Towards An Explanation of Size Sound Symbolism: Correspondences in Onomatopoeia

keywords: sound symbolism, iconicity, size, onomatopoeia, language evolution, kiki-bouba

Introduction

Analogy is a key function of language. It establishes links within the lexicon which allow speakers to scaffold meanings and acquire language. Likewise, analogy facilitates the linguistic encoding of non-linguistic features, such as sounds or gesture. Despite de Saussure’s (1916) well-cited dogma of language as arbitrary, a growing body of literature shows that some aspects of language are grounded in iconicity. That is, the structure of certain words and linguistic representations is determined by their physical referent. These linguistic representations fall under the umbrella term sound symbolism, which in turn encompasses several (often overlapping) feature/function-based categories: onomatopoeia, ideophones, and magnitude sound symbolism (cf. Hinton et al. 1994 for further classification), all of which are grounded in iconicity (Dingemanse et al. 2015). The scope of this paper focuses on magnitude sound symbolism (MSS), otherwise known as size sound symbolism, where the size (and arguably shape) of a referent determine its phonetic representation. This form of iconicity has often been investigated by teasing out analogical relationships between physical referents and language through behavioral experiments (Lockwood & Dingemanse 2015), namely kiki-bouba tests. These tests have demonstrated an iconic relationship cross-linguistically between physical appearance and novel linguistic representation: high vowels and front vowels are more likely to denote smallness while back vowels and low vowels typically denote bigness. Although some linguists have posited various avenues of explanation, mainly falling under neural (Ramachandran and Hubbard 2001; Westbury 2005; Ozturk et al. 2013), gestural (Oda 2000; Fort et al. 2013), and learned (Mauer et al. 2006; Thompson and Estes 2011; Nielsen and Rendall 2013) categories, none so far have been able to tease apart why one is more suitable than the next. Moreover, many such claims neglect to concretely discuss the motivation behind how such an explanation may have initially occurred or developed. This paper follows the learned approach where the robust MSS relationships exhibited by the kiki-bouba paradigm are extrapolated from pre-existing, but not innate, linguistic knowledge. Using a preliminary corpus of cross-linguistic onomatopoeia, this paper posits that extrapolation of acoustic information in size sensitive onomatopoeia could be the source of robustness in kiki-bouba findings.

Overview of Kiki-Bouba Tests

Kiki-bouba tests, originally introduced by Koehler (1929) as takete malouma tests, are conducted by asking participants to match novel words or syllables to novel objects. These objects are contrastive in that one is markedly angular or spiky and the other circular or round. Likewise, the novel words are contrastive in that one stimuli contains high or front vowels with voiceless consonants while the other contains low or back vowels with voiced consonants. Across different languages and studies, participants consistently match the angular object to the voiceless C +
high/front V word and the round object to the voiced C word + low/back V. Though kiki-bouba literature often refers to shape (angular vs. round) sound symbolism, this is essentially MSS as the angles of the visual stimuli vary according to size (acute vs. obtuse) and are the main contrastive features between them. Studies all follow this object-to-sound matching paradigm, but are often conducted along different variables.

**Presentation of Stimuli**

Novel words are either given orthographically (Ramachandran and Hubbard 2001; Westbury 2005; Thompson and Estes 2011; Nielsen and Rendall 2012; Nielsen and Rendall 2013) or auditorily (Mauer et al. 2006; Bremner et al. 2013; Fort et al. 2013; Ohtake and Haryu 2013; Ozturk et al. 2013; D’Onofrio 2014). Visual stimuli are not uniform across studies, however, they are consistent in their contrasts (angular vs obtuse). Most studies present visual stimuli simultaneously, which has been criticized as emphasizing the contrastive nature of the experiment, though Nielsen and Rendall (2013) and Thompson and Estes (2011) both presented them sequentially and still showed the MSS relationship to be significant. Nielsen and Rendall (2013) also allowed participants to choose two syllables from a selection of eight when naming an image, indicating a strong preference for MSS segments. Recent infant studies involved eye-tracking as images were displayed on screen (Ozturk et al. 2012; Fort et al. 2013), while the toddler task in Mauer et al. (2006) involved a series of forced choices during interactive play with a researcher.

**Kiki-Bouba Findings and Analyses**

The main findings of the literature are that high and front vowels as well as unvoiced consonants correspond to smallness, while low and back vowels as well as voiced consonants correspond to bigness. It has been shown that procedure does affect MSS but does not determine its realization (Nielsen and Rendall 2013; Thompson and Estes 2011). Bremner et al. (2013) has shown that orthography does not interfere with MSS association when testing Himba speakers in Namibia who do not use a written language. D’Onofrio (2014) further showed that when using everyday objects as visual stimuli participants preferred MSS naming.

Various approaches have been proposed to explain MSS. While none of them are officially defined or unified, they can be grouped roughly into three categories: gestural, neural, and learned. Gestural approaches claim that MSS is a result of using the oral cavity to mimic the size or shape of the referent in question (Oda 2000; Fort et al. 2013). While this sounds reasonable, given that gesture is an important feature of language and acquisitional processes, Ohtake and Haryu (2013) found that oral cavity size did not facilitate MSS in participants, and in fact found that pitch was a stronger correlate. Fernández-Prieto et al.’s (2015) findings corroborate this.

Neural approaches are considerably less grounded empirically. The large implications seem to be that MSS is innate or inherent to the activity of the brain, or how cognitive processes allow us to ‘perceive’ the world (Ramachandran and Hubbard 2001; Nielsen and Rendall 2012). The problem with the neural approach is a circular one, much like the classic conundrum “which came first the chicken or the egg?” It seems unlikely that parts of the brain are responsible for an inherent MSS instinct. They might facilitate the MSS process or may be active during the MSS process, but they might also do the same for other linguistic processes. Current research has failed to address this.
And even so, how does marked cognitive activity rule out that MSS is a learned process? Moreover, inconsistent age-based findings (see below) do not support the neural notion that MSS is an inherent cognitive process (Maurer et al. 2006; Ozturk et al. 2012; Fort et al. 2013; Fernández-Prieto et al. 2015).

The learned approach holds that MSS associations are, of course, learned or extrapolated from other linguistic information (e.g. other lexical items). Though how and why this extrapolation takes place has yet to be fleshed out in the current literature. The age-related studies mentioned below support the learned explanation, as some pre-lexical infants are unable to make MSS associations before 6 months. Implied that, the lexical information or learning is not yet sufficient enough for infants to make size–sound extrapolations.

The most important findings, from an acquisitional standpoint, are age-related. Maurer et al. (2006) have solidly shown that MSS is present in toddlers aged 2.5 years old. But findings for pre-lexical infants are more ambiguous, with results conflicting across languages. While Ozturk et al. (2012) found evidence of MSS in 4-month-old children of Turkish speaking parents, Fort et al. (2013) failed to find evidence of MSS associations in 4-month-old children of French speaking parents. Fort et al.'s (2013) findings are supported by Fernández-Prieto et al. (2015) who found that while 6-month-olds were able to make pitch-based MSS association, 4-month-olds were not. These findings implicate that MSS is precarious, the acquisition processes of certain languages (e.g. Turkish) may facilitate MSS extrapolations earlier than others. This, of course, requires further phonological and acquisitional investigation longitudinally to determine what is acquired first per language. Nevertheless these results clearly demonstrate that MSS is not the product or result of a specifically innate cognitive function.

Hypothesis

But what lexical information is sufficient to induce this analogical extrapolation of size and sound? The hypothesis here is that size and sound relationships evident in onomatopoeia are a starting point. That is to say, as small animals generally make higher pitch noises, while larger animals make lower pitch noises (Masuda 2002), size should be thus reflected accordingly in onomatopoeic interpretations cross-linguistically and along the lines of the kiki-bouba paradigm. In this way, it is supposed that the MSS relationship is learned during or after lexical onomatopoeic inventories are acquired. Empirical basis for this hypothesis is found in Masuda’s (2002: 141-166) phonetic analysis of English onomatopoeia, where the size of an animal often determines the quality of sound that it makes, this, in turn, can determine the vowel quality of some English onomatopoeia.

Aside from the natural physical size to sound correlation, and MSS associations noted by Masuda (2002), sound symbolic utterances (including onomatopoeia) are also notably present in Infant Directed Speech (Fernald and Morikawa 1993; Ogura 2006; Akita 2009; Saji & Imai 2013; Suzuki 2013). Moreover, sound symbolism has recently been shown to ease language acquisition (Imai et al. 2014; Asano et al. 2015; Lockwood et al. 2016) and is the basis of an acquisitional bootstrapping mechanism (Imai and Kita 2014). Given these findings and that the learned explanation is based on extrapolation from the infant lexicon, onomatopoeia is an obvious point of departure.

1 What could be argued is that the process of analogizing is cognitive, and possibly innate, but this does little to reveal the origins or evolution of MSS.
Methods

The corpus for this paper was compiled via an online survey in which participants were asked to answer the questions “What sound(s) does a(n) [animal name] typically make?” for a list of X different animals by filling in textboxes. A total of 23 questions were asked. Three pairs of size distinctions were overtly questioned (large dog vs. small dog, cat vs. kitten, and songbird vs. chick). 11 questions allowed for two answers. To narrow down the amount of onomatopoeia selected for the survey, questions were limited to 20 types of animals and insects: dog, cat, cow, horse, lion, crow, songbird, pigeon, duck, pig, bear, cricket, frog, mouse, housefly, mosquito, rooster, chicken, sheep, and bee. Participants were also asked to provide their age, native language, and the native languages of their parents and/or carer. Participants filled out the survey according to the lexical representations of onomatopoeia their native language. The following 26 languages were reported: Bahasa Indonesia, Hong Kong Cantonese (2), Shanghai Chinese, Taiwan Southern Min, Serbian, Tagalog, Thai, Faroese, German (4), Swedish, Portuguese, Latvian, French, Italian (3), Japanese (2), Spanish (2), Mandarin, Danish (2), English, Hungarian, Icelandic, Dutch, Norwegian, Ukrainian, and Polish. The average age of participants was 26.8 yrs with a range of 18-49 yrs. In all cases the participants’ native language was consistent with all their parents/carers, except for one Cantonese speaker who reported one parent/carer as an English speaker; one Thai speaker and one Faroese who reported one parent/carer as Chinese speakers respectively; and one Norwegian speaker who reported one parent/carer as an Arabic speaker.

Responses were asked to provide responses in their native prosody if possible. Responses were then converted into IPA based on phonological accounts. Where orthographic representations were unclear, participants were asked to clarify in order for IPA notation to be accurate. Results are thus preliminary as transcriptions were considerably broad and not phonetic. Though it should be noted that phonetic transcription would vary depending on prosodic context, something which would be challenging to keep consistent across participants.

Results

Cross-linguistically, there are many correspondences between animal sounds with respect to vowels and consonants, e.g. horse, frog, and pigeon. While these cross-linguistic correspondences provide a solid picture of onomatopoeic tendencies, they are not the focus here. The main concern in this study is the intra-linguistic (within a language) results and whether intra-linguistic systematic patterning is found cross-linguistically. The point of departure here is to see if animal sounds coincide with the MSS paradigm. That is to say, are small animal sounds encoded with high and front vowels and/or unvoiced consonants? Are big animal sounds encoded with low and back vowels and/or voiced consonants? Animal onomatopoeia in and of themselves cannot be governed by the MSS paradigm as they are derived through imitation of physical sounds and not physical size. However, the sound which an animal is physically capable to make can be determined by its size. This is what is hypothesized to set the stage for the MSS paradigms found in kiki-bouba tests.
large dog
small dog
cat
kitten
cow 1
cow 2
horse 1
horse 2
lion 1
lion 2
songbird
crow
pigeon
duck
pig 1
pig 2
bear 1
bear 2
chick 1
chick 2
cricket 1
cricket 2
frog 1
frog 2
mouse 1
mouse 2
mosquito
rooster
chicken
sheep 1
sheep 2
bee 1
bee 2
<table>
<thead>
<tr>
<th>Source Languages</th>
<th>Japanese</th>
<th>Japanese</th>
<th>Latvian</th>
<th>Mandarin</th>
<th>Norwegian</th>
<th>Polish</th>
<th>Portuguese (Portugal)</th>
<th>Serbian</th>
<th>Shanghai Wu</th>
<th>Southern Taiwan Min</th>
<th>Spanish 1</th>
<th>Spanish 2</th>
<th>Swedish</th>
<th>Tagalog</th>
<th>Thai</th>
<th>Ukrainian</th>
</tr>
</thead>
<tbody>
<tr>
<td>large dog</td>
<td>waɯ waɯ</td>
<td>waɯ waɯ</td>
<td>vaɯ vaɯ</td>
<td>waʃ waʃ</td>
<td>waʃ waʃ</td>
<td>wɐw wɐw</td>
<td>ab ab waʃ waʃ waʃ waʃ</td>
<td>uf uf uaf uaf uaf uaf</td>
<td>aw aw</td>
<td>hoŋ hoŋ</td>
<td>gaf</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Vowels provide a relatively navigable picture. A relationship between smallness and high vowels is found in songbird, chick, cricket, and mouse responses for 11 languages. And this smallness is often opposed by a relationship between bigness and low back vowels found in dog, cow, lion, bear, and pig responses in those same languages. The relationship between smallness and unvoiced consonants is found in chick, mouse, cricket, and songbird responses for 11 languages. And this smallness is often opposed by a relationship between bigness and voiced consonants found in dog, cow, lion, and crow responses in those languages.

None of the languages show absolutely zero correspondence to the MSS paradigm. Similarly, none of the results show an absolute MSS paradigm across all animal sounds. Responses associated to the horse sound are indicative of this, whereby a large animal is depicted using sounds associated with smallness in the MSS paradigm. In fact, each language has at least one response which partially (a mixture of small and big sounds are used, including diphthongs) or completely (small sounds are used for a big animal, e.g. horse) flouts the MSS paradigm in some dimension.

While all languages exhibit MSS associations in single responses, only some exhibit them within same-species size pairings. Bahasa Indonesia, Shanghai Wu, French, German, English, Dutch, and Japanese responses to cat and kitten coincide with the MSS paradigm. Portuguese, French, English, and Japanese responses to large dog and small dog coincide with the MSS paradigm. Interestingly, there is a (partial) MSS paradigm reflected in Finnish, French, Spanish, Portuguese, Icelandic, Japanese, Thai, German, and English responses to songbird and chick. No responses indicated an MSS paradigm for the rooster-chicken pairing.

Bahasa Indonesia, Cantonese, Italian, Thai, and English responses to mosquito and bee coincide with the MSS paradigm, where mosquito onomatopoeia are small sounds and bee sounds are big.

**Discussion**

The results of this paper demonstrate that there is a consistent and systematic MSS paradigm found across animal onomatopoeia within a language. This supports the learned approach that size and sound relationships are not innate but generalizable from learned linguistic information. Same-species animal pairings with size differences (songbird/chick, large dog/small dog, cat/kitten) also exhibited onomatopoeic differences along the MSS paradigm. Despite noticeable exceptions to the size-sound assignment, it should be noted that the learned approach does not require absolute consistency to be fulfilled as its main prediction is that even an imperfect contingency in the speaker’s environment (animals’ size and sound) can be extracted and extrapolated onto an unrelated domain (abstract shapes). Exceptions and how they might be navigated by the child while formulating the MSS paradigm will be discussed below. If anything, these exceptions are the most valuable insights into how onomatopoeia inform the analogical extrapolation needed to learn the MSS paradigm.

Exceptions to the MSS paradigm in the present corpus seem to be for the following reasons, either 1) syllabic contrast for rhythmical purposes, e.g. Finnish (horse) /hi: ha/, Hong Kong Cantonese (chick) /tsi: tsi: tsa:/; or 2) segment choice is deemed more archetypal for the animal sound and is thus preferred over segments which would fulfill the MSS paradigm, e.g. Mandarin (pigeon) /ku ku/ vs. *kiki or Polish (bee) /bz/ vs. *ps. The fact that infants have the potential to extrapolate MSS despite these confounds implies that either 1) the infants learn to systematically work around said confounds via some other means, such as extrapolation from non-onomatopoeic words, like
itsy-bitsy, little, teeny-weeny, mini etc.; or 2) the extrapolation of highness and frontness or voicelessness to smallness is learned first, and other associations are the default follow-up, e.g. anything not high and front or voiceless is more likely to be associated with bigness; or 3) a combination of (1) and (2). All options allow for speakers to navigate confounds evidenced in this corpora analysis while still producing the MSS results accumulated in the literature.

But why or how might the extrapolation of highness be learned first? A possible explanation, not tested in this paper, is that the MSS association of highness to smallness is actually that of pitch derived through prosody. Indeed, high pitch is prevalent throughout Infant Directed Speech (IDS) (Herold et al. 2011; Jesse & Johnson 2012; Reinisch et al. 2013). This possibility is supported by Fernández-Prieto et al.’s (2015) findings which showed that high tones were preferred by 6-month-olds for smaller visual stimuli and low tones for larger. Moreover, tone sandhi exclusive to IDS in Cantonese change lexically non-high tones to high regardless of vowel quality (Matthews and Yip 2011). It could be that vowel quality and prosody combined are what allow for extrapolation which leads to an MSS paradigm. Vowel quality must have some role given the results of previous kiki-bouba tests. Perhaps an association of smallness to high level prosody would first emerge through one extrapolation followed by subsequent extrapolation of the high pitch perceived in high and front vowels or unvoiced consonants. Whether vowel quality is more prominent than prosody in the formation of MSS would require more testing and investigation into the infant lexicon. As it stands, kiki-bouba tests in the literature have neglected to test prosody.

As a preliminary study, this survey-generated corpus analysis is not without limitations. The number of animal sounds taken into account, while sufficient for discussion, is not exhaustive. By using real-world examples, in this case animals, it is difficult to account for how size should be ranked comparatively (see D’Onofrio 2014). However, Thompson and Estes (2011) have shown that MSS is a gradable function. More participants should be recruited and data collection should be phonetic rather than orthographic. Languages such as Faroese, Latvian, and Tagalog require more responses in order to engage in further comparison. Future studies seeking to identify MSS relationships should examine the infant lexicon for size and sound quality, including prosody, to see how great the potential for extrapolation is beyond the imitiveness of onomatopoeia. For example, a longitudinal study might show whether there are MSS parallels in the infant lexicon or whether parents use high vowels or high pitch to convey smallness in IDS. Further kiki-bouba tests should design responses that do not resemble onomatopoeia, combinations of open and closed syllables should be included to see whether codas have any effect. Finally, more languages need to undergo kiki-bouba testing as many of the languages in this survey have not.

Conclusion

The MSS paradigm is not as straightforward in onomatopoeia as it is in kiki-bouba tests. Nonetheless, the study has shown that onomatopoeia have the potential to inform the learned approach as the imitative information therein coincides with MSS both intra-linguistically and cross-linguistically. The age-based results in the literature imply that once infants reach six months of age, they have learned to extrapolate size and sound. That is to say, by six months of age, the infant lexicon is developed enough to allow for size-sound association. Given the prevalence of onomatopoeia in IDS (Fernald and Morikawa 1993; Ogura 2006; Akita 2009; Saji & Imai 2013; Suzuki 2013), the corpus analysis here confirms the presence of inherent size-sound information
in onomatopoeia which can be the driving force behind learning MSS. This paper recognizes that prosody may play a role in learning MSS alongside onomatopoeia. A parallel of MSS in onomatopoeia allows for further links to be drawn in later lexical development (MSS in other lexical items). Language has already been shown to exhibit lexical crossover with regards to sound symbolism, where sound symbolic utterances lose their imitative connotations and shift to eventually become fully functional, unmarked lexemes (Thompson 2016). This in and of itself is evidence how non-arbitrary analogies, such as those of MSS, can occur and diverge.

References


