Community- and individual-level variation in Japanese compound loanword formation

REMI LAMARQUE & JAMES SNEED GERMAN
Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France[*]

1. Introduction

An important issue for the theory of language change concerns the emergence of regularity from variation. Developments in the analysis of community-level sociological processes have highlighted the importance of considering networks of language users to address this issue, though the relationship between individual behavior and community-level behaviors remains poorly understood. This study explores how those two levels of grammar can behave differently in the evolution of new morphological processes from variation to regularity.

This work takes as an example a specific morphological process that occurs in the part of the Japanese lexicon that consists of loanwords. In Japanese, studying the lexicon constituted by recent loanwords can provide interesting keys to understanding the phonology of that language. Indeed, the phonological particularities that are observable in loanwords represent violations to the language's phonological constraints that would otherwise apply to Japanese words. The study of loanwords thus allows us to better understand the extremes of variation that are permissible in order to be considered part of the Japanese lexicon. In other words, loanwords give us an overview of the Japanese lexicon's boundaries as well as the phonological constraints that it must follow. It is not surprising then that Japanese loanwords have been the focus of many works, especially in an Optimality Theoretical framework (Itô & Mester 1995, 2008; Kubozono, Ito & Mester 2008; Lamarque 2013, 2015a; Lovins 1973; Ohso 1971; Shinohara 1996, 1997).

The morphological process that is the focus of this study involves the creation of an abbreviation out of two words of foreign origin. A well-known example is the word *pokemon*, which originates from the two English words 'pocket' and 'monster', pronounced *poketto* and *monsutaa* when adapted to Japanese phonology (Shinohara 1997). This abbreviation process is very productive in modern Japanese and gives rise to so called Compound Abbreviated

Loanwords (hereafter 'CALs', Labrune 2006). CALs are derived from a compound base typically consisting of two constituents of a foreign origin. The final derived CAL always includes elements from both constituents, and its formation is subject to prosodic constraints. While the ordering of these constraints was initially variable in Japanese, it has stabilized in recent decades, such that more recent attested forms tend to conform to a single grammar (Lamarque 2015b). Here, we explore how this stabilization process relates to community-level processes by examining the results of two studies using different methodological approaches: the first involves an analysis of lexicons of attested CALs used by the community of Japanese speakers, while the second explores the abbreviation habits of individuals in the creation of new CALs using an experimental approach.

1.1 Outline of the paper

In section 2, we will review some of the previous studies on CALs. The first part of this section focuses on Labrune's contribution to the study of CALs through the creation and analysis of a large lexicon of abbreviations of Japanese loanwords (Labrune 2002, 2006, 2007, 2008). The second part of the section presents a previous study that builds on Labrune's research to explore the evolution of CALs by comparing lexicons from different time periods (Lamarque 2015b). By examining attested forms collected in different lexicons, this section focuses on the regularization of CALs at a community level. This section also provides an Optimality Theoretic (Prince & Smolensky 1993) analysis of the CAL derivation process.

Section 3 reports on the methodology and findings of a new study. The main objective of this study is to directly test several observations from the lexicon-based studies using a more experimental approach. The experiment takes the form of an online survey in which Japanese speakers are asked to create new abbreviations. The results suggest that in an isolated situation, individuals do not show the regular abbreviation patterns that are observed at a community level. In the discussion, we provide an analysis of the results of the survey and attempt to explain their apparent contradiction with data found in the lexicons. The last part of this section provides an analysis of CALs based on stochastic Optimality Theory (Anttila 1997; Boersma & Hayes 2001; inter alia). This framework turns out to be better suited to modelling highly variable processes such as those observed in the derivation of CALs at an individual level.

2. Previous studies on CALs

In order to understand how CALs are formed, it is crucial to understand the role of the mora as a prosodic unit in Japanese. A mora is a minor prosodic unit whose size ranges between a phoneme and an entire syllable. Morae are used to measure the weight of syllables (Davis 2011). A light syllable (i.e. a syllable without a coda) weighs one mora; whereas a heavy syllable (i.e. a syllable with a coda) weighs two morae. This is illustrated in figure 1 below, where syllables are represented by 'σ' and morae are represented by 'μ' (from Labrune 2005).

Figure 1: Syllable structure for a light syllable (/ta/) and two heavy syllables (/ta:/ and /tap/)
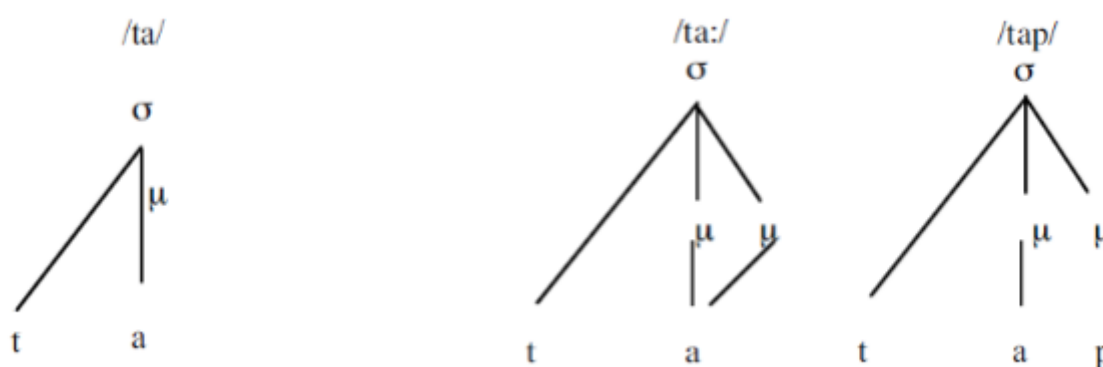
The presence of a coda is what determines the weight of a syllable. In Japanese, only three types of segments can appear in coda position, and these are all underspecified in some way:

- /R/: represents a vowel coda and corresponds to the lengthening of the preceding vowel (for example: /ta.**R**/ is pronounced [ta.**a**])

- /Q/: represents a consonant coda and corresponds to the first part of a geminate consonant (for example: /ta.**Q**.ta/ is pronounced [ta.**t**.ta])

- /N/: represents a homorganic nasal consonant coda, which assimilates to the place of articulation of the following consonant (for example: /ta.**N**.ku/ is pronounced [ta.**ŋ**.ku])

In terms of prosodic typology, Japanese is classified as a mora-timed language (Ramus 1999), which means that the mora is its fundamental rhythmic unit (Labrune 2005). This implies that in Japanese, morae play an important role in a wide range of phonological processes. Of course, this is also the case for the formation of CALs.

2.1 The general derivation principle for CALs

The formation process for CALs follows a general derivation principle which superficially appears to be quite simple. Labrune (2007) describes it as follows: 'the two first morae of the two constituents of the base (hereafter 'C1' and 'C2') are maintained'.[1] The derivation process for the word *pokemon* is illustrated in Figure 2 below.

Figure 2: Derivation process for the word *pokemon*
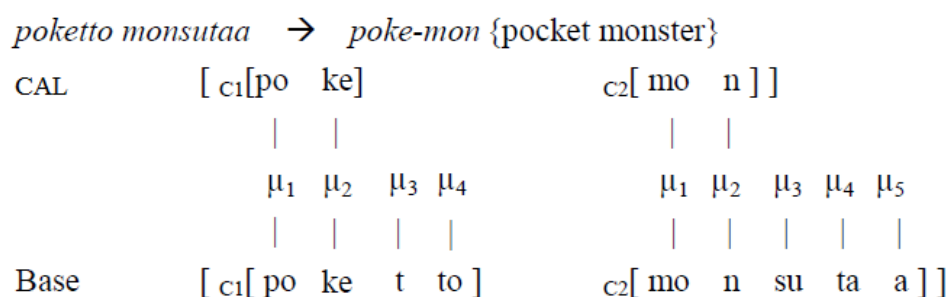
In this example, the two first morae of ***po.ke**.t.to* and ***mo.n**.su.ta.a* are maintained to form the four-mora long abbreviation ***po.ke.mo.n*** However, while this general derivation principle represents the most typical way of creating CALs, there are a number of abbreviations that deviate from it in some way. This suggests that the CAL formation process is rather complex, and indeed, a statistical analysis of the phonology of CALs sheds light on the underlying regularity of a number of apparently exceptional abbreviations.

This was in fact the aim of Labrune's work on Japanese loanword abbreviations (Labrune 2002, 2006, 2007, 2008). By collecting and analyzing a lexicon of 1314 abbreviations, she was able, firstly, to make a distinction between Simple Abbreviated Loanwords (hereafter SALs) and CALs. While the formation of CALs, as already mentioned, relies on the conservation of the initial morae of several constituents, SALs' formation relies on the accent pattern of a single constituent. Since those two abbreviation types follow

---

[1] Translated from French (Labrune 2007:3).

entirely different rules, they deserve to be considered and studied as two different morphological processes.

A large proportion of the literature on Japanese loanwords has been grounded in an Optimality Theoretic approach (Prince & Smolensky 1993), which has proven to be an appropriate framework for this subject (Itô & Mester 1995, 2008; Shinohara 1997). In her work, Labrune also provided an analysis in Correspondence Theory (Benua 1995, 1997; McCarthy & Prince 1995). While this theoretical framework grew out of Optimality Theory, it proved to be better suited for the description of linguistic processes involving the reduplication of some already existing output (in our case: an abbreviation process). Whereas for standard OT studies, faithfulness constraints are involved in an input-output relation, Correspondence Theory focuses on identity constraints involved in an output-output relation, the first output being the base and the second output its abbreviation. This is illustrated in Figure 3 below (Benua 1995):

Figure 3: Differences between input-output and output-output relations

Base-Abbreviation - Identity

Base ⬅====➡ Abbreviation

Input-Output
Faithfulness ⬍

Input
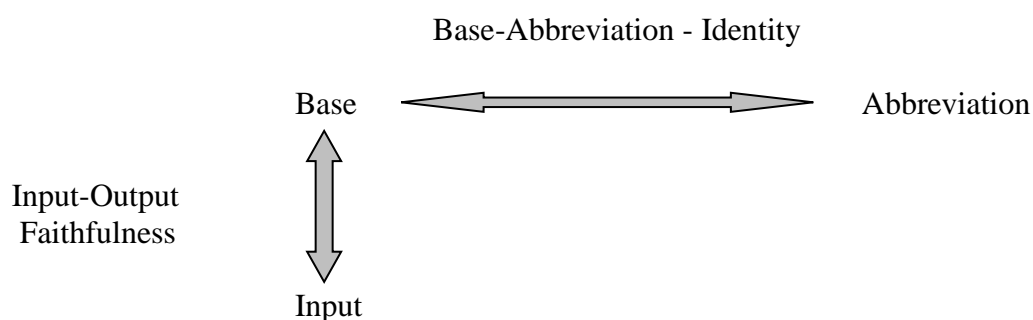
Building on the general derivation principle cited above, Labrune (2007) identifies the following identity constraints that play a role in the formation of CALs:

- BINARYFEET: all feet must be binary in morae.
- BINARYPROSODICWORD: the prosodic word (i.e. the abbreviation) must be binary in feet.
- MAXLEX: every lexeme of the base must be present in the abbreviation.
- ANCHORLEFT: every foot of the abbreviation must be anchored to the left border of the base lexeme it comes from.
- CONTIG: the maintained morae in each foot must be contiguous in the base.
- DEP: all segments in the abbreviation must be present in the base (no epenthesis).

- IDENTFEATURES: distinctive features are identical between corresponding segments in the base and the abbreviation (no modification of distinctive feature).

A statistical analysis of Labrune's lexicon suggests the following ranking for these constraints:

BINARYPROSODICWORD >> CONTIG, ANCHORLEFT >> BINARYFEET, MAXLEX, DEP, IDENTFEATURES

Labrune (2007:5) also suggests that in Japanese, all morae must be parsed in a foot. This in enforced by an inviolable constraint PARSEµ.

BINARYPROSODICWORD enforces that all CALs are binary in feet, thus, one foot for the first constituent (C1) and another one for the second constituent (C2). Even if some CALs contain one or three-mora long constituents, there is however no four-mora long constituent in any CAL. Since violation of BINARYPROSODICWORD apparently never occurs, it seems logical that this constraint appears at the top of the hierarchy.

The fact that the two initial morae of each constituent are maintained in a CAL can be explained by the action of three constraints: CONTIG, ANCHORLEFT and BINARYFEET. CONTIG requires that the maintained morae are contiguous in the base, ANCHORLEFT that they are initial, and BINARYFEET that they are exactly two in number.

The three remaining constraints (i.e MAXLEX, DEP and IDENTFEATURES) provide some conformity between the base and its abbreviation. MAXLEX enforces that the abbreviation contains at least one element from the two constituents of the base, DEP prevents any element that is not present in the base to be added to the abbreviation, and IDENTFEATURES excludes eventual modifications that segments could undergo during the abbreviation process.

The tableau 1 below shows the derivation process for a canonical CAL [meta-kiN], from 'metal king' in English, pronounced [metaɾu-kiNgu]. This illustrates which types of output candidates are ruled out by which constraint.

Tableau 1: [metaɾu-kiNgu] => [meta-kiN] 'metal king'

	[metaɾu-kiNgu]	PARSEμ	BINPW	CONTIG	ANCRL	BINFT	MAXLEX	DEP	IDENTF
1.	(meta)		*!				*		
2.	(me-ki)		*!	*					
3.	(me)-(ki)					**!			
4.	(meta)-(ki)					*!			
5.	(me)-(kiN)					*!			
6. =>	(meta)-(kiN)								
7.	(meta)-(kiN)-(gu)		*!			*			
8.	(meta)-(kiN)-gu	*!							
9.	(meɾu)-(kiN)			*!					
10.	(taɾu)-(kiN)				*!				
11.	(meR)-(kiN)							*!	
12.	(meta)-(giN)								*!

2.2 Exceptions to the general derivation principle

The derivation principles for CALs described above represent only general tendencies, and there is a wide range of exceptions, which involve infractions to one or more of constraints mentioned above. Labrune (2007) notes certain phonological contexts that favor the creation of exceptional CALs. In a majority of cases, for example, the creation of an exceptional CAL is due to the presence of either /R/ (i.e. vowel lengthening) or /Q/ (i.e. first part of a geminate consonant) as the second mora of C2. For example, the name of the famous actor Brad Pitt, pronounced *bu.ra.d.do pi.t.to*, in which the second mora of C2 is the first part of a geminate consonant (*.t.*), is abbreviated as *bu.ra.pi*.

It appears then that the presence of one of those special segments in the final position of a CAL is somehow problematic for the Japanese phonological grammar. This is understandable if we consider, on the one hand, that /Q/ needs a consonant at its right which it can geminate in order to be pronounceable, and on the other hand, that /R/ is generally disfavored in the final position of Japanese words.

In those specific phonological contexts, there are four possible treatments for the special segments /R/ and /Q/:

#1: Maintained

The special segment is maintained as any other mora. CALs thus created are not exceptions to the general derivation principle but regular abbreviations. Note that it is impossible for /Q/ to be maintained, since there is no possible pronunciation for /Q/ if there is no consonant at its right.

Ex.1: In English: Mario   kart
       In Japanese: マリオ[**ma.ri**.o]   カート[**ka.R**.to]
       CAL: マリカー[**ma.ri.ka.a**]

#2: Followed (by another mora)

The special segment is maintained, but it is followed by another mora (usually, the one that follows in the base constituent). This prevents /R/ or /Q/ to be in final position but creates a three-mora long C2 in violation of BINARYFEET.

Ex.2: In English: plastic   case
       In Japanese: プラスチック[**pu.ra**.su.ti.Q.ku]   ケース[**ke.R.su**]
       CAL: プラケース[**pu.ra.ke.R.su**]

Ex.3: In English: magazine   book
       In Japanese: マガジン[ma.ga.zi.N]   ブック[**bu.Q.ku**]
       CAL: マガブック[**ma.ga.bu.Q.ku**]

#3: Replaced

The special segment is replaced by another mora (usually, the one that follows in the base constituent). /R/ and /Q/ do not appear in the abbreviation, but this violates CONTIG.

Ex.4: In English: king   carnival
       In Japanese: キング[**ki.N**.gu]   カーニバル[**ka.R.ni**.ba.ru]
       CAL: キンカニ[**ki.N.ka.ni**]

Ex.5: In English: american   football

In Japanese: アメリカン[**a.me**.ɾi.ka.N]   フットボール[**ɸu**.Q.**to**.bo.R.ɾu]

CAL: アメフト[**a.me.ɸu.to**]

#4: Deleted

The special segment is simply deleted. This creates a one mora C2, which is also a violation to BINARYFEET.

Ex.6: In English: mister   donuts

In Japanese: ミスター[**mi.su**.ta.R]   ドーナツ [**do**.R.na.t͡su]

CAL: ミスド[**mi.su.do**]

Ex.7: In English: potato   chips

In Japanese: ポテート[**po.te**.R.to]   チップス[**t͡ɕi**.Q.pu.su]

CAL: ポテチ[**po.te.t͡ɕi**]

Labrune's studies reveal many important facts about the main constraints that play a role in the derivation of CALs as well as the phonological contexts that favor the creation of exceptions. However, CAL creation as a productive morphophonological process entered into the Japanese grammar only relatively recently (post World War II). Studies based on a chronologically fixed lexicon, such as Labrune's, are not able to capture changes which may be taking place as this relatively new process diffuses through the community. Building on Labrune's research, the next section explores how CAL derivation has evolved over time, using as a first method the comparison of two lexicons from different time periods.

2.3 Comparison of corpora

The fourth possible treatment presented above, Deletion, is particularly interesting in that it results in a three-mora long CAL. Indeed, it seems that this kind of exception, even in the cases when it is not due to the presence of /R/ or /Q/ as second mora of C2, is frequent enough to be noticed by naive native speakers of Japanese. A few speakers we interviewed suggested that the tendency for three-mora CALs is a relatively recent trend. In other words, recent

abbreviations would tend to be cut to three morae more often than older CALs. In a previous study (Lamarque 2015b) we tested the hypothesis that three-mora CALs are more frequent among recent abbreviations. Particular attention was paid to CALs involving Deletion of /R/ and /Q/ when these appear as the second mora of C2.

In order to compare CALs from different time periods, we created a second lexicon of more recent CALs. This new lexicon (hereafter the 'video game lexicon') contained 300 CALs found on websites concerning video games. Considering that video games are a relatively recent media mostly used by younger individuals, we thus intended to observe the newest tendencies regarding the formation of CALs. The video game lexicon was then compared to the sub-lexicon of 711 CALs (hereafter 'Labrune's lexicon') collected by Labrune. Note that Labrune's lexicon was collected around 2005 and contains CALs from diverse sources (dictionaries, newspapers, conversations, etc.). For the purposes of this study, we therefore take it as representative of 'older' tendencies in the formation of CALs. To assure that the comparison between the two lexicons was meaningful, we created the video game lexicon using the same criteria as Labrune:

- A single iteration of a CAL was sufficient for inclusion
- Only one entry for each CAL, even when attested several times.
- If two CALs were created from the same base, then both CALs were included.
- Certain types of abbreviations similar to CALs were excluded: acronyms (ex: *OL* (*oo-eeru*) < *ofisu redii* 'office lady'); abbreviations mixing a Japanese word with a foreign one (ex: *kara-oke* < *kara* 'empty' (Japanese) *ookesutora* 'orchestra' (English)); abbreviations involving the deletion of a whole lexical word (ex: *amerikan* < *amerikan koohii* 'American coffee'); etc.).

If we first compare the proportion of CALs in each lexicon by length in morae, it appears that the video game lexicon indeed contains a larger proportion of three-mora abbreviations. This is illustrated in Figure 4 below:

Figure 4: Proportion of CALs by number of morae in both lexicons
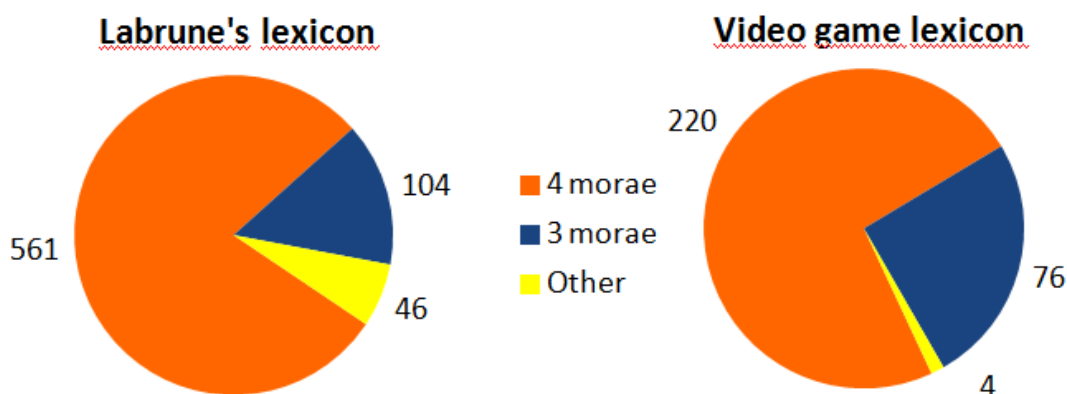
Overall, these results support the hypothesis that three-mora CALs are more frequent in recent abbreviations. A closer look reveals possible phonological motivations for this increased proportion of irregular CALs. As mentioned above, one of the major triggers for irregular CALs is the presence of /R/ and /Q/ as the second mora of C2. Figure 5 therefore shows the proportion of the four treatments of /R/ and /Q/ in each lexicon:

Figure 5: Treatment of /R/ and /Q/ as second mora of C2 in each lexicon

		Labrune's lexicon	Video game lexicon
/R/	Maintained	**38%**	11%
	Followed	10%	0%
	Replaced	7%	8%
	Deleted	**45%**	**81%**

/Q/	Maintained	0%	0%
	Followed	**40%**	0%
	Replaced	15%	14%
	Deleted	**45%**	**86%**

In the older lexicon, it is not possible to identify one treatment that is preferred over the others. It appears that at least two treatments are in competition both for /R/ and /Q/. In the case of /R/, it is Maintained almost as often as it is Deleted; and in the case of /Q/, it is Followed almost as often as Deleted. However, the video game lexicon does not show such a balanced competition. Indeed, the Deletion is the preferred solution in every case, with a proportion of more than 80%. This shows that the deletion of the special segments /R/ and /Q/

when they appear as second mora of C2 has become systematic with time. These results therefore suggest an evolution of the treatment of /R/ and /Q/ with time.

This new regularity should be included in the analysis of the CAL derivation process. From an OT point of view, this requires the inclusion or modification of certain constraints. The new analysis we present here builds on Labrune's (2007) though it is slightly different regarding the inclusion of the treatment of /R/ and /Q/.[2] First, it is necessary to include two constraints that express the dispreference for /R/ and /Q/ in final position. Such constraints are violated whenever /R/ or /Q/ are Maintained:

- *V:#: no long vowel in final position.
- LICENSE/Q/: the special segment /Q/ must be correctly licensed by a consonant at its right.

It is then necessary to distinguish between the Following and Deletion treatments, which are both in violation to the same constraint, *BINARYFEET, though in different ways. Without such a distinction, it would be impossible, for example, for the grammar to differentiate between the two candidates in Tableau 2 below:

Tableau 2:

	[misutaa doRnat͡su]	*BINARYFEET
1.?	(mi.su)(do.R.na)	*
2.?	(mi.su)(do)	*

To address this problem, we proposed in Lamarque (2016) that *BINARYFEET be divided into two different constraints. The first one disfavors the Following treatment whereas the second one disfavors the Deletion treatment:

- *FEET>2: feet must not be longer than two morae.
- *FEET<2: feet must not be shorter than two morae.

---

[2] Another difference is that we exclude the MAXLEX constraint from our analysis. We consider CALs to be abbreviations that maintain exactly two lexemes out of the base constituents. On the same basis that one-lexeme abbreviations (Simple Abbreviate Loanwords) follow a different derivation process, we consider that three or more lexeme abbreviations have to follow slightly different rules that take into consideration those extra lexemes. In other words, an abbreviation considered to be a CAL in violation of MAXLEX in Labrune's analysis is not considered a CAL at all in ours. In any case, such abbreviations are very rare and would not have a significant influence on the observations presented here.

Tableaux 3 to 6 below illustrate how the constraints involved in the treatment of /R/ and /Q/ should be ranked, according to the data from the video game lexicon.

Tableau 3:

[misutaa doRnats͡u]	*V:#	*FEET<2
*(mi.su)(do.R)	*	
(mi.su)(do)		*

Tableau 4:

|poteRto t͡ɕiQpusu]	*LICENSE/Q/	*FEET<2
*(po.te)(t͡ɕi.Q)	*	
(po.te)(t͡ɕi)		*

Tableau 5:

[misutaa doRnat͡su]	CONTIG	*FEET<2
*(mi.su)(do.na)	*	
(mi.su)(do)		*

Tableau 6:

[misutaa doRnats͡u]	*FEET>2	*FEET<2
*(mi.su)(do.R.na)	*	
(mi.su)(do)		*

Note that in each case, the lower candidate is the one that is selected by the grammar. This means that the constraint *FEET<2 is dominated by all the other constraints. Regarding the ranking of the other constraints, since LICENSE/Q/ and *FEET>2 are never violated in the video game corpus, they must be placed above CONTIG and *V:#. The slight preference for the Maintaining of /R/ over the Replacement treatment indicates that *V:# is ranked higher than CONTIG. The hierarchy of all constraints is thus as follows[3]:

---

[3] This is a simplified version that does not take into consideration, for example, the treatment of /R/ and /Q/ when they appear in final position of C1, which is also a source of irregularities in CALs. For a more detailed description, see Lamarque (2015b and 2016).

BINARYPROSODICWORD >> LICENSE/Q/, *FEET>2 >> *V:# >> CONTIG, ANCHORLEFT >> DEP, IDENTFEATURES >> *FEET<2

This new hierarchy illustrates how some irregularities in CALs could actually be included in the analysis of the derivation process. The general derivation principle mentioned above was in fact too simple to encompass rules conditioned by specific phonological contexts such as the presence of /R/ or /Q/ as second mora of C2. The treatment of those special segments seems to have evolved towards a single optimal option: Deletion. Among the 76 tri-moraic CALs in the video game lexicon (see figure 4), 60 are due to the deletion of either /R/ or /Q/ in final position. Therefore, the recent regularization of the treatment of those special segments seems to be linked to the overall increase in the number of three-mora long CALs that native speakers perceived. The comparison of two lexicons of different time periods allowed not only confirming this hypothesis but also identifying phonological explanations for it.

## 3. Influence of age on abbreviation preferences

Through the analysis of lexicons of attested abbreviations, the results of the previous studies on CALs provided important insights regarding the phonological constraints governing their formation. Attested forms from the existing lexicons reflect well-formedness for CALs with respect to what is accepted and used by the community of speakers as a whole. Those studies also provided a general picture of the evolution of the derivation process over time. However, the associated attested forms do not necessarily reflect the grammatical preferences of individual language users based on their everyday experience with their language. One major issue with this methodology is that each entry in the lexicons has the same weight regardless of how frequently it occurred. In other words, a CAL that has been uttered once, in a very specific situation is represented with the same importance as a CAL that is used often and in a wide variety of situations. This makes it difficult to draw clear conclusions regarding the actual use of CALs, as well as how usage relates to the internal states of individuals. In order to obtain broad-based corroboration for the observations from the lexicon studies, we therefore conducted a new study using an experimental approach that explores the behavioral tendencies of individuals.

3.1 Methodology

The main hypothesis for the present study follows from the prediction that if the grammar governing the derivation process for CALs has recently undergone an evolutionary change, then there should be an observable change in the abbreviation preferences of individual speakers. We further assume that such a change-in-progress would manifest as a difference in behavior across speakers of different ages.

In order to test this hypothesis, we conducted an experiment in the form of an online survey in which Japanese speakers were asked to create new CALs from specific base constituents. This survey was created using SurveyMonkey, an online survey development cloud-based software. Participation was solicited by advertising through e-mails and social networks for about a month (from 03/08/2016 to 10/04/2016) and 118 participants answered over that period. Figure 6 below shows the distribution of participants by age.

Figure 6: Repartition of participants to the survey in age groups
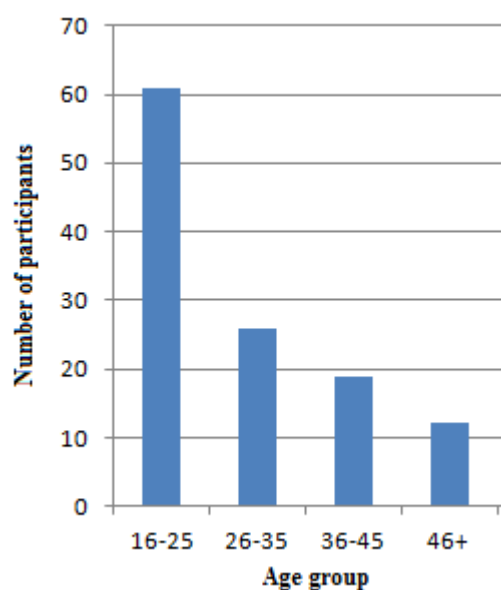
The task consisted in creating new CALs from 100 pairs of English words. Both the English and Japanese *katakana* writing forms were visible to the participants. They were asked to enter their answer in an empty text box before proceeding to the next pair of words. However, they could not go back to modify a previous answer. Figure 7 below shows the example of the pair 'master' + 'cool'.

Figure 7: Example of a pair of words to abbreviate

Most of the loanwords that were selected for this study were taken from the two lexicons used in previous studies, so that each constituent was already attested in a CAL. However, for some specific phonological contexts, the lexicons did not contain enough different items, so loanwords from different sources (online specialized dictionaries) had to be used. When extracted from a lexicon, loanwords were paired with a different word than the one they appeared with, so that participants were creating new abbreviations and were not simply using ones they already knew. Since the focus of the experiment was the treatment of /R/ and /Q/ in final position of C2, particular attention was paid to minimize the possible influence of C1 on the outcome. All C1s were selected to be as neutral as possible (i.e. containing no phonological characteristic that favors the creation of irregular CALs) and two versions of the survey were created with the same C1s associated with different C2s. Participants answered one of the two different versions at random.

3.2 Results

Figure 8 below presents the results for all items having /R/ as second mora of C2, while figures 9 and 10 show the results for /Q/. It appeared that the number of morae of the base constituent of C2 had a significant influence on the treatment of /Q/, so the data is presented separately according to whether C2 is tri-moraic (figure 10) or non tri-moraic (figure 9).
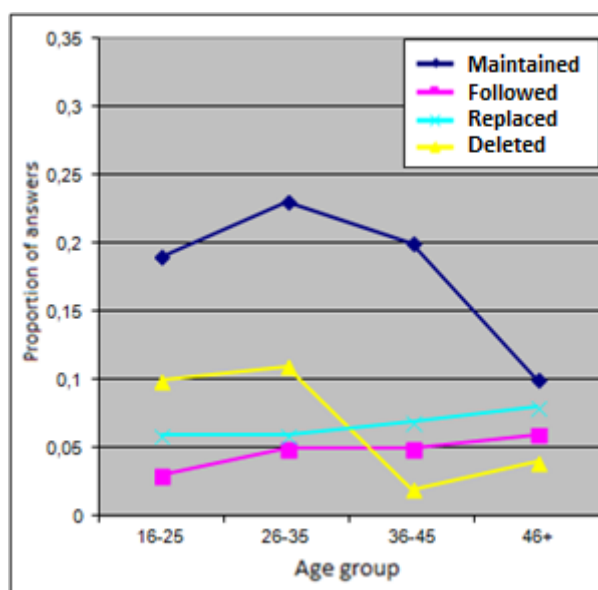
Figure 8: Treatment of /R/ as second mora of C2

Figure 9 : Treatment of /Q/ as second mora of a non tri-moraic C2[4]
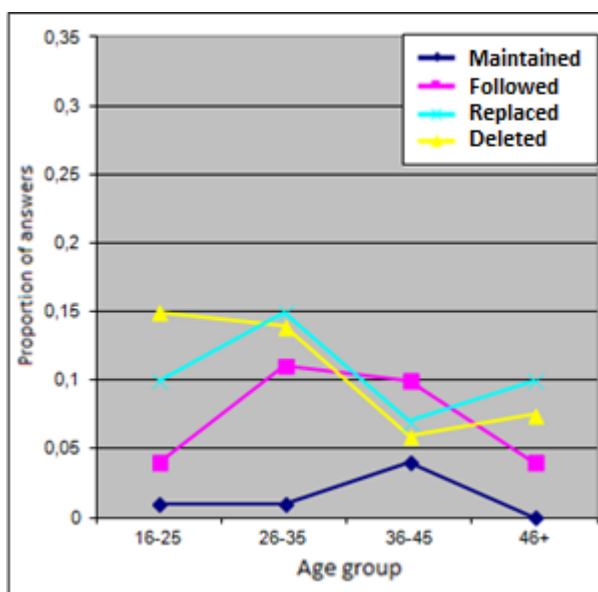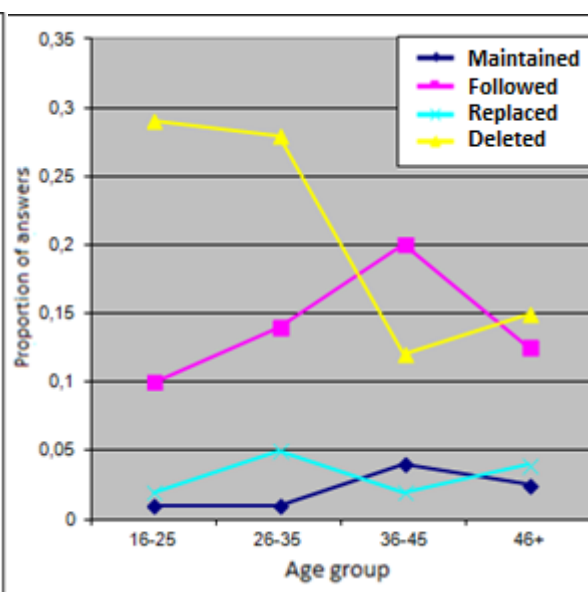

Figure 10 : Treatment of /Q/ as second mora of a tri-moraic C2

Note, first of all, that the percentage of responses corresponding to one of the four usual treatments for /R/ and /Q/ does not sum to 100%. This is due to the presence of responses involving either a CAL with very irregular formation or some other abbreviation type that could not be considered a true CAL. The first kind includes, for example, not maintaining the initial mora of C2 but the second and third morae instead (in violation of

---

[4] The results present a few cases of maintaining of /Q/ which is supposedly impossible. This corresponds to cases where /Q/ was changed to /t͡su/ because of the relation of those two morae in the Japanese writing system. Indeed, <Q>ッ is spelled with a smaller version of the character for <t͡su>ツ. Since /t͡su/ does not appear in the base, this is a violation of DEP. As those cases were very rare, they will not be discussed further here.

ANCHORLEFT), as well as violations of other constraints with no apparent phonological motivation. The second kind includes abbreviations such as acronyms or portmanteaux. Overall, the presence of such unexpected forms fits with a more general observation: the results of the survey show a high degree of intra-speaker variation. Indeed, regardless of the age of speakers, all possible treatments were used and in most cases there is not a single treatment that clearly dominates the others. Moreover, the participants explored in their responses some unexpected possibilities for deriving CALs.

Figure 8 shows that Maintaining is the preferred treatment of /R/ as the second mora of C2 regardless of the age of the speaker. However, this is not as clear in the case of /Q/. Indeed, it is clear from figure 9 that there is not one treatment that dominates the others, and figure 10 shows that the Deletion and Following treatments are preferred in the case of a tri-moraic C2, though age does seem to have an influence on which of the two is more frequent.

In all three figures, it appears that Deletion of /R/ and /Q/ is more frequently used by younger people (yellow line). The statistical analyses were run using a generalized linear mixed model on R with the lme4 package (Bates et al. 2014). In every case, the fixed factor is age and the dependent variable is the kind of treatment chosen. Even though age is divided in only four groups in the figures, for visual representation, it was considered as a continuum in the statistical analysis. This analysis revealed that the effect of age on proportion of Deletion is significant ($p < 0.001$ for figure 8; $p < 0.01$ for figure 9; and $p < 0.001$ for figure 10). The increase in the proportion of Following treatment with age in figure 10 is also significant ($p < 0.01$). This is consistent with the general pattern of the plots which suggests that, with age, Deletion decreases in frequency while Following increases.

3.3 Discussion

The results presented above partly corroborate the findings of the lexicon-based studies. In spite of some surprising aspects, for example, the survey results overall support the hypothesis that age has a significant influence on the abbreviation habits of Japanese speakers. Moreover, regarding the treatment of /R/ and /Q/, young people actually use Deletion more frequently than older people, which is also consistent with the results of previous lexicon-based studies. Additionally, as with the lexicon-based studies, the size of C2 had an influence on the treatment of /Q/. Looking back at figure 5, it can be noted that /Q/ was Followed in 40% of the cases in Labrune's lexicon while this treatment never occurred once in the video game

lexicon. In fact, the cases of /Q/ being Followed in Labrune's lexicon always occur when /Q/ is the second mora of a tri-moraic constituent. This kind of constituent usually arises from the adaptation of English words that consist of just one heavy syllable. For example, 'bed', bag' or 'pit' become respectively /be.Q.do/, /ba.Q.gu/ and /pi.Q.to/ in Japanese. In Labrune's lexicon, this kind of constituent would be kept as a whole in the abbreviation, whereas in the video game lexicon, there is not a single occurrence of such a treatment. This phenomenon is reflected in the results of the survey shown in figure 10, whereby young speakers have a strong preference for the Deletion of /Q/ in a tri-moraic constituent whereas older people tend to prefer the Following treatment. The crossing of the two curves underlines the change of preference from Following to Deletion with age.

In spite of some similarities between the results of the lexicon-based studies and those of the new experiment, they also differ in several important respects. Most saliently, the results of the survey show a high degree of variation regardless of the age of the speakers. While the data from the lexicons suggest that an already regular derivation process became more regular with time, the survey suggests that CAL formation at the individual level remains a highly variable process. Participants in the survey made use of all four possible treatments for /R/ and /Q/, whereas even in Labrune's lexicon, only two of these were dominant for each type of special segment. Moreover, Deletion, which was the preferred treatment in both lexicons and for both segments, is not the most common treatment for all cases in the survey. This includes the fact that it loses to the Maintaining treatment for /R/ across all age groups (figure 8), and to various other possible treatments for older age groups for both /R/ and /Q/.

Another important difference lies in the fact that the total proportion of responses involving the four main treatments for /R/ and /Q/ is much less than 100%. The use of a wide range of different abbreviation strategies and various unexpected violations to otherwise active constraints is surprising considering how productive and overall regular the CAL formation process appears to be among attested forms. In the end, the Deletion treatment represents less than 30% of survey responses (see figure 10), which is even less than in Labrune's lexicon. In that sense, it is difficult to say that this treatment represents a new regularity.

Since the present study sought to corroborate that CAL formation has regularized, the large amount of variation observed in the results is somewhat puzzling. The divergence between the results of lexicon studies and those of the survey may be due to differences in the nature of the collected data. On the one hand, the CALs in the lexicons represent actual

attested forms. Those abbreviations have been used and accepted by the community, at least to the point that their use does not seem completely out of place in their associated contexts (video game websites, dictionaries, newspapers, conversations, etc.). The lexicons thus represent the abbreviation habits in the community of Japanese speakers. On the other hand, in the survey responses, community-level considerations presumably played less of a role, and only abbreviation preferences of individuals were assessed. Consider that in the isolated setting of answering a survey on a computer, speakers do not have access to any kind of community feedback, and thus may not consider community norms when choosing their responses.

Given these considerations, it appears that the CAL derivation process is indeed undergoing regularization at the community level, but individuals are lagging in the assimilation of the newly emerged rules. The fact that the survey-based study shows similar overall tendencies to the lexicon data but to a smaller extent suggests that individuals actually follow the same path of regularization as the community, only at a slower pace.

With respect to the community-level patterns, which exhibited relatively less variation, a standard Optimality Theoretical analysis proved to be very useful for characterizing the results. But this kind of framework, without some adjustments, is not as well-suited for characterizing the inter- and intra-individual variation found in the survey study. Stochastic Optimality Theory (Anttila 1997; Boersma & Hayes 2001; inter alia), by comparison, can capture not only individual-level variation, but also subtle gradient changes to the pattern of variation over time or across generations. In this framework, constraints are not considered to be strictly ranked in a hierarchy but instead occupy overlapping distributions on a harmony scale, and the ordering on any given occasion is obtained by sampling from those distributions.

We will here examine only the five constraints that enter in consideration in the treatment of /R/ and /Q/ as second mora of C2: LICENSE/Q/, *V#, CONTIG, *FEET<2, *FEET>2. Since LICENSE/Q/ is never violated, it is not necessary to consider its overlap with other constraints. The figures below illustrate the overlapping distributions for the four remaining constraints. LICENSE/Q/ is not shown here; it would be out of frame on the left.

Figure 11: Stochastic OT analysis of older people's preferences
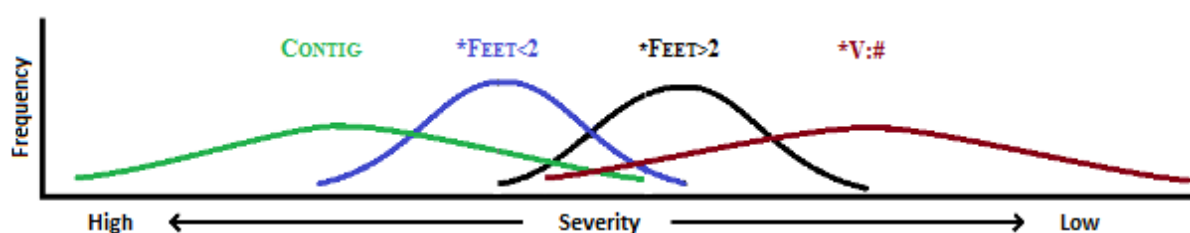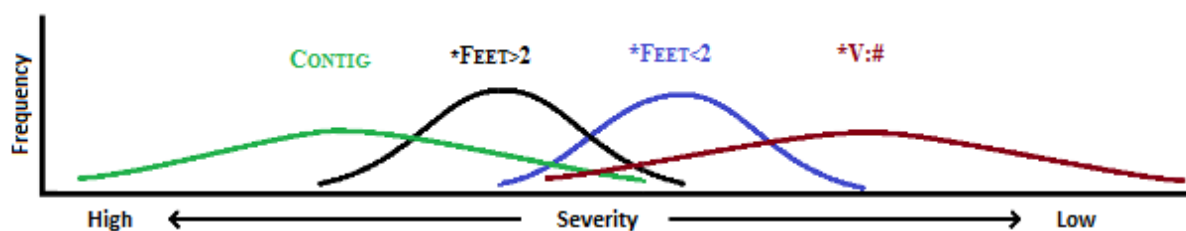

Figure 12: Stochastic OT analysis of younger people's preferences

The overall placement of the constraints on the harmony scale reflects the results presented in figures 8, 9 and 10 above. Since /R/ is more often Maintained than any of the other treatments, *V:# has to be the lowest-ranked constraint. By contrast, Replacement seems to be the least favored treatment[5], so CONTIG appears toward the upper end of the scale. The constraints associated with Deletion and Following would then appear toward the middle. In both figures, the distribution of all four constraints distributions overlap with every other constraint around the center of the figure. This reflects the possibility for every treatment (i.e., ranking) to be selected stochastically while also expressing the different level of preference for each one.

The influence of age on the abbreviation habits can be explained in terms of gradual shift of the distribution of *FEET<2 and *FEET>2 over time. As shown in the transition from figure 11 to figure 12, the shifting preference from Following to Deletion from older to younger speakers is represented by the reversal of those two constraints' positions on the harmony scale. It was observed in the lexicons that Deletion became systematic at a community level, which lead to the ranking of *FEET<2 at the very bottom of the hierarchy. It appears that this constraint is undergoing the same down-ranking at the individual level too, though the process is still ongoing.

---

[5] This may not appear clearly on those figures, but the survey study explores several other phonological contexts in which replacement is largely disfavored (Lamarque, 2016). We thus considered CONTIG to be the highest ranked of those four constraints. In any case, CONTIG's ranking does not have any effect on the following arguments.

4. Conclusion

The CAL derivation process, being a recent and productive process in modern Japanese, provides interesting insights into the evolution of that language's phonology. Since CAL formation is based on loanwords, it can, on the one hand, be viewed as relatively peripheral in in the lexicon of Japanese (Ito & Mester 1995, 2008). At the same time, CALs are formed according to a typically Japanese process, and therefore tend to be more respectful of the central phonological constraints in Japanese phonology. It seems that CALs have been well adopted by the community and this is reflected in the quantity of such abbreviations that already exists and which are constantly being created. Moreover, the community tends towards more and more regularity in the creation of new CALs, thus improving further their utility and acceptability as a means for appropriating words of foreign origin.

However, given the recent increase in the number of CALs, it appears that newly emerged preferences, especially those regarding the treatment of /R/ and /Q/, have not yet been fully assimilated by individuals. The results of our experiment show that the evolution observed in the lexicons is correlated with the abbreviation preferences of individuals, with younger individuals being more inclined to use the Deletion treatment for /R/ and /Q/. Nevertheless, there is still a strong discrepancy between the highly regular behavior of the community and the highly variable abbreviation patterns of individuals.

In that sense, the CAL derivation process represents an excellent case study for exploring community-level behaviors and their influence on the emergence of new linguistic regularities. In our future research we will explore the relationship between community-level and individual-level patterns directly by studying how individuals behave in the presence of real-time community feedback, and in the context of different types of community network organization.

REFERENCES

ANTTILA, ARTO. 1997. *Variation in Finnish phonology and morphology*. Stanford, CA: Stanford University dissertation.

BATES, DOUGLAS; MARTIN MÄCHLER; BEN BOLKER; and STEVE WALKER. 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

BENUA, LAURA. 1995. Identity effects in morphological truncation. *University of Massachusetts occasional papers in linguistics* 18.77–136.

BENUA, LAURA. 1997. *Transderivational Identity, Phonological Relations between Words*. Amherst, MA : University of Massachussetts dissertation.

BOERSMA, PAUL, and BRUCE HAYES. 2001. Empirical tests of the gradual learning algorithm. *Linguistic inquiry* 32.1.45–86.

DAVIS, STUART. 2011. Quantity. *The handbook of phonological theory (2nd ed.)*, 103–140. Wiley-Blackwell.

ITÔ, JUNKO, and ARMIN MESTER. 1995. Japanese Phonology. *The handbook of phonological theory*, ed. by John A. Goldsmith, 817–838. Oxford: Blackwell Publishers.

ITÔ, JUNKO, and ARMIN MESTER. 2008. *Lexical classes in phonology*. Oxford: Oxford University Press.

KUBOZONO, HARUO; JUNKO ITO; and ARMIN MESTER. 2008. Consonant gemination in Japanese loanword phonology. *18th International Congress of Linguistics, Seoul*. Citeseer.

LABRUNE, LAURENCE. 2002. The prosodic structure of simple abbreviated loanwords in Japanese: a constraint-based account. *Journal of the Phonetic Society of Japan* 6.1. 98–120.

LABRUNE, LAURENCE. 2005. Autour de la syllabe : les constituants prosodiques mineurs en phonologie. *Phonétique et phonologie, approches contemporaines*, ed. by Noël Nguyen, Sophie Wauquiers and Jacques Durand, 95–116. Hermès.

LABRUNE, LAURENCE. 2006. Patterns of Phonemic Preferences in Japanese non-headed Binary Compounds: What waa-puro, are-kore and mecha-kucha Have in Common. *Gengo Kenkyu* 129.3–41.

LABRUNE, LAURENCE. 2007. Morphophonologie des emprunts abrégés en japonais : aspects prosodiques et mélodiques. *Des sons et des sens : données et modèles en phonologie et en morphologie*. ed. by E. Delais-Roussarie and Laurence Labrune, Paris: Hermès.

LABRUNE, LAURENCE. 2008. Principes d'organisation phonémique des emprunts occidentaux composés abrégés. *Revue d'Etudes Japonaises*. 107–121.

LAMARQUE, REMI. 2013. *Théories des contraintes et phonologie des mots d'emprunt en japonais.* Aix-en-Provence, France: Aix-Marseille Université dissertation, 64p.

LAMARQUE, REMI. 2015a. Contraintes phonologiques et mots d'emprunt en japonais. *Actes de la 5e édition des JéTou*, 21–30. Toulouse: Université de Toulouse - Jean Jaurès.

LAMARQUE, REMI. 2015b. *Morphophonologie des emprunts abrégés japonais dans les jeux vidéo : évolution du processus d'abréviation*. Aix-en-Provence, France: Aix-Marseille Université dissertation, 132p.

LAMARQUE, REMI. 2016. *La variation en phonologie : le cas des emprunts abrégés composés en japonais.* Aix-en-Provence, France: Aix-Marseille Université dissertation, 74p.

LOVINS, JULIE. 1973. *Loanwords and the phonological structure of Japanese.* Chicago, IL: University of Chicago dissertation.

MCCARTHY, JOHN J.; and ALAN PRINCE. 1995. Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers in Linguistics,* ed. by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 249-384. Amherst, MA: GLSA publications. ROA-60.

OHSO, MIEKO. 1971. *A Phonological Study of Some English Loan Words in Japanese*. Columbus, OH: Ohio State University dissertation.

PRINCE, ALAN, and PAUL SMOLENSKY. 1993. *Optimality Theory: Constraint interaction in generative grammar*. Rutgers University Center for Cognitive Science.

RAMUS, FRANCK. 1999. La discrimination des langues par la prosodie: Modélisation linguistique et études comportementales. *De la caractérisation à l'identification des langues*. 131.

SHINOHARA, SHIGEKO. 1996. The roles of the syllable and the mora in Japanese: Adaptation of French words. *Cahiers de linguistique - Asie orientale* 25.1.87–112.

SHINOHARA, SHIGEKO. 1997. *Analyse phonologique de l'adaptation japonaise de mots étrangers*. Paris, France: Université de la Sorbonne Nouvelle - Paris III dissertation, 219p.