# Applications of Machine Learning for Modeling Sound Symbolic Systems in Japanese and Korean Ideophones


by

Jacob J. Fernandes


presented to the Department of Linguistics
and the Department of Computer Science

in partial fulfillment of the requirements
for the degree with honors
of Bachelor of Arts

Harvard College
March 2024

# Abstract

Ideophones, or 'onomatopoeia', are a special class of adverbial modifiers present in Japanese, Korean, and many other languages across Asia and Sub-Saharan Africa that are strongly associated with imagery and certain audio-visual and other sensory traits. Ideophones are an oft neglected linguistic anomaly with research largely restricted to the kind of phonesthetic phenomena in English and other Indo-European languages (Azari and Sharififar 2017).

Yet, evidence suggests that ideophones play a large role in the discourse of Japanese, Korean, and many other languages (ibid.). As such, research into ideophones has been of particular focus in Japanese linguists, with Hirose (1981) and Hamano (1998) publishing landmark research on ideophones, although much of the existing research into ideophones has been restricted to their typology (Kita 1997). This study expands on Hamano's sound-symbolic system, investigating the links between sound and meaning for mimetics and their significant verbal collocations, with particular reference to Akita and Usuki (2016)'s finding that the semantic content of ideophones are strongly connected to complement information. This connection, along with the promise of Frame Semantics of complements as an effective way of evaluating adverbs in Natural Language Processing (Nikolaev et al. 2023), suggests that a model correlating the phonological form of the mimetic with the semantic frame of significant collocations may lead to insights into the semantic behavior of ideophones.

Using three classifiers—a Naïve Bayes, Random Forest, and Recurrent Neural Network (RNN)—I construct a model trained on distinctive featural representations of the phonological form on a large set of mimetics across a variety of vocabulary sources to predict the semantic content of ver-

bal collocations in the form of Semantic Frames, assisted by FrameNet as used by Akita (2013). I found that vowels in particular had a significant effect on the semantic frame of a complement, with [±High] as the most significant feature in both Korean and Japanese mimetics. While these models particularly struggled with the polysemy present in Korean and Japanese mimetics and verbal elements, this thesis found that Random Forest classifiers in particular are very effective learners of the relationships between sound and meaning. I found further that mimetic adverbials indeed performed better than their non-mimetic adverbial counterparts, indicating real iconicity in the so-called ideophones that is not present in regular manner adverbials in either language.

# Acknowledgements

Thank you, to my advisor Kevin Ryan who met with me and helped me talk through my thesis despite my incoherent ramblings: I could not have done this without your astute advice and encouraging words. Your guidance has been instrumental in completing this process.

To Stuart Shieber and Wei Wei Pan for providing me with invaluable advice in how to structure the computational aspects of my thesis, especially when I wasn't sure how to proceed. To Jack Rabinovitch, Yağmur Sağ, Tanya Bondarenko, and the entire Linguistics faculty for helping me through the writing process, for teaching me how to collect Linguistic data, and for providing me with resources and ideas to make my thesis the best it could be.

To the Harvard Asia Center, for providing me with funding to travel to South Korea and Japan to collect language data for verification of my experiments: I cannot thank you enough for providing me with the tools and means to really practice what it means to do research. To my anonymous consultants who provided me with the invaluable data even when my questions were a bit funky. Thank you for giving your best judgements. All errors are my own.

To June Park, for practicing Japanese with me for three years, personally connecting me with consultants in South Korea, and helping me plan this thesis since before I even knew what the topic would be.

To Matt, John, Yana, Alisa, and Sam, who put up with my constant questions and asking for advice, who supported me in my hardest times, and who uplifted and encouraged me to see the process through.

And finally, to my parents who instilled in me a love of language and a curiosity to understand the world around me. Without them, I would have accomplished nothing.

Thank you. I could not have completed this without any of you.

# Contents

# List of Figures

# List of Tables

# Glossing Conventions

| | |
|---|---|
| ACC | Accusative |
| ADV | Adverbial |
| CAUS | Causative |
| CONJ | Conjunctive |
| COP | Copula |
| DAT | Dative |
| EXPL | Explanatory |
| FORM | Formal Ending |
| IE | Informal Ending |
| LOC | Locative |
| MIM | Mimetic (ideophone) |
| NEG | Negative |
| NMLZ | Nominalization |
| NOM | Nominative |
| NPST | Nonpast |
| PASS | Passive |
| POL | Polite |
| PST | Past |
| QUOT | Quotative |
| TOP | Topic |

一を聞いて、十を知る

*Ask one thing, learn ten.*

– Japanese Proverb

# Chapter 1

# Introduction

시작이 반이다

*Starting is the half.*

– Korean Proverb

For both learners and translators of Japanese and Korean, few other classes of words cause as much confusion or difficulty than mimetic modifiers. Whereas the relationship between sounds and their semantic meaning is often considered arbitrary, onomatopoeic words are an exception (Azari and Sharififar 2017), and often subject to the effects of sound symbolism. In the speech and literature of languages across the world, onomatopoeia are used to indicate feelings, emotions, experiences, and (especially) auditory phenomena. These **mimetic words** are very common in Japanese and Korean. While also present in English, their role is more restricted. Bredin (1996) noted among the distinct types of onomatopoeia in English *direct onomatopoeia* and *associative onomatopoeia*[1]. *Direct onomatopoeia* refers to those words which resembles the word that it means (e.g, **hiss** and **moan**, although **moan** is not etymologically onomatopoeic in origin ("moan, v." 2023)). Included in this category are terms for animal sounds, which are "at the very lowest threshold for direct onomatopoeic resemblance", and so universal among human lan-

[1] Bredin also considered a third type of onomatopoeia, *exemplary onomatopoeia* which refers to the amount of effort a speaker makes to say a word. This type of onomatopoeia is not relevant for the purposes of this thesis due to it not having a direct analogue in either Japanese or Korean

guage as to be "common human instinct" to derive onomatopoeic words for their sounds (Bredin 1996). *Associative Onomatopoeia* refers to those words which resemble a sound of something concurrent with what it denotes: these words describe actions or nouns, but resemble in sound that which is associated with their semantic meaning (ibid.). Bredin posits **cuckoo** as an example of a word resembling the sound which the bird itself makes. In fact, this type of onomatopoeia is relatively common cross-linguistically (c.f, Mandarin 猫 *māo* "cat", Arabic *ʾiwazz* "goose").

Beyond these 'lexical' onomatopoeia, a large amount of onomatopoeic words describe sounds, feelings, emotions, or other phenomena made up by speakers *ad hoc*. These onomatopoeia often vary between speaker, although some can be encoded in the language. These words, like '**pow**' have strong imagery associated with them, but they are not grammaticalized to the extent that mimetic modifiers are in Japanese and other languages, and are largely restricted in their usage. Due to this less-important role of onomatopoeia in languages like English and other languages which make up the bulk of modern linguistic research, onomatopoeia are sometimes considered a "second class" amongst words and rarely subject to research (Azari and Sharififar 2017).

In stark contrast, in Japanese speech and literature, onomatopoeia play an incredibly important role beyond mere mimetic purposes (ibid.). In Japanese, a subset of onomatopoeic words makes up a special class of lexically encoded and phonemically unique modifiers. Some of these modifiers feature phonemes in contrastive positions that are non-phonemic elsewhere in the language's native vocabulary and can encode additional psycho-affective meaning in the form of auditory, visual, or sensory imagery (Kita 1997). This fact, compounded with their primary use as manner-adverbials makes them unique, and necessitates comparison to other languages under the Davidsonian model.

Existing research into onomatopoeia is often motivated by the difficulties onomatopoeia place onto second language(L2) learners. In fact, the very way that L2 speakers process onomatopoeia differs from the way that native speakers do: A study from Osawa (2019) found that even native Korean speakers were more likely to understand unfamiliar Japanese onomatopoeia "analytically" by considering the form, sound and meaning of a word, whereas Japanese speakers

were more likely to understand them "generally". Osawa assigns a handful of reasons to this phenomenon. The first is a general judgement on the different structures of Korean and Japanese onomatopoeia: perhaps the Korean sound symbolic system has stronger connections between sound and meaning, due to the rich variations in onomatopoeia that exist in Korean, when compared to Japanese (Osawa 2019). The second is the notion that as learners of a foreign language, these native Korean speakers were forced to piece together meaning section by section, processing the meaning of the word in a 'top-down' approach in order to better their understanding of the mimetic phrase, whereas Japanese native speakers did not need to (ibid.).

Regardless of whether both suggestions are compounding each other (as Osawa posits), or one particular solution is the root of the issue, processing and understanding of onomatopoeia for second language learners and descriptive linguists alike is a difficult task. Oftentimes, native speakers are unable to describe or even paraphrase the meaning of these words, simply repeating them accompanied with hand motions or sound effects when asked to pin down the semantic content or paraphrase (Kita 1997). Yet, their markedly systematic form described by Hirose (1981), Hamano (1998) and others makes them an interesting subject of research. By better understanding exactly how the phonology of these supposedly sound-symbolic words contributes to its meaning, we can better understand the semantics of ideophones in general.

As two languages which make particular use of these onomatopoeia in their literature, [2] Japanese and Korean have mimetic adverbials that are of particular interest because of their proposed semantic connection to phonological form. There is evidence that, at least in Japanese, the semantic connection between mimetic adverbial and predicate is so strong, that any verbal complement can often be omitted (Akita and Usuki 2016), and that certain mimetics collocate significantly with one specific host predicate (Akita 2012). Hamano (1998)'s system of sound symbolism in Japanese ideophones seems to further suggest that the relationship between phonological form and semantic context is further influenced not only by pure phonemic content, but the length of and order of phonemes within a lexeme.

[2] Termed as GIONGO and GITAIGO in Japanese, and ŭITŏNGŏ and ŭITCHAEŏ in Korean

This understanding is critical not only for theoretical linguistics and language instruction, but also for the world of Natural Language Processing and machine translation. Ideophones and other onomatopoeic structures often pose significant difficulties for machine translation algorithms. A study by Reenald and Marion (2022), found that onomatopoeia translations between Indonesian and Japanese often missed important semantic context or were outright incorrect due to ambiguity and lack of accurate training data. This was confounded because of the failures of existing algorithms to translate onomatopoeia, which tend to do so on a word-for-word basis (ibid.). Reenald and Marion (ibid.)'s findings indicate a clear need to consider and better understand the semantics of onomatopoeia beyond a word-by-word basis, and better compare them in the context of their environments.

As such, it becomes imperative that research be conducted to further quantify these relationships between sound and meaning from an objective, nonlinear standard, moving beyond the labels of consonant and vowel, to one which more accurately represents the differences between individual phonemes themselves: distinctive features which strongly contribute semantic qualities. In this thesis, I propose a model which investigates the symbolic link between the phonological character of a word and the semantic quality of its complements, rather than previous works which have investigated the semantic quality of the mimetic itself. I develop two experiments trained on large corpus data to connect a phonological feature representation of the sounds in a given Japanese or Korean mimetic word with the semantic content of the complement with which it is used. Using a Naïve Bayesian, Random Forest, and Recurrent Neural Network (RNN) Classifier, I determine which model is best able to model sound symbolic relationships. I then evaluate the predictions of these models against the results of two human validation experiments, comparing the predictions of computational models with native Japanese and Korean speakers' judgements on both real and nonce words.

I found that in both Korean and Japanese mimetic adverbials, vowels stood out as a particularly significant phonological component. [±High] was the most statistically significant feature contributing to semantic content of verbal collocates in both the composite set of mimetics in

Japanese and Korean with mean importance of 1.9% and 4.0% respectively. Among the classification models I trained, the Random Forest model was the most efficient and effective across both languages, although RNN classifiers were similarly promising for use in predicting semantic content in Japanese. While Korean mimetics appear to be more iconic, with stronger recall and precision across models in semantic frame-based experiments, mimetics in both languages clearly outperform nonmimetic adverbials in forming strong collocations with specific semantic frames, validating my predictions that a model based on distinctive feature data, rather than merely phoneme labels, is particularly effective in isolating specific categories contributing to the semantic quality of a mimetic and its complement.

Elicitation further revealed that native speakers of Korean and Japanese differed in their ability to handle nonce vocabulary. While distinctive features acssociated with vowels were found the be the most significant predictor of semantic frame, speakers expressed that changes in vowel quality did not meaningfully change either speaker perception of meaning for mimetics or collocates used in most cases. Instead, voicing for Japanese and 'tenseness' ([±C.G]) for Korean proved to be the most significant contrastive features.

# Chapter 2

# Background

*If I have seen further, it is by standing on the shoulders of giants.*

– Sir Isaac Newton

Ideophones as a special class of words remain understudied phenomena across language families, with much of the research remaining descriptive on how the phenomena interact with the phonological systems of the languages which they belong to. This research is particularly sparse among research done on English and other Western languages, particularly because these languages lack such systems found in Japanese and Korean, as well as many African languages (Azari and Sharififar 2017). This makes ideophones an important subject of research, especially with computational models making use of the "systematicity" (Haslett and Cai 2023) particularly present in the robust Japanese and Korean sound symbolic systems.

Systematicity, which Haslett and Cai (ibid.) define as the phenomenon whereby "semantically related words share a phonological or orthographic feature, helps both human cognition and computational processing—especially in the case of rare words, being able to rely on systematic phonological features and word shapes allows for psychological plausibility on the generated output (ibid.). Indeed, as Hamano (1998) and Akita and Pardeshi (2019) make clear, the systematicity in Japanese (and Korean) ideophones is largely based on phonotactic structures, with

secondary syntactic, prosodic, and morphological systems as well. This thesis aims only to investigate the phonological systematicity, its interactions with the semantic structures of complements, and the effectiveness of representing such interactions in computational models, but for comprehensiveness, the other structures will be briefly mentioned in Section 2.3 and 2.4.

This chapter aims to describe existing attempts to map this systematicity. Further, I describe existing structures in Japanese, Korean, and other languages for describing ideophones and the links between phonological form and meaning. I begin with a brief description of Japanese and Korean phonology to create a better understanding of possible contrastive features and their roles in the sound-symbolic systems of the two languages. Then, I give a brief description of the formation and basic use of ideophones in both languages, followed by a discussion on challenges in researching and working with ideophones. I conclude by looking at what existing body of literature exists researching the sound-symbolic systems in Japanese and Korean, particularly from a computational lens.

## 2.1 A Description of Japanese Phonology

Despite having only a 'Moderately-Small' consonant inventory (Maddieson 2013a) and an 'Average' vowel quality inventory (Maddieson 2013b), Japanese is host to a rich array of phonotactic questions to be answered, especially in regards to ideophones, which some unique phonemes or phonotactic patterns ideophones are restricted to. The consonant inventory of Japanese, including phonemes exclusive to ideophones, is summarized in Table 1, with allophones in brackets and marginal or contested phonemes in parentheses (Labrune 2012).

Notes on Table 1:

(1) The bilabial plosive /p/ occurs only in a limited scope, appearing typically only in mimetic native strata or in Sino-Japanese loans (ibid.)—words which were borrowed from Middle Chinese. In native vocabulary, this phoneme almost appears exclusively as a geminate at word boundaries as in 子供っぽい *kodomoppoi* (*kodomo* + *-poi*) (ibid.). Even within mimetic

Table 1: Japanese Consonant Inventory

| | Labials | Alveolars | Palatals | Velars | Glottal |
|---|---|---|---|---|---|
| Plosives | p[1] b | t d | | k g | |
| Fricatives | {ɸ β}[2] | s z | (ɕ ʑ)[4] | | h |
| Affricates | | {ts}[2] | (tɕ)[3] | | |
| Nasals | m | n | | (ŋ)[5] | |
| Glides | | | j | w | |
| Liquid | | r | | | |

roots it is seemingly limited, occurring only infrequently in initial position (whereas /p/ never appears initially in non-mimetic native Japanese vocabulary), and typically occurs as a geminate as a reinforced form of /h/ (Labrune 2012) (c.f やっぱり *yappari* "indeed" < やはり *yahari*, さっぱり *sappari* "refreshed", たっぷり *tappuri* "plentifully, abundantly").

Elsewhere, in Sino-Japanese strata, it appears almost exclussively in the geminate form [pp] as a realization of the cluster /Qh/ or /Nh/. (c.f 一本 *ippon* </iQhon/).

/b/ is the voiced counterpart of /h/ and /p/ in *rendaku*, the process by which voicing occurs in compounding structures (c.f 花火 *hanabi* < *hana* + *hi*).

(2) /ɸ/ and /ts/ are the realizations of /h/ and /t/ respectively before the high back compressed vowel /ɯᵝ/. /ɸ/ is phonemic only in recent loanwords and in ideophones.

(3) /t/ is realized as [tɕ] before the high front vowel /i/.

(4) /s/ and /z/ are realized as as [ɕ] and [ʑ] before the high front vowel /i/.

(5) /ŋ/ only surfaces in internal cluster of /N/ and a velar consonant due to place assimilation.

Japanese has a 5-vowel system with relative symmetry, containing the vowels /a/, /i/, /ɯᵝ/, /o/, and /e/ (ibid.). Vowels also contrast in length, with both short and long vowels phonemic. The Japanese vowels can be summarized featurally below (ibid.).

Of these, /e/ is perhaps the most unique: it remains by far the least common vowel in the Japanese language, occurring only infrequently in native vocabulary, and once ever within a single root morpheme. It also has negative connotations when used in mimetic adverbs (ibid.).

Table 2: Japanese Vowel Features

| | a | i | ɯ$^{\beta}$ | e | o |
|---|---|---|---|---|---|
| high | - | + | + | - | - |
| low | + | - | - | - | - |
| front | - | + | - | + | - |

In addition to having a moderately small phoneme inventory (Maddieson 2013a), Japanese also has a relatively restricted syllable structure of $(C)(y)V(\{N, Q, R\})$ (Labrune 2012). Labrune lists the archiphonemes N, Q, and R as representing moraic [1] codas, including the moraic nasal(N), which assimilates in place of articulation to successive consonants, the archiphoneme Q, which causes gemination to occur in the successive plosive, and R, which lengthens previous vowels.

## 2.2 A Description of Korean Phonology

Standard Korean is one of only two members of the Koreanic language family along with Jeju-eo, a Koreanic language spoken on Jeju Island (*Spoken L1 Language: Korean* n.d.). Once considered linked to Japanese and other languages of East and Central Asia as a part of the hypothesized Altaic language family, Korean is nowadays considered a true language isolate. Still, because of geographic proximity, Korean and Japanese have many similarities in phonology, morphology, and grammar (Robbeets 2008).

The Korean dialect of focus in this thesis will be the Seoul variant, and that which has been studied the most in theoretical and comparative literature. As such, all consultants used for data collection in this thesis will be speakers of the Seoul variant. The Seoul Korean phonological system is one of much linguistic study due to several curiosities in its allophony, consonant alterations, and vowel harmony system. This section will summarize descriptively a state of the language for analysis and use in this thesis. The Korean consonants can be summarized as follows.

Notes on Table 3

[1] A mora is a unit of syllabic information often used in Japanese. A mora is a constant length full syllable, such that the moraic nasal ん *n* is pronounced identically long as any CV syllable

Table 3: Korean Consonants and Approximates (Cho 2016)

| | | Bilabial | Alveolar | Alveo-Palatal | Velar | Glottal |
|---|---|---|---|---|---|---|
| Plosives | Lenis | p (ㅂ) | t (ㄷ) | c (ㅈ) | k (ㄱ) | |
| | Aspirated | $p^h$ (ㅍ) | $t^h$ (ㅌ) | $c^h$ (ㅊ) | $k^h$ (ㅋ) | |
| | Tense[1] | p' (ㅃ) | t' (ㄸ) | c' (ㅉ) | k' (ㄲ) | |
| Fricatives | Lax | | s (ㅅ) | | | h (ㅎ) |
| | Tense[1] | | s' (ㅆ) | | | |
| Nasals | | m (ㅁ) | n (ㄴ) | | ŋ[2] (ㅇ) | |
| Liquid | | | l (ㄹ) | | | |
| Glides | | w | | j | | |

(1) The so-called 'tense' consonants, alternatively termed 'fortis' (Cho 2016), are contested in their representation under a distinctive feature framework among Korean literature. They are often treated as geminates: indeed, in Yale Romanization and standard Hangul they are written as with a doubling of the consonant: c.f 빨 *ppal* and 발 *bal*. However, they are perhaps more accurately treated as [+constricted glottis] or [+stiff vocal] (ibid.).

(2) Only surfaces in coda position.

### 2.2.1 Vowel Harmony

Modern Standard Korean inherits a vowel harmony system, a common trait for "Altaic languages"[2] from a far more robust one that existed in Middle Korean (Ko 2022). This system has been collapsing among many Korean varieties since at least the fifteenth century CE (Larsen and Heinz 2012), and in modern speech is typically constrained to specific morphological environments (Ko 2022).

Contemporary varieties have two distinct systems of vowel harmony still in use: one system used for verbal suffixes, and another used exclusively in sound symbolic words (also known as ideophones – see Section 2.3). Korean's system contrasts with the phonetic bases that many other 'Altaic' languages present, which include Tongue Root Harmony, Backness, Rounding, and Height(Larsen and Heinz 2012). Instead, both systems of vowel harmony divide their vowel

[2]Both the existence of the Altaic language family and the placement of Korean in it are controversial. See Robbeets (2008) and Vovin (2005).

space into categories labelled 'dark' and 'light' phonemes (Also labeled YIN and YANG in some literature, and categorized as 'masculine' and 'feminine' in others). The system used for verbal suffixes is unique in not being based in "natural classes" described in traditional phonology, although some attempts to describe the system treat the distinguishing feature as [±Low] (Larsen and Heinz 2012). The names 'dark' and 'light' should perhaps not be taken as qualitative descriptions of the vowel quality themselves: the label as *yin* or *yang* appears referential to the direction the vowel line faces when written in Hangul: up or right for *yang* vowels (오 *o* 아 *a* 외 *ø* and 애 *æ*), and down or left for *yin* vowels (우 *u* 으 ɨ 위 *y* 에 *e* and 어 ʌ); the vowels nonetheless are not arbitrarily placed in their respective categories. An important exception to this system is the phoneme /i/, which often behaves as a neutral vowel relative to the dark and light classes (Ko 2022). Larsen and Heinz (2012)'s descriptions of these classes are shown in Tables 4 and 5.

Table 4: Vowel Harmony in Korean Verbal Suffixes

| | front | front round | central | back round | |
|---|---|---|---|---|---|
| high | i | y | ɨ | u | dark |
| mid | e | ø | ʌ | o | light |
| low | æ | | a | | |

The first system is restricted to the form of a suffix which must attach to the verb root before other suffixes, including the past allomorph. This suffix changes its realization depending on the final vowel of the verbal root. After dark vowels /a/ and /o/, it is realized as [-a], whereas after light vowels it is realized as [-ʌ] (Larsen and Heinz 2012). This harmony is further restricted by locality and does not apply to suffixes with initial consonants (Ko 2022).

| | front | front round | central | back round | |
|---|---|---|---|---|---|
| high | (i) | y | ɨ | u | dark |
| mid | e | ø | ʌ | o | light |
| low | æ | | a | | |

Table 5: Vowel Harmony in Korean Ideophones

The second system features a set of pairs of one dark vowel and one light vowel each, whose

Figure 1: Correspondences in the Korean Ideophone Vowel Harmony System (Ko 2022), examples from Sohn (1994)

| | | | |
|---|---|---|---|
| [i]∼[æ] | 맨둥 ∼ 민둥 | mæntung ∼ mintung | 'bald, deforested' |
| [e]∼[æ] | 때각 ∼ 떼각 | t'ækak-t'ekak | 'cracking' |
| [y]∼[ø] | 휘휘 ∼ 회회 | hy-hy ∼ hø-hø | 'round about' |
| [ɨ]∼[a] | 살살 ∼ 슬슬 | sal-sal ∼ seul-seul | 'gently, lightly' |
| [ʌ]∼[a] | 반짝 ∼ 번짝 | panc'ak ∼ pʌnc'ʌk | 'glittering' |
| [u]∼[o] | 퐁당 ∼ 풍덩 | $p^h$oŋtaŋ ∼ $p^h$untʌŋ | 'with a plop' |

alternation changes the semantic information of the ideophone. An ideophone containing light vowels may be termed as a diminutive, whereas one containing a dark vowel is an augmentative (Ko (2022) and Larsen and Heinz (2012)). These are, however, not the only nuances which might be reflected by an alternation in harmony group: other metrics which are subject to alternation for a light-dark pair include brightness-darkness, lightness-heaviness, sharpness-dullness, quickness-slowness, smallness-deepness, and thinness-thickness (Sohn 1999 as cited in Larsen and Heinz (2012)). The correspondences are shown in Figure 1.

In addition to these, [i]∼[a] and [u]∼[a] may also alternate in initial syllables (Sohn 1994), although Ko (2022) posits that the former can be attributed to fronting effects from palatal initial consonants and the latter attributed to rounding effects from initial consonants compounded with a merger of [ʌ] and [a] in such syllables.

Complicating this system (and only the system used in sound-symbolic words) are three major factors: the defectiveness of some ideophones, with many lacking either a 'dark form' or a 'light form', coöcurrence of dark and light vowels in other ideophones, and the ability of high vowels to behave as 'transparent' in non-initial position, coöccurring with either dark or light vowels (ibid.).

## 2.3 Ideophones

Despite Japanese and Korean's phononological similarities, and the systematic research into ideophones individually in Japanese and Korean, such a systematic comparison of Japanese and Ko-

rean ideophones has not been thoroughly conducted due to apparent dissimilarity, although an inspection at monosyllabic ideophones reveals a "meaningful comparison" between the two systems (Hamano 2020). What follows in this section is such a comparison of the two systems and their similarities.

### 2.3.1 Japanese Mimetics - *Giongo and Gitaigo*

Japanese ideophones can be split largely into three major categories, of which two will be the focus of this thesis. *Giongo* represent 'sound-mimicking words' (Hamano 1998). In addition to these *giongo*, a second subset of ideophones, *gitaigo*, is differentiated by Hirose (1981) as 'mimetic' words, while Hirose labels *giongo* 'onomatopoeic' words. The English translation of these labels vary by author, but Hirose's perhaps best captures the similarity of *giongo* to English sound-symbolic words, with *gitaigo* representing something different altogether that happens to be represented by the same phonological and morphological system in Japanese. Hirose further divides both categories into sub categories based on the nature of the sound and mimed action, although those categories will not be explored in this thesis.

Ideophones in Japanese are marked by a strong semantic connection between predicate and mimetic (Hamano (1998); Akita and Usuki (2016))—so strong, that the "meaning of the mimetic inherits that of its host predicate" (ibid.). As such, the use of an ideophone often limits or narrows what possible predicates may be used, although such predicates are not fixed or restricted (Hamano 1998).

#### Giongo

Giongo represent auditory phenomena, representing physical actions by quoting them iconically (ibid.). This is accomplished with the use of the quotative particle *to*, along with the quotative verb *iu* when modifying a nominal attributively in a relative clause (ibid.). This is a large and diverse class of vocabulary, and can be used with verb nominalization to describe additional auditory information about the action, or auditory information that accompanies the action.

(1) a. ざあっと言うこぼれ方

zaaQ-to iu kobore-kata
MIM-QUOT say spill-manner
"The manner of spilling *with the sound of zaaQ*, in which a large quantity of liquid gushes out when spilled"(Hamano 1998)

b. ポタポタと言う落ち方

potapota-to iu ochi-kata
MIM-QUOT say fall-manner
"The manner of falling *accompanied by the sound of pota-pota*"(ibid.)

**Gitaigo**

While similar in both sound and form, *gitaigo* represent a different phenomena entirely. Rather than representing auditory phenomena, they often describe sensations, visual impressions, or emotional responses (Hirose 1981). Unlike *giongo*, *gitaigo* vary in which light verb they select in attributive use, with some *gitaigo* selecting the light verb *iu*, and others the light verb *shita* (Hamano 1998). While *shita* is morphologically the past tense of the verb *suru*, 'to do', in attributive use with ideophones, it is not possible to replace it with *suru*'s synonym, *yaru*, although this would be possible for attributive use in other nominals or adverbials (ibid.)[3]. Examples of *gitaigo* are listed in example (4).

(2)

| | | |
|---|---|---|
| フワフワ | fuwafuwa | "fluffily (like cake), bouncily, springily, " |
| ジーン | jīn | "going numb, tingling" |
| ピカピカ | pikapika | "with a sparkle" (ibid.) |

### 2.3.2 Korean Mimetics - *Ŭitŏngŏ and Ŭitchaeŏ*

Korean ideophones can be split into two categories which neatly map to those used in Japanese. *Ŭitŏngŏ*, the Korean equivalent of *giongo* represent auditory phenomena, while *ŭitchaeŏ* are words that imitate 'appearance' (Ferklova 2023).

[3] See Section 2.4.2

Compared to Japanese, the number of ideophonic roots in Korean is somewhat lower, although auditory mimetics make up a larger proportion of the total amount of mimetics. Only 802 manner mimetics and 394 auditory mimetics have distinct roots in Korean, compared to Japanese with 1265 manner mimetics and 225 auditory mimetics (Osawa (Itou) 2006 as cited in Osawa 2019). However, because of consonant alternation and vowel alternation that occurs for these roots, in reality the number of Korean ideophones in use is much larger than this figure suggest (ibid.).

## 2.4 Previous Research on the Topic

Onomatopoeia remain severely understudied in linguistics, although several attempts have been made to understand their origin in both cross-linguistic and single language surveys. While literature on mimetics in Japanese number in the thousands, historically the literature has focused on descriptive research of morphophonology (Akita and Pardeshi 2019). Yet, Vinson et al. (2015) cite that "Iconicity is [instead] argued to play a foundational role in providing a link between linguistic forms and physical experience with the world, both in terms of processing and language learning". If iconicity is truly so foundational to the human connection between language and the world, why is it not studied more?

In much literature on Western or European languages, onomatopoeia are often grouped with baby talk or imitative words meant to mimic the sounds created by or associated with a certain object or action (Bredin 1996). This is so much so, that Bredin (1996) considers many onomatopoeia to be 'human instinct'. In fact, in general linguistics, sound symbolism is a domain of more interest to the general public than to serious researchers; Saussure raises the arbitrariness of language as one of its principal features, such that the meaning of a word is unrelated to the sounds which comprise it (Kawahara and Shinohara 2012; Kita 1997). Compounding on this historical analysis is the small role that onomatopoeia play in English and other Western languages. The Saussurian focus on abstract semantics and not on iconicity is the fault of the

historical focus of linguistics on Indo-European languages and others "impoverished (in) iconic vocabularies" (Vinson et al. 2015).

That's not to say that iconicity is completely devoid in Indo-European language focused linguistics: perhaps one of the most well-known examples of research into iconicity is the BOUBA-KIKI EFFECT. Under this effect, it is said that "Humans robustly associate spiky shapes to words like 'Kiki' and round shapes to words like 'Bouba.'" (Passi and Arun 2022). This phenomenon seems to be markedly universal. Testing on Himba participants in northern Namibia with little Western interaction and no written language, Bremner et al. (2013) found that a majority of participants mapped the name "Bouba" to a round, blob-like shape and the name "Kiki" to a sharp, angular one. While this was one of the first study of its type to test the effect on a language with no written language, there is some strong influential effect of orthography and written language on the effect. Cuskley et al. (2017) found that letter curvature was a strong influence in participants' selection of 'kiki' or 'bouba', "significantly influenc(ing) word-shape associations *even* in auditory tasks". As such, when conducting studies on iconic words, it is important to keep in mind possible influences of orthography, which have remained under-specified in previous literature (ibid.).

Unlike the bouba-kiki effect which focuses on a single round-sharp contrast present in nonce vocabulary, evidence from ideophones in East-Asian and African linguistics point to them as forming a unique class of their own, worthy of linguistic research (Akita 2009; Hirose 1981; Kawahara and Shinohara 2012), even though these systems have not been treated fully by non-native linguistics (Hamano 1998). Much of the research research on onomatopoeia in Japanese and other East-Asian languages that does exist focuses on the phonological or morphological typology of mimetics[4], with research on the semantics of ideophones largely limited in scope and frequency (Kita 1997). Despite this, ideophones have been shown to be a class of vocabulary that depict a broad range of sensory experiences and with marked phonotactics, morphology, prosody, and syntax (Akita and Pardeshi 2019).

[4]Referred to in general linguistics literature and henceforth in this thesis as **ideophones**: words with symbolic marking which may or may not be directly referent to sounds associated with their meanings

Ideophones have also been linked in literature to "baby talk[5]". Kubozono (2019) found that both ideophones and motherese are morphologically similar, preferring reduplication structures of bimoraic bases and being restricted in length to three-mora and four-mora forms, in addition to sharing prosodic structures with HEAVY-HEAVY or HEAVY-LIGHT disyllabic structures as opposed to LIGHT-HEAVY ones. Where these differ is in their syntactic structures. While words from Motherese behave almost exclusively as nouns, ideophones may behave as either adverbials, nouns, adjectives, or as verbal nouns taking the light verb する *suru* "to do" to achieve predicative meaning, although there exists a strong preference for a role as an adverb (Kubozono 2019). For example, compare (1)-(3) from Kubozono.

(3) 腕にぶつぶつができた
ude-ni butsubutsu-ga deki-ta
arm-LOC MIM-NOM emerge-PST
'I have a rash on my arm'

(4) 肌がぶつぶつしている
hada-ga butsubutsu shite-iru
skin-NOM MIM do.CONJ-PROG
'My skin has broken out into a rash'

(5) 人がワンワン泣いた
hito-ga wanwan nai-ta
person-NOM MIM cry-PST
'The person cried out textitWah-wah!)'

(6) ワンワンが泣いた
wanwan-ga nai-ta
doggie-NOM cry-PST
'The doggie cried'

(7) * 人がワンワン泣いた
hito-ga wanwan nai-ta
person-NOM doggie cry-PST
'The person cried (like) a doggie'

Note that in the above examples, the meaning of a single ideophone can often vary slightly between its various syntactic roles. While the ideophones in examples (4)-(6) were able to occur as either a noun, verbal noun, or as an adverb, the "motherese" word *wanwan* (derived from the

[5] In Japanese linguistics, often referred to as 母親語 *hahaoya-go*, lit. "Mother-ese". This refers to words like ワンワン *wanwan* "doggie", itself derived from a *giongo* referring to the sound a dog makes, うんこ *poop*, クック *kukku* "shoes" < Adult Vocabulary 靴 *kutsu* "shoes"

same ideophone used in (6)) can only be used as a noun as in (7). An attempt to use it in (8) as an adverb fails when trying to keep its meaning consistent as the motherese one.

### 2.4.1 Translation of Ideophones

As can be seen in (6), translation of ideophones can often be a very difficult matter: Azari and Sharififar 2017 discuss the most common methods of translating onomatopoeia in Persian-English translations, many of which are often used in translating between Japanese and English.

- Translation with non-onomatopoeic words
- Translation with onomatopoeic words
- Translation using paraphrasing
- No translation (omission)
- Translation using loan word or no explanation

Azari and Sharififar suggest as the most common strategy using translation with onomatopoeic words when there exists similar onomatopoeic words between the source and target language. This is common in Japanese translation examples like in (8).

(8) 白いタイルがピカピカ輝く
shiroi tairu-ga pikapika kagayaku
white tile-NOM MIM$_{with.a.sparkle}$ sparkle
"White tile catching the light (ting! ting!)" (Yoshimoto 1992)[6]

When an equivalent cannot be found, as is common with Japanese ideophones, Azari recommends using paraphrasing. This is especially common with certain classes of ideophones which can be used to specify manners of activity. Hirose (1981) cites the case of the verb 歩く *aruku* "to walk", which can be paired with a host of mimetics to translate motion verbs either phrasally or with single words in English. Compare in example (9).

[6]Translation taken from the official English translation of the novel, while Japanese sentence taken from the original publication.

(9)

| | | |
|---|---|---|
| どさどさ歩く | *dosadosa aruku* | 'to walk with a loud noise' |
| ドシンドシン歩く | *doshindoshin aruku* | 'to lumber' |
| ヨタヨタ歩く | *yotayota aruku* | 'to waddle' |
| ぶらぶら歩く | *burabura aruku* | 'to stroll leisurely' |
| てくてく歩く | *tekuteku aruku* | 'to hike' |
| ウロウロ歩く | *urouro aruku* | 'to loiter' (Hirose 1981) |

### 2.4.2 Semantics and Sound Symbolism

This ability to convey a broad range of semantic information is due to the property of ideophones as representing the AFFECTO-SEMANTIC DOMAIN — the domain in which language interacts with sensory, motor, and psychological information (Kita 1997) — rather than an imitation of sound: few ideophones are purely auditory (*giongo*) (Hamano 1998). This domain is also termed the "expressive mode" (Diffloth 1972, as cited in Dingemanse 2012) or "performative mode" (Nuckolls 1995) Such words are unique in their ability to involve the affecto-semantic domain in their semantics as opposed to solely the ANALYTIC DOMAIN: that of decontextualized predication (Kita 1997). Kita argues that one such proof that ideophones exist in the affecto-semantic domain, separate from non-iconistic vocabulary, lies in their ability to not create redundancy in an otherwise redundant sentence. Consider the following examples Kita gives.

(10) *太郎は急足で早歩きをした

tarō-ha isogi-ashi-de haya-aruki-wo shita
Tarō-TOP hurried-feet-INSTR hasty-walk-ACC do.PST
"Taro walked hastily hurriedly"

(11) 太郎はそたそたと早歩きをした

tarō-ha sotasota-to haya-aruki-wo shita
Tarō-TOP MIM$_{with.a.hurried.walk}$-QUOT hasty-walk-ACC do.PST
"Taro walked hastily in a hurried walk"

Kita attributes this feature to the ability of mimetics to enhance the vividness and tone of a verb or sentence through their secondary visual, spacial, and psychological information. This

trait is not unique to Japanese: iconicity and the use of the affecto-semantic domain is also shared with sign languages, which are able to utilize a broader range of iconic lexical ranges to encode information both through abstract analytic information as well as iconicity in hand shapes, motion, position, and other methods (Vinson et al. 2015). Vinson et al. found the iconicity present in British Sign Language to facilitate pattern-sign matching, lexical processing, and phonological decision making. In their study, Vinson et al. argue for equal application of their findings beyond the field of sign linguistics to general spoken linguistics and the research into the semantics of iconicity and serious treatment of studies in onomatopoeia.

Ideophones in Japanese follow a seemingly codified system, with different features of a morpheme impacting the meaning of the resultant word in predictable patterns (Akita 2009; Hamano 1998, 2013). These patterns differ depending on the length of the morpheme for a given ideophone. Generally, ideophones are split in the literature between CV and CVCV morphemic structures, with those ending in the archiphonemes /N/ and /Q/, a diphthong, or a long vowel considered extensions of the basic CV or CVCV structure, rather than their own unique ideophonic form (Akita 2009; Hamano 1998, 2013). Some evidence points to CV ideophones as being more iconic than CVCV, although the latter is far more represented in mimetic dictionaries and often considered to be more conventional than CV mimetic stems (Akita 2009). To this extent, Hamano (1998) lists the features involved in the sound-symbolic system of CV ideophones. The four primary semantic factors conveyed by consonants in morphemes are described in (12).

(12)

| | |
|---|---|
| Initial consonant: | Tactile Information |
| Obstruent Voicing: | Weight |
| Palatalization: | Energy |
| Final Consonant Presence: | Final aspect |

Vowels seem to contribute a lesser role in the iconic meaning of an ideophone, a phenomenon not unique to Japanese ideophones, but present even among such examples as the bouba-kiki

effect[7] (Fort et al. 2015; Hamano 1998, 2013). According to Hamano (1998), the features present among a vowel contribute as follows.

(13)

| | |
|---|---|
| Vowel Quality: | Size and Shape of Movement |
| Length: | Length |
| Diphthongization: | Circular Movement |

Finally, repetition of a morpheme contributes to the so-called 'phase' of a word, mirroring the number of acoustic signals or actions with the number of forms (ibid.). While often resulting in an iconic word corresponding to the insensity or length of its event, this connection can often be less obvious (ibid.).

(14) 演説者は拳でテーブルをどんとたたいた
enzetsusha-ha kobushi-de tēburu-wo don-to tatai-ta
speaker-TOP fist-INSTR table-ACC MIM$_{with.a.bang}$ knock-PST
"The speaker banged on the table with his fist"

(15) 泥棒は拳でドアをどんどんたたいた
dorobō-ha kobushi-de doa-wo dondon tatai-ta
thief-TOP fist-INSTR door-ACC MIM$_{with.repeated.banging}$ knock-PST
"The thief pounded on the door with his fist"

However this repetition does not necessarily incur the progressive aspect, typically associated with continuous or ongoing events as in (16). This is possibly related to Kita's (1997) analysis of ideophones existing in a separate domain from the lexical verb, which inhabits the analytic domain.

(16) 長い間ドアをたたいていた
nagai aida doa-wo tatai-te-i-ta
long period door-ACC knock-CONJ-PROG-PST
"I was knocking on the door for a long time"

Bimoraic mimetic stems (ideophones whose base morpheme is formed from a CVCV sequence) complicate slightly the sound symbolic framework. Within a CVCV stem, Hamano (2013) breaks down each phoneme into a constituent part which has its own role in composing

[7] See Fort et al. (2015)

the meaning of the overall morpheme: $C_1$, $C_2$, $V_1$, and $V_2$. These are summarized below as in Hamano (1998).

(17)

| | |
|---|---|
| $C_1$: | Tactile information |
| $C_1$ Obstruent Voicing: | Weight and mass |
| $V_1$: | Initial shape of movement/object |
| $C_2$: | Type or manner of movement |
| $V_2$: | Resultant shape and size of movement |
| Medial C: | Intensification |

In addition to these, some features behave identically to monomoraic mimetic stems: palatalization, extended elements (/N/ and /Q/), and vowel length. Unique to bimoraic morphemes is the /-ri/ final element which indicates a quiet ending of movement (ibid.). Hamano cites /r/ as symbolizing smooth movement here with /i/ as epenthetic without any kind of obvious symbolic ties between phonological and semantic information.

Like in monomoraic stems, bimoraic stems have less iconic information encoded on their vowels. This is so much so that Hamano (ibid.) describes those words with /p/ realizations of $C_2$ as having little to no iconicity for vocalic elements, wheras those with /p/ in initial position have some iconicity reported (though this may be due to the rarity of medial /p/ even among ideophones).

One strong association of a specific phoneme shared among both types of mimetic stems are those of the archiphoneme /N/ and /Q/. In those mimetic stems which are extended with /N/, a sense of resonance and 'lingering' is felt (ibid.; Hirose 1981; Kita 1997), while those extended with /Q/ relate an "abrupt or forceful movement" (Hamano 1998). Hirose (1981) lists the following examples in (19).

(18)

| | | |
|---|---|---|
| かん | kan | the sound a large bell in a clock tower |
| ごん | gon | the sound of a huge bell in a Buddhist temple |
| ずきんずきん | zukin-zukin | a throbbing pain (with lingering impulses) |
| ざぶん | zabun | the sound of waves crashing against a cliff |

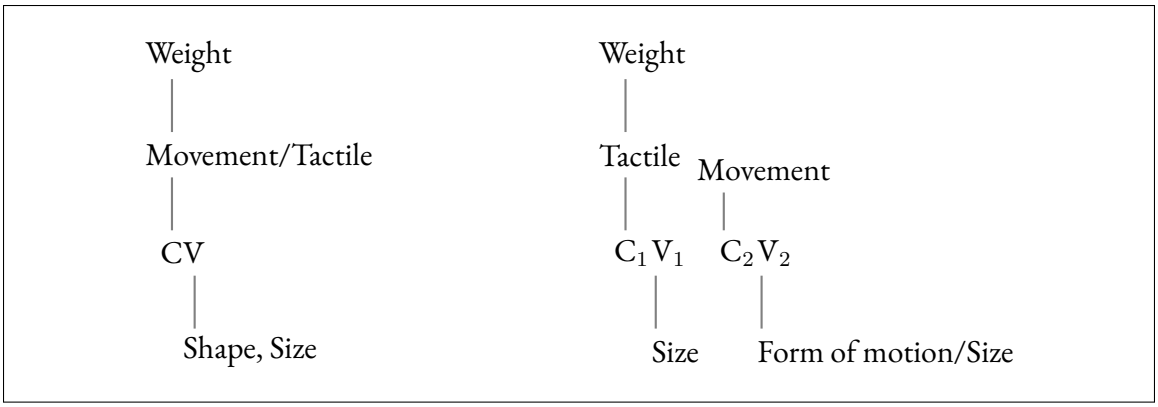



Figure 2: Hamano's Sound Symbolism System for Ideophone Stems

All together, Hamano's associations of sound symbolism can be summarized in the chart in Figure 2 (Hamano 2013), noting that CV and CVCV type syllable-roots behave mostly similarly in terms of sound symbolism.

The semantics of an ideophone are related not only to the phonology and construction of the word, but also to its syntax and the way that it is used in a sentence. As mentioned in reference to motherese, certain ideophones are capable of undergoing quasi-incorporation with the light verb する *suru* (Kubozono 2019; Akita and Usuki 2016). This is a true incorporation, as evidenced by the inability of such compounds to replace the suffixed *suru* with it's synonym やる *yaru* "to do" in their non-incorporated counterparts (Hamano 1998).

(19) 研究をした

kenkyū-wo shita
research-ACC do.PST
"(I) did research"

(20) 研究をやった

kenkyū-wo yat-ta
research-ACC do-PST
"(I) did research"

(21) ピカピカしたイワシ

pikapika shita iwashi
MIM$_{with.a.sparkle}$ do.PST herring
"shiny herrings"

(22) *ピカピカやったイワシ

pikapika yatta iwashi
MIM$_{with.a.sparkle}$ do.PST herring
"shiny herrings"

This is not true of all ideophones: many, especially *giongo*, take the quotative verb いう *iu* "to say" instead, with Hamano (1998) arguing that the ideophones which make use of *suru* share more of the characteristics of stative verbs than those which do otherwise. Further, these verbs incorporating *suru* are argued to be less iconic, with more abstract meanings (Akita and Usuki 2016; Hamano 1998). Some ideophones may make use of both verbs to form attributive forms, although in such cases, the use of *iu* is more expressive (ibid.). Which verb a mimetic takes for attributive use has strong correlation with rules regarding its use of the quotative particle と in adverbial modification as well (Akita and Usuki 2016).

### 2.4.3 Restraints

As a large portion of literature on the topic has focused on morphophonology of ideophones, the typological restrictions placed on the class of words are well known. Despite this, classification and even definition of a word as an ideophone remains a challenge (Nuckolls 2019; Akita 2009). Akita (ibid.) describes the difficulty with categorizing words as ideophones (particularly in Japanese) due to their similarities in behavior with regular native vocabulary. There seems to be a sense of "I know it when I see it" when it comes to the classification of a word as mimetic. While native speakers can easily identify such words, attempts to classify ideophones as a separate strata from the regular lexicon have mostly failed on morphological, historical, and certain phonological senses due to a lack of unique markers identifying such a strata (ibid.). In fact, not all words diachronistically considered ideophones were always considered such, with words like のんびり *nonbiri* "leisurely" being considered today an ideophones due to its form, morphological, and syntactic behavior, despite originating from the lexical verb *nobu* "to stretch" (Hamano 1998).

Further, unlike many English onomatopoeias, ideophones are socially constituted in their

associations between mean and form, and new mimetics cannot be easily and created extemporaneously (Kita 1997; Hirose 1981).

### 2.4.4 Computational Research

While most research has centered on descriptive morphophonology, recent research in the past few decades has seen a rise in computational analyses of ideophones, especially in regards to sound symbolism. However, such research remains sparse in Korean and other East Asian languages with similar phenomena. Indeed, sound symbolism remains an under-studied phenomena within **Natural Language Processing** (NLP) cross-linguistically. In a study conducted on Japanese participants with Japanese, Chinese, and Korean Pokémon[8] names, Kilpatrick et al. (2023) found that RANDOM FORESTS were "efficient learners of systematic sound-meaning correspondence patterns and [could] classify samples with greater accuracy than the human participants." Training a random forest on the names of Pokémon before and after evolution[9], the most important features among the Japanese names were the phonemes /m/ and /N/ at 0.78% and 0.63% respectively, while phoneme length (both of vowels and consonants) was third at 0.45% (ibid.). Kilpatrick et al. found that such NLP algorithms were not only successful at classifying sound-symbolic properties, but were more accurate on test classification than human participants despite having fewer samples (ibid.). While Kilpatrick et al.'s finding is inspiring for future sound-symbolism work in NLP, it was limited to a unique set of words—Pokémon names, across three langauges with very different naming conventions. Unlike traditional ideophones, Pokémon names are exclusively nouns, and do not carry the same affecto-semantic information(sensory, auditory, tactile, and psychological) as mimetics do.

Computational methods for categorizing ideophones in Japanese have also been developed by those like Takayama et al. (2019) designed a classification algorithm to categorize ideophones in Japanese into certain domains based on semantic use, focusing on sound symbolism. Takayama

[8] A fictional character from the Pokémon video game series. The word Pokémon originates as a shortening of the Japanese word ポケットモンスター: *pokettomonsutā*, literally "pocket monster"

[9] The process under which a Pokémon changes form and name and increases in power; not to be confused with Darwinian evolution.

et al. suggests using a model that maps a feature vector onto a given onomatopoeic word, weighting certain binary factors (clear versus cloudy, neat versus dull, brilliant versus blurred, etc.). Choosing 116 words from onomatopoeic words describing nature published in a Japanese onomatopoeic dictionary for which there were only one usage rule, Takayama et al. used a Random Forest model and found that precision and recall of categorization were higher than previous studies which attempted to categorize sound symbolism for ideophones. When repeated with human experiments, Takayama et al. found that the accuracy for the algorithmic model were more accurate both for precision and recall than the human participants. While this is promising, Takayama et al.'s model was extremely limited in scope, investigating only binary choices for semantic meanings using a very small subset of onomatopoeic words which were themselves limited in semantic variety..

### 2.4.5 Ideophones in Other Languages

Studies into similar sound symbolic systems exist in linguistic research into other languages as well. Ibarretxe-Antuñano (2019), focusing on motion ideophones cross linguistically, notes that many languages' phonemes that occur in ideophones do not occur elsewhere. Additionally, while they are cannot be "homogeneously characterized", they share a number of typological characteristics among which they can be compared (ibid.).

Cross-linguistically, ideophones carry an ability to mark vivid sensory information in addition to plain declarative or informational information. Nuckolls (1995) remarks on the ability of ideophones in Quechua to enable a speaker and listener of a symbolic statement to imagine a process, action, or event in a sound symbolic performance. In this sense, sound symbolic vocabulary are both objectively truthful and aesthetically vivid (ibid.). Likewise, (Ibarretxe-Antuñano 2019) argues that ideophones are able to "economically" capture an enourmous amount of vivid depictive information though difficult to translate, partially explaining a tendency for ideophones to be produced "together with gestures and prosody".

Attempts to codify ideophones in cross-linguistic typology have found that the occurence of

ideophones in a language tends to follow a subtype hierarchy, where

> If a language has ideophones at all it will have at least ideophones for sound (i.e. onomatopoeia). If a language has ideophones for movement it will also have ideophones for sounds. If a language has ideophones for visual patterns (e.g. spatial configuration or surface appearance), it will also have ideophones for movements and sounds, et cetera. Conversely, a language that does not have ideophones for sounds or movements will not have ideophones for cognitive states (Dingemanse 2012)

That is,

(23) SOUND < MOVEMENT < VISUAL PATTERNS < OTHER SENSORY PERCEPTIONS < INNER FEELINGS AND COGNITIVE STATES

Hinton et al. (1994) also describes a cross-linguistic system of categorization for ideophones, though not hierarchically. They describe a series of types of sound symbolism including CORPOREAL sound symbolism (the internal state of the speaker: emotional or physical), IMITATIVE sound symbolism (those words which imitate the sounds of the environment around oneself), SYNESTHETIC sound symbolism (visual, tactile or proprioceptive properties of objects), and CONVENTIONAL sound symbolism (phonesthemes, as in *gl-* in English words like *glitter*, *glisten*, etc.)

From the aforementioned hierarchy and Hinton et al.'s set of sound symbolism categories, Ibarretxe-Antuñano (2019) develops a so-called semantic grid for motion ideophones, building off of the work of Cifuentes-Féres (2010), Nuckols (2014), Toratani (2012) and in collaboration with Kimi Akita. This grid consists of seven first-level semantic components with 26 second-level subcomponents and 121 third level sub-subcomponents.

The seven first-level semantic categories categorize motion ideophones on the basis of FIGURE, GROUND, PATH, MOTION, MANNER (the largest component), CAUSE, and EVENT EXTENSION. Applying this grid to ideophones from Basque, Ibarretxe-Antuñano drew ideophones from the *Orotariko Euskal Hiztegia* (*General Dictionary of the Basque Language*) to create a 453

lemma-large corpus upon which categorization of the semantic grid's components were applied. Ibarretxe-Antuñano (2019) found that the MOTION component was found in all 453 tokens, with MANNER the next most frequent, in 440 tokens. While Ibarretxe-Antuñano found it not possible to "draw sound conclusions on the prominence of a certain (sub)component or component combination", they argue that it is possible to highlight general tendencies of component prominence and complexity through such an analysis.

Indeed, a common thread between various research on ideophones is that ideophones very rarely carry only a single iconic component (Nuckolls 2019). Nuckolls found in a survey of 500 ideophone tokens in Pastaza Quichua that only 7 ideophones encoded for a single category, all of which encoded a SOUND meaning. Moreover, MOVEMENT overrepresented for categorization, while VISUAL and SOUND formed a "supercategory" which were underspecified for their category (ibid.).

## 2.5 Conclusion

In this chapter, I have investigated the frameworks under which later analysis can be conducted on ideophones for both Standard Japanese and Standard Korean, as well as the status of research on ideophone-like systems in other non-Indo-European langauges.

Despite the sizeable body of recent research into the field of phonosemantics and theoretical topics on ideophones, little comparative work cross-linguistically exists. Further, research on ideophones in all languages considered remain largely descriptive, with ideophonic research in Korean and Japanese nearly exclusively limited to descriptions of phonetic and morphological systems. Research into ideophones is indeed becoming more mainstream despite protest from Saussurian thought, however as I have shown remains largely restricted to individual languages, and in the case of Korean, almost exclusively done in Korean language literature.

As such, this thesis attempts to contribute a work comparing Korean and Japanese using the models described above in hopes of enlightening comparisons in ideophonic categorization

and structure of sound symbolic systems between the two languages, utilizing such structures as created by Ibarretxe et al. Future chapters will investigate the relationship between ideophones and the semantic content of their verbal complements, often termed 'sound symbolism' from a computational standpoint, building on the work described in Section 2.4.4 with the goal of ascertaining which models work best to predict such relationships and to what extent phonetic content is predictive of semantic information through comparison of Korean and Japanese.

# Chapter 3

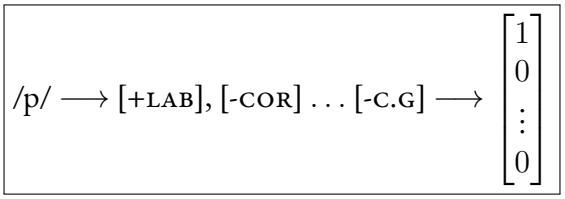
# Methods

This chapter describes the application of classification algorithms in parsing Japanese linguistic data for the replication of previous studies on ideophones and the classification of those words overlooked in prior studies. As seen in Chapter 2, computational studies on ideophones looking at relationship between phonological form and semantic content have restrictions in scope, either focusing on binary features, or restricted to a small, special class of vocabulary. In this chapter, I describe the methodology of two experiments and two human validation studies conducted on both corpus data and mimetic dictionaries, as has been the standard in cross-linguistic ideophonic research.

Experiments I and II are controlled supervised learning experiments utilizing an application of Akita's FrameNet assisted lexical framework. These experiments compared the performance of different classification models on predicting the semantic categories of verb complements in both auditory and non-auditory mimetics, as well as in nonmimetic adverbials as a control.This thesis finally concludes with a human validation experiment which will be labeled Experiment Ia and IIa for validation on Modern Japanese and Modern Korean to compare the results of the classification models with human judgements. Ideophones were modeled as a one-hot encoding of the distinctive features for each of their constituent phonemes, such that every contrastive feature present in Japanese and Korean phonology is represented as either 0 (not present) or 1

(present). Compare figure 3 which shows this process for a phoneme /p/.

Figure 3: Binary Encoding of Phonetic Features in /p/




The minimum set of contrastive features used for analysis in Japanese experiments are described below.

(24)

| | | |
|---|---|---|
| 1. [±LAB] | 6. [±SILIBANT] | 11. [±COMPRESSED][1] |
| 2. [±COR] | 7. [±SPREAD.GLOTTIS] | 12. [±HIGH] |
| 3. [±DOR] | 8. [±VOICE] | |
| 4. [±NASAL] | 9. [±DELAYED.RELEASE] | 13. [±LOW] |
| 5. [±SONORANT] | 10. [±SYLLABIC] | 14. [±FRONT] |

For Japanese experiments, in addition to these binary values which encode purely phonological data, each vector also contained a field to represent the presence or absence of *to*, a quotative marker whose ability to be omitted is not always predictable.

Korean experiments used the minimal pairs described below.

(25)

| | | |
|---|---|---|
| 1. [±LAB] | 6. [±SILIBANT] | 11. [±LOW] |
| 2. [±COR] | 7. [±SPREAD.GLOTTIS] | 12. [±FRONT] |
| 3. [±DOR] | 8. [±ASPIRATED] | |
| 4. [±NASAL] | 9. [±TENSE[2]] | 13. [±BACK] |
| 5. [±SONORANT] | 10. [±HIGH] | 14. [±ROUND] |

[1] This is not typically included in lists of features, but as the realization of /w/ in Japanese is with compressed lips, represented in IPA as /ɯᵝ/ (Okada 1999). Unlike a true /w/, the lips are not rounded.

[2] See 2.2

## 3.1 Experiment I

Research into ideophones has repeatedly shown that bare mimetics bear a strong semantic association with verbal collocates that they modify (Akita and Usuki 2016), with some evidence indicating that under a Semantic Framing analysis, mimetic frames are in fact an "elaboration of verb frames" (Akita 2013). As such, one possible method for ascertaining sound symbolism in ideophones is to investigate the significance of certain phonemic features in determining the semantic frame of their collocate arguments. This is further motivated by (Nikolaev et al. 2023)'s suggestion for analysis of adverbials in NLP contexts through the use of Frame Semantics because of the ability to capture lexical ambiguity and other inferences characterized by the semantic frame element.

As this experiment aimed to predict the most likely Semantic Frame an adverbial might take as complement given its phonological form, the use of a classification algorithm is especially prudent. In this experiment, I used three CLASSIFICATION algorithms to determine a model which most effectively made such predictions based on word phonological form. Models took vectorized phonological forms as described above and semantic frames of collocates as input. This was done with the goal of building a model that could either confirm or contradict Hamano (1998)'s system of sound symbolism in Japanese. In accordance with Hamano's model, we should expect different phonemes within a sound symbolic root to contribute together to the overall meaning of the word, but their dependence is unclear. The three algorithms in use in this experiment were NAÏVE BAYES, RANDOM FORESTS, and RECURRENT NEURAL NETWORKS.

A Bayesian model provides a classification model that assumes that attributes are conditionally independence, maximizing the posterior probability $P(y|\mathrm{x})$ of each class $y$ given object $\mathbf{x}$ (Webb 2017). Naïve Bayes models also tend to have high computational efficiency and low variance (ibid.), which along with their simplicity makes them an attractive possibility for a first model to investigate. The effectiveness of this model in comparison to other models may further uncover any dependence between factors: indeed due to phonemes having multiple distinct features, some dependence and relation appears to exist between distinctive features. As such,

I predict that the Naïve Bayesian model will not perform as well as the other models described below.

Random Forests are an ensemble learning method often used in classification, where a number of random decision trees are constructed and then combined to be used for classification—such trees can even be used to deep neural network-like structures while still being easily interpretable(Wang et al. 2018). Because of their cross-domain applications, Random Forests have become largely popular as a tool for classification tasks, being used in other linguistic studies on ideophones like in Takayama et al. (2019) and Kilpatrick et al. (2023), which found them to be a particularly efficient and effective classifier in predicting a Pokémon's evolutionary state. However, their instability and lack of smoothness remains concerning (Altmann et al. 2010)

Recurrent Neural Networks present themselves as a third, potentially interesting for this NLP problem. A type of deep-learning Neural Network (NN), RNNs utilize time-unit delays to deal with sequential data and as such are highly useful for tasks in NLP like speech and audio recognition, and translation (Rezk et al. 2020). With impressive computational ability (Siegelman 1999, as cited in Sakurai 2008), RNNs provide a helpful model for the ideophone problem, since ideophones almost always precede their verbal complements (Akita 2013) and as such form a sequential unit. As a deep learning algorithm, they have the added ability of making predictions based on complex and non-linearly related data, like the sound-symbolic systems Japanese and Korean appear to have.

In this experiment, adverbials were grouped into three categories: non-auditory mimetics, auditory mimetics, and non-mimetic adverbials. I further grouped mimetics into two classes: *giongo* and *gitaigo*. *Giongo*, reflective of their frequent English translation as 'onomatopoeia', have higher iconicity than their non-auditory counterparts (Akita 2016). I predict that experimenting solely on auditory phenomenon describing ideophones should yield a larger correlation between word and semantic frame, and experimentation on non-auditory phenomenon describing ideophones should yield lower, but non-negligible correlations. Where appropriate, non-auditory and auditory mimetics were instead grouped together under a composite category and

compared with non-mimetic adverbials.

### 3.1.1 Materials

Only CVCV root mimetics were considered in this study, which (Hamano 1998) describes as the typical root form of a mimetic adverbial. This was done as in Akita (2012), with small modifications: if such mimetic was particularly common or contributed to a phonological contrast poorly attested in the corpus then it was included for consideration in this study.

To obtain these mimetics, this thesis sourced data from Kimi Akita's 2012 conference paper, obtained through personal correspondence. The data investigated verbal and nominal framing of various ideophones and included a set of 56 ideophones and 57 non-ideophonic vocabulary items that exhibited the strongest correlation to a specific semantic frame. In addition to the ideophones sourced from Akita, I collected a total of 606 other unique mimetics, exclusive of orthographic variations, though these were included in the parsing algorithm as well.

The remainder of the 606 mimetic adverbials were sourced from one of two other sources. Both of these were Japanese onomatopoeia dictionaries: *Nihongo Onomatope Jiten: Giongo, Gitaigo 4500* (Japanese Onomatopoeia Dictionary: *Giongo, Gitaigo* 4500) and the *Dictionary of Iconic Expressions in Japanese* in Japanese. *Nihongo Onomatope Jiten* is a dictionary containing thousands of examples of Onomatopoeia in Japanese, sorted alphabetically and by usage type, with mimetics listed under the verb or predicate with which they may carry related meanings (*Nihongo onomatope jiten : giongo, gitaigo 4500* 2007). Approximately 103 mimetics were sourced from this dictionary. The *Dictionary of Iconic Expressions in Japanese* is a two volume set of several orders of thousands of mimetics in Japanese (*Dictionary of iconic expressions in Japanese* 1996), including their semantic and morphological variations. All remaining mimetics were sourced from this dictionary, with those which maximized the number of contrastive features included in the set used for analysis.

Mimetics from *Nihongo Onomatope Jiten* and the *Dictionary of Iconic Expressions in Japanese* were additionally sorted into two sets, based on their classification as *giongo* or *gitaigo*. Since

no comprehensive list or dictionary exists portioning Japanese mimetics into these categories, I manually sorted the entries in accordance with their listings in either dictionaries. Onomatopoeia listed in *Nihongo Onomatope Jiten* under the meanings *naku* 'Cry' and *iu* 'Say' (the 'Make_Noise' and 'Statement' Frame, respectively) were treates as *giongo* while others as *gitaigo*. Vocabulary sourced from the *Dictionary of Iconic Expressions* were classified *giongo* if their definition contained a description of a "sound", while entries with a description of "manner" were classified as *gitaigo*. Any polysemous mimetics—that is, containing multiple, not clearly connected definitions—were included in both sets of mimetics if at least one definitions would have been classified as *giongo* and another as *gitaigo* were they independent entries.

Non mimetic adverbials were primarily sourced from Akita (2013)'s list of 57 adverbials with the highest correlation with a specific verbal or nominal correlate. An additional 25 non-mimetic adverbials were selected from the list of most frequent occurring adverbials in the National Institute for Japanese Language and Linguistics Balanced Corpus of Contemporary Written Japanese (NINJAL-BCCW) (「現代日本語書き言葉均衡コーパス（*BCCWJ*）」 *Balanced Corpus of Modern Written Japanese* 2024). This corpus includes data sourced from several subcorpora, and included several useful metrics for comparison, including morphological patterns.

All vocabulary-collocate tokens collected were restricted to those adverbials with at least one significant collocate with which they coöccurred more than twice. Collocates were split between verbal and nominal collocates. Finally, each collocate was labelled with the semantic frame that it was identified with on Japanese FrameNet. Example frames are described in Figure 4 for verbal collocates and as in Figure 5 for nominal collocates.[3]

Figure 4: Verbal Collocate Frames

1. Motion Directional
2. Ingestion
3. Self-Motion
4. Possession
5. Body-Movement
6. Statement/Communication
7. Text Creation
8. Moving in Place

[3] See Appendix A for the full set of frame categories considered in this study.

Figure 5: Nominal Collocate Frames

1. People
2. Sounds
3. Observable
4. Excretion
5. Direction
6. Natural Features

### 3.1.2 Procedure

All ideophones and non-ideophones were transcribed, taking account for orthographic variations. Each ideophone was transcribed into Hepburn romanization in order to best match IPA phonemic value to character input. Subsequently, each romanization was transformed into the one-hot encoding schema described in Section 3.1.1.

Then, I conducted a corpus study. This study utilized the Aozora Bunko Corpus (ABC) and the Japanese Speech Corpus of of Saruwatari Labaratory, the University of Tokyo (JSUT) to search for mimetic tokens. ABC is a large, free, corpus comprehensive of available Japanese public domain literature including translations(Morse 2016). At the time of writing, ABC contains 17483 entries, and this amount is updated regularly (*Aozora Bunko Corpus* n.d.). Corpus data is separated morphemically, which simplified mimetic parsing. JSUT is a free, large-scale speech corpus which includes 10 hours of transcribed speech data as well as both vocabulary and sentences from several different domains (Sonobe et al. 2017). The selection of a novel-based corpus as opposed to a newspaper corpus is based data which suggests corpora based in novels perform better in providing a diversity of collocations for mimetics and verbs (Tamaoka et al. 2011 as cited in Akita 2013).

With reference to Akita (2012), collocates were only considered if they were within a certain proximity of the host adverbial. However, unlike Akita's study, this study only considered verb complements following any adverbial: in alignment with Shibasaki (2009)'s finding that mimetics have a markedly strong tendency to be immediately preverbal, or sometimes in ditransitive sentences scrambled before another adverbial. As such, this study only investigated verbal elements promptly following a mimetic, not considering any verb complements in successive clauses, or

those more than three words after the adverbial.

This experiment utilized FrameNet to assist in the categorization of semantic meaning in verbal elements associated with any given adverbial element. Where verbal elements were only associated with a single semantic frame, these were labeled as such. Otherwise, in the case of some ideophones which had multiple collocates, lemmas were only labeled with their most significant semantic frames data, as cited from (Akita 2013). If such data couldn't be ascertained, then a verb was labeled with whichever collocate appeared the most in the set of other extant collocates. In cases with orthographic variations, all variants were labeled with the same frame.

Data was split into an 80-20 train-test split using sklearn's `test_train_split` functionality. The train set of semantic frames were passed into an sklearn label encoder. Data was trained separately on each of a Naïve Bayes, Random Forest (max_ depth=11, n_ estimators=25), and Recurrent Neural Network. Data was fit to each model, and test data was scored on three performance metrics.

1. Accuracy. This metric measured the success rate of correctly labeling a given semantic frame. The formula equivalent of accuracy is written as such.

$$\frac{\#\text{of true positives} + \#\text{of true negatives}}{\#\text{of all predictions}}$$

2. Recall. This metric measured the ability of a model to find all positive instances of a semantic frame—those which actually belong to semantic frame $x$ within the actual label dataset (*Open-Source ML Monitoring and Observability* n.d.). The formula equivalent of recall is written as such.

$$\frac{\#\text{of true positives}}{\#\text{of true positives} + \#\text{of false negatives}}$$

3. F1-Score. This metric evaluates the ratio of recall to precision (how well a model is able to predict positive predictions correctly). (ibid.). The formula equivalent of F1-Score is

written as such.

$$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Finally, for the Random Forest model, I ranked feature importance based on of Altmann et al. (2010)'s Permutation Importance (PIMP) algorithm as used in Kilpatrick et al. (2023). Altmann's method, calculates the increase due to randomization in Out-of-Bag (OOB) Error (ibid.), a measure of "observations in the original data set that do not occur in a bootstrap sample" (Cutler et al. 2007). This algorithm involved normalizing $p$-values based on tests of random permutations of original data before finally a significance result for each feature is returned rather than for the entire model (Kilpatrick et al. 2023) This is a helpful metric for understanding exactly what features are contributing to the classification and are important to the model. However, feature importance does not communicate directionality (ibid.). On models with large numbers of categories, performing a permutation algorithm to determine importance reduces alleviates bias and improves interpretability (Altmann et al. 2010), a helpful trait for nonlinearly correlated dataset with many features like this one.

### 3.1.3 Experiment Ia

Following analysis done in the three models in Section 3.1, I conducted a human validation experiment on Japanese Onomatopoeia with native Japanese speakers for the purposes of comparing machine predictions with human judgements. This experiment was considered "not human research" by Harvard University's Institutional Review Board (IRB) and therefore the approval requirement was waved.

#### Participants

Participants were selected through social media and street interviews. Before conducting street interviews, a small number of pre-arranged participants were recruited via story posts on Instagram. Two native speakers in Japan responded to my invitation to participate in a study, and one

further speaker was recruited to be interviewed through personal correspondence. In addition to the three preselected participants, I conducted a series of street interviews on Japanese speakers across several cities in the *Tōhoku* region of Japan as well as Sapporo and Tokyo. In total, 15 participants were selected for interview. Personal, identifying characteristics of participants were not collected.

**Materials**

The features which contributed most to the semantic categorization of a word's collocates' frame were compiled using the model constructed in Experiment I. An equal set of ideophones and non-ideophones which exemplified these features were compiled including 1 nonce ideophones and 1 nonce non-ideophone for each feature. Interviews were conducted in person and recorded for audio for later analysis, although all audio recordings were later discarded.

**Procedure**

Interviews were conducted with consultants between January 9, 2024, and January 21, 2024. Scheduled interviews with prearranged consultants were held at a local café, while impromptu interviews with street-intercepted consultants took place in various public spaces during daylight hours, particularly within public parks. To ensure participant suitability, each individual completed a pre-study validation survey, confirming their native proficiency in Japanese and literacy. If potential participants were either non-native speakers or illiterate in Japanese, they were excluded from this study. Consent was obtained for each participant verbally, in Japanese. All participants were told that their participation was completely voluntary and that they could stop the interview at any time.

Each participant was presented with a series of items from the test set of vocabulary, and asked to present a verb and/or nominal which most naturally extended the meaning of the mimetic word. Participants were then asked to present a sentence utilizing both the mimetic and the verb complement. For nonce vocabulary, participants were asked to imagine the most likely meaning

of such a word and what word might be modified. These words were annotated using Japanese FrameNet resources to determine the semantic frame of the complement each participant most associated with the mimetic. Following the interview, each participant was paid 1000 JPY in compensation.

## 3.2 Experiment II

Experiment II aimed to replicate the results of experiment IV with Korean, focusing primarily on Korean's extant vowel harmony system and quality contrasts in plosives.

### 3.2.1 Materials

Since an existing set of significant Semantic Frames used in collocations did not exist for Korean language data, prior to data analysis it was necessary to use the Korean FrameNet database to tag verbal collocates with their respective semantic Frames. Based on the findings from Experiment I, I only investigated verbal complements in this experiment.

Mimetics were primarily sourced from a list of 350 common ideophone sets in Sohn (1994)'s descriptive grammar of the Korean Language. Ideophones remain understudied in Korean linguistics, and as such I was unable to find comprehensive onomatopoeia dictionaries from which to extract ideophone entries. Including variations due to vowel harmony and consonant alternation (described in Section 2.3.2), I transcribed a total 226 auditory and 819 non-auditory phenomenon describing mimetics when taking into account variations for consonant alteration and vowel harmony effects.

Classification into auditory or non-auditory proved a harder task with Korean data than with Japanese data. Without a clear metric to separate ideophones into auditory and non-auditory events, in Sohn (ibid.)'s list of mimetics, I inferenced the type of an adverbial from its definition. Definitions which were translated into English primarily as English onomatopoeia (e.g 'bang', 'pop', etc.) or which included a description as 'the sound of' something, were sorted into the

auditory list of onomatopoeia. Other phenomena whose descriptions described sounds in English (e.g 'crashing', 'clanging', 'screaming') were also sorted into the auditory mimetics category. Mimetics which described an action *with the accompaniment* of a sound, were sorted into the non-auditory category as the primary focus of the mimetic was the manner of action. This subgroup of mimetics and any other mimetic which was difficult to place into one category or the other were manually confirmed with at least two Korean consultants about whether the word primarily described an auditory event, or a manner of action. Words that were polysemous were included in both lists of mimetics.

I also investigated 100 non-mimetic adverbials, selected correspond to similar definitions already included in both sets of mimetic adverbials. I focused on including manner adverbials, with frequency, degree, or duration adverbials excluded because of the difficulty of existing models in understanding their nuances (Nikolaev et al. 2023), and their ability to combine with a wide range of complements. Such adverbials were only included when necessary to account for particularly uncommon feature contrasts.

I searched monolingual Korean corpora for tokens in mimetic adverbials. However, Korean had no large-scale, available corpus of transcribed literature to perform analysis on when compared with projects like the Aozora Bunko Corpus used in Experiment I. Some large-scale corpora have had development in the past, like the Korean National Corpus (Sejong Corpus), consisting of 60 million words, half of which were sourced from sentences in books (H.-g. Kim et al. 2007). However, the Sejong Corpus, and several other novel-based corpora are no longer available or accessible online. This experiment resorted instead to three web-based corpora from the Leipzig Corpora Collection, a project maintained by Leipzig University. The major text sources for these two corpora were news websites and randomly accessed websites from the world wide web, with "further sources [...] used as input material for new corpora. This includes corpora based on movie subtitles, chat room texts, (and) Twitter messages" (Eckart and Quasthoff 2013). This experiment used the 2019 Korean Web Corpus, 2021 Wikipedia Corpus, and 2022 Korean News Corpus. In sum, the number of sentences used as corpus search data was three million, with each

individual corpus contributing exactly one million sentences.

### 3.2.2 Procedure

Each mimetic was manually transcribed from Sohn (1994)'s common ideophone list converted into Hangul for direct comparison between corpora tokens and ideophones., taking into account all variations for harmony or plosive quality were transcribed. Using regular expressions, I searched the corpus for every instance of a given adverbial, limiting myself to those instances which followed a whitespace or were the initial part of a string to avoid selecting strings which were actually embed components of a larger word. This was especially needed for the few necessary-to-consider monosyllabic mimetic adverbials like *thok* which was misselected unless limited to those cases following a whitespace. Then, any morphological endings were tagged alongside the mimetic itself. Corpus data from the Leipzig Corpora is tokenized such that whitespaces separate units termed INTEGRATED EOJEOLS—A lexical unit used in Korean text processing that represents syntactic units which may be separated by whitespaces (Choi et al. 2021). *Eojeols* represent a 'syntactic word' or 'polymorphemic word' which can be separated into a lemma stem and postpositional or other morphemic endings(J. Kim and Nam 2018).

In this experiment, I collated *eojeols* that began with any member of the set of investigated adverbials. I stored these *eojeols* along with collocate verbs in the immediate complement, as done in Experiment I. I separated the stem from any morphological endings, such that I could train a classification model to predict what morphological ending a given mimetic is mostly likely to use.

Using Open-Korean-Text, an open-source tokenizer for Korean originaly developed for use on Twitter posts, I extracted the underlying morphemes for each of the endings collected. Because Korean is a highly inflecting language, this was needed to normalize the possible forms of morphological endings and make classification both easier and more consistent.

I followed the process described in Experiment I to create a phonological feature vector encoding the binary presence of features in each of the phonemes within the adverbial. Finally, verbal collocates were tagged with their most significant or frequent semantic frame using Ko-

rean FrameNet. Once again, a Naïve Bayesian, Random Forest, and RNN classifier was trained on the phonological data with morphological ending and collate as labels in two independent sub-experiments. For the Random Forest model only I also considered any secondary semantic frames for verbal collocates in a multi-labelling classification problem. I concluded by once again calculating feature significance.

### 3.2.3 Experiment IIa

After training the models in Experiment II, I conducted a similar human validation and elicitation experiment to that done in Experiment Ia. The requirement for approval from Harvard University's Institutional Review Board (IRB) was waved for this experiment.

#### Participants

Participants were selected through advertisements and word of mouth at Seoul National University in Seoul, South Korea. Flyers were distributed advertising linguistic consultation for bilingual Korean-English speakers at Seoul National University. Ten native Korean speakers were selected, and no personal information about these speakers was collected except for contact information for followup on the interview if necessary.

#### Materials

Once again, the features which contributed most to the semantic categorization of a word's collocates' frame were compiled. An equal set of ideophones and non-ideophones which exemplified these features were compiled including 1 nonce ideophones and 1 nonce non-ideophone for each feature. Interviews were conducted in person and recorded for audio for later analysis, although all audio recordings were later discarded.

**Procedure**

Consent was obtained for each participant verbally, in English, prior to conducting the study. All participants were told that their participation was completely voluntary and that they could stop the interview at any time. Participants were compensated with ₩10,000 South Korean Won for a 30 minute interview at the completion of the interview.

Participants were given one of three randomized questionnaires aimed at eliciting a specific feature for comparison. These questionnaires isolated one of three variables, controlling all other phonological features of the words included in them.

- Plosive quality: Tenuis/Plain, Aspirate, and "Tense"
- YIN-YANG quality of vowels
- Place of articulation: Labial, Alveolar, Velar, and Alveo-Palatal

Participants were given two tasks. The first task involved sentence making where they were told to use an ideophone in a sentence. Any verbal or nominal collocates were written down and classified in accordance with the FrameNet classification system. If participants were unable to use the word in a sentence or could not think of a context in which it made sense, the word was skipped. Both real and nonce vocabulary were considered, with real ideophones sourced from the set of vocabulary used to train or test the model in Experiment III.

# Chapter 4

# Experiment I: Japanese Ideophones

Of the three models, the Recurrent Neural Network (RNN) and Random Forest performed best, perhaps unsurprisingly as previous experiments on mimetic adverbials in Japanese have proven Random Forests to be effective learners of relationships between sound and meaning.

Bayes performed particularly poorly on the composite set of mimetics, with much lower accuracy, precision, and recall than either of the other two classifiers. Indeed, in the confusion matrix in Figure 6, the model suffered from high entropy: the model was most likely to assign a seemingly random label to one of the rarer frames (in the category labeled 'other'), and little clear pattern could be ascertained. This is made clear when comparing performance metrics between the three classification models on the composite mimetic dataset, as in Table 6.

Table 6: Comparison of Classifier Model Performance on the Composite Set of Mimetics

| | Naïve Bayes | Random Forest | RNN |
|---|---|---|---|
| Accuracy | 0.13 | 0.50 | 0.51 |
| Recall | 0.25 | 0.20 | 0.16 |
| F1-Score | 0.12 | 0.22 | 0.18 |

Figures 6 and 7 make clear that RNN and Random Forest classifiers performed the best of the three. Bayesian models in particular struggled with Accuracy and Recall, with large amounts of noise in the predictions made, and infrequent frames. To make comparisons between adverbial categories in the context of different collocates more consistent, for the remainder of this chapter,

Figure 6: Experiment I: Naïve Bayes Confusion Matrix on Composite Set of Mimetics

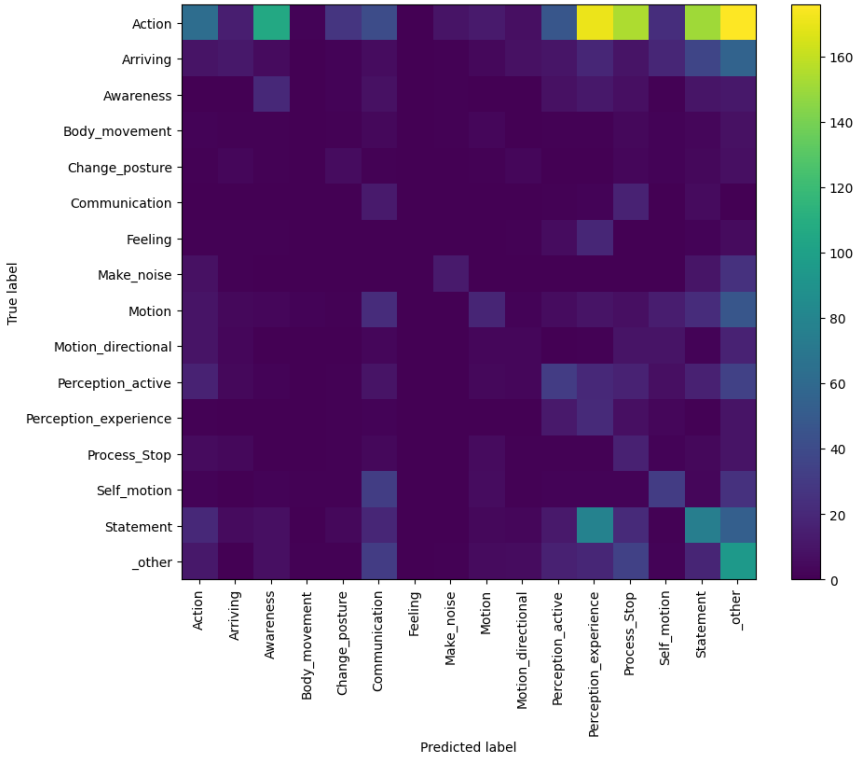
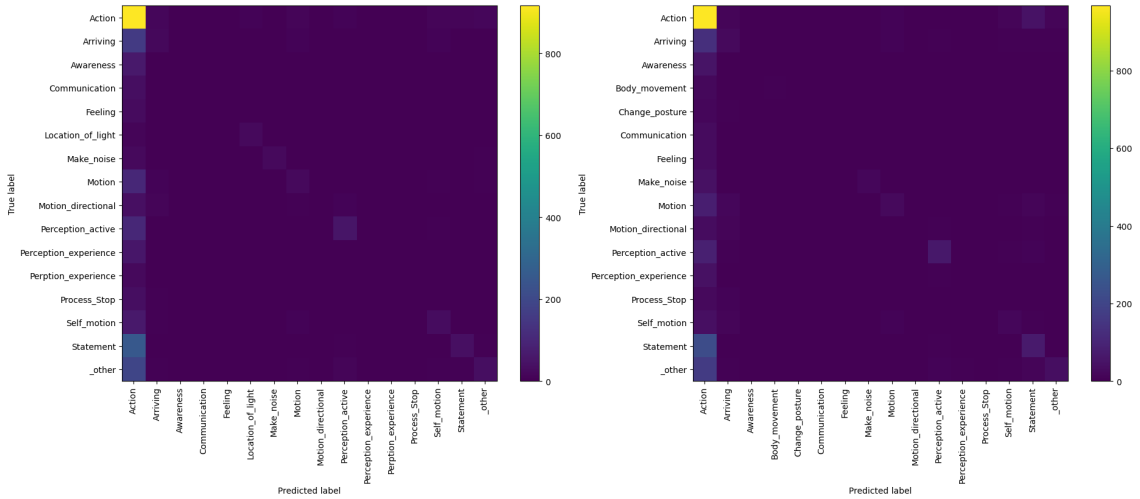



Figure 7: Experiment I: Random Forest(Left) and RNN(Right) Confusion Matrices On Composite Set of Mimetics



data will be limited to being sourced from the Random Forest model unless expressly mentioned in order to compare model performance on certain metrics.

## Hyperparameters

Accuracy and loss calculations on the initial RNN experiment ran with 20 epochs for training, yielding mixed results. Accuracy remained low for data from the written corpora, although with an increase in epochs improved. This initial version of the experiment raised several questions about the iconicity of the ideophones used in the experiment.

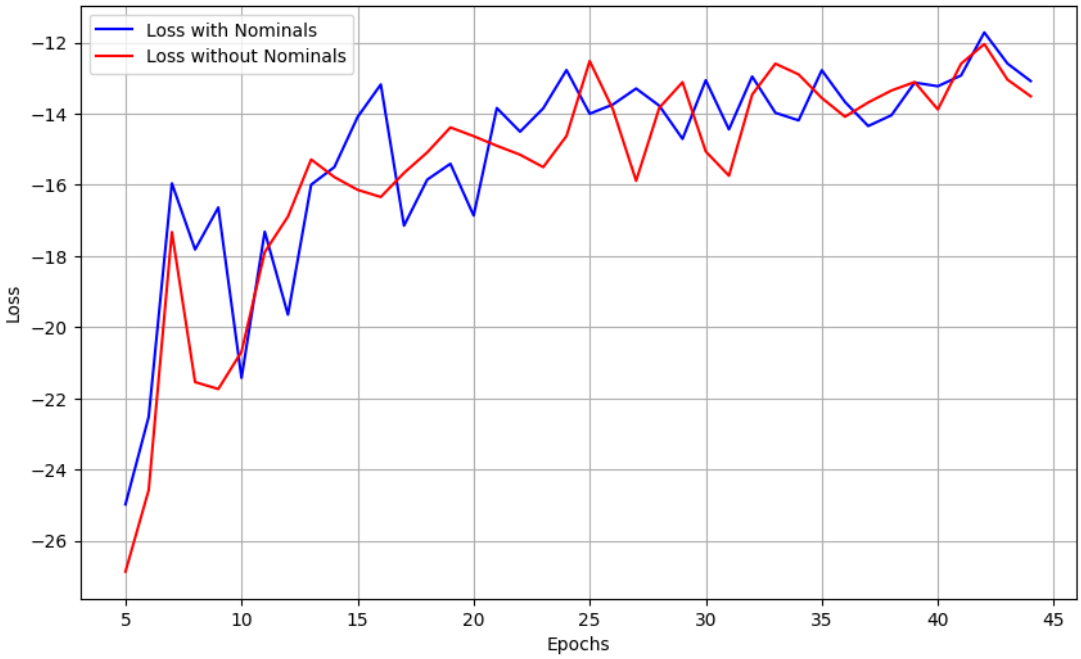
Metrics were further constrained by the presence of the mimetic tokens collected from the two corpora. Of the 664 searched mimetic adverbials, only 279 unique mimetics were collocated with at least a single semantic frame more than three times. Table 7 shows the most frequent collocates found in the corpus search conducted. 'Action' was by far the most frequent semantic frame, with more than three times the number of occurrences as the second most frequently occurring frame, 'Statement'. This is likely from the frequent use of *suru*, labeled 'Action' in Japanese FrameNet, as a light verb for both attributive and predicative use with mimetics and other adverbials and nominals in the abence of a specific verb. Only 63 of the total unique mimetics read did not have a single instance collocated with 'Action', whereas every other semantic frame only collocated with at most one-half of the total read mimetics.

Table 7: Most Frequest Semantic Frames in the Combined Corpora

| Semantic Frame | Number of Occurences | Unique Mimetic Correlates |
|---|---|---|
| 'Action' | 5100 | 216 |
| 'Statement' | 1565 | 138 |
| 'Arrival' | 977 | 114 |

Nominal collocates did not seem to meaningfully contribute to the performance of the model: Neither accuracy nor loss were significantly affect even when not considering nominals. This finding is partially inline with Akita (2013), who found that mimetics were especially likely to "form significant collocates with verbs, although they do have significant nominal collocates as

Figure 8: Percent Change in Loss Between First and Final Epoch




well". Given this finding, for subsequent trials in this experiment, nominal collocates are excluded from the set of semantic frames to be considered.

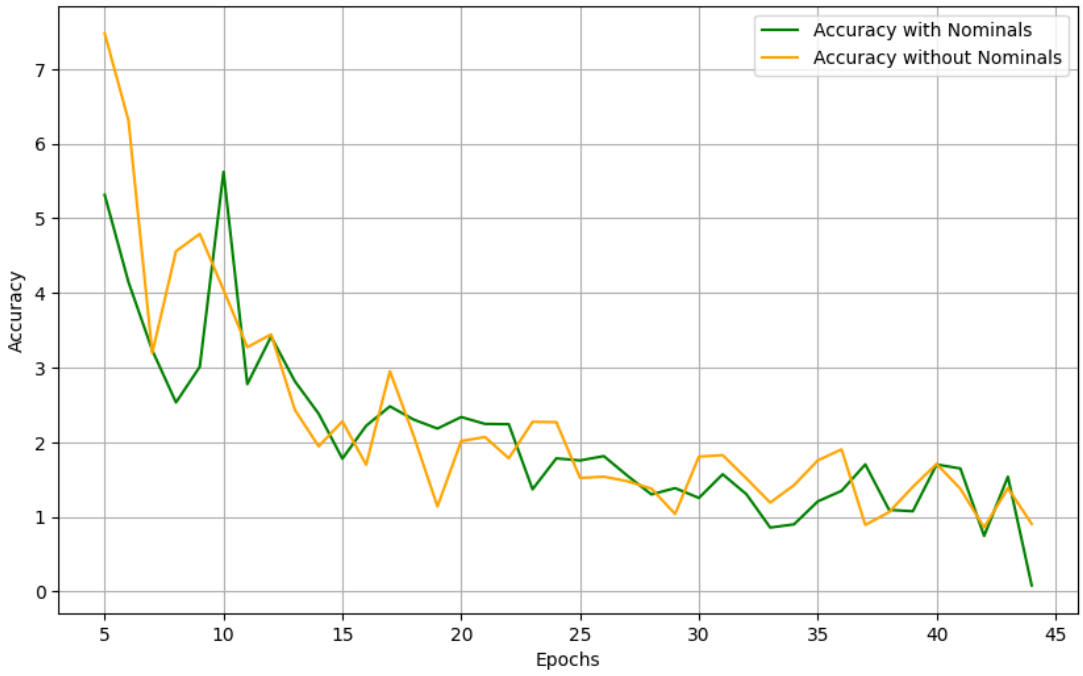
Figures 8 and 9 shows the change in Loss and Accuracy from first to final epochs for an increasing number of epochs with and without nominal collocates. As expected, both accuracy and loss improve sharply until the 15th epoch before tapering off with little improvement in accuracy, although loss continues to improve for a subsequent small increase in epoch count.

Mimetics were further separates by even type, and metrics considered for both subclasses of mimetics. Auditory ideophones (*giongo*) showed a much stronger correlation with a particular semantic frame, with improved accuracy and loss when compared with that of the composite set.

Consistently across all improvements to the RNN, accuracy on mimetic adverbials remained higher than that of non-mimetic adverbials, while recall remained low for both categories. Breaking the mimetic category into auditory versus non-auditory revealed a significantly improved accuracy, but a worse recall: This recall was shared with non-mimetic adverbials which had the lowest accuracy and recall across all three categories.

All three categories scored relatively low in F1-Score, although the F1-Score for non-auditory

Figure 9: Percent Increase in Accuracy Between First and Final Epoch




mimetic adverbials was more several times larger than that for non-mimetic adverbials indicating that while all three classes struggled with either recall or precision, the non-auditory mimetics performed best in correctly labelling relevant elements. Tables 8 - 10 display these metrics measured in aggregate as well as for the two best performing frames, with all values given rounded to the nearest one-hundredth.

Table 8: Preliminary Metrics of Different Adverbial Categories in Japanese

| | Accuracy | Recall | F1-Score |
|---|---|---|---|
| Auditory Mimetics | 0.47 | 0.12 | 0.11 |
| Non-Auditory Mimetics | 0.43 | 0.14 | 0.13 |
| All Mimetics | 0.44 | 0.15 | 0.15 |
| Non-Mimetics | 0.27 | 0.03 | 0.02 |

## Overfitting to 'Action' and 'Statement'

One frame in particular performed well in comparison to other semantic frames. The 'Statement' semantic frame, associated with the verb *iu* and other verbs of communication had a particu-

larly strong accuracy and recall in all three groups of adverbials. F1-score was higher for auditory ideophones, but still low for all three types of adverbials, reflective of the value in the composite calculation.

Across all three adverbial types, accuracy was much higher than in the composite metric, with the auditory mimetics performing the worst. However, auditory mimetics performed better in recall, and the best among the three in F1-Score, indicating higher precision and better ratio between recall and F1-score than the other two types. Non-Auditory Mimetics performed slightly better than non-mimetics in accuracy, but far outperformed non-mimetics in recall, and F1-score, which non-mimetics both performed particularly poorly compared to the other metrics.

Table 9: Experiment I Evaluation Metrics for 'Statement'

| | Accuracy | Recall | F1-Score |
|---|---|---|---|
| Auditory Mimetics | 0.81 | 0.26 | 0.33 |
| Non-Auditory Mimetics | 0.90 | 0.14 | 0.22 |
| All Mimetics | 0.89 | 0.24 | 0.30 |
| Non-Mimetics | 0.87 | 0.08 | 0.10 |

Likwise, the 'Action' semantic frame, associated in this study solely with the light verb *suru* and its orthographic variants also particularly strong. Recall and F1-Score were both higher for both mimetics when compared to non-mimetics. However, the F1-score was still low enough for the mimetic adverbials, given their near 1.0 recall, that it indicates a somewhat low precision, compared to the recall. None of the three types of adverbials had a particularly high accuracy for 'Action' Semantic Frame objects, reasserting the low-precision of model for predicting the 'Action' frame, and a potentially overfitting to 'Action' in the train set.

Table 10: Experiment I Evaluation Metrics for 'Action'

| | Accuracy | Recall | F1-Score |
|---|---|---|---|
| Auditory Mimetics | 0.28 | 0.92 | 0.63 |
| Non-Auditory Mimetics | 0.37 | 0.86 | 0.61 |
| All Mimetics | 0.35 | 0.89 | 0.58 |
| Non-Mimetics | 0.21 | 0.93 | 0.52 |

Without the 'Action' frame, the number of tokens read from the dataset was greatly reduced in the composite dataset: 8,070 instances were read without including 'Action' as a semantic frame, while 13,170 were read with 'Action' included in the dataset. Removing 'Action' from the set of collates to read largely improved the precision of the trained model for all three classification algorithms. Further, the metrics for mimetics, both auditory and non-auditory substantially exceeded that of non-mimetics with 'Action' semantic frames excluded. Mimetic adverbials still under-performed in accuracy measurements compared to their performance when the 'Action' frame was included in the train data set, however.

Table 11: Experiment I Composite Evaluation Metrics Without 'Action' Labelled (RNN)

| | Accuracy | Recall | F1-Score |
|---|---|---|---|
| Composite Mimetics | 0.47 | 0.26 | 0.22 |
| Non-Mimetic | 0.24 | 0.06 | 0.04 |

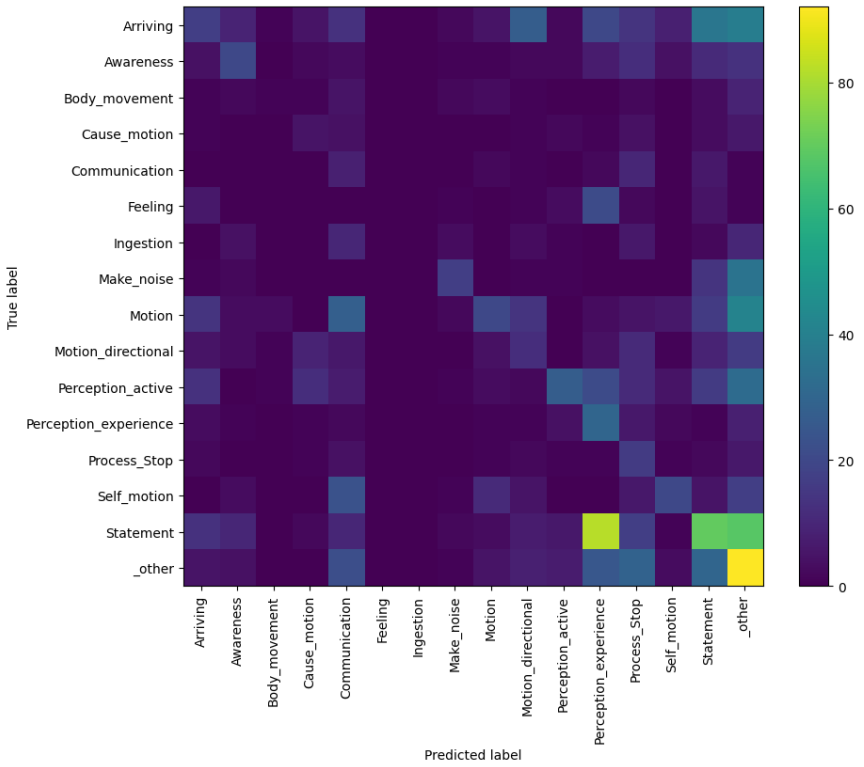
All three metrics were particularly high for the 'Statement' semantic frame when the 'Action' frame was excluded from the dataset. Across all three categories of adverbials, accuracy was greater than 50%. Mimetics were effective in Recall and F1-Score, while non-mimetic adverbials struggled with precision.

Table 12: Experiment I 'Statement' Evaluation Metrics Without 'Action' Labelled

| | Accuracy | Recall | F1-Score |
|---|---|---|---|
| Auditory Mimetics | 0.51 | 0.90 | 0.63 |
| Non-Auditory Mimetics | 0.69 | 0.73 | 0.44 |
| Non-Mimetics | 0.52 | 0.70 | 0.36 |

Furthermore, removing 'Action' from the set of semantic frames reduced the amount of 'noise' in predicted-vs-true labeled frames and overfitting to a single frame—that is, the ratio of correctly-labelled and incorrectly-labelled data improved, at least for the top 15 most common semantic frames in the set of searched collocates. Noise remained somewhat high for infrequent frames. This tendency did not hold true in the Bayesian model, where overfitting remained a problem, and significant noise due to mislabeling for even the top 15 most frequent frames occurred. Compare the confusion matrix in Figure 10 to that of Figure 11.

Figure 10: Naïve Bayes Confusion Matrix for Top 15 Semantic Frames, Excluding 'Action'




Only when removing 'Action' from the set of searchable collocates did the RNN and Random Forest differ meaningfully in any metric. As seen in the confusion matrices in 11, infrequent mimetics categorized as 'other' were mislabelled less often, and only one label, 'Statement' appeared overfit in the RNN model when compared to the Random Forest Model. This translated to performance metrics, where the accuracy and recall were all improved by a non-trivial amount in the RNN model, with consistent improvements of about 0.03 and 0.02 respectively, when compared to the Random Forest.

Perhaps most interesting is what resulted from the investigation of only the two most common Semantic Frames: 'Action' and 'Statement'. In doing so, found marked improvement of the model in terms of both precision and accuracy for the RNN model.

Bayes showed promise in its large count of true 'Action' judgements, while also correctly labelling 'Statement' frames at a relatively high rate. Misclassification remained a problem, with the number of true-label 'Action' frames predicted to be of 'Statement' nearly equalling half that of

Figure 11: Confusion Matrices for Random Forest(Left) and RNN(Right) Classifiers for Top 15 Semantic Frames, Excluding 'Action'

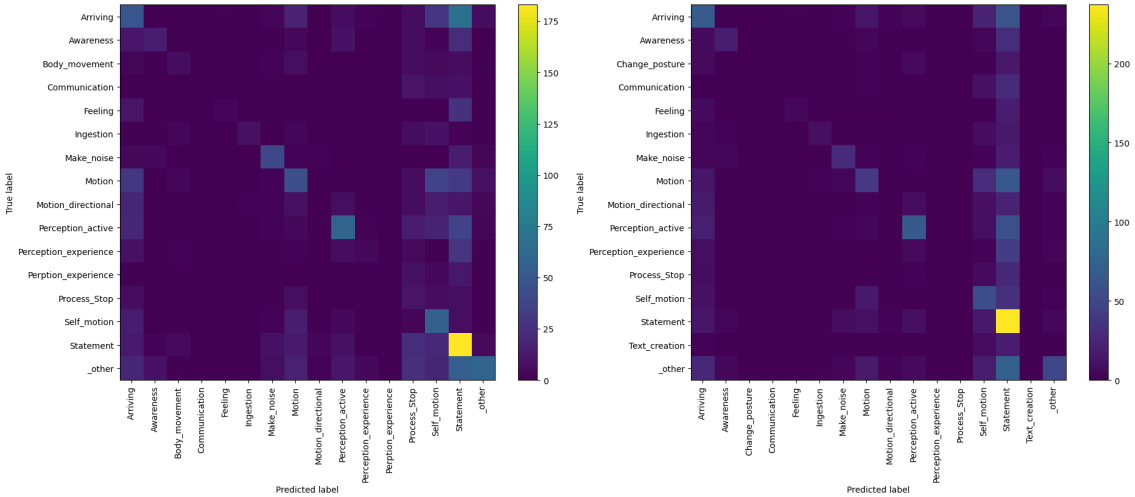



Figure 12: Naïve Bayes Confusion Matrix on Statement-vs-Action

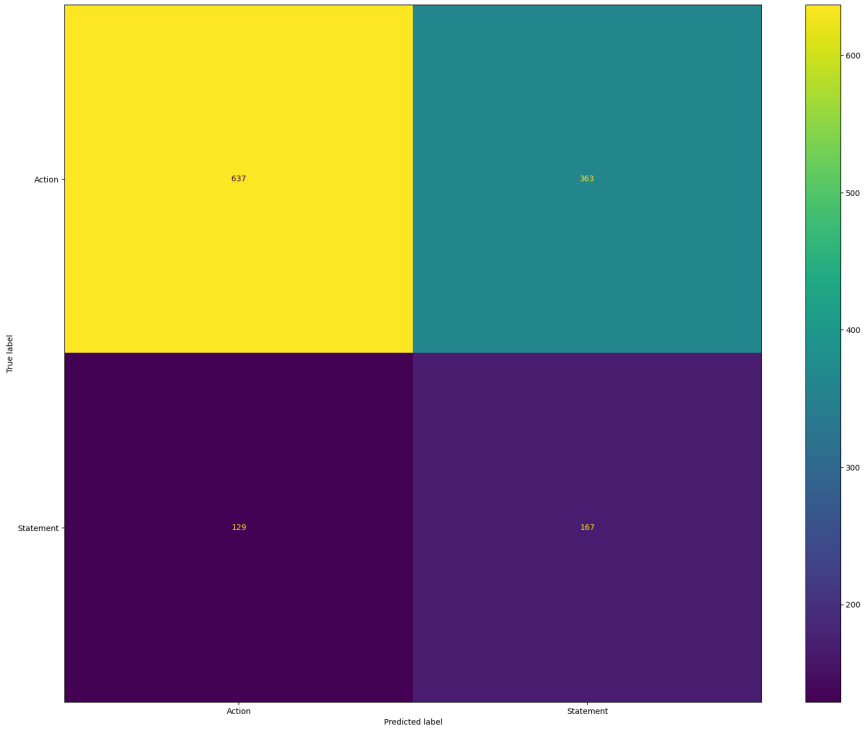
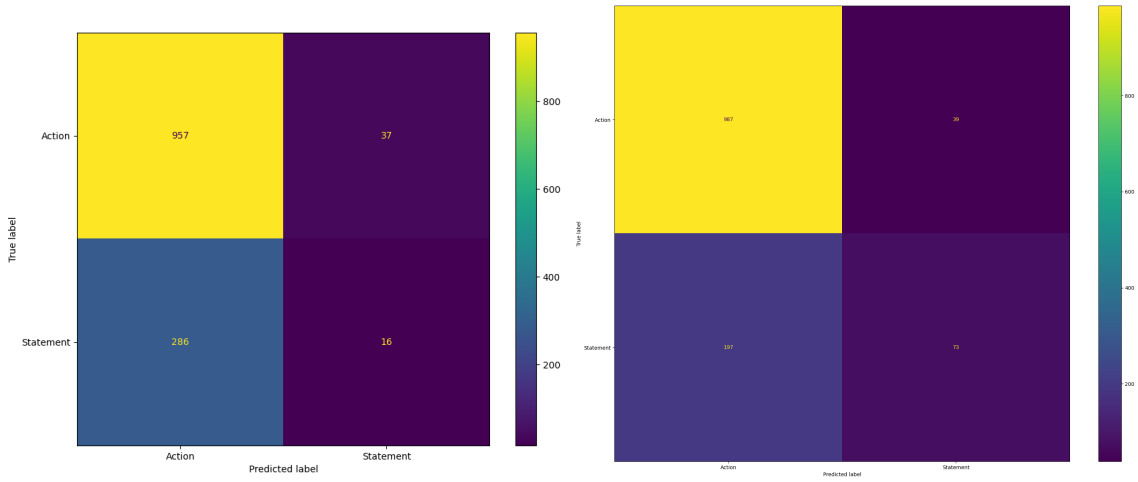


Figure 13: Random Forest(Left) and RNN(Right) Confusion Matrices on Statement-vs-Confusion




the correctly predicted 'Action' frames. The Random Forest model in Figure 13 didn't perform much better. While nearly eliminating the number of incorrectly labeled 'Action Frames', this model defaulted to labeling anything as 'Action', with nearly all true-label 'Statement' Frames predicted as 'Action'. The best performing model in this trial was the RNN, which performed equally well as the Random Forest in false 'Statement' label predictions, while also greatly improving the classification of true-label 'Statement' predictions.

## Feature Importance and Distribution

Each distinctive feature was collected and its distribution modeled with a stacked bar graph. Most features were either equally distributed, or distributed in favor of the non-auditory mimetics, possibly because of the size of the auditory mimetic dataset. The distribution of the distinctive features between auditory and non-auditory mimetics is shown in Figure 14[1].

Using a permutation importance algorithm as opposed to an impurity-based one when calculative feature importance served to alleviate bias and handle the large number of features on which the model was trained[2]. In Japanese Auditory Mimetics, only a few features were signifi-

[1] For space considerations the following nonstandard abbreviations have been included. V: 'Voiced', D: 'Dorsal', Lab: 'Labial', H: 'High', F: 'Front', Comp: 'Compressed', N: 'Nasal', Lo: 'Low'

[2] See Altmann et al. (2010) and Kilpatrick et al. (2023) for benefits of using a permutation importance algorithm for many-featured classifiers

Figure 14: Feature Distribution of Distinctive Features in Japanese Mimetics

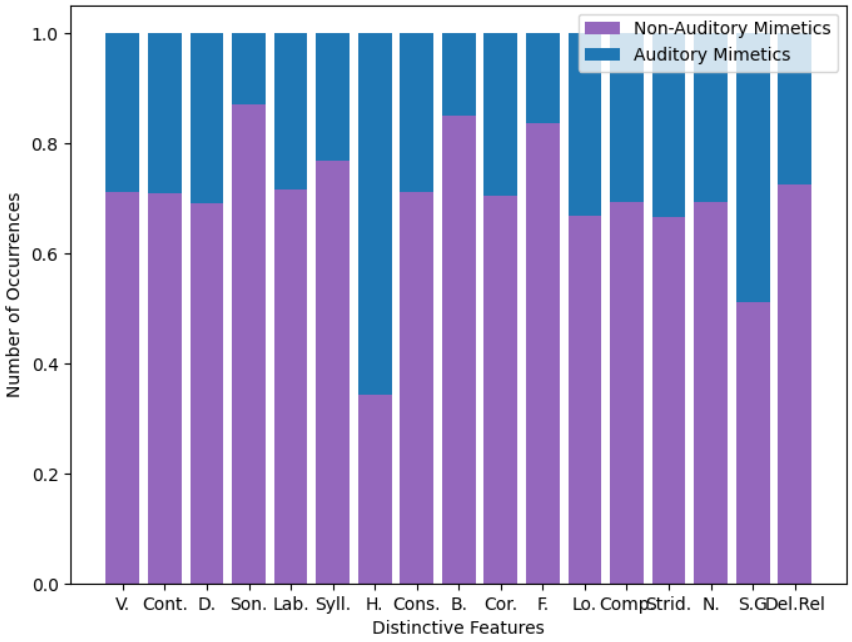



Table 13: Experiment I: Permutated Feature Significance on Train Mimetics, Auditory Mimetics ($n\_repeats = 100$)

| Feature | Significance | Standard Deviation | Example Phoneme |
|---|---|---|---|
| [±Low] | 0.019 | 0.002 | /a/ |
| [±Son] | 0.003 | 0.002 | /r/ |
| [±Cont] | 0.003 | 0.002 | /s/ |
| [±Voice] | 0.002 | 0.002 | /b/ |

cantly correlated, with [±Low] being the only feature with statistically significant correlation to semantic frame classification in auditory event describing mimetics at 1.9%. For non-auditory mimetics, the most significant mimetics with permutated importance within two standard deviations of the mean were [±High], [±Back], [±Son], [±Voice], [±Low] and [±Compressed], the last of which is exclusively associated with the phonemes /ɯ̟/ and /ɯ/.

No specific position within a word contributed to the overall meaning of that adverbial more than another, with feature significance for all phoneme positions no more than one standard deviation from 0.01% in all cases. The only position with importance at least one standard deviation from 0.01% was $V_2$(Position 3) with feature significance 0.61% and standard deviation 0.39%. The *to*-flag representing the presence of the quotative particle *to* had higher feature importance

Table 14: Experiment I: Permutated Feature Significance on Train Mimetics, Composite Mimetics ($n\_repeats = 100$)

| Feature | Significance | Standard Deviation | Example Phoneme |
|---|---|---|---|
| [±High] | 0.019 | 0.003 | /i/ |
| [±Back] | 0.010 | 0.003 | /u/ |
| [±Son] | 0.010 | 0.003 | /m/ |
| [±Voice] | 0.009 | 0.002 | /b/ |
| [±Low] | 0.008 | 0.002 | /a/ |
| [±Compressed] | 0.007 | 0.002 | /w/ |
| [±Continuant] | 0.005 | 0.003 | /s/ |
| [±Dorsal] | 0.005 | 0.003 | /k/ |
| [±Front] | 0.005 | 0.002 | /e/ |
| [±Syllabic] | 0.005 | 0.002 | /a/ |
| [±Labial] | 0.005 | 0.003 | /p/ |
| [±Consonantal] | 0.004 | 0.002 | /t/ |

than any single position at 0.66% but was only one standard deviation away from 0.01% with a standard deviation of 0.36%. This may be implementation error, with *to* represented as a vector of ones if *to* was present and zeros if not. As such, this result may be statistically insignificant.

## 4.1 Experiment Ia

Validation with Japanese speakers revealed a clear difference in the meaning of a mimetic from voicing distinctions. All speakers expressed a easily distinguishable difference in meaning between a voiced-unvoiced pair, with many such differences only being that of intensity, while others expressed entirely different meanings. No other contrast appeared to cause as much clear distinction in the speakers interviewed, although speakers consistently described mimetics with /i/ as describing a smaller, sharper action.

One consultant in particular exclusively paired the light verb *suru* with every single mimetic proposed to them. When asked whether *suru* was the only verb possible for those mimetics, the consultant explained that doing so was perhaps a more casual way of formulating a context for the mimetic, before offering an alternative for some of the mimetics.

Speakers in this validation experiment were very receptive to the nonce words provided to

them. Although sometimes confused initially, several consultants gave rich description of words which they claimed to have never heard before. In one case, a speaker asked to describe a nonce word **bikabika*, described to them as a "fake word, (that they were) told that it had a similar meaning to *pikapika*, but with a slight difference". However, the speaker, while saying they had never heard of this word, gave a rich description of the word's meaning, and arguing that it had a "completely different feeling" to *pikapika*.

In some cases, the most natural verbal collocate used by participants was similar to that used by a mimetic with slight feature differences, but reflective of the intensity shift in mimetic. One consultant offered the examples in (26)

(26) a. 浜辺で太陽は私たちをギラギラと照り付けた

hamabe-de taiyō-ha watashitachi-wo giragira-to teritsuke-ta
beach-LOC sun-TOP 1.PL-ACC MIM-QUOT blaze-PST
"At the beach, the sun was blazing down on us (*giragira*)"

b. キラキラと目を輝かせている

kirakira-to me-wo kagayak-ase-te-iru
MIM-QUOT eyes-ACC shine-CAUSCONJ-PROG
"They made their eyes all sparkly like *kirakira*"

For this speaker, the difference in meaning was only slight: *giragira* represented a "more sparkly" version of *kirakira*. The speaker presented the images in Figure 15 as an example of how the difference in pronunciation affected their perception of "sparklyness". Despite the apparent similarity in meaning, the two mimetics used completely distinct verbs in all the examples provided by this consultant.

## 4.2 Discussion

Results from this experiment confirm several preconceptions of the data regarding relationships between adverbial type and semantic frame of verbal complements. Both auditory and non-auditory ideophones show a stronger correlation with verbal collocates than non-mimetic adverbials. The low recall and precision via F1-score in the composite calculation suggest the iconicity

Figure 15: Speaker Perceptions of *Kirakira*(left) and *Giragira*(right)

of mimetics is lower than previously expected: however inspecting the most frequent semantic frames reveals that a particular frame is over-represented in the data: *suru*. Removing *suru* to reduce bias in the dataset indeed proves that while potentially not purely iconic, mimetic adverbials have a strong correlation between their phonological form and their semantics that is not held by non-mimetic adverbials.

The data presented a classic issue of CLASS IMBALANCE: one class, 'Action' far outnumbered any other frame, with 'Statement' similarly outbalanced. This is inline with expectations: Both *suru* and *iu*, typical verbs for the 'Action' and 'Statement' frames, are used as semantically vacuous extensions of ideophones in attributive use. Surprisingly, even auditory mimetics displayed this imbalance. Although Hamano (1998) indicated that for auditory mimetics, only *iu* can be used in attributive senses, in predicative use *iu* is not paired exclusively with mimetics especially in non-indicative contexts. NINJAL-LWP for BCCWJ provides examples of both *iu* and *suru* as complements for the same auditory mimetic. Compare the examples in (27).

(27) a. 風邪気味らしく、鼻をぐすぐすさせながら、北川は声を潜めた。

kaze-gimi-rashi-ku, hana-wo gusugusu sase-nagara, kitakawa-ha koe-wo
cold-slight-seem-ADV nose-ACC MIM do.CAUS-while Kitakawa-TOP voice-ACC

hisomeru
become.quiet
"Kitakawa, like they had a slight cold lowered their voice, makes their nose go *gusugusu*"

b. 家人は鼻をグスグスいわせ、腰が痛いと泣き言を言い、私はいそいそと台所に立ったのであった。

kajin-ha hana-wo gusugusu iwa-se koshi-ga itai-to nakigon-wo
family-TOP nose-ACC MIM say.IRR-CAUS back-NOM hurt-QUOT complaint-ACC
ii watashi-ha isoiso-to daidokoro-ni taQ-ta no de-aQ-ta
say 1.SG-TOP MIM-QUOT kitchen-DAT set.off-PST EXPL LOC-exist-PST
"(My) family members were making their nose sound *gusugusu* and complaining their backs hurt, and (so) I set off for the kitchen."

Examples like this suggest a strong connection between mimetic and their nominal collocate (in this case, a semantic connection between 'nose' and the sound of sniffling). This is merely coincidence: none of the other corpus examples in NINJAL-LWP for BCCWJ include the morpheme 'nose'. This fact, coupled with my finding that removing nominal collocates from the set to investigate did not significantly affect RNN performance indicates that whatever semantic relationships between mimetics and nominals are not significant.

*Iu* remained abundant as a morphologically distinct verbal complement as the second most frequent complement in all subsets of adverbials. At the same time, *suru* remains steadfastly a more productive way of enabling predicative use for mimetics without introducing a verb that would further limit or modify the semantic content of the mimetic.

In both considering and not considering *suru*, the Random Forest classifier and the RNN performed consistently better. This is in agreement with Kilpatrick et al. (2023) and Takayama et al. (2019)'s findings that Random Forests were "Efficient Learners" of the relationships between sound and meaning. The discovery of RNN as a model that performed similarly well in particular indicates that the use of neural networks may indeed merit research as a potentially effective model for use in ideophone studies. While the Random Forest model consistently performed slightly better across all validation metrics when compared to the RNN, the RNN in particular

was able to eliminate overfitting to many other semantic frames, with the exception of the 'Statement' label. The RNN performed best on data with fewer labels, robustly outperforming the Random Forest model and far outperforming the Bayesian classifier on Statement-versus-Action classification. However, while the RNN may be an apt model for sound-symbolic problems with few categories, a Random Forest model remains one of the most effective models for data with many categories.

When using the Random Forest model, several features stood out as particularly correlated with semantic content in a verbal collocation: the [±High] feature, associated with the high vowels /i/ and /ɯ/ stood out as the most significant for the composite set with significance of 1.9%, potentially due to overpresence of [±High] features in auditory mimetics. Yet this was not a significant feature in auditory mimetics. This suggests that while high vowels are indicative of auditory information in their complements, they themselves do not contribute meaningfully to differentiation between auditory events. One feature stood out for being associated with only two phonemes: [±Compressed], associated solely with /ɰ/ and /ɯ/. Although this feature's significance was just 0.7%, it stood out as one of the most important and significant features in determining the semantic quality of a verbal collate. All of the top six most significant features were those associated with vowels or semivowels, underscoring the importance of vowels in Japanese's sound-symbolic system, although consonants have often been the focus of much research. Successive studies should be conducted to validate this finding, to determine if vowels truly are playing as large of a role in determining meaning as they appear. Unsurprisingly, voicing was also one of the most significant features in determining semantic frame content. This is largely in line with Hamano (1998)'s sound-symbolic system, which posited specifically voicing as a feature contributing to the meaning of the mimetic. This finding suggests that this meaning contribution extends even to collocates.

Yet, no position within a word contributed more to the meaning to the mimetic than another with roughly equal significance across all positions. The presence of absence of a quotative particle *to* had a larger effect on the semantic of the frame of the verbal collocate than any

single phoneme position. This is largely in line with Hamano (1998)'s model which posits every phoneme in a mimetic as carrying a somewhat equal contribution to the overall meaning of the mimetic. However, it suggests that these contributions are not very significant in predicting meaning of complements. This may indicate that any such roles carried out by specific ordering effects contribute instead to particular connotations inherent to a mimetic and not shared across complement structures. Further investigations should be done replicating this kind of investigation with semantic classification of mimetics themselves to verify this finding.

This study was in particular held back by the limited availability of FrameNet data. While the Japanese FrameNet tagging is available online for free, a readily available corpus of the size necessary for this kind of regression is not, and as such dealing with polysemy in particular arose as challenging: many verbal collocates had several possible frames to which they could be assigned to. Without a larger corpus, this experiment could not reliably tell the sense of a given word verb or mimetic. This led to this experiment only investigating a single sense for each verbal element: the most 'significant' frame, as described in Section 3.1.1

Overall, this experiment indicated that mimetics behave differently from nonmimetic adverbials in forming collocations with verbal complements, and that their meanings could be partially predicted from the distinctive features of a mimetic adverbial. Vowels in particular seem to be important in this prediction, with Voicing secondarily affecting semantic content.

# Chapter 5

# Experiment II: Korean Ideophones

I once again investigated the effectiveness of each of Naïve Bayes, Random Forest, and Recurrent Neural Network (RNN) classification algorithms to predict relationships between phonological qualities and morphological endings as well as the semantic frame of any verb complement. Modeling the first type of relationships yielded much better results than the other experiments performed in this thesis.

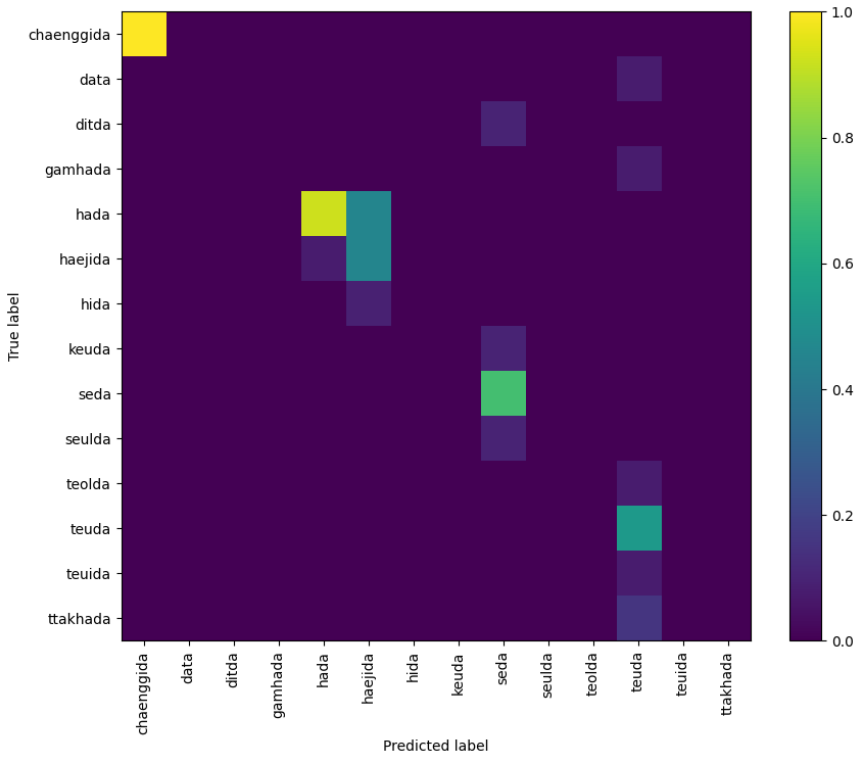
Of the three models that I investigated, the Random Forest Classifiers performed better on all three metrics that I investigated: acccuracy, recall, and f1-Score. The RNN model performed similarly well on recall and f1-Score, indicating that these two classifiers had particularly better precision when compared to the Bayesian model. However, the Bayesian model still performed better than in Experiment I. Table 15 visualizes the differences in tested metrics across the three algorithms.

Table 15: Experiment II: Comparison of Classifier Model Performance on the Composite Set of Mimetics and Morphological Endings

| | Naïve Bayes | Random Forest | RNN |
|---|---|---|---|
| Accuracy | 0.67 | 0.82 | 0.81 |
| Recall | 0.25 | 0.45 | 0.35 |
| F1-Score | 0.2 | 0.36 | 0.33 |

Figures 16 and 17 reflect these metrics in a confusion matrix for endings which occurred more

Figure 16: Experiment II: Confusion Matrix for Top 15 Frequent Verbal Endings in the Composite Set of Ideophones, Naïve Bayes Classifier (Normalized)

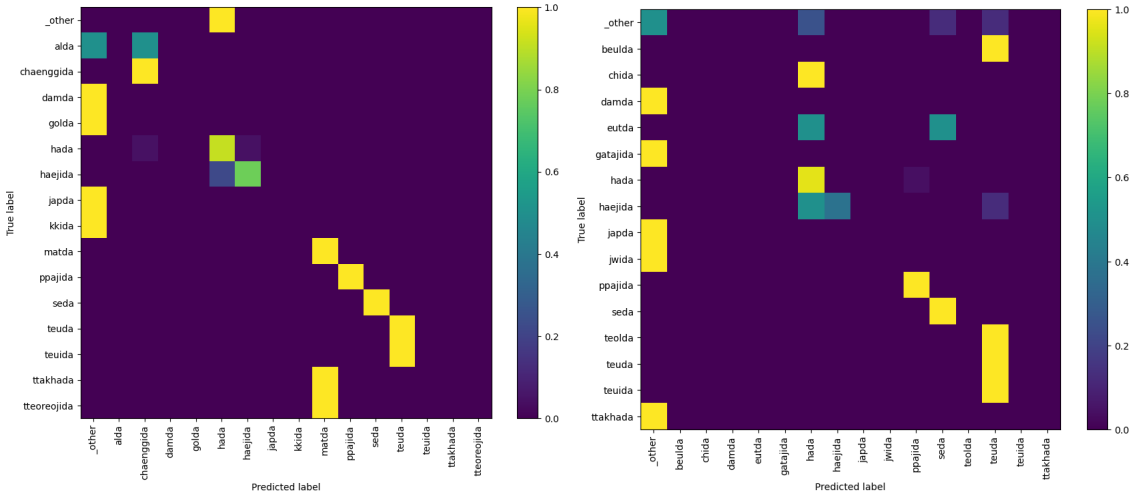



than a single time in the data set. While the Bayes classifier largely was able to limit mislabels, some still persisted, largely restricted to those with *true* labels with a single occurrence. Because of class-imbalance of *hada* data presented in the confusion matrix in the figures in this section are normalized with respect to the *true* values.

Several of these verbal elements were dubious in origin. *Ttakhada* was of particular concern: itself decomposable into the elements 딱 *ttak* 'tightly, resolutely' and 하다 *hada*, in the set of *true* labels it only occurred in a few example sentences, in all of which it was exclusively coupled with the mimetic *ttak*. This was indicative that the underlying verbal element was instead *hada*, affixed to the mimetic 딱딱 *ttakttak* 'difficulty, hard; very tight' which was also included in the set of mimetics to investigate.

As such, following the initial training, I restricted myself to the top seven verbal endings which could be shown to themselves not be made up of any other morphemes, and which were taggable using the Korean FrameNet as containing at least one specific semantic frame. Doing

Figure 17: Experiment II: Confusion Matrix for Top 15 Frequent Verbal Endings in the Composite Set of Ideophones, Random Forest (Left) and RNN (Right) Classifiers (Normalized)




so revealed a much higher precision across all verbal endings.

For the set of investigated non-mimetics, the dataset was extremely unbalanced, with nearly all tokens considered taking *hada* as a morphological ending. This is unsurprising, given that many considered non-mimetic adverbs contained the suffix *-hi*, itself derived from *hada*, as an etymon. The set of adverbs was so unbalanced, that when only considering those morphological ending collocates with more than a single coöcurrence, no verb other than *hada* occurred more than 50 times. Table 16 shows the full set of verbal endings which occur more than a single time with one of the considered non-mimetic adverbials. Note that non-mimetics were particularly unlikely to be followed by a verbal morphological ending, with only 426 instances out of the total 13,769 tokens parsed.

Table 16: Number of Coöccurrences of Verbal Endings in Non-Mimetic Adverbials (Korean)

| Adverbial ending | Number of Coöccurrences |
|---|---|
| 하다 *'hada'* | 335 |
| 차다 *'chada'* | 35 |
| 서다 *'seoda'* | 11 |
| 돌다 *'dolda'* | 9 |
| 오다 *'oda'* | 4 |
| 해지다 *'haejida'* | 4 |
| 이기다 *'igida'* | 4 |

Figure 18: Experiment II: Confusion Matrix for Top 7 Frequent Verbal Endings in Composite Set of Ideophones, Bayesian Classifier (Normalized)

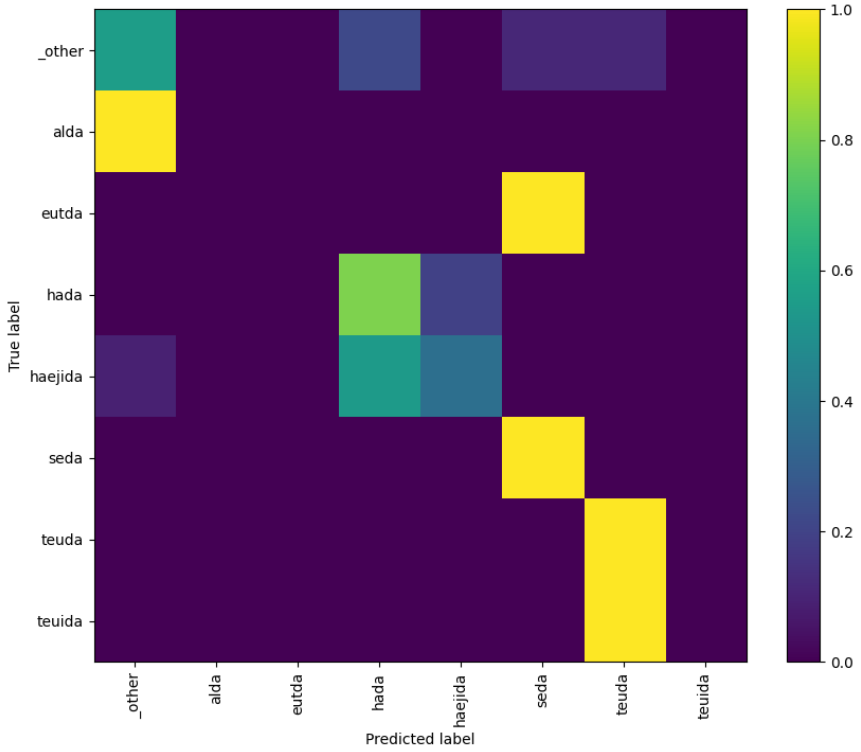



Figure 19: Experiment II: Confusion Matrix for Top 7 Frequent Verbal Endings in Composite Set of Ideophones, Random Forest (Left) and RNN (Right) Classifiers (Normalized)

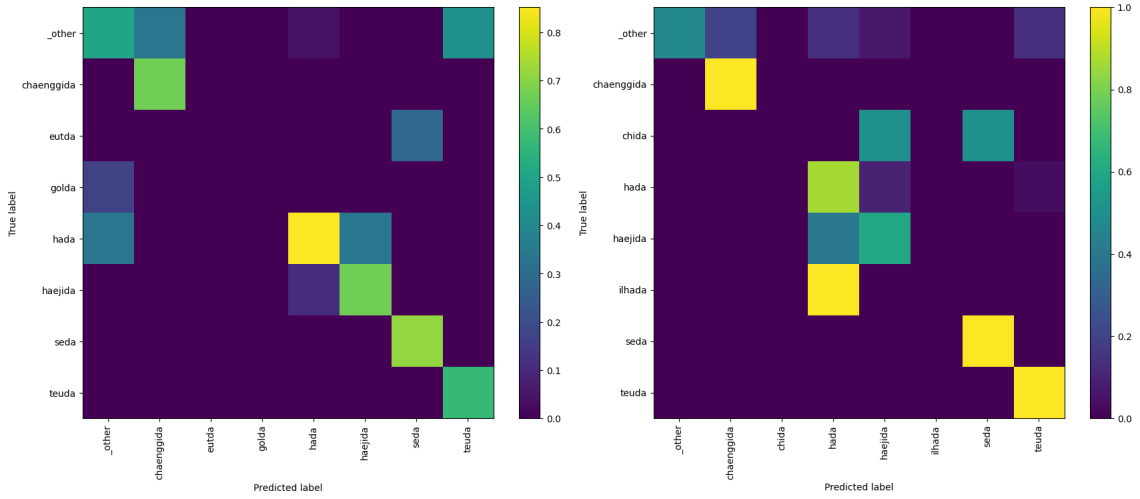


Figure 20: Random Forest Confusion Matrix for Non-Mimetic Adverbials in Korean, Normalized

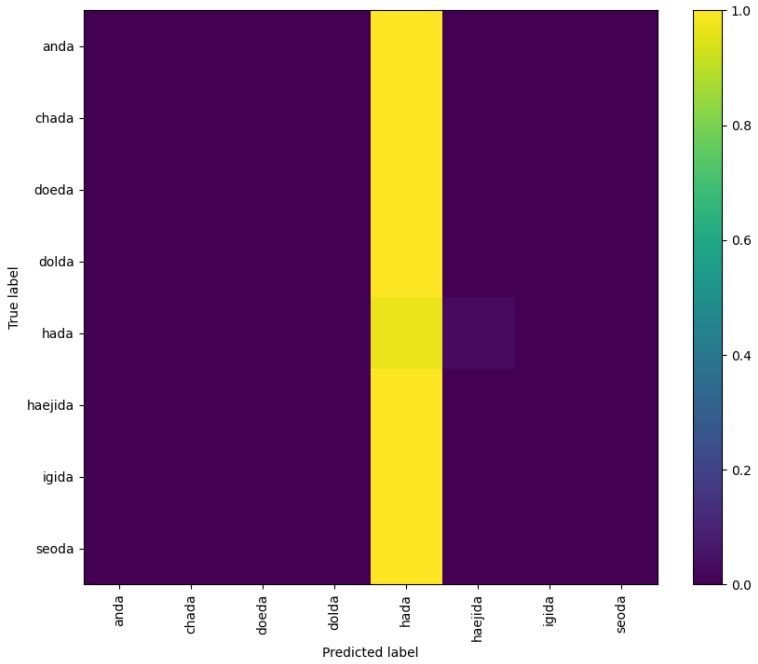



Table 17: Performance Metrics for Non-Mimetic Adverbs in Korean

| | Bayes | Random Forest | RNN |
|---|---|---|---|
| Accuracy | 0.80 | 0.81 | 0.85 |
| Recall | 0.09 | 0.16 | 0.13 |
| F1-Score | 0.11 | 0.15 | 0.10 |

When tested on the three classification algorithms, non-mimetics adverbials produced a deceptively impressive accuracy. However, inspecting the recall and precision inherent to the F1-score revealed that non-mimetics were performing significantly poorer than any mimetic adverbial subset. Even accounting for class imbalance and applying a normalization to the *true* labels, no model was likely to predict any other verbal ending other than *hada* for a single non-mimetic adverbial. This indicates that non-mimetics may behave differently not only in terms of the phonological and sound-symbolic systems which formulate mimetic vocabulary, but are also much more restricted in word choice than mimetic vocabulary with regards to morphological endings.

Within the mimetic adverbials, there was little difference in performance between auditory

Table 18: Performance Metrics for Auditory Mimetic Adverbs and Morphological Endings in Korean

| | Bayes | Random Forest | RNN |
|---|---|---|---|
| Accuracy | 0.58 | 0.73 | 0.69 |
| Recall | 0.19 | 0.37 | 0.29 |
| F1-Score | 0.13 | 0.40 | 0.30 |

Table 19: Performance Metrics on the Random Forest Multi-Label Classifier (Random Forest)

| | Composite Mimetics | Non Mimetics |
|---|---|---|
| Accuracy | 0.49 | 0.34 |
| Recall | 0.30 | 0.09 |
| F1-Score | 0.27 | 0.10 |

and non-auditory mimetics. The few differences suggested auditory mimetics performed *worse*, relative to the composite dataset. Table 18 summarizes these results: see table 15 for comparison with the composite set of mimetics.

## Frame Semantics

Investigating collocates rather than direct verbal morphological endings yielded a much higher amount of tokens to train on, though lower than in Experiment I. This is due to the size difference between the corpora used and the tendency for ideophones to be conspicuously absent from the body of newspapers.[1]

Semantic Frame analysis of verbal collocates once again revealed relatively high recall and f1-score for the classification problem when considering all possible labels for a verbal collocate, but with reduced accuracy. Overall, this experiment produced results resembling that of Experiment I. Table 19 shows the evaluation metrics for a Random Forest classifier trained on multilabel data for collocates with at least three coöcurrences in the set of parsed tokens. Mimetic adverbials outperformed non-mimetics on all metrics by a significant margin. This was increased even more so when evaluating non-mimetics on a classifier trained on mimetic adverbials: In such cases, accuracy hovered near 0.10, with recall and F1-score both below 1.0%.

[1] For this thesis' discussion on the topic, see Section 5.3. See also, Park (2022)

Table 20: Composite Mimetic Performance Metrics on Single-Labelled Semantic Frame Collocates

| | Naïve Bayes | Random Forest | RNN |
|---|---|---|---|
| Accuracy | 0.25 | 0.62 | 0.46 |
| Recall | 0.27 | 0.35 | 0.11 |
| F1-Score | 0.16 | 0.32 | 0.07 |

Table 21: Non-Mimetic Performance Metrics on Single-Labelled Semantic Frame Collocates

| | Naïve Bayes | Random Forest | RNN |
|---|---|---|---|
| Accuracy | 0.03 | 0.51 | 0.46 |
| Recall | 0.14 | 0.16 | 0.11 |
| F1-Score | 0.06 | 0.14 | 0.05 |

The RNN classifier performed particularly poorly in this experiment, with non-mimetics nearly equivalent to mimetic adverbials. The Bayesian model also performed poorly in comparison to the Random Forest Model, being far more likely to mislabel a given mimetic's semantic frame collocate as an infrequently occurring frame. Meanwhile, the Random Forest Model performed exceedingly well in this experiment, with high accuracy, recall, and precision for the composite mimetic set, and low recall and precision for non-mimetics. Tables 20 and 21 summarize these results, with Figures 21 and 22 visualize these findings in the confusion matrices for the best and worst performing models, respectively.

## 5.1 Feature Importance and Distribution

Phonological features in the Korean dataset were well-distributed: While some features were overrepresented in the set, the majority of the features were evenly distributed. [+Dorsal] was the most frequent present feature at 7718 occurrences, and [+Front] was the least frequent present feature, with 620 occurrences. Table 22 shows the set of most common features[2].

Distribution was mostly skewed towards non-auditory mimetics except for in a few specific features. For Korean, this was the [+Nasal] feature, which occurred in auditory mimetics slightly more often than in non-auditory mimetics. This distribution is displayed in Figure 23

[2] See Appendix A for the full set of feature occurrences.

Figure 21: Random Forest Confusion Matrix for Top 5 Most Frequent Non-Mimetic Adverbials, Normalized

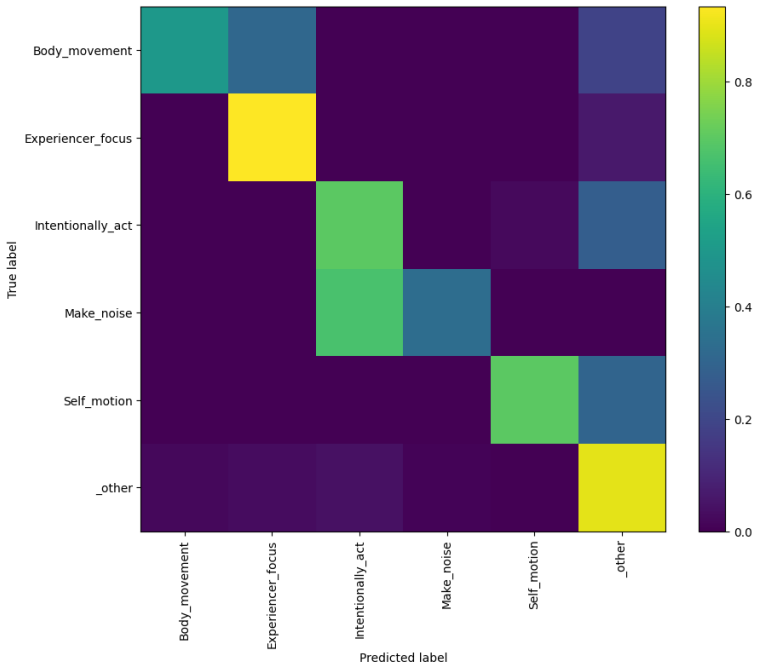



Figure 22: Naïve Bayes Confusion Matrix for Top 5 Most Frequent Non-Mimetic Adverbials, Normalized

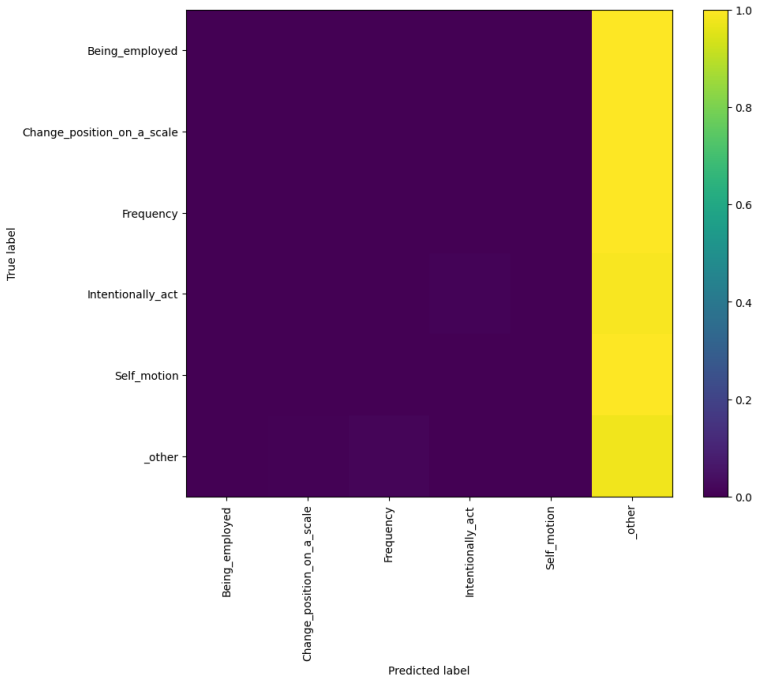


Table 22: Most Frequent Phonemes in Mimetic With at Least 1 Significant Collocation

| Feature | Occurrences | Example Phoneme |
|---|---|---|
| [+Dorsal] | 7718 | /k/ |
| [+Consonantal] | 6491 | /l/ |
| [+High] | 5323 | /i/ |
| [+Back] | 4953 | /u/ |
| [+Son] | 4863 | /m/ |

Figure 23: Feature Distribution in Korean Mimetic Adverbials

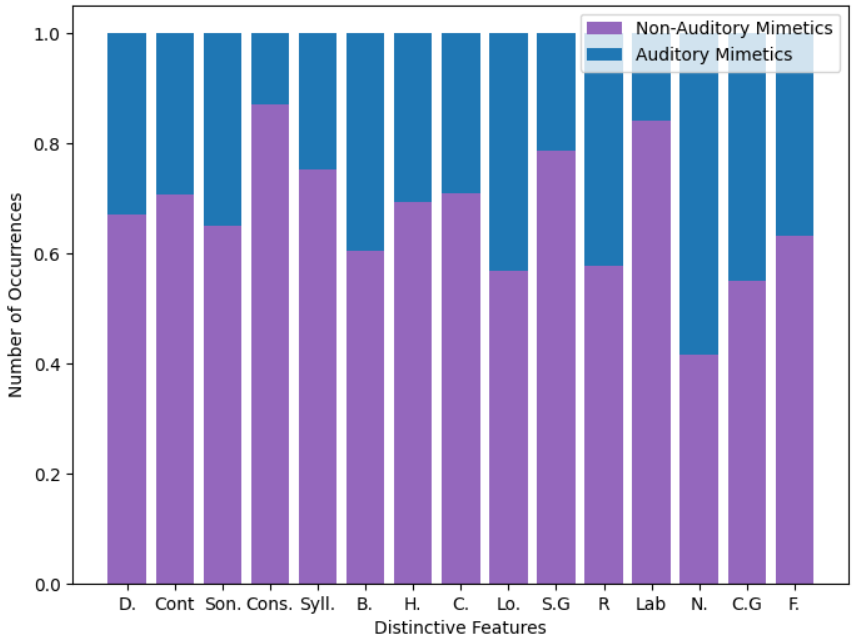


Table 23: Experiment II: Permutated Feature Significance on Test Mimetics ($n\_repeats = 100$)

| Feature | Significance | Standard Deviation | Example Phoneme |
|---|---|---|---|
| [±High] | 0.033 | 0.009 | /i/ |
| [±Labial] | 0.019 | 0.007 | /p/ |
| [±Round] | 0.006 | 0.004 | /o/ |
| [±Nasal] | 0.005 | 0.005 | /m/ |

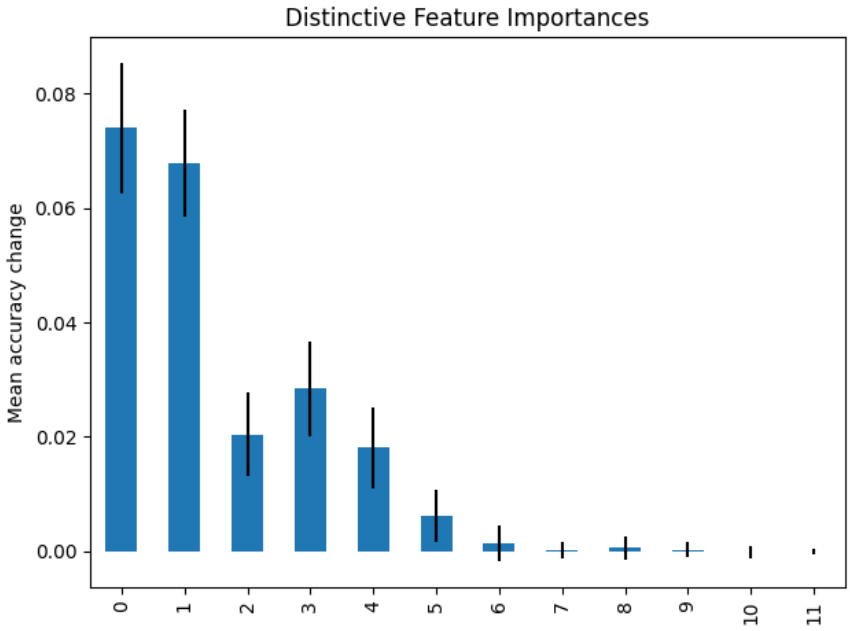
Table 24: Experiment II: Permutated Feature Significance on Train Mimetics ($n\_repeats = 100$)

| Feature | Significance | Standard Deviation | Example Phoneme |
|---|---|---|---|
| [±High] | 0.040 | 0.006 | /i/ |
| [±Lab] | 0.032 | 0.004 | /p/ |
| [±Round] | 0.011 | 0.003 | /o/ |
| [±Cor] | 0.008 | 0.004 | /t/ |
| [±Low] | 0.004 | 0.003 | /a/ |
| [±C.G] | 0.003 | 0.003 | /t'/ |
| [±S.G] | 0.003 | 0.002 | /$t^h$/ |

As in Experiment I, feature importance of distinctive phonological features in predicting the semantic frame of verbal collocations was collected using sklearn's permutation importance algorithm on both training and testing subsets. This was done on composite mimetics in order to determine which of these features had the most impact on the classification of semantic frame for verbal collocations. I performed the permutation algorithm 100 times as done in Kilpatrick et al. (2023) for calculation and took the mean and standard deviation for each distinctive feature. Table 23 summarizes the distinct phonological features with mean significance $> 0.1\%$ for data within at least one standard deviation on the set of data excluded from training. For Korean, the most significant feature was [±High], followed by [±Labial] and [±Round] in both the train and test data. Surprisingly, [±Nasal] was the next most significant in the train set, although it did not appear at all in the list of significant features for the train dataset. The train dataset was mostly equivalent to the test dataset for the most significant features, but indicates some degree of overfitting to the [±Coronal], and [±Low].

I also collected the aggregate feature importances for phoneme positions within a mimetic adverbial. Unlike the experiment on Japanese, Korean mimetics displayed a strong frontward

Figure 24: Feature Importance of Phoneme Positions in Korean




bias, with earlier positions nearly unilaterally having higher importance scores than those in later positions. The first two positions of the word by far contribute more than any other position to the semantics of the word, with a combined importance of nearly 15.0%. This is shown in Figure 24.

## 5.2 Experiment IIa

A survey on human speakers of Korean revealed that the perception of differences between different forms of ideophones was not nearly as significant as the results of the Random Forest model suggests. Moreso than in the experiment in Ia, consultants struggled to provide example contexts or descriptions for nonce vocabulary and even for particularly rare ideophones which were a part of the real Korean lexicon.

When presented with a pair of vocabulary that alternated between a *yin* and *yang* form, one speaker described the contrasted vowel as contributing "almost no difference in the meaning" of the mimetic, using them in almost the exact same contexts. The speaker offered the sentences in (28) to illustrate, describing them as having basically the same meaning, and being unable to put

into words any differences that may have existed. Interestingly the model trained in Experiment II revealed similar results when tested on these mimetic adverbials, predicting the same Semantic Frame for several ideophone pairs with vowel-harmony alterations.

(28) a. 새가 훌훌 날아갔다.

sae-ga hulhul naraga-tda
bird-NOM MIM fly-FORM.PST
"'*hulhul*', the bird flew away."

b. 새가 홀홀 날아갔다.

sae-ga holhol naraga-tda
bird-NOM MIM fly-FORM.PST
"'*holhol*', the bird flew away."

However, not all speakers expressed this similarity. In fact, another speaker described a different *yin-yang* pair as having a strikingly large difference in strength, with the *yang* form 우슬우슬 being described as a much stronger. Both the Random Forest model and elicited data suggested that consonant quality alterations in ideophones were more likely to be subject to alterations, although most such cases favored one common complement for at least two of the three possible consonant alterations.

When presented with nonce words, Korean consultants struggled much more in providing descriptions of meaning or providing contexts in which to place the nonce adverbial. Many would instead suggest a different word that sounded similar to them. One consultant, asked to provide a definition and verbal context for the nonce word **baen'gil*, provided a sentence but justified their answer by offering that "it sounds like *ban'gil* and this is the kind of sentence you'd use for that." While this may suggest that the sound-symbolic system used in Korean is not as productive as Japanese and therefore unreceptive to the creation of new words, it is also possible that the consultants, conducting an interview with me in English, thought I was just mispronouncing what was actually a real word.

## 5.3 Discussion

This experiment affirmed Experiment I, in finding Random Forest Classifiers and RNN classifiers to be most effective in predicting associations between mimetics of a given phonological form and significant collocations which may be attached with them. However, Random Forest classifiers performed particularly well in comparison to the RNN in terms of recall and F1-score in this experiment. This finding affirms previous studies which have found Random Forest classifiers to be 'efficient learners' of the relationship between sound and meaning in Japanese, Korean, and Chinese vocabulary (Kilpatrick et al. 2023). Indeed, in this experiment, Random Forest classifiers performed the best on every metric considered, and largely eliminated erroneous classifications except for morphological collocates which occurred only a single time.

Perhaps the most significant finding was the ability of the Random Forest Model's ability to distinguish between usage of *hada* and *haejida*. While *hada* remains for mimetics and non-mimetics alike an extremely productive light verb which may be used, *haejida* is a similarly productive marker, often used to indicate a passive sense to the predicative form, with a meaning similar to 'become X'. Compare the examples in (29), sourced from the Leipzig corpora and translated by one of my consultants. This indicates that there is some relationship between the phonological content of a mimetic and the valency required by its morphological affixes with some mimetics.

(29) a. 이후에 팔아서 지금 마음이 씁쓸하다

ihu-e par-aseo jigeum maeum-i sseubsseul-hada
henceforth-LOC sell-IE.PST now mind-NMNLZ MIM-do
"After selling (it), I feel bitter"

b. (?) 너무나 비현실적이어서 빠져나오는 즉시 씁쓸해지기도 하지만 말이다

neomuna bihyeonsiljeog-ieoseo ppajyeonao-neun jeugsi sseubsseul-haeji-gi-do
too.much unrealistic-next escape-that immediately MIM-become-NMNLZ-also
hajiman mal-i-da
however word-NOM-COP
"It's so unrealistic that I couldn't even be bittersweet about it as I was leaving it"

With regards to Semantic Frames, Experiment II provided solid proof that Random Forest Classifiers stand as one of the most efficient and effective classifiers for modeling sound-symbolism in ideophones. Random Forests were also capable of modeling the multi-labeling of semantic frames. This feature helps account for polysemy in verbal complements, and the possibility of a complement being interpretable in multiple senses and therefore multiple semantic frames.

Similar to the findings in (Kilpatrick et al. 2023), this experiment found that features typically associated with vowels were the most significant features for both test and train data when predicting the semantic content of a verbal collocation. [±High], [±Round], and [±Low], all of which are typically restricted to differentiating vowels were among the most significant features, with [±High] outperforming any other feature by a large margin. [±Labial] was likewise associated with the phoneme /w/, a mandatory onglide of several of Korean's diphthongs. This data altogether suggests that before any consonantal feature, vowel quality is one of the primary factors in predicting semantic content. Two of the three most significant consonant features were [±C.G] and [±S.G], the features associated with the 'Tense' and 'Aspirate' versions of plosives, reflective of existing consonant alteration systems in Korean ideophones.

It is perhaps unsurprising that [±Round] and [±Labial] were two of the most significant features: This is reflective of the vowel harmony system in Korean, described in Section 2.2. While it is known that both vowel harmony and consonant alteration contribute to the specific meaning of an ideophone, this data suggests that vowel harmony and quality is far more significant in determining the likely semantic condition of any complements.

Beyond specific feature presences, the position of a phoneme within a mimetic adverbial had significant effects on the predicted semantic frame for collocates. Korean mimetics displayed a striking bias for earlier appearing phonemes with the first two phonemes contributing more in accuracy than all other positions in the word combined. This is perhaps not surprising: many of the ideophones under consideration were reduplicative of a monosyllabic morpheme. Yet, as the many others were not, this suggests that indeed the first two phonemes within a word are incredibly important in determining semantic content of a mimetic, and therefore its verbal

complements.

Elicitation experiments affirmed that these features were significant in their contributions to the meaning of the mimetic and therefore the context in which it can be used. While in human elicitation, vowel contrasts did not appear to contribute much to the overall meaning of a mimetic, consonant feature contrasts did, with [±Labial] changing the meaning of the same mimetic much more so than other features: One consultant predicted the *same* collocate for every mimetic with initial /p/, and nowhere else. While certainly the vowel contrasts in Korean mimetics have some function and operation in determining the meaning, the Random Forest model was more effective at clearly denoting these differences with regards to complements. Elicitation suggested vowel quality as the sole contrast between two mimetics was not an effective measure of semantic difference, but in tandem with other features contributed meaningfully to the semantic frame of collocates. Considering the poor performance of Naïve Bayes and successful performance of the Random Forest and RNN model in predicting both morphological endings and semantic frame of collocates, This affirmed previous predictions of the sound-symbolic system of Korean as varied and non-linear.

One large caveat to working with Korean mimetic data was the infrequency of certain mimetic forms within the corpora. In Korean, ideophones are particularly unlikely to occur within the body of most news articles, but rather are more likely to occur within news headlines. Park (2022) found that ideophones were increasingly likely to be found in headlines of Korean newspapers, especially as news headlines localized and shifted away from the HANJA (Chinese characters) standard. In fact, ideophones were so much more likely to appear in headlines that a corpus study of Korean newspapers revealed that zero of the top 100 most common ideophones were found in the bodies of the newspapers sampled, while 60 of the 100 were found in headlines (Jo and Kang 2013, as cited in ibid.).

Further, unlike in Japanese, Korean uses a script which is almost entirely phonetic (with the exception of very infrequent use of *hanja*) and does not encode semantic information as the text in Japanese does. This in particular created a difficulty for the tokenization algorithm that I used,

which led to several 'mimetic + verb' parsed directly as a verb, and the first syllable of the mimetic misparsed as a separate morpheme of its own. This was particularly a problem for some mimetics which were common syllables used in non-mimetic Sino-Korean compounds. Because of this, I excluded monosyllabic mimetics from the dataset except for in cases where absolutely necessary in order to account for phonological contrasts not accounted for by bisyllabic or longer morphemes. For the full set of mimetics considered, see the dataset linked in the Bibliography of this thesis.

Still, the results largely imply a greater iconic dependency between the phonological form and the morphological ending in mimetic vocabulary than in non-mimetic vocabulary, although there does not appear to be a significant different between auditory mimetics and non-auditory mimetics as there was in the Japanese examples. In particular the poor performance and overprediction of *hada* for non-mimetics suggests that there indeed is a significant iconicity present in mimetic adverbials.

Polysemy[3] in particular stood out as one of the challenges in conducting this experiment for both Japanese and Korean language data: as languages with high degrees of polysemy, many possible collocates contained several possible Semantic Frames depending on context. For example, *hada* had no less than 16 semantic frames associated with it and its various connotations. Without a large-scale tagged corpus, it is difficult to determine the exact sense that a given word might have from a verb or its morphologically related forms: The Korean Framenet corpus only contains 22,935 instances on which to train on: Even assuming a similar distribution of ideophones in the FrameNet corpus to that of the the Leipzig corpora used in this experiment, one could only expect 13 instances of ideophones to occur, far from enough to provide accurate training and modelling.

Polysemy was further an issue when it came to classifying an adverbial as auditory mimetic or non-auditory mimetic or sometimes even as mimetic or non-mimetic! There exists a large host of ideophones with several senses in Korean, with many ideophones describing both a sound and manner in different senses. A smaller amount of ideophones had further gained other senses

[3] The ability for a word to have multiple different, not necessarily related, senses. C.f 'right'

and taken on other non-mimetic meanings that described neither an auditory event, a manner of action, or even a psychological state, as would be expected by the third class of ideophones common to Japanese and Korean but not included in this investigation. (30) is one such example. In such cases, the mimetic will naturally be associated with different complements for different senses in meaning. Such ideophones were included in both sets unless a consultant indicated that a given word was strongly associated with either a sound or meaning.

(30) a. 두통 때문에 골골하다

dutong ttaemune golgol-hada
headache because MIM-do
"Because of my headache I am sickly."

b. 칠면조가 골골 운다

chilmyeonjo-ga golgol unda
turkey-NOM MIM cry.FOR.NPST
"The turkey is crying *golgol*"

The ability for a mimetic with the same exact phonological form to have multiple senses, and therefore, multiple types of complements which are highly associative with only a single sense poses serious issues for the proposal of this thesis: indeed comparing auditory mimetics with non-auditory mimetics yielded minimal comparable results. Yet, the Random Forest model was markedly capable of predicting collocations in the composite mimetic set. Further research should be done to determine exactly how this polysemy can be affected by phonological form, and especially the role that reduplication and length plays in affecto-semantic quality.

# Chapter 6

# Conclusion

*End? No, the journey doesn't end here.*

Gandalf the White

This thesis set out to determine the effectiveness of using different classification models in representing sound symbolism on mimetics in a larger scale than has been previously done before. Aided by the FrameNet framework, I classified the semantic quality and content of verbal and nominal complements, based on the phonological feature representation of a given mimetic adverbial. This thesis was largely successful in its goal, constructing several models and determining Random Forests as a model with particular promise for represent sound symbolism in ideophones in both Korean and Japanese. It likewise found Recurrent Neural Networks as promising in modelling such relationships, although primarily for cases with less possible categories. Further, This study found that both Korean and Japanese mimetics far outperformed non-mimetic adverbials in predicting the semantic content of their verbal complement, suggesting a link between the phonological form of a mimetic and the context in which it may be used. This link was particularly stronger in Korean, especially for predicting endings in the formation of attributive forms of mimetics. Symbolic links were statistically significant in both languages, with a Random Forest classifier able to outperform a naïve and random algorithm in all experiments.

The findings from all experiments affirm that the phonological form of a mimetic is strongly

related to the semantic content of its complements, with certain distinctive phonological features like [±High] being particularly likely to describe an auditory event. In fact, for Japanese, [±High] was one of the only features on set-aside data whose presence or absence was statistically significant with its importance more than three standard deviations away from 0.01%. Significant features themselves in Japanese were largely consistent between trained and set-aside data: all of the top three significant features ([±High], [±Labial], and [±Round]) from set-aside data were shared with train data. Some consonantal features were only significant on train data, like [±Coronal] and [±C.G], suggesting some weak overtraining effects, although because their significance was either within a single standard of 0%, this was not a strong enough effect to note in this thesis. Ordering effects revealed that the first position in a word contributed the most to the semantic content of the mimetic for Korean mimetic adverbials, although no such relationship was apparent in Japanese mimetics. For Korean adverbials this bias towards phonemes appearing earlier in the word extended to the first two phonemes regardless of distinctive feature content.

This suggest that vowels in particular remain an area which merits consideration for their effect on sound-symbolic vocabulary. Despite some consultants considerations that "the vowel sound didn't really change the meaning of the sentence", many consultants used wildly differing complement structures depending on the vowel quality of these mimetics. However, this belief only surfaced when consultants were presented multiple forms of particularly uncommon ideophones. Some models indeed place vowels as important in determining meaning for mimetic words—Hamano (1998)'s model in particular attributes Shape and Size to the first and second vowel in a bimoraic mimetic root respectively. Yet the findings of this thesis indicate that they are more involved in the sound-symbolic system, especially for Korean, than previously thought.

This thesis only focused on the absolute presence versus non-presence of all possible distinctive features for a mimetic adverbial, taking account for ordering effects and did not investigate how these features interact with each other. That is, I did not consider all possible pairs, triads, or other combinations of specific features in a possible mimetic. However, because of the nonlinearity of the sound-symbolic system in Korean and Japanese shown in this thesis and previous

works, it is apparent that a model which investigates how the significant features interact with each other is needed to further understand exactly the relationship between phonological form of an adverbial, and its semantic qualities. Still, this is no small task. With 18 and 17 contrastive features for Korean and Japanese respectively, the number of feature combinations to investigate would quickly rise into unfeasible levels, with $\sum_{r=2}^{36} \frac{36!}{r!(n-r)!}$ possible combinations just for Japanese alone. However, further studies may find it prudect to investigate at least those which are pre-determined to be of interest (such as [+High, -Consonantal] vs [-High, -Consonantal]). Doing so may validate the findings in this thesis and disambiguate whether the significant features are truly attributable to vowels, or simply obfuscated by velar and palatal consonants like /k/ and /c/ which also are analyzable with some features typically reserved for vowels, like [±High] and [±Back]. Further opportunities for optimization on this study include the possibility of using random split and validation on training data, or k-fold cross-validation where training data is split into $k$-number subsamples, which are then used as validation for multiple trainings of a model. However, because of the relatively small sample sizes of data for my Korean experiment, such a method was not feasible in this thesis.

Semantic research on ideophones remains sparsely researched across langauges. While general research into ideophones has certainly increased in recent years, it is still lacking especially in the still growing field of Korean NLP and is particularly restricted for English-language literature, a problem this thesis attempts to rectify. While this thesis focused on sound-symbolic effects of verbal complements, evidence from Hamano (1998) as well as data from the elicitation surveys done in this thesis suggest that certain distinctive features contribute to the meaning of the mimetics in meaningful ways that are not reflected in the semantic frame of the verbal collocate. However, due to a lack of a large-scale semantically tagged corpus and without a determistic way of classifying adverbials, such research will be difficult to conduct.

Ideophone research still has a long way to go: syntactic and semantic roles of mimetics in sentences are still as of yet unclear, and the data from this thesis implies that there is still a lot we don't yet know about the other influences on semantic quality and sentence structure of ut-

terances involving ideophones. But as research continues, ideophone by ideophone, we will no doubt learn more about this mysterious, overlooked linguistic puzzle.

# Bibliography

(Mar. 2024). URL: https://osf.io/py43z/?view_only=b6d0eee440184c91bfb53b8ec8dcda94.

Akita, Kimi (2009). "A Grammar of Sound-Symbolic Words in Japanese: Theoretical Approaches to Iconic and Lexical Properties of Mimetics". PhD thesis. Kobe University.

— (2012). "Toward a frame-semantic definition of sound-symbolic words: A collocational analysis of Japanese mimetics". In: *Cognitive Linguistics* 23.1, pp. 67–90.

— (2013). "Frame phonosemantics: A frame-semantic approach to Japanese mimetics".

— (2016). "A Multimedia Encyclopedia of Japanese Mimetics: A Frame-Semantic Approach to L2 Sound-Symbolic Words". In: *Cognitive-Functional Approaches to the Study of Japanese as a Second Language*. Ed. by Kaori Kabata and Kiyoko Toratani. Vol. 46. Studies on Language Acquisition. Berlin/New York: De Gruyter Mouton, pp. 139–168.

Akita, Kimi and Prashant Pardeshi (2019). "Ideophones, mimetics, and expressives". In: John Benjamins Publishing Company. Chap. Introduction, pp. 1, 9.

Akita, Kimi and Takeshi Usuki (2016). "A constructional account of the 'optional' quotative marking on Japanese mimetics". eng. In: *Journal of linguistics* 52.2, pp. 245–275. ISSN: 0022-2267.

Altmann, André, Laura Toloşi, Oliver Sander, and Thomas Lengauer (Apr. 2010). "Permutation importance: a corrected feature importance measure". In: *Bioinformatics* 26.10, pp. 1340–1347. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btq134. eprint: https://academic.oup.com/bioinformatics/article-pdf/26/10/1340/48851160/bioinformatics_26_10_1340.pdf. URL: https://doi.org/10.1093/bioinformatics/btq134.

Azari, Raziyeh and Masoud Sharififar (2017). "Translating onomatopoeia: An attempt toward translation strategies". In: *Review of Applied Linguistics Research* 3.3, pp. 72–92.

Bredin, Hugh (1996). "Onomatopoeia as a Figure and a Linguistic Principle". In: *New Literary History* 27.3, pp. 555–569. ISSN: 00286087, 1080661X. URL: http://www.jstor.org/stable/20057371 (visited on 10/02/2023).

Bremner, Andrew J., Serge Caparos, Jules Davidoff, Jan de Fockert, Karina J. Linnell, and Charles Spence (2013). " "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners". eng. In: *Cognition* 126.2, pp. 165–172. ISSN: 0010-0277.

Cho, Young-mee Yu (Oct. 2016). *Korean Phonetics and Phonology*. DOI: 10.1093/acrefore/9780199384655.013.176. URL: https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-176.

Choi, DongHyun, Ilnam Park, Myeongcheol Shin, Eunggyun Kim, and Dong Ryeol Shin (2021). "Korean Erroneous Sentence Classification With Integrated Eojeol Embedding". In: *IEEE Access* 9, pp. 81778–81785. ISSN: 2169-3536. DOI: 10.1109/access.2021.3085864. URL: http://dx.doi.org/10.1109/ACCESS.2021.3085864.

Cuskley, Christine, Julia Simner, and Simon Kirby (2017). "Phonological and orthographic influences in the bouba-kiki effect". eng. In: *Psychological research* 81.1, pp. 119–130. ISSN: 0340-0727.

Cutler, D. Richard, Thomas C. Edwards, Karen H. Beard, Adele Cutler, Kyle T. Hess, Jacob Gibson, and Joshua J. Lawler (2007). "Random Forests for Classification in Ecology". In: *Ecology* 88.11, pp. 2783–2792. ISSN: 00129658, 19399170. URL: http://www.jstor.org/stable/27651436 (visited on 03/07/2024).

*Dictionary of iconic expressions in Japanese* (1996). eng;jpn. Trends in linguistics. Documentation 12. Berlin ; New York: Mouton de Gruyter. ISBN: 3110128101.

Dingemanse, Mark (2012). "Advances in the Cross-Linguistic Study of Ideophones". eng. In: *Language and linguistics compass* 6.10, pp. 654–672. ISSN: 1749-818X.

Ferklova, Blanka (2023). "Textual distribution of Korean onomatopoeias and ideophones and its specific features". eng. In: *SKASE journal of theoretical linguistics* 20.1, p. 113. ISSN: 1336-782X.

Fort, Mathilde, Alexander Martin, and Sharon Peperkamp (2015). "Consonants are More Important than Vowels in the Bouba-kiki Effect". In: *Language and Speech* 58.2. PMID: 26677645, pp. 247–266. DOI: 10.1177/0023830914534951. eprint: https://doi.org/10.1177/0023830914534951. URL: https://doi.org/10.1177/0023830914534951.

Hamano, Shoko (1998). *The Sound-Symbolic System of Japanese*. CSLI Publications and Kurosio Shuppan.

— (2013). "「方言における擬音語・擬態語の体系的研究の意義」Significance of Systematic Research of Phonomimes and Phenomimes in Japanese Dialects". In: 「オノマトペ研究の射程：近づく音と意味」*(The Scope of Onomotopoeia Research: Nearing Sound and Meaning)*. 株式ひつじ書房. Chap. 8, pp. 133, 147.

— (July 2020). "The Development of Japanese Ideophones: What We Learn from Korean Monosyllabic Ideophones". In: *Japanese/Korean Linguistics* 27. URL: https://web.stanford.edu/group/cslipublications/cslipublications/ja-ko-contents/JK27/jako27-posters.shtml.

Haslett, David A. and Zhenguang G. Cai (2023). "Systematic mappings of sound to meaning: A theoretical review". eng. In: *Psychonomic bulletin & review*. ISSN: 1069-9384.

Hinton, Leanne, Johanna Nichols, and John Ohala (1994). *Sound symbolism*. eng. Cambridge [England] ; New York, NY: Cambridge University Press. ISBN: 0521452198.

Hirose, Masayoshi (1981). *JAPANESE AND ENGLISH CONTRASTIVE LEXICOLOGY: THE ROLE OF JAPANESE "MIMETIC ADVERBS"*. eng.

Ibarretxe-Antuñano, Iraide (2019). "Towards a semantic typological classification of motion ideophones". In: *Ideophones, Mimetics, and Expressives*. Ed. by Kimi Akita and Prashant Pardeshi. John Benjamins Publishing Company. Chap. 6.

Kawahara, Shigeto and Kazuko Shinohara (2012). "A tripartite trans-modal relationship among sounds, shapes and emotions: A case of abrupt modulation". In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 34.

Kilpatrick, Alexander James, Aleksandra Ćwiek, and Shigeto Kawahara (2023). "Random forests, sound symbolism and Pokémon evolution". eng. In: *PloS one* 18.1, e0279350–e0279350. ISSN: 1932-6203.

Kim, Hung-gyu, Beom-mo Kang, and Jungha Hong (2007). "21st Century Sejong Corpora (to be) Completed". In: *The Korean Language in America* 12, pp. 31–42. ISSN: 23320346, 2374670X. URL: http://www.jstor.org/stable/42922169 (visited on 02/23/2024).

Kim, Jeahong and Kichun Nam (2018). "Lexical Factors that Influence the Korean Eojeol Recognition". eng. In: *Korean Journal of Cognitive and Biological Psychology* 30.4, pp. 373–390. ISSN: 1226-9654.

Kita, Sotaro (1997). "Two-dimensional semantic analysis of Japanese mimetics". eng. In: *Linguistics* 35.2, pp. 379–416. ISSN: 0024-3949.

Ko, Seongyeon (2022). "Vowel Harmony". In: *The Cambridge Handbook of Korean Linguistics*. Ed. by Sungdai Cho and JohnEditors Whitman. Cambridge Handbooks in Language and Linguistics. Cambridge University Press, pp. 177–214.

Kubozono, Haruo (2019). "The phonological structure of Japanese mimetics and motherese". In: John Benjamins Publishing Company. Chap. 2, pp. 35, 56.

Labrune, Laurence (2012). *The Phonology of Japanese*. eng. The Phonology of the World's Languages. Oxford: Oxford University Press. ISBN: 0199545839.

Larsen, Darrell and Jeffrey Heinz (2012). "Neutral vowels in sound-symbolic vowel harmony in Korean". In: *Phonology* 29.3, pp. 433–464. DOI: 10.1017/S095267571200022X.

Maddieson, Ian (2013a). "Consonant Inventories (v2020.3)". In: *The World Atlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Zenodo. DOI: 10.5281/zenodo.7385533. URL: https://doi.org/10.5281/zenodo.7385533.

Maddieson, Ian (2013b). "Vowel Quality Inventories (v2020.3)". In: *The World Atlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Zenodo. DOI: 10.5281/zenodo.7385533. URL: https://doi.org/10.5281/zenodo.7385533.

"moan, v." (July 2023). In: *Oxford English Dictionary*. Oxford University Press. DOI: https://doi.org/10.1093/OED/6678953490.

Morse, Christopher (Dec. 2016). *Digital Japanese Literature: Aozora Bunko*. URL: https://digitalhumanities.fas.harvard.edu/digital-japanese-literature-aozora-bunko/.

*Nihongo onomatope jiten : giongo, gitaigo 4500* (2007). jpn. 初版. Tōkyō: Shōgakkan. ISBN: 9784095041742.

Nikolaev, Dmitry, Collin F Baker, Miriam R L Petruck, and Sebastian Padó (2023). "Adverbs, Surprisingly". eng. In: *arXiv.org*. ISSN: 2331-8422.

Nuckolls, Janis B. (1995). "Quechua texts of perception". eng. In: *Semiotica* 103.1-2, pp. 145–170. ISSN: 0037-1998.

— (2019). "The sensori-semantic clustering of ideophonic meaning in Pastaza Quichua". In: *Ideophones, Mimetics, and Expressives*. Ed. by Kimi Akita and Prashant Pardeshi. John Benjamins Publishing Company. Chap. 7.

Okada, Hideo (1999). "Handbook of the International Phonetic Association : a guide to the use of the International Phonetic Alphabet". eng. In: Cambridge, UK ; New York, NY: Cambridge University Press, pp. 117–119. ISBN: 0521652367.

*Open-Source ML Monitoring and Observability* (n.d.). URL: https://www.evidentlyai.com/classification-metrics/accuracy-precision-recall#what-is-recall.

Osawa, Rie (2019). "「日本語オノマトペ有契性に関する印象調査とその分析：韓国語母語話者である日本語学習者と日本語母語話者を対象に」(*Analysis of an Impression Survey of Motivation in Japanese Onomatopoeia : A Case Study of Korean Learners of Japanese and Native Speaker*)". jpn. In: 研究論集 *(Journal of Inquiry and Research)* 109, pp. 187–204. ISSN: 0388-1067.

Osawa (Itou), Rie (2006). "「オノマトペにおける～スル形の考察」(*Treatment of the -suru Form in Onomatopoeia*)". jpn. In: 日本認知言語学会論文集 6, pp. 310–320. ISSN: 1346-7964.

Park, Jiyeon (2022). "The Use of Korean Ideophones in Newspaper Headlines". In: *Japanese/Korean Linguistics* 29, pp. 425–434.

Passi, Ananya and S. P. Arun (2022). "The Bouba-Kiki effect is predicted by sound properties but not speech properties". eng. In: *Attention, perception & psychophysics*. ISSN: 1943-3921.

Reenald, Richard and Elisa Carolina Marion (Sept. 2022). "Analysis of Onomatopeia Translation Result on "Komi Can't Communicate" Comic Using Machine Translation". In: *Proceedings of the 3rd Asia Pacific International Conference on Industrial Engineering and Operations Management*. IEOM Society International. Johor Bahru, Malaysia.

Rezk, Nesma M., Madhura Purnaprajna, Tomas Nordstrom, and Zain Ul-Abdin (2020). "Recurrent Neural Networks: An Embedded Computing Perspective". eng. In: *IEEE access* 8.1, pp. 57967–57996. ISSN: 2169-3536.

Robbeets, Martine (2008). "The Historical Comparison of Japanese, Korean and the Trans-Eurasian Languages". In: *Rivista degli studi orientali* 81.1/4, pp. 261–287. ISSN: 03924866, 17241863. URL: http://www.jstor.org/stable/41913342 (visited on 02/12/2024).

Sakurai, Akito (2008). "Linguistic Productivity and Recurrent Neural Networks". eng. In: *In-Tech eBooks*. IntechOpen. ISBN: 9789537619084.

Shibasaki, Reijirou (2009). "Semantic Constraints on the Diachronic Productivity of Japanese Reduplication". eng. In: *Grazer Linguistische Studien* 71.spring, pp. 79–97. ISSN: 1015-0498.

Sohn, Ho-min (1994). *Korean*. eng. Descriptive grammars. London ; New York: Routledge. ISBN: 0415003180.

*Spoken L1 Language: Korean* (n.d.). URL: https://glottolog.org/resource/languoid/id/kore1280.

Takayama, Hiromu, Nastumi Takeuchi, Daiki Urata, Tsuyoshi Nakamura, Ahmed Moustafa, and Takayuki Ito (2019). "Feature Extraction of Onomatopoeia Considering Phonotactics

for Automatic Semantic Usage Classification of Onomatopoeia". In: 人工知能学会全国大会論文集 JSAI2019, 2J1E502–2J1E502. DOI: 10.11517/pjsai.JSAI2019.0_2J1E502.

Tamaoka, Katsuo, Sachiko Kiyama, and Yayoi Miyaoka (2011). "「新聞と小説のコーパスにおけるオノマトペと動詞の共起パターン」(*Coöccurence Patterns for Verbs and Onomatopoeia in Newspaper and Novel Corpora*)". jpn. In: 言語研究 139, pp. 57–84. ISSN: 2185-6710.

Vinson, David, Robin L. Thompson, Robert Skinner, and Gabriella Vigliocco (2015). "A faster path between meaning and form? Iconicity facilitates sign recognition and production in British Sign Language". In: *Journal of Memory and Language* 82, pp. 56–85. ISSN: 0749-596X. DOI: https://doi.org/10.1016/j.jml.2015.03.002. URL: https://www.sciencedirect.com/science/article/pii/S0749596X1500025X.

Vovin, Alexander (2005). "The End of the Altaic Controversy In Memory of Gerhard Doerfer". In: *Central Asiatic Journal* 49.1, pp. 71–132. ISSN: 00089192. URL: http://www.jstor.org/stable/41928378 (visited on 02/12/2024).

Wang, Yisen, Shu-Tao Xia, Qingtao Tang, Jia Wu, and Xingquan Zhu (2018). "A Novel Consistent Random Forest Framework: Bernoulli Random Forests". eng. In: *IEEE transaction on neural networks and learning systems* 29.8, pp. 3510–3523. ISSN: 2162-237X.

Webb, Geoffrey I. (2017). "Naïve Bayes". In: *Encyclopedia of Machine Learning and Data Mining*. Ed. by Claude Sammut and Geoffrey I. Webb. Boston, MA: Springer US, pp. 895–896. ISBN: 978-1-4899-7687-1. DOI: 10.1007/978-1-4899-7687-1_581. URL: https://doi.org/10.1007/978-1-4899-7687-1_581.

Yoshimoto, Banana (1992). *Kitchen*. Fukutake Shoten.

# Data Repository

(Mar. 2024). URL: https://osf.io/py43z/?view_only=b6d0eee440184c91bfb53b8ec8dcda94.

# Corpora

「現代日本語書き言葉均衡コーパス（*BCCWJ*）」*Balanced Corpus of Modern Written Japanese* (2024). URL: https://clrd.ninjal.ac.jp/bccwj/.

*Aozora Bunko Corpus* (n.d.). URL: https://www.aozora.gr.jp/.

Eckart, T. and U. Quasthoff (2013). *Statistical Corpus and Language Comparison on Comparable Corpora*.

Sonobe, Ryosuke, Shinnosuke Takamichi, and Hiroshi Saruwatari (2017). *"JSUT corpus: free large-scale Japanese speech corpus for end-to-end speech synthesis,"*.

# Appendices

# Appendix A

# Other Experimental Data

Table 25: Experiment I: List of Semantic Frames Considered

| | | |
|---|---|---|
| 'Motion_directional' | 'Ingestion' | 'Self_motion' |
| 'Possession' | 'Body_movement' | 'Text_creation' |
| 'Location_of_light' | 'Moving_in_place' | 'Motion' |
| 'Communication' | 'Make_noise' | 'Impact' |
| 'Perception_active' | 'Sleep' | 'Cause_motion' |
| 'Fluidic_motion' | 'Feeling' | 'Departing' |
| 'Emotion_directed' | 'Arriving' | 'Precipitation' |
| 'Awareness' | 'Statement' | 'Being_employed' |
| 'Become_silent' | 'Cure' | 'Giving' |
| 'Change_posture' | 'Perption_experience' | 'Communication_manner' |
| 'Taking' | 'Perception_experience' | 'Grooming' |
| 'Process_continue' | 'Posture' | 'Process_Stop' |

Table 26: Experiment II: Frequency of Distinctive Features in Mimetics with at Least One Significant Collocate

| Feature | Occurrences | Example Phoneme |
|---|---|---|
| [+Dorsal] | 7718 | /k/ |
| [+Consonantal] | 6491 | /l/ |
| [+High] | 5323 | /i/ |
| [+Back] | 4953 | /u/ |
| [+Son] | 4863 | /m/ |
| [+Continuant] | 4489 | /s/ |
| [+Syllabic] | 3491 | /a/ |
| [+Coronal] | 2675 | /t/ |
| [+Constricted Glottis] | 1467 | /t’/ |
| [+Round] | 1334 | /ø/ |
| [+Low] | 1333 | /æ/ |
| [+Spread Glottis] | 1242 | /$p^h$/ |
| [+Labial] | 785 | /p/ |
| [+Nasal] | 629 | /n/ |
| [+Front] | 620 | /ɛ/ |