How Regular is Japanese Loanword Adaptation? A Computational Study

Lingshuang Jack Mao and **Mans Hulden**
Department of Linguistics
University of Colorado
{lima4664,mans.hulden}@colorado.edu

Abstract

The modifications that foreign loanwords undergo when adapted into Japanese have been the subject of much study in linguistics. The scholarly interest of the topic can be attributed to the fact that Japanese loanwords undergo a complex series of phonological adaptations, something which has been puzzling scholars for decades. While previous studies of Japanese loanword accommodation have focused on specific phonological phenomena of limited scope, the current study leverages computational methods to provide a more complete description of all the sound changes that occur when adopting English words into Japanese. To investigate this, we have developed a parallel corpus of 250 English transcriptions and their respective Japanese equivalents. These words were then used to develop a wide-coverage finite state transducer based phonological grammar that mimics the behavior of the Japanese adaptation process, mapping e.g **cream** [kɹiːm] to [kɯ.ɾiː.mɯ]. By developing rules with the goal of accounting completely for a large number of borrowings, and analyzing forms mistakenly generated by the system, we discover an internal inconsistency within the loanword phonology of the Japanese language, something arguably underestimated by previous studies. The result of the investigation suggests that there are multiple **dimensions** that shape the output form of the current Japanese loanwords. These dimensions include orthography, phonetics, and historical changes.

1 Introduction

Borrowing lexical items from one language to another is a common linguistic phenomenon. For example, someone can wear a **beret** /bəˈreɪ/ while enjoying **sushi** /ˈsuʃi/ next to a pet **alpaca** /ælˈpækə/. Loanword adaptation refers to the process of how foreign sounds or phonological structures are made to conform to the sounds and structures of the recipient language (Goldsmith et al., 2011). Loanword phonology, as the study of the systematicity of such adaptation, attracts scholarly attention because it arguably 'pushes the limit' of the phonological system of the borrowing language in order to preserve the phonetic quality and phonological structure of the source language. This study aims to investigate how 'systematic' or 'unsystematic' a loanword phonology can be. We choose Japanese as our specific research context.

Loanwords in Japanese are ubiquitous. According to Tsunoda (1988), loanwords account for 10 to 25 percent of the lexicon in nationally circulated, news-oriented weeklies, and up to 70 percent in professional journals, medicine and science in particular. However, any attempt to provide a general account for the Japanese loanword adaptation mechanism is an ambitious enterprise because the loanword lexicon contains many phonological exceptions whose formation processes have become opaque. Furthermore, most earlier works in Japanese loanword phonology focus only on particulars or a handful of specific arguments revolving around a limited number of constraints, or assume that a set of rules or constraints discussed have covered all the phenomena; few works have actually given a complete overview of the process. Therefore, a need for a fuller description of the loanword system has motivated this study.

Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 847–856, Osaka, Japan, December 11-17 2016.

2 Related work

2.1 Lexical stratification of the Japanese language

It is well-known that the Japanese lexicon is stratified along the lines of etymology, and Itō and Mester (1993) formalizes such a stratification on a phonological level. That work claims that the Japanese lexicon can be categorized into the strata *Yamato* (native), *Sino-Japanese* (lexicon historically imported from China), *Mimetic* (various onomatopoeic expressions), and *Foreign* (modern loanwords, usually from western languages), based on how prototypical or peripheral the words are with respect to the phonological constraints of the language. The main argument is that each stratum has a different degree of strictness in applying certain phonological constraints. The following example illustrates this claim further with three well-known constraints on Japanese words:

1. ***P**: /p/ is only tolerated in geminates or partial geminates
2. ***NT**: post-nasal obstruents must be voiced
3. **CODACOND**: coda consonants may not have a place feature

The following table adapted from Itō and Mester (1993) shows how each constraint applies to each lexical stratum:

Lexical stratum	***P**	***NT**	**CODACOND**
Yamato	Yes	Yes	Yes
Sino-Japanese	Yes	No	Yes
Mimetic	No	Yes	Yes
Foreign	No	No	No

Table 1: Coverage of three constraints for each lexical stratum in stratified models of the Japanese lexicon.

In Table 1 we can see that in the *Yamato* stratum, all three constraints are all expected to be satisfied; in *Sino-Japanese* and *Mimetic*, only two are mandatory; while in *Foreign*, none are mandatory. Thus, Itō and Mester (1993) hypothesize that the *Yamato* stratum forms the core of the Japanese lexicon, because it violates the fewest constraints, while the *Foreign* stratum forms the outermost periphery of the lexicon because it violates the largest number of these constraints. Figure 1 shows a visual representation of the phonological stratification of Japanese lexicon.

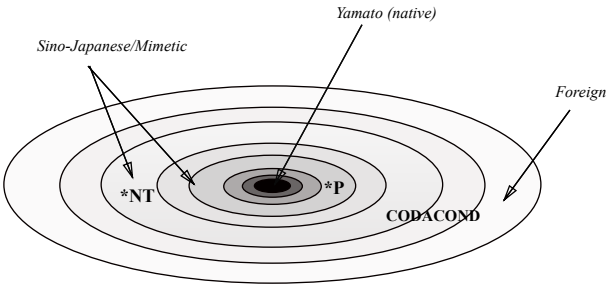


Figure 1: The core-periphery structure, adapted from Itō and Mester (1993)

This model of stratification of the Japanese lexicon has informed us that when approaching the periphery from the core, more and more forms are considered 'grammatical' because constraints becoming weaker and weaker until abolished altogether. Since the *Foreign* stratum sits on the periphery, we can assume that the loanword phonology actually **contains** the core phonology. Therefore, studying the loanword phonology matters to linguists who wish to study essential characteristics of the entire phonological system of a language and the tension between the native and foreign lexical strata.

2.2 Generative Phonology vs. Optimality Theory

As in most areas of phonology, the majority of previous work on Japanese loanword phonology has focused on analyzing its various aspects in the framework of rule-based Generative Phonology (Chomsky and Halle, 1968) and/or the newer and more fashionable constraint-based framework of Optimality Theory (OT) (Prince and Smolensky, 1993). Generative phonology regards a grammar as a set of phonological calculations (i.e. rules) that map an *underlying representation* (UR) into a *surface representation* (SR). OT on the other hand does not model the grammar of a language as a sequential algorithm; rather, a grammar is a set of defined structural limitations to be obeyed, each with a certain ranking of importance, known as *constraints*, and these constrains also interact with each other through their ranking. From a set of possible candidates of a certain form, the grammar chooses the candidate with the fewest violations of the constraints (i.e the 'optimal' one). Earlier works on Japanese loanwords are mainly based on generative phonology where the source language is the UR and the adapted Japanese loanword is the SR; with OT, a similar setup is assumed, but the SR is dictated by interaction of constraints.

2.3 Previous work and current study

Works on Japanese loanwords are substantial (e.g. Ichikawa (1929), Lovins (1975), Shirai (1999), Kubozono et al. (2013)) As a representative of earlier work in this field, Lovins (1975), in her dissertation *Loanwords and the phonological structure of Japanese*, takes a rule-based approach to investigate the loanword adaptation mechanisms from Western languages to Japanese. Her analysis is very detailed and the coverage of phenomena is also quite extensive, but, because the analysis and statistics are done manually, the whole endeavor is inevitably cumbersome. Some exceptions and corner cases are not clearly stated and explained, and the relations between loanwords and the rest of the Japanese language remains obscure. Such an effort of attempting to account for various linguistic forms has extended into recent work. For instance, Shoji and Shoji (2014) have proposed an OT account for vowel epenthesis and consonant deletion in Japanese loanwords adapted from English. However, such a framework predicts numerous Japanese outputs for an English word.

From those studies, not only did we see a limited scope of the investigation, but also a general underlying assumption that there exists a phonological system such that it can accurately transform the source language input into the borrowing language outputs, and that such a phonological system can potentially account for all the exceptions encountered in previous studies. In order to test the validity of this assumption, we decided to conduct a corpus-based computational study on Japanese loanword phonology. We decided to adopt a rewrite rule approach instead of OT because, as Karttunen (2005) points out, OT grammars are very difficult and tedious to debug; on the other hand, grammars based on rewrite rules built using finite-state transducer (FST) techniques are straightforward to construct and easily debugged.

What is different between our approach and previous ones is that we do not test or intuit the result of the rule compilation manually; rather, we utilize a more rigorous method: composing finite-state transducers to model phonological rules for mimicking the loanword adaptation processes, and automatically calculating coverage and adjusting the rules to maximize coverage. The benefit of the computational approach is obvious: we can immediately obtain the accuracy of the system output during each round of the rule development and modification process, which is nearly impossible when resorting to paper-and-pencil approaches.

3 Japanese loanword corpus compilation

We manually compiled a word list of 250 tokens of Japanese loanwords from a Japanese-Multilingual Dictionary—essentially a multilingual lexical database with Japanese as the pivot language (Breen, 2004). The word list contains parallel phonetic transcriptions of both English (source) and Japanese (borrowing) entries, and we designed it carefully in the following ways:

1. The word list includes all the phonemes in English and Japanese.
2. The word list reflects as many different phonological interactions as possible collected from previous studies.

3. The word list intentionally includes 'peculiar instances' discussed in previous studies (e.g. /pɪk.nɪk/ ↦ [pi.kɯ.nik.kɯ] in which the word-final /k/ is geminated while the word-medial /k/ is not.) (Kubozono et al., 2008).
4. We only included words with an English origin although some of them were often mistakenly attributed to other languages by some scholars. For example, Energie /ɛnɛʁgi/ ↦ [e.ne.rɯ.giː] (German) vs. energy /ɛnəʤi/ ↦ [e.na.jiː] (British English).

For majority of the words, we used standard British English transcription as the input instead of American English. The differences between the two varieties are largely irrelevant for the purpose of this study. But assuming British English to be the source is slightly preferred because it lacks the postvocalic /ɹ/ which is easier to adapt to Japanese as the language does not permit consonant codas except for the moraic nasal /N/ (e.g. /gɪtɑ(ɹ)/ ↦ [gitaː], and /fɔ(ɹ)k/ ↦ [foːkɯ]). However, some words were obviously borrowed from American English, and in such cases, American English transcription was used. For example, the Japanese SR for 'schedule' is [sɯ.ke.jɯɯ.rɯ] which is obviously derived from the American variant /skɛʤʊl/ as opposed to the British /ʃɛdjuːl/. In addition, the phonetic transcription is broad enough to reflect phonological processes rather than phonetic details. For instance, the Japanese adaptation of egg /ɛg/ was transcribed as [eg.gɯ] although voiced geminates are often devoiced in actual speech.

Some Japanese loanwords have two forms, as shown in the study of truncation/epenthesis loan doublets by Arakawa (1977):

	Truncation form	**Epenthesis form**	**English**
(1)	[pok.ke]	[po.ket.to]	pocket /pɒkɪt/
	[pɯ.ɾiN]	[pɯ.diŋ.gɯ]	pudding /pʊdɪŋ/
	[ɾa.mɯ.ne]	[ɾe.mo.neː.do]	lemonade /lɛməneɪd/

In these cases, only the epenthesis form was selected to be included in the corpus for the following reasons. First, the truncation form largely depends on the phonetics and phonology of the source language instead of the borrowing language. For example, because English phonology permits unreleased coda stops, such as /t/ in the source word pocket /pɒkɪt/, it provides a base for the Japanese loanword system to adapt the input also without the final consonant [t], namely [pok.ke]. Second, the epenthesis form usually keeps all the original consonants (or their adapted version) in the output, which contains more complete linguistic information. Third, the truncation form is not productive in the Japanese loanword system, and many loanwords do not have such forms. For example, gadget /gæʤɪt/ can only be adapted as [ga.jet.to] even though the English input ends with a /t/ that need not to be released.

We primarily relied on previous literature to decide which words to include in the corpus and how many instances to include for each process. Because we focused on compiling a computational grammar by composition of finite-state transducers, we believe that our corpus comprehensively reflects the Japanese loanword adaptation patterns gathered from previous literature. For a single segment, either a vowel or a consonant, previous literature and observations suggest that the adaptation is mostly predictable, so a few examples of each segment was considered sufficient. For instance, /θ/ ↦ [s] and /ð/ ↦ [z]. When dealing with unpredictable cases where a single input can be adapted into multiple output forms, more examples were included to ensure the generality of the rules. For example, /i/ is adapted as [iː] in comedy /kɒmɪdi/ ↦ [ko.me.d**iː**]; however, /i/ is adapted as [i] in celery /sɛləri/ ↦ [se.ro.r**i**]. In such cases, we included numerous examples of both adaptations and chose the rule that gave the better overall coverage. For more complex phonological phenomena such as gemination and vowel epenthesis, generally a large sample was included to ensure the coverage of all the phenomena described by previous literature.

In total, we discovered that the majority of the adaptation phenomena could be accounted for by a relatively small number of rules, the number of tokens included in the corpus was deemed sufficient for the purpose of this study.

4 Implementation

4.1 Finite-state phonology

The loanword adaptation grammar is implemented as a sequence of phonological rewrite rules in the Xerox finite-state calculus (Beesley and Karttunen, 2003). This is the standard mode of implementing complex hand-written phonological grammars computationally, capturing phenomena such as phonological alternations (Karttunen and Beesley, 2005), syllabification processes (Hulden, 2006), optimality-theoretic constraints (Karttunen, 1998), and nonconcatenative morphophonological processes (Beesley, 1998). The system relies on the ability to convert phonological replacement (or 'rewrite') rules into individual finite state transducers. A collection of such transducers can then be composed in a certain order, yielding one monolithic transducer, the end result essentially replicating the effect of applying multiple phonological rules in a sequence. This grammar transducer can also be applied in the inverse direction, mapping from a loanword phonological form to its possible sources. Although we don't explore this possibility here, the invertibility of a transducer is useful for debugging our rules. While finite-state tools have also been used to model other prominent approaches to describing loanword adaptation such as OT grammars (Karttunen, 1998; Gerdemann and van Noord, 2000; Gerdemann and Hulden, 2012), we restrict ourselves to the rewrite-rule model in this study.

The grammar itself is developed using the *foma*-toolkit (Hulden, 2009). A summary of the core formalism is given in Table 2. Although the formalism offers a vast number of operations, we essentially only make use of a standard context-conditioned rewrite rule, which has the following appearance

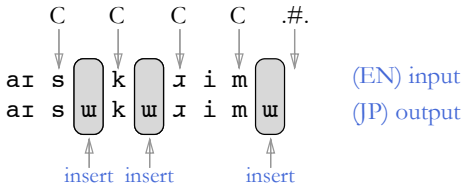
(2) `LHS -> RHS || LC _ RC`

which essentially reads as: replace all occurrences of the pattern `LHS` with the pattern `RHS`, whenever it occurs between the patterns `LC` and `RC`. The *patterns* in question can be expressed as *regular expressions*. Unlike the rewrite rules found in the phonological literature (Hayes, 2011), such a replacement rule is taken to operate simultaneously (and not iteratively in any directional manner) replacing all occurrences of `LHS` whenever warranted by the rule. For example, a rough approximation of accommodating the Japanese prohibition against complex consonant clusters, solved by epenthesis of [ɯ] between two adjacent consonants, or between a consonant and end-of-word, could be expressed as:

(3) `[..] -> ɯ || C _ [C|.#.]`

In effect, this says we insert a high unrounded back vowel [ɯ], whenever we have a consonant on the left-hand side and a consonant or end-of-word on the right. The rule itself compiles into an FST that gives rise to mappings such as the one below, translating the English phonetic representation of **ice cream** into a Japanese one (disregarding other relevant changes that also occur).

(4)



Note that this rule is simplified for the sake of exposition, and we will later provide a more accurate rule for the same phenomenon.

4.2 Developing and debugging rewrite rules

During development, we automatically checked the accuracy of our rules against the corpus, which contained not only the source forms (English) but also the attested output forms, i.e. the Japanese equivalents. This permitted us to gauge the coverage of each rule separately and iteratively address shortcomings of the adaptation grammar. Below is an example of three of our rules (here called R1, R2, and R3) that are active in transforming the English input **Christmas** /krɪsməs/ into the corresponding Japanese SR [kɯ.ri.sɯ.ma.sɯ].

`AB`	Concatenation
`A|B`	Union
`A & B`	Intersection
`A*`	Kleene Star
`~A`	Complement
`?`	Any symbol in alphabet
`0`	The empty string (epsilon)
`[` and `]`	Grouping brackets
`A -> B`	Change A to B
`[..] -> B`	Epenthesize B
`|| C _ D`	Context specifier
`.#.`	End or beginning of string
`def X ...`	Label a rewrite rule/regular expression

Table 2: An overview of the relevant regular expression and rewrite rule notation in *foma*.

R1: Epenthesize the vowel [ɯ][1] between two consonants or at the end of a word. The first consonant cannot be /n/, /t/, /d/, /tt/, or /dd/.

- `def insertu [..] -> u || C & ~[n|t|d|tt|dd] _ [C|.#.];`

R2: English input /ɪ/ becomes Japanese output /i/ in any phonological context.

- `def ichange ɪ -> i;`

R3: English input /ə/ becomes Japanese output /a/ in any phonological context.

- `def schwa ə -> a;`

Although they work, the context-free rule (R3) is too general. For example, for the input word, **violin** /vaɪəlɪn/ ↦ [bai.o.rin], it is necessary to apply another rule:

R4: English input /ə/ becomes Japanese output /o/ in any phonological context.

- `def schwa ə -> o;`

In order to reconcile the conflict between R3 and R4, new information has to be coded to differentiate versions of the input /ə/. We observed that the adaptation of /ə/ heavily depends on the original *orthographic* representation. Therefore, we coded each input /ə/ with the written orthographic symbol it originated from. For instance, 'əE' means that the input segment is /ə/ and it is represented as the letter **e** orthographically. After incorporating orthographic shape into the English transcription, we collapsed R3 and R4 into a single rule:

R5: The rule now can adapt /ə/ into different Japanese output forms.

- `def schwa əE -> e, əA -> a, əO -> o;`

Since the finite-state calculus allows us to view the input and output (domain and range) of a transducer as an automaton, we can compose our grammar with identity transducers that represent the source and target forms, respectively. This allows for quick identification of which rules fail and which overgenerate, which is helpful for rapid debugging and deployment—a standard technique in finite-state development (Beesley and Karttunen, 2003).

(5) `Sourcewords .o. Grammar .o. Targetwords`

After repeatedly refining our rules we arrive at the final Japanese loanword adaptation grammar.

5 Results

5.1 Phonological rules for adaptations from English to Japanese

By manual development of 34 phonological rules (3 insertion + 1 deletion + 30 alternation rules), the entire grammar correctly maps 215 out of 250 (86.0%) input forms to their Japanese counterparts.

[1]For convenience, we use [u] to represent [ɯ]—as there is no high back round vowel in Japanese, this cannot cause a conflict.

5.2 Error types

Before any rule is formally introduced into the system, we always test if the rule in question increases the overall coverage of the grammar. Although the final phonological system achieves high coverage for modeling loanword adaptation from English to Japanese, it is linguistically informative to analyze what kinds of mistakes the system makes. The mistakes arguably represent not only the shortcomings of this specific system, but also the characteristics of the Japanese loanword phonology as a whole.

Specifically, the Japanese SR forms which are mistakenly generated by the loanword adaptation grammar can be categorized into 3 types. Each type is presented in Table 3 with examples taken from the corpus for illustration.

English Input		System Output	Correct SR	Error Type	Percentage among mistaken outputs
diamond	/daɪəmənd/	dai.a.mon.do	dai.**y**a.mon.do		
steak	/steɪk/	sɯ.teː.kɯ	sɯ.teː.k**i**	Remainder of historical forms	28.6%
sport	/spɔːt/	sɯ.poː.to	sɯ.poː.**ts**ɯ		
gasoline	/gæsəlin/	gya.so.rin	**ga**.so.rin		
volunteer	/vɒlənˈtɪə/	bo.ran.chi.a	bo.ran.**t**i.a	Phonotactics relaxed	22.9%
teenager	/tineɪʤɜː/	chi.neː.jaː	**t**iː.neː.jaː		
bolt	/bəʊlt/	boː.rɯ.to	**bo**.rɯ.to		
tough	/tʌf/	taf.fɯ	ta.**f**ɯ	Unexpected segmental lengths	48.5%
sausage	/sɒsəʤ/	so.sej.ji	**soː.seː.**ji		

Table 3: Error types of the system output

From these mistaken forms, various types of linguistically useful information can be gleaned. First, historical forms, which comply more strictly with the native phonology of Japanese, and new adapted forms, which are more likely to be phonologically relaxed, exist simultaneously in the Japanese loanword lexicon. For example, the newer loanword **bike** /baɪk/ is adapted as [bai.k**ɯ**], while the older loanword **strike** /straɪk/ becomes [sɯ.to.rai.k**i]** due to a vowel harmony system in archaic Japanese (Shoji and Shoji, 2014). Secondly, some phonological processes, such as gemination, are unpredictable. It is difficult to explain why **bag** /bæg/ is adapted as /ba**g.g**ɯ/ but **bug** /bʌg/ becomes /ba.**gɯ**/ which forbids the gemination of the word-final voiced stop [g]. Therefore, the only way to model this type of ambiguous adaptation is to use segment-level statistics to help determine which adaptation has the highest probability to be the correct output form. This type of error also suggests that there might not be a single consistent way to generalize the loanword system of the Japanese language because the explanation for many corner cases and exceptions has been lost in history.

5.3 Rule modifications

According to these error types, we considered whether it was possible to modify the rules to eliminate these errors. Such a question can help us get closer to the ultimate goal: to construct a definitive phonological system which accurately produces adapted output forms. However, we noticed that when one rule is modified, it usually negatively affects the overall accuracy of the system output, because, although it 'fixes' a few current erroneous outputs, it also ruins previously correct forms. Table 4 shows how the number of mistaken outputs change when the current rules are modified according to various parameters.

Original number of erroneous forms	Parameter of modification	Current number of erroneous forms
35	Always inserting [y] between [i] and any vowel	35
35	No voiced geminates allowed	38
35	No velar stops allowed to be palatalized	38
35	No [t] allowed to be palatalized	38
35	First vowel of a trisyllabic word not allowed to be long	40
35	First vowel of any word not allowed to be long	71
35	No orthographic information coded into the lexicon	72-76 (depends one which uniform output is chosen, e.g. /ə/ ↦ [a], [o], or [e])
35	Only epenthesizing the default vowel [ɯ]	80

Table 4: Changes before and after rule modifications

As we see, none of the modifications given in Table 4 is really satisfactory since each of them increases the number of overall erroneous output forms of the system. This indicates that the essential 34 rules discussed in §5.1 have covered most of core phenomena in Japanese loanword phonology, and the remaining erroneous output forms are indeed exceptions outside of the set the core rules. This also means that there is clear limit to the predictability of the phonological system that drives the accommodation of loanwords in Japanese. In other words, after a certain point—corresponding to roughly 90% per type—there is a residue of internal inconsistency which is not capturable by any satisfactory phonological generalization.

6 Discussion

Because of such inconsistency within the system, if any analysis attempts to achieve a comprehensive account for Japanese loanword phonology, the model must involve external **dimensions** other than the one-to-one correspondence between the source tokens and borrowing language output. **Dimension** here refers to any domain which most significantly contributes to the nativization process of a certain group of loanwords. Based on the analysis, we incorporate four dimensions into our phonological system:

1. **Phonetic dimension**: phonetic characteristics of the source language heavily influences the form of output
   - Example: *pocket* /pɒkɪt/ ↦ [pok.ke] because English permits an unreleased /t/ at the end of the word, which is difficult for non-native speakers to perceive. As the result, Japanese adaptation also overlooks the final /t/ phoneme.
2. **Phonemic dimension**: The adapted form corresponds to all or almost all segments in the source language.
   - Examples: *pocket* /pɒkɪt/ ↦ [pok.ket.to] because all the segments in the source form are transformed into corresponding segments in the adapted form. In other words, the number of segments in the adapted form is no smaller than that in the source form.
3. **Orthographic dimension**: the orthography of the segment in the source language clearly determines or heavily influences the adaptation choice.
   - Example: *waffle* /wɒfəl/ ↦ [wa**f.f**ɯ.rɯ], where the gemination of [f] is clearly motivated by the orthography because the phonological environment where it occurs does not usually motivate gemination in Japanese. However, not every geminate reflects its original orthography. Compare *bottle* /bɒtəl/ ↦ [bo.**t**o.rɯ].
4. **Historical dimension**: For historical reasons, older forms survived diachronic changes and have become the only accepted adaptation in the current loanword system.
   - Example: The accepted Japanese loanword for the input *sport* /spɔːt/ is [sɯ.poː.**tsɯ**] with the affricate [ts] as the output instead of the newer adaptation mapping /t/ ↦ [t].

Notice that the orthographic and historical dimensions are not on the same hierarchical level as the phonetic and phonemic dimensions; rather, they exist as submembers of the phonemic dimension (see Figure 2 (a) and (b) for two examples). This hierarchical model successfully captures the fact that orthography and historical factors are only relevant when the phonemic dimension is chosen first.

7 Conclusion

Most previous studies on Japanese loanword phonology or on any phonology in general tend to take a 'paper-and-pencil' approach. Although doing phonology manually in fieldwork is considered traditional and long-standing, computational tools can arguably assist scholars to understand a new phonological system much more efficiently. In this study, we have investigated the phonological system of the Japanese loanwords using the finite-state tool, *foma*. By observing a carefully-compiled corpus and manually developing rewrite rules in *foma*, we obtained a holistic 'phonological grammar' of the adaptation process of Japanese. By analyzing the types of residual errors to which no good solution exists, modifying rules according to different linguistic parameters, we have discovered an internal inconsistency of the phonological system itself. This reveals linguistically important information about the phonological system:

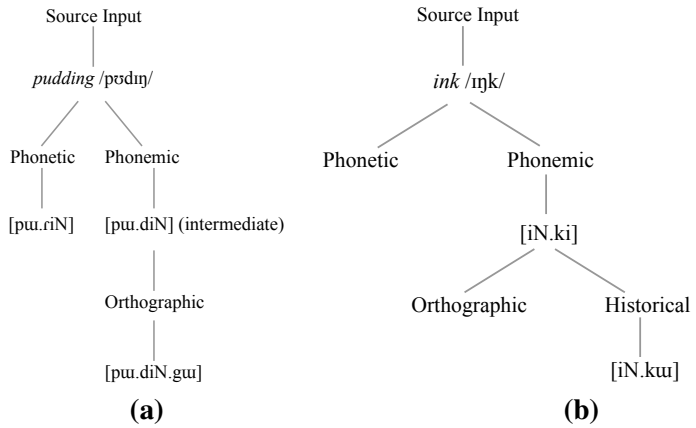


Figure 2: The hierarchical structure of phonetic, phonemic, and orthographic dimensions (a), and (b), the incorporation of a historical dimension into the hierarchical structure of dimensions.

multiple dimensions exist simultaneous within the Japanese loanword phonology that can represented in a tree-like hierarchical structure, and each foreign word has been adapted into Japanese through one specific dimension. All these discoveries would have been extremely difficult, if not impossible, to make without the computational tools available.

References

Sōbē Arakawa. 1977. *Gairaigo ziten*. Kadokawa, Tokyo, 2nd edition.

Kenneth R. Beesley and Lauri Karttunen. 2003. *Finite State Morphology*. CSLI Publications, Stanford, CA.

Kenneth R. Beesley. 1998. Arabic morphology using only finite-state operations. In *Proceedings of the Workshop on Computational Approaches to Semitic languages*, pages 50–57. Association for Computational Linguistics.

James Breen. 2004. JMDict: a Japanese-multilingual dictionary. In *Proceedings of the Workshop on Multilingual Linguistic Ressources*, pages 71–79. Association for Computational Linguistics.

Noam Chomsky and Morris Halle. 1968. *The sound pattern of English.* Harper & Row.

Dale Gerdemann and Mans Hulden. 2012. Practical finite state optimality theory. In *Proceedings of the 10th International Workshop on Finite State Methods and Natural Language Processing*, pages 10–19, Donostia–San Sebastián, July. Association for Computational Linguistics.

Dale Gerdemann and Gertjan van Noord. 2000. Approximation and exactness in finite state optimality theory. In *Proceedings of the Fifth Workshop of the ACL Special Interest Group in Computational Phonology*.

John A. Goldsmith, Jason Riggle, and Alan C. L. Yu, editors. 2011. *The handbook of phonological theory*, volume 75. John Wiley and Sons, New York.

Bruce Hayes. 2011. *Introductory Phonology*. John Wiley & Sons.

Mans Hulden. 2006. Finite-state syllabification. *Lecture Notes in Artificial Intelligence*, 4002:86–96.

Mans Hulden. 2009. Foma: a finite-state compiler and library. In *Proceedings of the 12th conference of the European Chapter of the Association for Computational Linguistics,*, pages 29–32.

Sanki Ichikawa. 1929. *Foreign Influences on the Japanese Language*. Japanese Council Institute of Pacific Relations, Tokyo.

Junko Itō and Armin Mester. 1993. Japanese phonology: constraint domains and structure preservation. *The handbook of phonological theory*, pages 871–838.

Lauri Karttunen and Kenneth R. Beesley. 2005. Twenty-five years of finite-state morphology. *Inquiries Into Words, a Festschrift for Kimmo Koskenniemi on his 60th Birthday*, pages 71–83.

Lauri Karttunen. 1998. The proper treatment of optimality theory in computational phonology. In *Proceedings of the International Workshop on Finite State Methods in Natural Language Processing (FSMNLP)*.

Lauri Karttunen. 2005. Finnish optimality-theoretic prosody. In *International Workshop on Finite-State Methods and Natural Language Processing*, pages 9–10. Springer.

Haruo Kubozono, Junko Ito, and Armin Mester. 2008. Consonant gemination in Japanese loanword phonology. In *18th International Congress of Linguistics, Seoul*.

Haruo Kubozono, Hajime Takeyasu, and Mikio Giriko. 2013. On the positional asymmetry of consonant gemination in Japanese loanwords. *Journal of East Asian Linguistics*, 22(4):339–371.

Julie B. Lovins. 1975. Loanwords and the phonological structure of Japanese. *Indiana University Linguistics Club*.

Alan Prince and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. *Ms. Rutgers University Cognitive Science Center*.

Setsuko Shirai. 1999. Gemination in loans from English to Japanese. *Master's thesis*.

Shinichi Shoji and Kazuko Shoji. 2014. Vowel epenthesis and consonant deletion in Japanese loanwords from English. *Proceedings of the Annual Meetings on Phonology*, 1(1).

Waka Tsunoda. 1988. The influx of English in Japanese language and literature. *World Literature Today*, 62(3):425–430.