Distribution Agreement

Signature:

Nihar M. Mathur 4/15/10

Phonetic Correlates of Sound Symbolism

by

Nihar M. Mathur

Adviser

Lynne Nygaard, PhD

Program in Linguistics

Lynne Nygaard, PhD
Adviser

Laura Namy, PhD
Committee Member

Susan Tamasi, PhD
Committee Member

Matthew Weinschenk, PhD
Committee Member

4/15/10

Phonetic Correlates of Sound Symbolism

By

Nihar M. Mathur

Adviser

Lynne Nygaard, PhD

An abstract of
A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Program in Linguistics

2010

Abstract

Phonetic Correlates of Sound Symbolism
By Nihar M. Mathur

The sound structure of spoken language is widely assumed to bear an exclusively arbitrary relationship to meaning. However, recent research into sound symbolism has shown that listeners are sensitive to sound-to-meaning correspondences that appear to occur cross-linguistically. The current study evaluated potential correspondences between inventories of certain types of phonemes and particular semantic domains. Antonyms from different semantic domains were recorded from native speakers of ten different languages. Participants with no prior knowledge of the languages were presented with the sets of antonyms and asked to guess their meanings. Broad phonetic transcriptions of words referring to size (*big/small*), contour (*round/pointy*), motion (*fast/slow*), speed (*still/moving*), and valence (*bad/good*) were analyzed to determine if the phonemic profiles would differ as a function of meaning. In general, vowel height and consonant voicing distinguished words sounding big from small, vowel roundedness and the amount of vowels distinguished words sounding round from pointy, vowel height distinguished words sounding still from moving, and vowel height, vowel roundedness, and consonant sonority distinguished words sounding slow from fast. No significant differences seemed to cue listeners to distinguish words sounding bad from good, however. These findings illustrate a reliable sensitivity to the mapping of certain speech sounds to certain semantic domains.

Keywords: sound symbolism, phonetic symbolism, arbitrariness, cross-modality

Phonetic Correlates of Sound Symbolism

By

Nihar M. Mathur

Adviser Lynne Nygaard, PhD

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Program in Linguistics

2010

Acknowledgments

Special thanks to Lauren Clepper, whose guidance, advice, and work were integral to the success of this project. Thanks to Dr. Lynne Nygaard and Dr. Laura Namy for advising and guiding this project. Thanks to Dr. Susan Tamasi and Dr. Matthew Weinschenk for their support on the defense committee. Thanks to the members of the Speech Perception Lab, Emory College Honors Program, and the Emory Program in Linguistics for their support. This work was funded in part by an Emory College Instrumentation, Bridge, Instruction, and Seed (IBIS) Grant awarded to Dr. Lynne Nygaard and Dr. Laura Namy.

Table of Contents

Phonetic Correlates of Sound Symbolism

Language acts as a conduit through which people can share information, emotions, and ideas with others. Although it varies greatly in method of transmission, written, spoken, or signed, and in form, from Albanian to Japanese, its basic function remains the same: to be communicative. Written language communicates through the use of alphabets or other types of orthographic conventions. Signed languages communicate using a complex, visuo-spatial signal, with hand gestures conveying meaning. Spoken language communicates meaning using a complex, time-varying acoustic signal. Adding to the complexity of the variety of communicative conduits that comprise language is the complex process of gleaning meaning from these various types of representational forms.

Across these myriad ways to convey meaning, there is substantial variation in the forms used to signal similar emotions, ideas, and information. Languages use different combinations or inventories of sounds to refer to particular objects or actions in the world. For example, Spanish *perro* and Hindi *kutha* both refer to the animal that, in English, is called *dog*. This variation illuminates a fundamental assumption regarding the nature of linguistic reference: the relationship between sound structure and meaning is arbitrary (de Saussure, 1916/1959). It is widely accepted that nothing about the specific sounds of a word bears any resemblance to, or inherently specifies, its referent. For example, consider a table. The prevailing assumption is that nothing about the specific sounds, or phonetic segments, used in the word *table* universally symbolize the properties of a table. The sounds of the word do not have an inherent connection to its meaning.

This arbitrariness assumption, however, has not proven to be all encompassing. Indeed, certain subsets of words are well known to have a non-arbitrary link between sound and semantic meaning. For example, onomatopoeia is by definition, a non-arbitrary connection between sound and meaning. Onomatopoetic words have sounds that reflect or resemble a sound produced by some thing, such as *boom* or *moo.* Onomatopoeia is seen cross-linguistically, even if with variations, such as *quack* in English compared with *coin coin* in French. Saussure's argument, however, posits that even these seemingly non-arbitrary forms are not sound-to-sound representations but simply reflect conventionalized correspondences rather than natural sound to meaning biases.

Phonesthemes, too, appear to exhibit seemingly non-arbitrary connections between sound and meaning. Phonesthemes are units of words, similar to morphemes, which seem to carry inherent meanings. The word-initial *gl-* cluster, for instance, has been hypothesized to have a meaning relating to flickering light when used in English words, as seen in the words *glitter, glow,* and *glisten.* Words with a meaning reflected in its sound structure—be it phonetic or suprasegmental—are considered to be symbolic of their referent. Although Saussure's sentiments are still well established among scholars, further investigation into the structure of language has shown that examples such as onomatopoeia and phonesthemes may not be the sole class of words that exhibit this sound-meaning connection. As such, the theory that there exist non-arbitrary connections between semantic meaning and aspects of the sound structure of words is known as sound symbolism.

Discussion concerning the existence of ubiquitous, cross-linguistic sound symbolism predates Ferdinand de Saussure's widely accepted theory. On the one hand, The Upanishads, a collection of ancient Hindu scriptures, contain references to relationships between certain classes of phones and nature (Müller, 1879). On the other hand, John Locke's "An Essay Concerning Human Understanding" argues against the existence of phonosemantics, positing that such a connection would necessarily require the existence of a single language (Locke, 1690). Until only recently, however, this debate has lacked systematic experimental evidence on which to base assumptions.

Indeed, modern research into sound symbolism has shown that sound-to-meaning mappings may be naturally occurring, unlike those that Saussure argues are conventionalized, such as onomatopoeia. A strong argument against the conventionalization hypothesis lies in the existence of natural phonosemantic biases cross-linguistically. By the process of their formation, words with meanings that are conventionally sound symbolic would only be recognized as symbolic by speakers of that specific language. A speaker of language A, for instance, would not judge a word in language B to be sound symbolic because they would not be sensitive to conventionalized sound symbolism relationships.

Sound-to-meaning correspondences

Recent work has provided evidence that listeners may be sensitive to certain universal sound-to-meaning correspondences. For example, studies have demonstrated that participants will label objects using novel nonwords with characteristics that resemble or correspond to features of the objects. One classic study investigated sound symbolic relationships between object labels and size of the object. Sapir's (1929) work used two tables of differing sizes as stimuli and asked participants to choose the label they felt was most appropriate for each, either *mal* or *mil.* Here, nonwords were used as stimuli so that the participants could not use any preconceived biases from their knowledge of or practice with language to influence their decisions; the words, although phonotactically probable, were seen as novel by the participants. If words are arbitrarily connected to their referents, then participants should not discriminate between the labels and should choose either *mil* or *mal* approximately equally for each of the tables. The participants, however, did appear to discriminate between the labels: 80% applied the label *mal* to the larger table. Sapir (1929) explained this result through the different phonemes used in the auditory stimuli. From his data, he deduced that the speakers could make judgments regarding an object's size based on the vowel used, with /a/ from *mal* acting as a cue for a large object. He further theorized that this magnitude sound symbolism was derived from a direct correlation between the size of the oral cavity and the size of the object: articulation of the vowel /a/ lowers the tongue, creating a larger oral cavity than that found from articulation of /i/, which raises the tongue. The speakers appeared to be sensitive to this relationship between sound, its accompanying articulatory movements, and meaning when making semantic decisions, providing evidence for the existence of sound symbolism for this visual dimension (Sapir, 1929).

Speakers also appear to be sensitive to sound symbolic relationships between sound and object contour. Originally, Köhler (1947) asked participants to label different round, smooth shapes and spiky, sharp shapes with the nonsense words *maluma* and *takete* (see Figure 1). The participants consistently judged *maluma* to be the label for the rounded shape and *takete* to be the label for the sharp shape. This study was extended by

Ramanchandran and Hubbard (2001), who used similar visual stimuli but included other novel words, such as *bouba* and *kiki.* Their study yielded similar results, with participants consistently choosing certain labels for each of the shapes. With multiple novel words, they were also able to speculate about which cues the listeners were using. They found that listeners mapped rounded vowels, such as the /u/ in *bouba*, to the smoother, rounded shapes and unrounded vowels, such as /i/ in *kiki*, to the sharper, spikier shapes at 95% consistency. Here, a connection was found between the acoustic (or phonetic) signal from the stimulus words and the visual stimuli. Ramachandran and Hubbard assert that the participants used the sound structure as the cue to make semantic judgments about the visual stimuli. The sudden, sharp changes in direction in the spikier shapes seem to imitate "the sharp phonemic inflections of the sound kiki, as well as the sharp inflection of the tongue on the palate" (Ramachandran & Hubbard, 2001).

The use of nonsense words, however, may not have entirely precluded participants from using their knowledge of and experience with English or any foreign languages to make semantic decisions. Although labels for each of the shapes included certain classes of vowels (and consonants), this did not necessarily mean that the participants did not generalize from their knowledge of English to these novel words. To address this issue, Maurer, Pathman, and Mondloch (2006) conducted a study with preschool-aged children to discover if these mappings are based on natural biases or, indeed, generalizations based on the listener's pre-existing knowledge of language. Preschoolers around 2.5 years old would not have the extensive vocabulary of adults that participated in the previous studies. Using similar auditory and visual stimuli, the study yielded similar results with the toddlers, compared to those found with adults.

Maurer et al's study also addressed the concern that participants employed orthographic representations of words to make judgments. They could, upon hearing the auditory stimulus, visualize a possible spelling of the word and use the shape of individual letters, such as the round “o,” to map to and match the shape of the visual stimulus. Thus, rather than the participants being sensitive to and using sound symbolic auditory-visual mappings to make judgments, they would be using visual-visual mappings. The preschool-aged children used in the study were pre-literate, however, and would not be able to visualize orthographic representations of the words using their knowledge of English. Interestingly, this work substantiates Sapir’s (1929) belief that sound symbolic mappings are engendered at an early age and are not entirely dependent on knowledge of orthography. Although further research into the role of orthographic knowledge in this context is necessary, it does not appear that listeners use it as a principal method of determining meaning. As such, the evidence points toward the existence of certain sound-to-meaning correlations to which listeners, even with limited language knowledge, are sensitive.

Cross-linguistic sound symbolism

Recent research in sound symbolism has begun to examine the existence of sound symbolism in extant languages as well. In one such study, Kunihira (1971) presented English-speaking participants with antonym pairs in Japanese and asked them to guess word meaning in a two alternative choice task. Participants were placed in one of three conditions, in which they either heard the Japanese stimuli produced by a native Japanese speaker in a monotone voice, the same stimuli with an expressive voice that mimicked how the words are used in normal speech, or were presented the stimuli written in Roman characters. Regardless of condition, the participants were significantly better than chance at choosing the correct word meanings, which suggests that listeners were sensitive to properties of the word forms that helped disambiguate meaning. Interestingly, however, participants who heard stimuli in expressive voice had significantly better performance than when the foreign words were presented in monotone or printed text, suggesting that participants can use various aspects of the sound structure of utterances to determine meaning in natural languages.

Other work has more directly determined the role that this expressive voice plays in providing information about words. This aspect of the sound structure is referred to as prosody. Prosody is a suprasegmental, or non-phonetic, component of the speech signal, referring to the rhythm and intonation patterns in speech. Nygaard, Herold, and Namy (2009) demonstrated that prosody carries semantic cues to meaning, using novel words in a two alternative choice decision task. Novel words, such as *blicket* and *seebow*, were recorded in infant-directed speech with either neutral or meaningful prosody. When recording the speech stimuli with meaningful prosody, the speakers were asked to produce the words as if they had a specific meaning, which was given in pairs of antonyms. For example, the word *blicket* was assumed to mean “hot” for one recording and then “cold” for another recording. Participants were shown two pictures side-by-side, each depicting one of a pair of antonyms, and the recorded stimuli were played in the frame “Can you get the _____ one?” The results indicate that the participants were significantly more likely to select pictures referring to the word’s intended meaning when meaningful prosody was used. The listeners could pick up on semantic cues embedded in the prosody intentionally produced by the speakers; the speakers, thus, were able to convey intended meaning through a potentially sound-symbolic property--prosody.

Interestingly, the existence of sound symbolic cues seems to lead naïve listeners to making correct judgments about object categories. Berlin (1994) asked individuals to judge meanings based on auditory labels for natural kind categories from an unfamiliar language. English-speaking participants were presented with words referring to birds or fish in the Peruvian language Huambisa and were asked to judge if each word was a bird or a fish name. None of the participants had any knowledge of Huambisa, yet they were better able than chance to correctly determine whether the word was a name for a bird or fish. Koriat and Levy (1979) found that Hebrew-speaking individuals were above chance at choosing the correct meaning in Hebrew for Chinese characters associated with different antonym pairs**.** Taken together, these studies suggest that sound symbolism not only exists in natural languages, but also that listeners are sensitive to these mappings cross-linguistically. The ability of listeners to correctly judge foreign word meanings above chance in several different, unrelated languages suggests that these sound-meaning biases, both segmental and prosodic, are universal and naturally-based (Berlin, 1994).

The psychological reality of sound symbolism

The existence of sound symbolism cross-linguistically and language users' sensitivity to these properties has raised the question of why non-arbitrary mappings exist in natural languages. Are sound symbolic words processed differently during language understanding, and does their existence have any implications for linguistic representation and processing? Interestingly, sound symbolism has been found to influence language processing and learning, suggesting that non-arbitrary correspondences between sound and meaning may have psychological functionality. In a vocabulary-learning study, Nygaard, Cook and Namy (2009) taught Japanese antonyms to native speakers of English. Each antonym was either matched during learning to its correct meaning, opposite meaning, or a random meaning in English. For example, the Japanese word *ue* would have been taught to participants matched to its correct meaning, "up," matched to its antonym meaning, "down," or to a randomly selected antonym meaning, such as "slow." At test, participants were more accurate and faster at identifying Japanese words that were matched during learning to the correct English meaning than for words that were randomly paired with an English meaning. Sensitivity to these non-arbitrary sound-to-meaning correspondences facilitated the processing of word meaning, influencing memory and retrieval during vocabulary learning.

The psychological functionality of sound symbolism appears to extend to children as well**.** Imai, Kita, Nagumo, and Okada (2008) found that Japanese children were sensitive to sound-symbolic mappings between actions and words and that this sensitivity aided in verb learning**.** Novel words were constructed to be similar to Japanese mimetics, a class of sound symbolic words in Japanese. These words were then determined by Japanese-speaking and English-speaking adults as well as 2- and 3-year old Japanese speakers to match a certain action, seen visually, more than other actions. At test, 3-year-old Japanese children were taught either sound symbolic or non-sound symbolic words matched to the appropriate action. When shown different actors performing the same actions, children taught sound symbolic words were able to generalize the word meaning to the different actor, while those taught non-sound symbolic words were not.

In Imai et al's study, sound symbolism was found to aid in word learning—specifically verb learning—in young Japanese-speaking children. It did not, however, confirm that the psychological functionality of the sound-meaning correlations extended to speakers of other languages. Japanese is known to have a large number of sound-symbolic elements, and if speakers of another genetically unrelated language were able to successfully utilize the sound symbolic elements to benefit word learning, then the functionality of sound symbolism would be cross-linguistic as well. Kantartzis, Kita, and Imai (2008) conducted a similar study with English-speaking children listening to Japanese mimetic-like stimuli to discover if their ability to infer word meaning from the sound structure of the words was, indeed, cross-linguistic. The results showed a benefit to verb generalization when the stimuli were sound symbolic, thereby confirming that listeners were *cross-linguistically* sensitive to and exhibited world learning benefits due to sound symbolism.

Word learning is not the sole processing benefit that has been shown for sound symbolism. Sound-symbolic words have been shown to facilitate categorization as well. In a recent study, Kovic, Plunkett, and Westermann (2010) presented participants with cartoon drawings of creatures, which were comprised of a body with four parts: a head, wing, tail, and legs. Each had the same body but varied between two choices in the shape of its head, wings, and tail, as well as in the number of legs. Participants were asked to classify the creatures based on their visual characteristics. The creature categories were indicated by two auditory labels, *mot* and *riff*. In the congruent condition, the creatures' most salient features, the head and tail, were sound-symbolically matched with the two labels. For example, the creatures with a round head, triangle wing, and round tail would be classified as a *mot* based on vowel-salient shape sound symbolism and the creatures with a triangle head and tail would be classified as a *riff*. In the incongruent condition, the labels were not sound-symbolically matched to the salient features of the creatures in a particular category.

The participants were trained with nine different creatures that comprised two categories whose labels were either congruent or incongruent and continued training until proficient in categorizing the creatures. At test, the participants were shown the set of creatures in random order and asked to determine if the label shown matched the type of creature. Interestingly, although participants in either condition spent similar amounts of time in training and were approximately equally accurate at test, those in the congruent condition were significantly faster to classify the creatures at test. Listeners not only appeared to be sensitive to certain sound-to-shape correspondences, but they were also able to utilize these correspondences to facilitate categorization.

Although much work has been done regarding the psychological reality of sound symbolism, only recently has the technology been available to investigate any neurological underpinnings for these sound structure-to-meaning biases. The neurological basis of language is not yet well understood and, as such, different theories of its structure and function exist, including discussion as to whether language is its own distinct process or the result of the integration of other cognitive processes (Christiansen & Chater, 2008). Electrophysiological technology has aided research into this basis, and its coupling with behavioral research in sound symbolism has allowed for a new viewpoint into the nature of language itself. These techniques provide evidence for the existence of cross-modal integration as a possible mechanism for sound symbolism. In

Kovic, Plunkett, and Westermann's (2009) study, event-related potentials (ERPs) were also recorded from participants during the classification task. When participants were presented with sound symbolically congruent auditory and visual stimuli, a strong, negatively peaking wave was detected. In incongruent conditions, however, this component was not found. Such data provide neural evidence that these sound-meaning correspondences may result from integration of auditory and visual processing.

These results substantiate Ramachandran and Hubbard's (2001) argument that the neural substrate of sound symbolism may exist as the result of synesthesia. The sound symbolism bias would, thus, be a neurological one, arising from an interesting cross-activation between a sensory system and a motor system or another sensory system. Such a cross-modal connection could provide the basis for sound symbolism. Ramachandran and Hubbard (2001) further postulate that sound symbolism lies in a mapping between lip shape and an object's appearance, an extension of the sensory-motor synesthesia. For instance, the phoneme /o/ is produced using rounded lips, so words with this phoneme would be used to label objects which have a rounded shape almost imitatively. This theory would also explain the consistency found in using rounded vowels of nonwords to label rounded shapes and unrounded vowels for pointed shapes in Köhler's (1947) original study and follow-up studies, including Maurer, Pathman, and Mondloch's (2006). Work done with mirror neurons in monkeys, which appear to fire both when the monkey performs an action as well as when it views another animal performing the same action, may provide further evidence for a natural neural basis to sound symbolism (Rizzolatti & Craighero, 2004). However, a human analog to mirror neurons has yet to be discovered and these findings are not widely accepted as the principle foundation for sound-meaning mappings (Ramachandran & Hubbard, 2001).

The idea that sound symbolism facilitates psychological processing has, in part, brought the role of sound structure-to-meaning correspondences in language evolution to the forefront (Ramachandran & Hubbard, 2001). Contact between different groups of people with different cultures and languages as well as technological advances have certainly been integral to historical language change, but Ramachandran and Hubbard (2001) postulate that non-arbitrary sound-meaning relationships have constrained and delineated the creation of object labels and other words. Sound symbolism, after all, has been shown to facilitate word learning (Imai, Kita, Nagumo, & Okada, 2008; Kantartzis, Kita, & Imai, 2008; Nygaard et al., 2009) and categorization (Kovic, Plunkett, & Westermann, 2009) cross-linguistically. Thus, it follows that words with sound symbolic relationships to their referents would be more easily retained by speakers because they would appear to be more intuitive (Ramachandran & Hubbard, 2001). Computational modeling theories, however, point to arbitrary sound-meaning mappings as being beneficial to the creation of language. Non-arbitrary mappings inherently restrict the available pairings between form and meaning, thus suiting them for a smaller lexicon; arbitrary mappings, though, would allow for much more freedom in pairing form to referent, allowing for a much larger lexicon size reflective of that used in natural languages (Gasser, 2004).

The present study

Although a growing body of research has explored the existence cross-linguistic sound symbolism, its effects on language processing, and the neural and evolutionary origins of sound to meaning correspondences, surprisingly little systematic work has investigated the specific cues embedded in the speech signal that listeners use to make these semantic judgments. In the present study, my aim was to examine the specific cues underlying phonetic symbolism by attempting to quantify the features of the phonetic segments that carry semantic information. My study used a corpus of previously collected 1,220 antonym pairs from native speakers of 10 different languages, Albanian, Dutch, Gujarati, Indonesian, Korean, Mandarin, Romanian, Tamil, Turkish, and Yoruba. Native speakers of each language were asked to nominate synonyms for pairs of dimensional adjectives (*big, small, pointy, round, dark, bright, loud, quiet, bad, good, slow, fast, still, moving, down, up, near, far*) and then record the nominated words. These foreign synonyms were then presented to monolingual English speakers for identification. A subset of the foreign words corresponding to each pair of adjectives (e.g., the Albanian word *dhembezuar)* was presented one at a time and listeners were asked to match each word to one of two meanings (*round or pointy*). These listener ratings were averaged to determine how well each foreign synonym was assigned to its correct meaning. Overall, native English-speaking listeners were above chance, suggesting that they were able to use the sound structure of the synonyms to determine word meaning.

For the current project, I phonetically transcribed a subset of the words in the foreign synonym corpus that corresponded to five different meaning dimensions in order to create an inventory of types of sounds that are related to particular meanings. The frequency of occurrence of particular phonetic classes, such as speech sounds with a particular manner of articulation (e.g., fricatives like /s/ or /z/ versus liquids like /l/ or /ɹ/), were computed and then used to predict native English-speaking listeners' ratings of meaning for those words. Frequency counts of certain types of phonetic classes for words judged to correspond to particular meanings were calculated. The goal was to determine the inherent meanings that speakers associate with certain classes of sounds.

If reliable connections exist between the sounds of language and their perceived semantic meanings, then we should find greater frequencies of certain phonetic classes depending on the semantic category that listeners predominantly assign to the foreign words. This result would suggest that certain sounds of languages—coming from multiple unrelated languages—may carry inherent meanings and that listeners are sensitive to these particular constellations of sound-to-meaning correspondences.

Method

Corpus Preparation

Stimulus words were elicited from native speakers of Albanian, Dutch, Gujarati, Indonesian, Korean, Mandarin, Romanian, Tamil, Turkish, and Yoruba. These languages were selected in part to sample from several different language families in an attempt to create a varied, cross-linguistic corpus. Native speakers of each language were recruited from Emory University and the Atlanta area. A different speaker was recruited for each language. There were six male speakers and four female speakers, with the Indonesian, Korean, Tamil, and Yoruba speakers being female. Each speaker was asked via email to nominate words in their native language for pairs of adjectives relating to object features (*big, small, round, pointy, dark, bright*), motion (*slow, fast, still, moving),* spatial location *(up, down, near, far*), sound (*loud, quiet*) and valence (*bad, good*). If relatively few synonyms were nominated for a particular set of dimensional adjectives, the speakers were then encouraged to nominate additional synonyms for each of those words. The speakers were compensated for their time.

Each speaker was then brought into the laboratory and recorded the set of nominated synonyms in their native language. Stimulus words were recorded in a sound-attenuated room using a SHURE 5155D microphone and an EMU 0202 USB external sound card into a digital file at a 44.100 kHz sampling rate using Audacity software version 1.2.6. The recorded stimuli were amplitude normalized and segmented into individual word files using Sound Studio version 3.5.3. Each stimulus recording was saved as a separate .wav file to create the word corpus. Table 1 shows the number of synonyms nominated broken down by adjective pair and language. The top dimensional adjective pairs were analyzed in the present study and the remainder represent the composition of the entire corpus.

Behavioral Ratings

Judgments of word meaning were collected for each item in the entire corpus to determine how sound symbolic listeners perceived the foreign synonyms.

Participants. Nine separate groups of fifteen participants each were presented with stimuli from the corpus. Participants were all native speakers of American English and had no reported history of speech or hearing disorders. They were self-reported to have not learned other languages during their childhood and screened for knowledge of any of the stimulus languages. Participants were recruited from the Emory University community and were either paid $15 or received course credit for an Introductory Psychology class for their participation.

Procedure. Participants were told that they would be hearing foreign words and were asked to judge whether the word sounded as if it meant, for instance, *round* or *pointy*. Each group of participants was presented with all synonyms across each language that corresponded to a dimensional adjective pair (e.g., *big/small, good/bad*). For example, one group of participants was presented with all nominated synonyms across languages for the antonym pair *big/small*. Another group was presented with the foreign word synonyms meaning *round/pointy,* and so on for each of the nine pairs of contrastive synonyms.

Each foreign synonym corresponding to an adjective pair (e.g., the Albanian word *dhembezuar* 'moving') was presented auditorily one at a time to the listeners using PC-based E-Prime version 2.0 through an EMU 0202 USB external sound card and Beyer Dynamic DT100 headphones. Listeners were then asked to match each word to one of two antonym meanings (e.g., *round* or *pointy*) using a button box. Listener response choices and reaction times were collected.

The current project focuses on response choices for foreign words referring to object size (*big/small*)*,* object contour (*round/pointy*)*,* object motion (*still/moving*)*,* object speed (*fast/slow*), and valence (*bad/good*). In general, across languages, native English-speaking listeners judged foreign words that were synonyms of each of these dimensions above chance, indicating that their judgments corresponded to the correct meaning of the words. For the analyses in this project, the percentage of participants that judged each item as meaning *big* for the *big/small* dimension was calculated. Percentage *round, still, fast,* and *good* were calculated for each respective dimension as well. These scores served as a perceptual metric for how sound symbolic of each meaning each item in the corpus was judged.

Broad Phonetic Transcription

All synonyms corresponding to each of five dimensions, *big/small, fast/slow, still/moving, round/pointy,* and *good/bad*, from the word corpus were broadly transcribed using the International Phonetic Alphabet (IPA). Words were transcribed using only extant phones in American English. Phones existing in the stimuli that did not exist in American English were transcribed as the most perceptually-similar American English phoneme. For example, the phone /y/, which is the vowel in the French word *tu* 'you,' is most likely to sound to an English listener as its closest perceptual relative in English, /u/. This allowed for a representation of the sounds closest to that which the listeners attuned to American English might perceive. A list of the phones used in the broad phonetic transcription are listed in Tables 2 and 3. A small selection of the transcriptions were checked by a second transcriber blind to the initial transcriptions.

Feature coding. The IPA transcriptions for each item were then coded by phone class. Each word was coded to derive the phonetic inventory that comprised the word. To describe each consonant segment, its place of articulation (either bilabial, dental, glottal, interdental, labiovelar, labiodental, palatal, postalveolar, or velar), manner of articulation (either affricate, approximant, fricative, lateral approximant, nasal, or stop), and voicing (either voiced or unvoiced) were coded. To describe each vowel segment, vowel height (either close, near-close, close-mid, mid, open-mid, near-open, or open), vowel backness (either back, near-back, central, near-front, or front), and its roundedness (rounded or unrounded) were coded. Together, these features can relatively unambiguously describe the articulatory gestures used to create each specific speech sound.

At the word level, counts were conducted for number of phones and number of syllables in each word. Information about the specific position of each phone within a word and its syllable location was also recorded. For example, the Albanian word *dhembezuar* was transcribed using IPA as /ðɛm.bə.zu.aɹ/. Each transcribed segment was described by its phonological features and position within the word and syllable. The phone /b/ from *dhembezuar* would thus be described as being the fourth segment of nine total in the word as well as being in the second syllable of four total. Its place of articulation would be coded as bilabial, its manner as stop, and its voicing as voiced. This analysis was conducted for each segment within each word. Counts for each of these features were then collapsed by word to show the numbers of each feature within the word.

These features were then collapsed into larger categories. The consonant manners of articulation were categorized as sonorants (which includes approximants, lateral approximants, and nasals) or obstruents (which includes affricates, fricatives, and stops). The places of articulation were categorized as labials (including bilabials, labiodentals, and labiovelars), coronals (including alveolars, dentals, interdentals, and postalveolars), dorsals (including palatals and velars), and glottals.

Vowel height was categorized as close (including close and near-close vowels), mid (including close-mid, mid, and open-mid vowels) or open (including open and near-open vowels). Vowel backness was categorized as back (including back and near-back vowels), central (including only central vowels), and front (including front and near-front vowels).

Proportions of each phonological feature were calculated with respect to either the total number of consonants, vowels, or phonemes within a word depending on which of type of feature it was. For instance, the proportion of back vowels for a given word would be calculated as the number of back vowels divided by the total number of vowels within the word, the proportion of labially-articulated consonants would be calculated as the number of labially-articulated consonants divided by the total number of consonants within the word, and the proportion of vowels within the word would be the total number of vowels divided by the total number of phonemes within the word.

Data analysis. The coded features were then compiled with behavioral data to combine listener judgments with specific phonetic information for each word. Listener judgments were recorded as a proportion of responses with respect to one of the words within the dimension. For example, for the *bad/good* dimension, behavioral data were reported as proportion of responses judging the foreign synonym to mean *bad*. A simple Pearson correlation was run between the behavioral responses and the proportions or total counts of the phonetic correlates to evaluate correlations between the phonological features and listener judgments. Next, to disentangle co-varying attributes, a multiple regression analysis was conducted between the phonetic correlates found to be significant in the simple correlations and the behavioral responses. The behavioral ratings (e.g., proportion of *big* responses) were regressed on the coded phonological features (e.g., consonant sonority). Results are reported as significant at the $p < 0.05$ and $p < 0.01$ levels.

Results

Simple Correlations

Simple Pearson correlations were conducted to assess potential relationships between judgments of foreign word meanings within each dimensions (e.g., *big/small)* and measures of phonetic content. Five sets of correlations were conducted, one set corresponding to each dimension. Correlations were conducted between the proportion of responses of a certain domain (e.g., proportion of *big* responses) and each of the tabulated phonetic correlates: total consonants, vowels, phonemes, and syllables; proportions of sonorant, labial, coronal, dorsal, glottal, and voiced consonants; and proportions of close, mid, open, back, central, front, and rounded vowels. Below, I report the correlations for each dimension. The correlations are also presented in Table 4.

Size (big/small), n =161. Correlations were conducted between the proportion of *big* responses and each of the 17 measures of phonetic content. Foreign words that were judged to sound more reliably like they meant *big* were significantly positively correlated with the proportion of voiced consonants ($r(159) = .250$, $p = .001$) and open vowels ($r(159) = .170$, $p < .05$), and negatively correlated with proportion of close vowels ($r(159) = -.266$, $p = .001$).

Contour (round/pointy), n =117. Correlations were conducted between the proportion of *round* responses and each of the measures of phonetic content. *Round*-sounding words had fewer total consonants ($r(115) = -.41$, $p < .001$), vowels ($r(115) = -.49$, $p < .001$), phonemes ($r(115) = -.47$, $p < .001$), syllables ($r(115) = -.46$, $p < .001$) and lower proportion of front vowels ($r(115) = -.29$, $p < .01$). *Round*-sounding words also had a significantly higher proportion of labial consonants ($r(115) = .19$, $p < .05$), voiced consonants ($r(115) = .23$, $p < .05$), mid vowels ($r(115) = .23$, p < .05), back vowels ($r(115) = .343$, p < .001), and rounded vowels ($r(115) =.414$, $p < .001$).

Speed (slow/fast), n = 124. Correlations were conducted between the proportion of *slow* responses and phonetic content. *Slow*-sounding words had fewer total consonants ($r(122) = -.29$, $p = .001$), vowels ($r(122) = -.31$, $p = .001$), phonemes ($r(122) = -.33$ $p < .001$), and syllables ($r(122) = -.33$, $p < .001$). *Slow*-sounding words also had a significantly higher proportion of sonorant consonants ($r(122) = .27$, $p < .01$) and were marginally significantly correlated with the proportion of rounded vowels ($r(122) = .17$, $p = .06$).

Motion (still/moving), n =135. Correlations were conducted between the proportion of *still* responses and phonetic content. *Still*-sounding words had fewer total consonants ($r(133) = -.47$, $p < .001$), vowels ($r(133) = -.64$, $p < .001$), phonemes ($r(133) = -.59$, $p < .001$), and syllables ($r(133) = -.64$, $p < .001$), as well as a lower proportion of voiced consonants ($r(133) = -.18$, $p < .05$) and open vowels ($r(133) = -.21$, $p < .05$). *Still*-sounding words also had a higher proportion of closed vowels ($r(133) = .27$. $p < .01$) and marginally lower proportion of glottal consonants ($r(133) = -.16$, $p = .06$).

Valence (bad/good), n = 177. Correlations were conducted between the proportion of *bad* responses and phonetic content. *Bad*-sounding words had fewer total vowels ($r(175) =-.15$, $p < .05$) and syllables ($r(175) = -.18$, $p < .05$) and a significantly lower proportion of voiced consonants ($r(175) = -.17$, $p < .05$).

Multiple Regression Analysis

To better elucidate the role that these sound symbolic phonetic correlates play in the variability in speaker judgments, multiple regression analyses were conducted. These analyses served to better determine those phonetic features that co-vary for particular meaning domains. The goal was to uncover which phonetic categories are the best predictors of speakers' judgments and the extent to which judgments can be attributed to each symbolic phonetic feature or category. Therefore, we regressed the proportion of responses for the meanings we used in the simple correlations (such as proportion of *big* responses) on total consonants, total vowels, total syllables, proportion sonorant consonants, proportion labial consonants, proportion voiced consonants, proportion closed vowels, proportion back vowels, and proportion rounded vowels. These predictor variables were chosen because each is representative of the phonological categories used in the simple correlations to describe consonants and vowels. For example, for the feature of consonant manner of articulation, a consonant must either be a sonorant or an obstruent and cannot be both; as such, using the variable proportion of sonorants inherently reflects the proportion of obstruents. For phonological categories in which there exist multiple values, such as place of articulation, the dimension's most significant value (as determined by simple correlation data) was used as a predictor variable. Significant and marginally significant predictors are also presented in Table 5.

Size (big/small). Proportion *big* responses were regressed on the predictor variables listed above. The predictor variables accounted for a significant proportion of the variance in meaning judgments ($R^2 = .17$, $F(9, 151) = 3.54$, $p = .001$). Of the variables, both the proportion of voiced consonants (ß = .355, $t(151)=3.71$, $p < .001$) and the proportion of close vowels (ß = -.255, $t(151) = -3.29$, $p = .001$) were significant predictors of the proportion of *big* responses.

Contour (round/pointy). The predictor variables accounted for a significant proportion of the variance in *round* responses (R^2 = .47, F(9, 107) = 10.69, p < .001). Of the variables examined, the proportion of rounded vowels (ß=.355, t(107) = 2.78, p < .01) and of total vowels (ß = -.780, t(107) = -2.56, p < .05) were significant predictors of the proportion of *round* judgments. In addition, the proportion of labial consonants was a marginally significant predictor (ß=.133, t(107)=1.80, p =.07).

Speed (slow/fast). The predictor variables accounted for a significant proportion of the variance for *slow* judgments in the *slow/fast* dimension (R^2 = .29, F(9, 114) = 5.14, p < .001). Of the variables, the proportion of sonorant consonants (ß=.280, t(114) = 2.44, p < .05), rounded vowels (ß = .304, t(114) = 2.62, p = .01), and close vowels (ß=-.238, t(114) = -2.45, p < .05) were significant predictors of participants deciding a word meant *slow.*

Motion (still/moving). The predictor variables accounted for a significant proportion of the variance for judgments in the *still/moving* dimension (R^2 = .47, F(9, 125) = 12.07, p < .001). Of the variables, only the proportion of closed vowels was a significant predictor of participants deciding a word meant *still* (ß=.174, t(125) = 2.52, p < .05).

Valence (bad/good). The predictor variables accounted for a marginally significant proportion of the variance for judgments in the *bad/good* dimension (R^2 = .09, F(9, 167) = 1.821, p = .07). Of the variables, total syllables was a marginally significant predictor of proportion of *bad* responses (ß =-.585, t(167) = -1.85, p = .07).

Summary of Results.

Across dimensions, listeners appeared to use collections of different cues when making judgments about word meaning. For the *big/small* dimension, words that were judged to mean *big* had more voiced consonants and fewer close vowels than words that were judged to mean *small*. For the *round/pointy* dimension, words that were judged to mean *round* had more rounded vowels and fewer total vowels than words judged to mean *pointy*. For the *slow/fast* dimension, words that were judged to mean *slow* had more sonorant consonants and rounded vowels and fewer close vowels than words judged to mean *fast*. For the *still/moving* dimension, words judged to mean *still* had more close vowels than words meaning *moving.* The regression equation was not significant for words meaning *bad* or *good,* suggesting a less strong relationship between phonological features and listener judgments of these meanings. However, words that were judged to mean *bad* had marginally fewer syllables than words judged to mean *good.*

Although different features played a role in listener judgments for different dimensions, some features acted as cues in multiple dimensions. Words judged to mean *round* and *slow* had more rounded vowels than those judged to mean *pointy* or *fast*. Words judged to mean *big, moving,* and *slow* had fewer close vowels than those meaning *small, still,* or *fast,* as well. No single consonantal feature was a cue in multiple dimensions, however.

General Discussion

The current study was designed to investigate the relationship between phonetic properties of spoken words and listeners' sensitivity to cross-language phonetic symbolism. Synonyms for five different dimensional adjective pairs from ten different genetically unrelated languages were used as stimuli for segmental analysis and for behavioral ratings of sound symbolic meaning. In order to eliminate any effects of specific language exposure in listeners' judgment and to determine if the cues listeners use depend on the language-specific properties, native English-speaking listeners who were unfamiliar with each language were used. The results of our phonetic analysis point to the existence of non-arbitrary connections to which listeners are sensitive between specific categories of speech sounds and meaning in natural languages. By combining naïve listeners' judgments about word meaning with a feature analysis of the foreign synonyms, we were able to discover by domain the types of sounds to which listeners are sensitive when making semantic judgments. Speakers were found to use different classes of phonetic cues to encode meaning for different dimensions, and listeners picked up on these cues even as they varied by semantic dimension.

Perhaps the most striking result comes from this sensitivity to certain phonetic cues in cross-linguistic stimuli. The ten languages are largely disparate, with Dutch being a Germanic language, Indonesian being an Austronesian language, Yoruba being a Niger-Congo language, Gujarati being an Indo-Iranian language, Tamil being a Dravidian language, Korean being a language isolate, Chinese being a Sino-Tibetan language, Romanian being an Italic language, Turkish being an Altaic language, and Albanian belonging to its own language family (Lewis, 2009). Such a wide variety of languages allows for any language-specific idiosyncrasies to be discounted, allowing for the findings to be based either in commonalities between languages or the perceptual system of the listeners.

If the cues were grounded in language-specific knowledge that did not overlap among languages, then no relationships between phonetic properties and meaning judgments should be apparent through our analyses. If, however, these cues are not tied to any specific language experience, then the cues should influence judgments regardless of stimulus language and should correlate to certain meanings across languages. If these cues were to be specific to American English and no sound-meaning connections were present, then the cues would not translate to other languages and we should not expect any sort of benefit cross-linguistically, such as correctly guessing the meanings of foreign words. This, however, was not the case. Listeners were above chance at guessing word meaning across dimensions, consistent with the idea that these cues are not part of specific language experience.

Stimulus words were selected from different dimensions representing different meaning and perceptual domains: object features, in which we investigated sound symbolism for object size and contour; motion, in which we investigated object motion and speed sound symbolism; and valence, in which we investigated whether sound symbolism might extend to or be based on affective or emotion properties. As such, we were able to investigate whether phonetic symbolism is confined to certain semantic domains—and, if so, which types of domains. Our results suggest that phonetic properties are related to meaning in words relating to object size, contour, motion, speed, and, to a lesser extent, valence. Interestingly, though, these features were specific to the particular dimensions.

Our findings do, also, substantiate earlier findings investigating phonetic symbolism in different dimensions. The most prolific and replicated studies, investigating *round* and *pointy* judgments, do provide similar results with rounded vowels occurring more frequently in *round*-sounding words (Maurer, Pathman, & Mondloch, 2006; Ramachandran & Hubbard, 2001). Our work, though, extends these findings, showing that the total number of vowels is also significant in this judgment. In magnitude sound symbolism, our work corroborates previous work finding that open vowels, such as /a/, appear more in words judged to sound *big* in relation to those judged to sound *small* (Newman, 1933; Sapir, 1929). Our work also extends these findings, showing that voiced consonants also predict *big* responses.

The analysis of the sounds comprising each word was done using the most common method of classifying segments, phonological features. This system breaks down the speech sound into components that describe the articulatory gestures that, when taken together, fully describe the articulation of the sound. Consonants are described by three values: their manner of articulation, referring to the way the articulators interact when producing a sound; their place of articulation, referring to the location of the articulators' interactions; and their voicing, generally referring to whether the vocal chords are vibrating or not during consonant production. Vowels are described by three values as well: vowel height and vowel backness, generally referring to the placement of the tongue's body during vowel production, and vowel roundedness, referring to whether the lips are rounded or not during vowel production. This methodology does assume that the correlates are, in fact, articulatory. Speech sounds can be studied in other ways—auditorily or acoustically—which, though interrelated, do describe the speech signal in different ways. The correlates may, in fact, be better described through one of the other viewpoints, perhaps even allowing for broader generalizations between the segmental correlates and meanings we describe here. For example, Ohala (1984) reasons that smaller animals tend to create higher frequency sounds and, as close vowels are characterized by higher frequencies, we should make correspondences between higher frequency sounds and smaller objects, which corresponds with the mappings found in the present study. Nevertheless, this study does find reliable correlations between certain articulatory gestures and semantic judgments.

In the size domain, correlations indicated that listeners were sensitive to consonant voicing, connecting voiced consonants and open vowels to big objects. Words meaning *big* were also found to have fewer close vowels. Correlations were also found in the contour domain, *round/pointy*, with *round*-sounding words shorter in length (having fewer consonants, vowels, total phonemes, and syllables) and having fewer front vowels. Labial consonants, voiced consonants, mid vowels, back vowels, and rounded vowels were also found in higher proportions with *round*-sounding words. *Slow*-sounding words were correlated with shorter length (having fewer consonants, vowels, total phonemes, and syllables), but did have more sonorant consonants, revealing sound symbolic correspondences in the speed dimension. For the *still/moving* dimension, correlations were found between *still*-sounding words and shorter word length (fewer consonants, vowels, total phonemes, and syllables), fewer voiced consonants, open vowels, and marginally fewer glottal consonants. *Still*-sounding words were correlated with more close vowels, however. Finally, correlations were found between the valence domain, instantiated as the *bad/good* dimension, but they were not as strong or as numerous as those found for the other dimensions. *Bad*-sounding words were correlated with fewer vowels, syllables, and voiced consonants.

The regression equations were significant for words in the object features and motion domains, but only marginally significant in the valence domain. In the size dimension, the most salient predictors of listener judgments were found in consonant voicing and vowel height, accounting for 17% of the variance in listener judgments. In the contour (*round/pointy*) dimension, the most salient predictors of speaker judgments were found in vowel roundedness and total vowels, with labial consonants marginally predicting judgments as well. Together, the phonetic correlates accounted for 47% of the variance in listeners' contour judgments. In the speed (*slow/fast*) dimension, consonant sonority and vowel roundedness and height were significant predictors of speaker judgments. The phonetic correlates accounted for 29% of the total variance in listener judgments in this domain. In the motion dimension, the sole strong predictor of listener judgments was found in vowel height and the correlates accounted for 47% percent of total variance in listener judgments. In the only affective dimension, the only predictor found was that of total syllables, a marginal predictor. The phonetic correlates accounted for only 9% of the total variability in listener judgment. Across dimension, these phonological correlates accounted for different proportions of variance in listeners' judgments. However, in each dimension except for valence, the phonetic correlates were responsible for a remarkably large portion of the total variance, accounting for almost half of the variability in the motion and contour dimensions.

The multiple regression analyses were conducted to elucidate which features accounted for significant individual variance in listeners' judgments of meaning. Co-variation could be explained in part by correlated features in English speech sounds. We assumed that listeners were interpreting phones in our stimuli through their English phonological perceptual system and as a consequence, would be perceived as the most perceptually similar English segments. Some of the features coded, however, were not necessarily independent of one another. Although this method of analysis using phonological features related to English does break down each speech sound into its component gestures, each component gesture is not neccesarily in free variation with other gestures; the existence of one component gesture may automatically preclude the existence of another. For example, if a consonant has a manner of articulation that is a plosive and a place of articulation that is glottal, it cannot be a voiced consonant due to our vocal tract's physiological constraints. Also, free variation can be limited by the specific sounds that occur in different languages. In English, for example, most back vowels are rounded as well, so these two features would co-vary in our analysis. The multiple regression, thus, aimed to account for co-variation in features to find the most salient predictors of listeners' judgments.

Overall, we can generally conclude that, within certain semantic dimensions, there exist speech sounds that distinguish certain meanings from their antonyms. This result corroborates the general hypothesis behind all studies in sound symbolism, further providing evidence against the all-encompassing nature of de Saussure's (1916/1959) assertion that language is completely arbitrary. Listeners are sensitive to these speech sounds when making semantic judgments, even if they are not aware of the relationship between sound and meaning. This appears to allow listeners to correctly judge foreign and novel words above chance without specific language experiences.

The distinguishing features of these speech sounds also are specific for different dimensions; listeners use different cues for different dimensions when making semantic judgments. This could be tied to the semantic dimension, with certain properties more closely related to certain meanings than others, suggesting a synesthetic or cross-modal explanation for the existence of sound symbolism. Also, these cues do not encompass the whole of the word. Each phonetic segment of a symbolic word is not a symbolic cue; rather, these cues are embedded within a framework of presumably phonetically non-symbolic segments. This, along with the variability in cues by dimension, may provide a synthesis between Gasser's (2004) demonstrations of the advantages of arbitrariness and non-arbitrary form-meaning views of the nature of language. Sound symbolism does exist, allowing for its psychological and processing benefits, but the variability in the cues chosen by dimension and the framework of potentially non-symbolic segments may allow for the larger lexicon from which an arbitrary form-meaning mapping benefits. This would substantiate Christiansen's (2010) hypothesis, that language is a balance of both sound symbolic and arbitrary mappings. Also, and most generally, rather than relying exclusively on suprasegmental and extra-linguistic cues to make these mappings, listeners are sensitive to cues at the level of phonological features as well.

When making semantic judgments about unknown words in certain domains, listeners must reason about word meaning without prior knowledge of the word's true meaning. Several components may play a role in the variability in listener judgment, including language experience, language-specific features, and language heritage. Our analysis finds that certain phonetic segments play a role in the variability in listener judgments as well. How large of a role the segments play is dictated by the semantic dimension of the word. The weakest effects of phonetic symbolism, interestingly, were found to lie in the valence dimension, while the strongest were found in the contour and motion dimensions. Here, the more tangible, imageable dimensions showed the strongest effects, while the intangible dimension investigated showed the weakest. Thus, there may exist a correlation between the imageability of the words referred to in each semantic dimension and the strength of the phonetic cues on listener judgments of meaning. Such a correlation would lend credence to a cross-modal explanation of sound symbolism, similar to that proposed by Ramachandran and Hubbard (2001), as the imageable words would be closely tied to visual perception, which, by cross-modal connections, would be tied to language production. Although phonetic symbolism does not appear to impart significant influence in the valence dimension, this does not preclude sound symbolism playing a role in this dimension. It is possible that another aspect of the sound structure, such as prosody, carries symbolic meaning in this dimension and perhaps in other less imageable dimensions.

At this point, any explanation of this result—and the reasoning behind why each phonetic correlate is symbolic for certain domains and not others—would be *post hoc*, but the stronger effects found in the contour and motion may substantiate claims about a synesthetic grounding for phonetic sound symbolism. Previous research, such as Maurer, Pathman, and Mondloch's (2006) study, has shown that limited and possibly no language experience underlies this sensitivity to phonetic symbolism. At the very least, this sensitivity appears not to be grounded in specific language experiences, as speakers of genetically unrelated languages exhibit similar processing advantages from sound symbolic stimuli (Imai, 2008; Kantartzis, Kita, & Imai, 2008).

Although this work has provided evidence for the existence of phonetic symbolism in unrelated natural languages, it opens up more doors for research. Sound symbolic nonwords have been found to facilitate psychological processing, and finding a similar effect with natural words cross-linguistically would further confirm that the sound-meaning relationship is non-arbitrary and perceptually-based. The current set of stimuli could also be presented to non-English-speaking participants to confirm that these cues are similar cross-linguistically, further substantiating the notion that phonetic symbolism is not tied to specific language experiences. Also, novel words could be created from the phonetic cues discovered in the study and presented to listeners to confirm their sensitivity to the cues and consistency of judgments. With the novel nonwords, as well, the number of cues—or conflicting cues—could be modified to see if listeners are sensitive to degrees of sound symbolism. For example, an object label containing a higher proportion of a certain cue, such as rounded vowels, may cause listeners to rate an object as more rounded than an object label with fewer rounded vowels. On the other hand, would a higher proportion of certain phonetic cues cause an object label to become a more typical member of a certain category?

The extent to which this symbolism exists in natural language could also be investigated, studying which semantic domains show phonetic symbolism and to what extent, which could further elucidate the roots of the symbolism. By discovering which domains contain sound symbolism in their words, we can understand where the connection lies between sound and cognition or perception. An acoustic analysis of these phones could also provide another viewpoint through which to view the data; the symbolism may not be best classified through articulatory gestures but through an acoustic signal, which may reveal another connection between these cues and their predictive meanings.

Investigating the connection between sound structure and meaning reveals a great deal about language. Such elementary acts as speaking and understanding speech do, in fact, occur through a complex series of processes which, when taken together, provide the conduit for information transfer. Only recently has the connection between what we mean to say and what we do say been investigated, and the results show how complex and rich the speech signal and our parsing of the information embedded within the signal are. Within this connection between sound and meaning, thus, may lie not only a deeper understanding of language, but a deeper understanding of cognition as well.

References

Berlin, B. (1994). Evidence for pervasive synaesthetic sound symbolism in ethnozoological nomenclature (pp. 77-93). In L. Hinton, J. Nichols and J. Ohala (Eds.), *Sound symbolism.* New York: Cambridge University Press.

Christiansen, M. H. (2010). *A possible division of labor between arbitrary and systematic sound-meaning mappings in language.* Emory Sound Symbolism Workshop. Atlanta, March 27, 2010.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31*, 489-509.

de Saussure, F. (1959). *Course in General Linguistics*. (W Baskin, Trans.). New York: McGraw-Hill. (Original work published 1916).

Gasser, M. (2004). The origins of arbitrariness in language. *Proceedings of the Cognitive Science Society Conference* (pp. 434-439). Hillsdale, NJ: LEA.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*, 54-65.

Kantartzis, K., Kita, S., Imai M. (2009). Japanese sound symbolism facilitates verb learning in English speaking children. *Proceedings of the Cognitive Science Society* (pp. 636-630). Hillsdale, NJ: LEA.

Kovic, V., Plunkett, K., & Westermann, G. (2009). The shape of words in the brain. *Cognition, 114*, 19-28.

Köhler, W. (1947). *Gestalt psychology* (2nd ed.). New York: Liveright Publishing.

Koriat, A., & Levy, I. (1977). The symbolic implications of vowels and of their orthographic representations in two natural languages. *Journal of Psycholinguistic Research, 6,* 93-103.

Kunihira, S. (1971). Effects of the expressive voice on phonetic symbolism. *Journal of Verbal Learning & Verbal Behavior, 10*, 427–429.

Lewis, M. Paul (ed.), 2009. *Ethnologue: Languages of the World,* Sixteenth edition. Dallas, Tex.: SIL International. Online version: http://www.ethnologue.com/.

Locke, J. (1690). *An Essay Concerning Human Understanding.*

Maurer, D., Pathman, T., & Mondloch, C. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science, 9,* 316 –322.

Müller, F.M. (1879). *The Sacred books of the East: Volume 1. The Upanishads. Part 1.* Oxford: Oxford UP.

Newman, S.S. (1933). Further experiments in phonetic symbolism. *American Journal of Psychology, 45*, 53-75.

Nygaard L.C., Cook A.E., & Namy L.L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition, 112,* 181-186.

Nygaard, L.C., Herold, D.S., & Namy, L.L. (2009)The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science, 33*, 127–146.

Ohala, J. J. (1984). An ethological perspective on common cross - language utilization of F0 of voice. *Phonetica, 41,* 1 - 16.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12,* 225-239.

Ramachandran, V.S. & Hubbard, E. M. (2001) Synaesthesia—A window into perception, thought, and language. *Journal of Consciousness Studies, 8,* 3-34.

Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system, *Annual Review of Neuroscience, 27,* 169–192.

Table 1. Numbers of nominated synonyms broken down by adjective pair and language. The top five dimensional adjective pairs (10 adjectives) were analyzed in the present study and the remaining four represent the composition of the rest of the corpus.

	Language										
Meaning	Albanian	Dutch	Gujarati	Indonesian	Korean	Mandarin	Romanian	Tamil	Turkish	Yoruba	TOTAL
big	8	7	5	6	9	10	11	5	7	6	74
small	8	7	6	7	11	11	10	9	8	10	87
round	10	9	6	7	6	7	7	5	8	3	68
pointy	7	3	2	4	7	5	5	4	9	3	49
slow	6	5	2	4	8	7	9	7	10	2	60
fast	6	7	2	5	7	7	11	6	11	3	65
still	8	6	5	5	7	5	11	7	8	3	65
moving	9	4	4	8	8	5	10	8	11	3	70
bad	12	7	4	12	14	8	8	7	10	5	67
good	11	8	7	8	13	11	8	8	11	5	90
dark	6	7	5	8	10	6	9	7	10	2	70
bright	5	7	8	6	8	9	10	10	9	2	74
loud	5	7	2	7	7	7	5	10	8	1	59
quiet	6	5	4	8	9	8	6	10	8	3	67
down	6	8	7	6	6	7	6	3	7	4	60
up	5	10	3	9	4	7	6	4	8	4	60
near	7	6	1	3	4	8	7	10	8	5	59
far	6	6	2	4	5	10	5	4	8	6	56

Table 2. Vowel feature categories for vowel height, vowel backness, and vowel roundedness.

Vowel	Height	Backness	Roundedness
i	close	front	unrounded
u	close	back	rounded
e	close-mid	front	unrounded
o	close-mid	back	rounded
ə	mid	central	unrounded
I	near-close	near-front	unrounded
ʊ	near-close	near-back	rounded
æ	near-open	front	unrounded
a	open	front	unrounded
ɔ	open-mid	back	rounded
ɛ	open-mid	front	unrounded
ʌ	open-mid	back	unrounded

Table 3. Consonant feature categorization of phones for place of articulation (PoA), manner of articulation (MoA), and voicing.

Consonant	PoA	MoA	Voicing
d	alveolar	stop	voiced
l	alveolar	approximant*	voiced
n	alveolar	nasal	voiced
ɹ	alveolar	approximant	voiced
s	alveolar	fricative	unvoiced
t	alveolar	stop	unvoiced
z	alveolar	fricative	voiced
b	bilabial	stop	voiced
m	bilabial	nasal	voiced
p	bilabial	stop	unvoiced
f	dental	fricative	unvoiced
θ	dental	fricative	unvoiced
h	glottal	fricative	unvoiced
ʔ	glottal	stop	unvoiced
ð	interdental	fricative	voiced
v	labiodental	fricative	voiced
w	labiovelar	approximant	voiced
j	palatal	approximant	voiced
dʒ	postalveolar	affricate	voiced
ʃ	postalveolar	fricative	unvoiced
tʃ	postalveolar	affricate	unvoiced
ʒ	postalveolar	fricative	voiced
g	velar	stop	voiced
k	velar	stop	unvoiced
ŋ	velar	nasal	voiced

*The phone /l/ is a lateral approximant but, due to the categories used, was coded as an approximant.

Table 4. Simple Pearson correlation between phonetic correlates and listener judgments. Significant correlations are highlighted in blue.

	Prop. Response				
	round	*slow*	*still*	*bad*	*big*
Total Consonants	**-.407****	**-.294****	**-.468****	-0.009	0.106
Total Vowels	**-.487****	**-.306****	**-.641****	**-.154***	0.135
Total Phonemes	**-.470****	**-.333****	**-.593****	-0.076	0.131
Total Syllables	**-.462****	**-.331****	**-.640****	**-.181***	0.126
Prop. Sonorant Cons.	0.164	**.273****	-0.125	-0.133	0.042
Prop. Labial Cons.	**.193***	-0.031	-0.07	0.08	0.019
Prop. Coronal Cons.	-0.02	0.04	0.091	0.098	-0.016
Prop. Dorsal Cons.	-0.15	-0.067	0.012	-0.129	-0.019
Prop. Glottal Cons.	-0.112	0.053	-0.161	-0.129	0.057
Prop. Voiced Cons.	**.231***	0.143	**-.178***	**-.166***	**.250****
Prop. Closed Vowels	-0.127	-0.136	**.269****	-0.025	**-.266****
Prop. Mid Vowels	**.231***	0.018	-0.06	0.055	0.119
Prop. Open Vowels	-0.147	0.106	**-.207***	-0.041	**.170***
Prop. Back Vowels	**.343****	0.116	0.123	0.037	-0.022
Prop. Front Vowels	**-.288****	-0.039	-0.085	-0.09	-0.002
Prop. Central Vowels	-0.124	-0.107	-0.052	0.092	0.036
Prop. Rounded Vowels	**.414****	0.173	0.095	0.016	-0.009

* $p < .05$

** $p < .01$

Table 5. Multiple regression of behavioral ratings on phonological features. Significant and marginally significant predictors are represented.

Dimension	Phonological Features			
Big (R^2 = .17, F(9, 151) = 3.54, p = .001)		close V (ß = -.255, t(151) = -3.29, p = .001)	voiced C (ß = .355, t(151)=3.71, p < .001)	
Round (R^2 = .47, F(9, 107) = 10.69, p < .001)	rounded V (ß =.355, t(107) = 2.78, p < .01)		labial C (ß =.133, t(107)=1.80, p =.07)	total V (ß = -.780, t(107) = -2.56, p < .05)
Still (R^2 = .47, F(9, 125) = 12.07, p < .001)		close V (ß = .174, t(125) = 2.52, p < .05)		
Slow (R^2 = .29, F(9, 114) = 5.14, p < .001)	rounded V (ß = .304, t(114) = 2.62, p = .01)	close V (ß = -.238, t(114) = -2.45, p < .05)	sonorants (ß = .280, t(114) = 2.44, p < .05)	
Bad (R^2 = .09, F(9, 167) = 1.821, p = .07)				total σ (ß = -.585, t(167) = -1.85, p = .07)

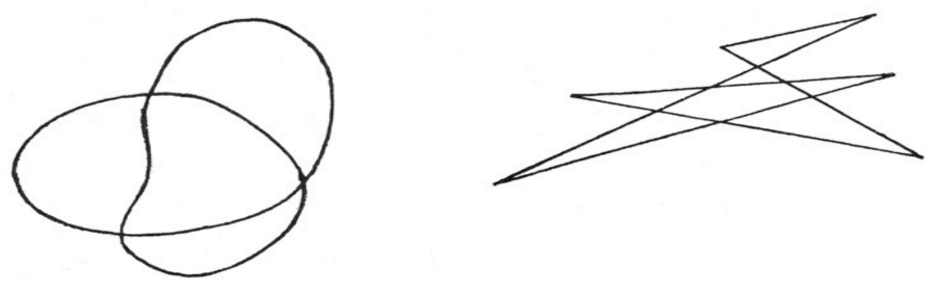

Figure 1. Adapted from Köhler’s (1947) original visual stimuli.