds.) *Cahiers de Grammaire 24 : Phonologie Théorie et Variation*. (pp. 183-219). Toulouse : Université Toulouse – Le Mirail / Erss.

Sibata, Takeshi (1994) “Gairaigo ni okeru akusentokaku no ichi,” in K. Satô (ed) *Kokugo Ronkyû 4, Gendaigo – Hôgen no Kenkyû*. (pp. 1/418-29/390). Tôkyô : Meiji Shoin.

Shinohara, Shigeiko (2000) “Default accentuation and foot structure in Japanese : evidence from Japanese adaptations of French words,” *Journal of East Asian Linguistics* 9, 55-96.

Tanomura, Tadaharu (1999) “Gairaigo akusento ni okeru gengo no hatsuon no kan’yo ni tsuite – 4 môra ika no go wo chûshin ni,” *Nihongo Kagaku* 5, 67-88.