Human language is not entirely arbitrary. Recent advances in the literature have shown sound symbolism to facilitate word-learning and acquisition cross-linguistically. Most research focuses on the segment level and not much is known about the prosody or the role it plays in the sound symbolism of tonal and non-tonal languages alike. By comparing sound-symbolic corpora with general lexicon corpora, this study shows that the sound-symbolic lexicons of three Sinitic varieties (Mandarin, Taiwanese Southern Min, Hong Kong Cantonese) are skewed so that the majority of sound-symbolic syllables are in high tone while the general lexicon is not. Building on the idea that gesture and eye movement facilitate word-learning, it is proposed that prosodic demarcation of descriptive and non-arbitrary words is analogous to a gesture in the speech stream which helps the listener (child) pinpoint its referent. This systematic mapping may be derived from the lexical embodiment of expressive voice like that found in Infant Directed Speech.

**Keywords:** sound symbolism, tone, iconicity, language acquisition, language evolution, Chinese

total page count: 20
1.0 Introduction

Sound symbolism defies de Saussure's (1916) commonly accepted dogma of language as arbitrary (where words bear no resemblance to their physical referent, e.g. 'dog'), demonstrating that features of the outside world are indeed replicated in the development and structure of certain linguistic features. How the features of the outside world are replicated in speech has been further organized according to (often overlapping) feature/function-based classifications: onomatopoeia, ideophones, size sound symbolism, etc. Recent studies have carried out experimental research into these classifications with phonological, language evolution, and acquisition-related claims in mind (see Lockwood & Dingemanse 2015 for a review). Most of these studies focus on the segment level, leaving much to be said about the prosodic level. By using corpora methods to examine Sinitic tonal inventories, this paper aims to explore the prosodic role in systematic organization and communicative strategies of sound symbolism in language acquisition and evolution.

1.1 Sound Symbolism

There are two major subcategories under the umbrella of sound symbolism. One fairly straightforward category is onomatopoeia. Hinton et al. (1994) classify this as imitative sound symbolism because onomatopoeia mimic sounds from the physical world. Take the following onomatopoeia for dogs barking: *woof woof* (English), *wàng wàng* (Mandarin), *wŏng wŏng* (Cantonese), *wan wan* (Japanese). These instances of sound symbolism are directly based on the linguistic interpretation of a physical referent (dog bark). The other major category is the ideophone, otherwise known as 'mimetic' or 'expressive'. Ideophones are based on a description of physical or emotional means (Dingemanse 2012). That is to say, they do not necessarily imitate a physical sound but a psychological one. The Cantonese ideophone *kàuh kàuh kêih kêih* describes an action done in a careless manner, it evokes a sensory or emotional depiction of its referent to both the listener and speaker, despite the lack of physical sound for acting in a careless way. Many languages, such as Japanese (Hamano 1998), Korean (Kim 1977), Mandarin (Li 2007), and Cantonese (Bodomo 2006), categorize ideophones and onomatopoeia into one sound-symbolic word class. Both onomatopoeia and ideophones are structured by how language encodes iconicity and systematicity.

1.2 Iconic Properties

Iconicity is the use of perceptuomotor analogy to connect a word to its referent (Dingemanse et al. 2015). In this way, iconicity is not arbitrary because inherent properties of the referent play a major role in determining the word's form and can be expressed through a number of modes, most of which are language-specific. Vowel quality has been shown, notably in kiki-bouba tests, to depict referent size (Bremner et al. 2013; Thompson & Estes 2011; Maurer et al. 2006) through analogy by widened or contracted dimensions of the oral cavity (Oda 2000) or pitch association (Ohtake & Haryu 2013). Repetition and reduplication has been analogically used to refer to the duration or iterative state of an action (Li 2007; Elders 2002). These are but a few examples showing that iconicity partnered

---

1 Abbreviations: HKC = Hong Kong Cantonese; TSM = Taiwan Southern Min; SS = Sound Symbolism, sound-symbolic; GL = General Lexicon; HRS = High Register Systematicity.
with phonology determines the structure of sound-symbolic words but the application and extension thereof is guided by systematic means.

### 1.3 Systematic Properties

While iconicity comprises the content of sound-symbolic words, systematicity is a guiding principle or method which demarcates a class of words from others in the lexicon (Perniss & Vigliocco 2014). Though applicable to iconicity, systematicity is grounded in arbitrariness since its function and form are not necessarily determined by perceptumotor analogies based on the referent. Systematicity aids language acquisition by providing a regular pattern for learners to create categorical correspondences between sound and meaning (Dingemanse et al. 2015: 608). How systematicity is applied seems largely language dependent. It is apparent in morphological phenomena such as verbal suffixes (-ić, -yć Polish verbs), sub-morphemic phenomena such as phonaesthemes (English: gl- gleam, glow, glisten, and sn- snigger, sniffle, snore) (Bergen 2004), and perhaps even regularities in phonological patterning (Monaghan et al. 2007). Non-segmental, prosodic, distinctions have also been shown to distinguish grammatical class (Kelly 1992; Monaghan et al. 2005). As explained later, prosody also plays a major role in the systematicity of Sinitic sound symbolism.

### 1.4 Recent Studies

Establishing iconic relationships between sound and different physical referents has been carried out in previous studies cross-linguistically (Oda 2000; Ramachadran & Hubbard 2001; Hirata et al. 2011; Nielsen & Rendall 2012, 2013; Ohtake & Haryu 2013). Recent studies take a more evolutionary and acquisitional approach by proposing how sound symbolism facilitates language acquisition and communication (Kantartzis et al. 2011; Yoshida 2012; Laing 2014; Imai & Kita 2014; Imai et al. 2015). In Sinitic, sound symbolic research is largely Mandarin-centric, descriptive, and restricted to orthographical means with emphasis on multi-syllabic phrasal structures (Wang 1992; Chan 1996; Li 2007; Chu 2012; Lai 2015; Thompson 2016). Prosody is thus a new direction for both general and Sinitic sound-symbolic research.

### 2.0 Methods

The methods in this study are based on existing corpora. One corpus of sound-symbolic words was consulted per variety of Sinitic: Mandarin, Hong Kong Cantonese (HKC), Taiwan Southern Min (TSM). See Table 1 for a description of tones in each of the Sinitic varieties examined. Corpus data for each variety’s general lexicon was used for comparative purposes. In each corpus, tokens were counted per tone, separately for the sound-symbolic class and the general lexicon. The goal of counting tokens was to examine whether the tonal distribution is different between the sound-symbolic class and general lexicon per Sinitic variety. A chi-squared test was used to compare tonal distribution between sound-symbolic and general lexicon datasets. Statistical inference was corrected for multiple comparisons using Bonferroni correction across 20 tests (1 test of general distribution per variety, as well as 4, 6, and 7 single-tone tests for Mandarin, HKC, and TSM respectively) at a threshold \( \alpha = 0.05. \)
Table 1: Descriptions of Tonal Inventory per Variety of Sinitic (Bauer & Benedict 1997: 49; Wu 2000: 2691)

<table>
<thead>
<tr>
<th>Variety</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
<th>T6</th>
<th>T7</th>
<th>T8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td>High level</td>
<td>High rising</td>
<td>Dipping</td>
<td>High falling</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hong Kong Cantonese</td>
<td>High level</td>
<td>Mid rising</td>
<td>Mid level</td>
<td>Low falling</td>
<td>Low rising</td>
<td>Low level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taiwan Southern Min (Taipei)</td>
<td>High level</td>
<td>High falling²</td>
<td>Low falling</td>
<td>Mid stopped</td>
<td>Rising</td>
<td>High falling</td>
<td>Mid level</td>
<td>High stopped</td>
</tr>
</tbody>
</table>

The Mandarin onomatopoeic corpus is Gong’s (1991) dictionary of Mandarin Onomatopoeia which consists of 188 different syllable types. The Mandarin general lexicon data is from van de Weijier and Sloos’s (2014) corpus-based survey of the 500 most common words, totalling to 759 syllables. The HKC sound-symbolic corpus was Bodomo’s (2008) A Corpus of Cantonese Ideophones which consists of 174 different syllable types. The HKC general lexicon data is taken from the Hong Kong Cantonese Adult Language Corpus cited in Leung et al. (2004: 503), consisting of 128,345 syllables. The TSM onomatopoeic corpus was Hung’s (2007) survey of TSM onomatopoeia which consists of 824 different syllable types. TSM general lexicon data is from the Taiwan Southern Min (TSM) Corpus 1.0, an open access source, compiled by Ching Chu Sun and John Newman at the University of Alberta, which consists of 9,131 syllables.

The sound-symbolic corpora cited above were designed only to show sound-symbolic utterances available per variety of Sinitic (i.e. the number of contexts in which a single given sound-symbolic word might be used). In this study, sound-symbolic tokens were counted according to syllable type instead of syllable frequency in order to avoid redundancy. For example, no matter how many times ma1 appears in a sound-symbolic corpus, it was only counted as one point in the Sound-symbolic Tone 1 category tally. In this way, the sound-symbolic corpora have been counted according to the tonal content of a variety’s sound-symbolic syllable inventory.

Conversely, counting the general lexicon per syllable type would lead to a uniform tonal distribution, since almost all tonal categories can map onto every syllable type. In this study, general lexicon tokens were counted according to frequency (i.e. how many times a syllable appeared throughout a corpus). For example, if ma1 occurs only twice in a general lexicon corpus, then only two points are awarded to the General Lexicon Tone 1 category tally. Hence, two different counting methods (sound-symbolic tokens = syllable type; general lexicon = syllable frequency) were implemented in

² Due to historical reasons T2 and T6 of Taiwan Southern Min (Taipei) have merged.
³ 73 additional syllable types were excluded. These are not sound symbolic words but adjectives and/or adverbs which exhibit a similar multi-syllabic structure to sound symbolic utterances.
⁴ Since Bodomo (2008) cites Cantonese checked tones according to their underlying form, T7 – T9 were not taken into account for this study.
As a control analysis, counting syllable frequency in Bodomo’s (2008) HKC ideophone corpus as well as Hung’s (2007) TSM onomatopoeia corpus was carried out in order to test whether the difference in counting methods skews the tonal distribution between datasets. The control analysis consisted of 538 syllables for HKC and 4,197 syllables for TSM. The same method of correcting for multiple comparisons was used in the control as well as main analysis. In the control analysis 15 tests have been corrected for (tonal distribution in HKC and TSM; additionally 6 and 7 single tone tests for HKC and TSM respectively).

### 3.0 Results

For Mandarin, the overall distribution of tones within the sound-symbolic (SS) corpus is significantly different from the distribution of tones within the general lexicon (GL) corpus (Figure 1). The relative frequency of Mandarin T1, T3, and T4 is significantly different between the SS and GL corpora (see Table 2 for $\chi^2$ test results) after correcting for multiple comparisons. Asterisks mark tests surviving correction for multiple comparisons (see also results for HKC and TSM).

**Figure 1, Tonal Distribution in Mandarin:** the overall tonal distribution is significantly different between corpora, with the majority of the SS corpus skewed to T1.

<table>
<thead>
<tr>
<th></th>
<th>SS tone count</th>
<th>GL tone count</th>
<th>$\chi^2$</th>
<th>DF</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Tones</td>
<td>188</td>
<td>759</td>
<td>147.4</td>
<td>3</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>T1</td>
<td>107</td>
<td>106</td>
<td>111.8</td>
<td>1</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>T2</td>
<td>44</td>
<td>91</td>
<td>6.84</td>
<td>1</td>
<td>n.s.</td>
</tr>
<tr>
<td>T3</td>
<td>10</td>
<td>190</td>
<td>52.47</td>
<td>1</td>
<td>&lt;.001*</td>
</tr>
<tr>
<td>T4</td>
<td>27</td>
<td>213</td>
<td>30.2</td>
<td>1</td>
<td>&lt;.001*</td>
</tr>
</tbody>
</table>

For HKC, the overall distribution of tones within the SS corpus is significantly different from the distribution of tones within the GL corpus (Figure 2). The relative frequency of HKC T1, T2, T4, T5, and T6 is significantly different between the SS and GL corpora (see Table 3 for $\chi^2$ test results). After correcting for multiple comparisons, the relative frequency of T1, T2, and T4 remains...
significantly different between the SS and GL corpora (see Table 3).

**Figure 2, Tonal Distribution in HKC**: the overall tonal distribution is significantly different between corpora, with the majority of the sound-symbolic corpus skewed to T1 and T4.

<table>
<thead>
<tr>
<th></th>
<th>SS tone count</th>
<th>GL tone count</th>
<th>$\chi^2$</th>
<th>DF</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Tones</td>
<td>174</td>
<td>128345</td>
<td>114.89</td>
<td>5</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T1</td>
<td>74</td>
<td>25193</td>
<td>57.69</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T2</td>
<td>9</td>
<td>23279</td>
<td>19.69</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T3</td>
<td>19</td>
<td>23693</td>
<td>6.57</td>
<td>1</td>
<td>n.s.</td>
</tr>
<tr>
<td>T4</td>
<td>47</td>
<td>15869</td>
<td>34.36</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T5</td>
<td>6</td>
<td>15093</td>
<td>11.58</td>
<td>1</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>T6</td>
<td>19</td>
<td>25218</td>
<td>8.39</td>
<td>1</td>
<td>&lt;0.05</td>
</tr>
</tbody>
</table>

For TSM, the overall distribution of tones within the SS corpus is significantly different from the distribution of tones within the GL corpus (Figure 3). The relative frequency of TSM T2, T4, T5, and T8 is significantly different between the SS and GL corpora (see Table 4). All these tests survive correction for multiple comparisons.

**Figure 3, Tonal Distribution in TSM**: the overall tonal distribution is significantly different between corpora.
Table 4: $\chi^2$ test result for TSM datasets

<table>
<thead>
<tr>
<th></th>
<th>SS tone count</th>
<th>GL tone count</th>
<th>$\chi^2$</th>
<th>DF</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Tones</td>
<td>824</td>
<td>9131</td>
<td>296.67</td>
<td>6</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T1</td>
<td>164</td>
<td>1512</td>
<td>6.04</td>
<td>1</td>
<td>n.s.</td>
</tr>
<tr>
<td>T2</td>
<td>92</td>
<td>2402</td>
<td>92.28</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T3</td>
<td>102</td>
<td>977</td>
<td>2.2</td>
<td>1</td>
<td>n.s.</td>
</tr>
<tr>
<td>T4</td>
<td>150</td>
<td>1156</td>
<td>20.38</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T5</td>
<td>70</td>
<td>1405</td>
<td>28.44</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T7</td>
<td>98</td>
<td>1174</td>
<td>0.63</td>
<td>1</td>
<td>n.s.</td>
</tr>
<tr>
<td>T8</td>
<td>148</td>
<td>505</td>
<td>190.54</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

3.1 Control Analyses

For the HKC control analysis, tonal frequency plus syllable frequency was counted for SS syllables rather than tonal frequency according to syllable type. The overall distribution of tones within the SS corpus is significantly different from the distribution of tones within the GL corpus (Figure 4). The relative frequency of HKC T1, T2, T3, T4, T5, and T6 is significantly different between the SS and GL corpora (see Table 5). After correcting for multiple comparisons, the relative frequency of T1, T2, T4, T5 and T6 remains significantly different between the SS and GL corpora (see Table 5). T3 is only significant at an uncorrected threshold of $\alpha=0.05$. Asterisks mark tests surviving correction for multiple comparisons.

Figure 4, Control Analysis in HKC: results show that the two different counting methods implemented do not affect the significance of results. As with Figure 2, the overall tonal distribution here is significantly different between corpora, with the majority of the SS corpus skewed to T1 and T4.

Table 5: $\chi^2$ test result for HKC control analysis

<table>
<thead>
<tr>
<th></th>
<th>SS tone frequency count</th>
<th>GL tone count</th>
<th>$\chi^2$</th>
<th>DF</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Tones</td>
<td>538</td>
<td>128345</td>
<td>264.22</td>
<td>5</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T1</td>
<td>201</td>
<td>25193</td>
<td>106.47</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T2</td>
<td>41</td>
<td>23279</td>
<td>39.99</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>
Similarly, for the TSM control analysis, tonal frequency plus syllable frequency was counted for SS syllables rather than tonal frequency according to syllable type. The overall distribution of tones within the SS corpus is significantly different from the distribution of tones within the GL corpus (Figure 5). The relative frequency of TSM T1, T2, T4, T5, T7, and T8 is significantly different between the SS and GL corpora (see Table 6). Thus, the significance of T1 and T7 is dependent on the counting method implemented. After correcting for multiple comparisons, the relative frequency of T2, T4, T5, and T8 remains significantly different between the SS and GL corpora (see Table 6). Asterisks mark tests surviving correction for multiple comparisons.

**Figure 5, Control Analysis in TSM:** results show that the two different counting methods implemented do affect the significance of results. As with Figure 3, the overall tonal distribution here is significantly different between corpora, however T1 and T7 are now significant after counting the SS corpus according to syllable frequency.

**Table 6:** $\chi^2$ test result for TSM control analysis

<table>
<thead>
<tr>
<th></th>
<th>SS tone frequency count</th>
<th>GL tone count</th>
<th>$\chi^2$</th>
<th>DF</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Tones</td>
<td>4197</td>
<td>9131</td>
<td>1720.33</td>
<td>6</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T1</td>
<td>1113</td>
<td>1512</td>
<td>180.34</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T2</td>
<td>317</td>
<td>2402</td>
<td>622.7</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T3</td>
<td>462</td>
<td>977</td>
<td>0.28</td>
<td>1</td>
<td>n.s.</td>
</tr>
<tr>
<td>T4</td>
<td>804</td>
<td>1156</td>
<td>96.74</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T5</td>
<td>181</td>
<td>1405</td>
<td>336.38</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T7</td>
<td>425</td>
<td>1174</td>
<td>20.31</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>T8</td>
<td>895</td>
<td>505</td>
<td>762.99</td>
<td>1</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>
Figure 6, Relative Percentage Across Corpora: Tone registers scoring at 1 indicate that the percentage of said tone register is equal for both SS and GL corpora. Tone registers scoring above 1 indicate that the percentage of a given tone register is higher in the SS corpus than GL corpus. Scoring below 1 indicates that said tone register has a lower percentage in the SS corpus than GL corpus. The high tone register is the only register which is consistently more common among SS corpora than GL corpora cross-linguistically.

Table 7: Percentage of Tonal Distribution across Corpora According to Register

<table>
<thead>
<tr>
<th>Tone (register)</th>
<th>General Lexicon</th>
<th>Sound-symbolic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mandarin</td>
<td>HKC</td>
</tr>
<tr>
<td>High</td>
<td>18%</td>
<td>20%</td>
</tr>
<tr>
<td>Rising</td>
<td>15%</td>
<td>18%</td>
</tr>
<tr>
<td>Mid</td>
<td>N/A</td>
<td>18%</td>
</tr>
<tr>
<td>Falling</td>
<td>36%</td>
<td>12%</td>
</tr>
<tr>
<td>Low</td>
<td>32%</td>
<td>32%</td>
</tr>
</tbody>
</table>

4.0 Discussion

Across three major Sinitic varieties, the majority of SS syllables are expressed in high register tones (Figure 6 and Table 7). This skewed tonal distribution is a product of systematicity. That is to say, the high register acts as identifying marker, or a pattern, systematically demarcating SS words from non-sound-symbolic words. Through iterated learning, this pattern thus signals the SS status of the utterance. This is not to say that systematicity is absolute or exclusive to phonology. This will also be addressed in the following sections. In addition, the origins of this high register systematicity (HRS) will be posited from both a phonological and evolutionary perspective. This prosodic systematicity, separate from segment-level systematicity, has a strong referential potential, much like gesture and eye movement, in the Bootstrapping Hypothesis for language acquisition and evolution (Imai & Kita 2014). The evolution of HRS itself reveals what linguistic stages exist between iconic and arbitrary on the continuum of arbitrariness. Finally, how prosodic systematicity, and what I call the Orthographic Opacity problem, could influence results of (past and future) SS research will be also discussed.
4.1 Tiers of Sound-symbolic Structure

Sinitic SS structure is not limited to phonology because it generally conforms to the syllabic structure of the variety in which it is found (Bauer 1985; Li 2007; Chu 2012). The segmental level supplies most of the iconic information associated with the referent while the prosodic level further conveys its SS status through HRS (Figure 7). Potential for expressivity is maximal here, in that Sinitic is using all available tiers (segmental, suprasegmental) to convey meaning.

![Figure 7, Sound-symbolic Tiers: The upper rectangle represents the prosodic level, which conveys systematicity, and the lower rectangle represents the segmental level, which conveys the iconic information determined by the referent. In non-SS words, both tiers are used for lexical purposes only.](image)

Prosody is ideal for systematic purposes as it neither interferes with nor compromises iconic information at the segmental level. Further subdivision of systematic domains is left free to occur at the segmental level. For example, SS words to do with unobstructed airflow (wind, breathing, laughter) are often /x/ initial in Mandarin (Li 2007). Size SS, like that of kiki-bouba tests in the literature, is also free to be extrapolated. Hence, this dual tier structure (Figure 7) allows for efficiency in signal which alerts the listener to the status of an utterance as SS through two simultaneous linguistic channels of processing.

There are, of course, exceptions. Systematicity is not absolute and not all SS words show HRS. The falling tone makes up the other majority of SS words in HKC. There are two potential reasons why this is the case: either 1) these prosodic exceptions possess iconic properties that somehow exempt them from HRS, or 2) the prosodic exception is not iconic but lexical and used to distinguish one word from another which is segmentally similar but semantically different.

---

5The author proposes that cross-linguistic size SS is not derived from iconic relationships expressed through oral cavity size but a systematic extrapolation of acoustical (vowel) associations previously acquired in onomatopoeia. Often smaller animals are assigned onomatopoeia with higher vowels (mice, cats, birds) than larger animals (dogs, cows, lions). Ohtake & Haryu (2013) findings support this as pitch was shown to be a stronger correlate than oral cavity size for size SS. Moreover Fernández-Prieto et al. (2015) found that six-month-olds make this pitch size association while four-month-olds do not.
### 4.2 Bootstrapping

Based on a number of acquisitional and perceptual SS studies, Imai and Kita (2014: 4) outline five claims in their Bootstrapping Hypothesis, the fourth and fifth of which are most relevant here: "(4) Sound symbolism helps infants associate speech sounds and their referents and establish a lexical representation. (5) Sound symbolism helps toddlers identify referents embedded in a complex scene." The present study shows that three varieties of Sinitic have maximalized said referential bootstrapping process by systematically lexicalizing the majority of SS words in high tone, known here as HRS, which is not attested in the general lexicon of these Sinitic varieties. This cross-linguistic trend shows that certain prosodic features can be systematically lexicalized to aid language acquisition. The consistent application of high tone may help the infant connect (SS) segments of the speech stream to the referent. Similar comprehension-oriented prosodic phenomena, though not lexicalized, have been attested in studies examining Infant Directed Speech (Nygaard et al. 2009; Herold 2011a, 2011b) as well as adult word learning (Reinisch et al. 2013). It seems unlikely that HRS alone would allow infants to make the referential connection of the Bootstrapping Hypothesis. Thus bootstrapping is a multi-modal approach: gesture (pointing, motion) and eye movement work together with HRS to aid in the process of language acquisition. HRS can be likened to a gesture presented in the speech signal which, by prosodically highlighting an utterance, aids in the connection the learner (i.e. child) must make to the referent. But where could HRS have originated from?

### 4.3 HRS and Language Evolution

SS and the use of SS cross-modal mapping has been proposed as a first step in the process of language evolution (Kita 2008; Kita et al. 2010). Studies have also shown that iconicity undergoes systematic patterning after acquisition and subsequent transmission to new speakers (Verhoef et al. 2015). Therefore one might assume that the iconic information, at the segment level, came before the prosodic systematicity of HRS. Even so, how did HRS come into play? There are at least two potential answers to this question: 1) HRS was initially iconic (e.g. high register depicts smallness) before eventually being regularized to its current systematic status, or 2) HRS was once a feature of expressive voice, with no iconic connotations, before eventually fossilizing into its current systematic status.

Given that HRS is present in diachronically divergent and distinct varieties of Sinitic, it seems unlikely that HRS evolved from some (universal) association between SS and tone. The absence of SS at the current prosodic level also fails to support this hypothesis. Moreover, this assumes that at one point the prosodic tier detached from the iconicity of the whole SS word and evolved into its current systematic tier. Though the systematic application of iconic features in language is well attested (Thiesen et al. 2010; Imai and Kita 2014; Imai et al. 2015; Dingemanse et al. 2015; Verhoef et al. 2015), it is worth noting that this split would have had to occur very early on, before Sinitic would have branched into fully-fledged and independent varieties, to remain diachronically consistent. Otherwise how could one explain how such a split might have occurred across three varieties separately without minimal interaction?
A more likely answer to the development of HRS lies in our second choice, that which is driven by an emphatic or expressive voice. Indeed, Dingemanse (2013) notes that most ideophones are prosodically emphasized in natural speech. Considering that SS is found in mother-to-child speech (Suzuki 2013; Saji & Imai 2013; Akita 2009; Ogura 2006; Fernald & Morikawa 1993), let us consider Infant Directed Speech (IDS) as catalyst for HRS. IDS is known to consist of higher pitch (Herold et al. 2011a). In order to discuss how HRS evolved from IDS, the hypothetical first generation of speakers to develop HRS will be dubbed proto-speakers. Likewise, the SS utterances used prior to the development of HRS will be proto-SS. For the first proto-speakers to develop HRS, proto-SS words were already sufficiently iconic in their segmental features alone, requiring no further prosodic distinction. In this way, their tonal composition was secondary to their segmental composition. Proto-SS tone did not referentially associate or contribute to meaning in the way that the segmental level did. However, because Sinitic languages are tonal, over time infants assigned a stable (or default) tone to this flexible SS prosodic tier. This process was gradual, happening as said proto-infants grew attuned to the phonological properties of their first language. Supposing that frequency has a hand in language acquisition, the assigned default tone would be one frequently associated with SS utterances. IDS is the vehicle that propelled this frequent association to SS words since IDS consists of higher pitch. High register assignment became regularized as the infant continues to grow and acquire language. Eventually this high register fossilises as more non-SS words enter the child’s lexicon and SS words fall out of regular use. Due to the once tonally ambiguous nature of these proto-SS utterances, adult speakers would find no reason to correct the systematic application of high register to proto-SS words in proto-child speech. As the language was transmitted, HRS is fossilized and applied as a prosodically gestural tool used to facilitate the bootstrapping process.

IDS seems more likely the catalyst (as opposed to an iconicity-based trigger) behind this systematic and uniform patterning across three varieties of Sinitic given that Sinitic is tonal and IDS is integral to language acquisition (Herold et al. 2011a, 2011b). Following on from the IDS-driven hypothesis above, it looks as though prosody and IDS go hand in hand when it comes to bootstrapping and SS. As SS in language processing is an emerging field of research, it remains to be seen whether the prosody of IDS could influence conclusions drawn from experimental findings in the literature. Future studies should take into account IDS prosody when conducting SS tests. Finally, given that prosodic patterning is evidenced in Sinitic SS, future studies should also investigate whether or not such patterning exists in other languages, as such phenomena could influence perceptual responses in future experiment-driven papers.

Discrepancies between SS judgement results cited in recent studies (Forte et al. 2013; Ozturk et al. 2012) might be down how participants perceive the prosodic differences in these trials. How would judgements change if pitch were altered? The prosodic presentation of stimuli could sway participants’ judgements as to whether or not stimuli are SS (Mitterer et al. 2012). With speakers of Sinitic, stimuli presented in high register seem likely to influence SS judgements. Future studies should investigate this as well as the potential for prosodic systematicity in other languages. To do this, the effect prosodic variation has on SS judgements in other tonal and non-tonal languages should be investigated, making sure to provide detailed descriptions of prosodic stimuli used. This will deepen the current understanding of iconicity, systematicity, and SS alike.
As other studies have shown, prosodic phenomena such as expressive voice (Herold et al. 2011b; Jesse & Johnson 2012; Reinisch et al. 2013) as well as IDS (Herold et al. 2011a) are used to facilitate language acquisition and comprehension. As SS has recently been shown to be an important feature of language acquisition (Imai et al. 2014; Asano et al. 2015; Lockwood et al. 2016), future studies should investigate the formative or acquisitional relationship between IDS, expressive voice, and SS, as the findings of the present study implicate. Not only will this advance the subfield of SS, but also open up new directions for research into language evolution, as IDS, expressive voice, and SS support an interface for language development.

4.4 Note on Orthographic Opacity

Studies indicate that prosodic phenomena, not accounted for by orthography, facilitate word learning (Nygaard et al. 2009; Reinisch et al. 2013; Herold et al. 2011a, 2011b). Other studies (Kelly 1992; Monaghan et al. 2005) show that prosody is used to systematically class grammatical categories. Thus, linguistic research cannot draw firm conclusions from orthography alone. Sinitic orthography, likewise, is opaque as the segmental and prosodic levels are not apparent. Moreover, certain Sinitic SS have no orthographic equivalent (Bauer 1985). The present study has worked around this obstacle through less opaque (Romanization) systems which reveal both segmental and prosodic information. The HRS phenomenon reported here is, however, not the end for systematicity in Sinitic SS. While Romanization systems (Yale, Pinyin, Peh-oe-jí) gloss high tone as orthographically identical to non-sound-symbolic high tones, it is possible that the SS instances exhibit marked differences in the natural speech stream. These differences could further contribute to acquisitional and bootstrapping process (e.g. HRS is louder or articulated at a higher frequency than non-sound-symbolic high tones). Subsequent investigation is needed to bypass opacity by referring directly to acoustic measurements of the speech stream.

5.0 Conclusion

The results of the present corpora analyses show that high register tone systematically marks a majority of SS words in Mandarin, Taiwan Southern Min, and Cantonese. This high register systematicity (HRS) is thought to facilitate language processing by prosodically alerting the listener to a word’s SS status while simultaneously allowing iconic information to be conveyed at the segmental level. HRS thus divides a SS word into two tiers where the systematic information is supplied at the prosodic level (systematic tier) and the iconic or SS information at the segment level (iconic tier). Moreover, through the Sound Symbolism Bootstrapping Hypothesis (Imai and Kita 2014), HRS has been posited as a prosodically gestural tool which helps learners (children) make the referential connection between a SS word and its meaning. Infant Directed Speech (IDS) is posited here as the evolutionary catalyst for such a uniform patterning, calling into question the influence of IDS on SS development cross-linguistically.

SS research is a key point of departure in language cognition for several reasons: it is prevalent in the acquisition process, it facilitates word-learning, and is known to have cross-linguistic tendencies. Given its groundings in language development, empirical SS research has the potential to lead the
discussion of how language evolves. Further research into the iconic relationships between language and referents will shed light on how SS shapes fundamental language development. Moreover, by harnessing the systematic relationships present in SS, we should come closer to developing an understanding of the underlying mechanisms which drive communication towards arbitrariness.
Acknowledgements:
Many thanks to Dr Kirsty Rowan, Dr Ryszard Auksztulewicz, Emily Lake, and Yang Li.
References


