Understanding without Babblefish: Reviewing the Evidence for Universal Sound Symbolism in Natural Languages

Sarah Hunter-Smith
Swarthmore College

Abstract

Sound symbolism refers to a non-arbitrary relationship between sound and meaning. Language-specific forms of sound symbolism are well documented, but many scholars have also been interested in whether some sound symbolic patterns are universal. If humans have common intuitions about how sound should represent meaning, these intuitions could have facilitated the origin of language. If humans share sound symbolic intuitions, and they were influential in the origin of language, then we would expect to see evidence of these patterns in natural languages. Some evidence has been found in size ablauting systems, deictic pronouns, and ethnozoological nomenclature, however, one study in particular, Brown *et al*. (1955), suggests that sound symbolic patterns may be evident even among sensory adjectives. Subjects in this study were able to correctly match a pair of sensory antonyms in a foreign language to their English translations at rates significantly above chance. By conducting a similar study using a well-described sound symbolic pattern to create a "symbolic" and "non-symbolic" list of the word pairs, I show that subjects need symbolic cues to perform at levels above chance. I further try to show that this was likely true of the subjects in the Brown *et al.* study as well. This suggests that sensory adjectives reflect human intuitions about how sound should represent meaning. A rough assay to determine the extent to which *big-small* in many languages conforms to a sound symbolic pattern, however, failed to find the pattern represented more often than would be expected by chance. This suggests that sound symbolic intuitions place at most a subtle constraint on sensory adjectives. In general, sound symbolic patterns seem to be represented inconsistently in language, perhaps because there is an opposing selective advantage of arbitrariness in language.*

1 Introduction

Sound symbolism is generally defined as a direct connection between sound and meaning. This connection may be imitative, as when we use *boom* to mimic the sound of an explosion. Alternatively it can be metaphoric, as one would find in an explanation

* I would like to gratefully acknowledge my thesis advisor, T. Fernald, and my second faculty reader K.D. Harrison for their helpful comments and guidance. I would also like to thank T. Johnson and B. Woods for proofreading and offering insightful comments. A special thanks to S. Byrd-Krueger and M.K. McInerney for their help distributing the survey, and to A. Gillis and N. Peters for their assistance in its design. Finally I would like to express my gratitude to the 56 people who took the time to participate in the survey, and also to the many friends who have patiently listened and acted as a sounding board for the last three months.

for sound patterns used in a poem. For example, we might say that the frequent use of words containing voiced stops in a certain poem evokes a sense of heaviness or power. Most people have at one time or another expressed the idea that the way a certain word sounds is particularly apt given its meaning, so the idea that certain sounds have certain meanings is not foreign. At the same time, this notion is in conflict with a basic observation about human language. Not all humans speak the same language. The group of sounds which mean *dog* in English are different and unrecognizable as having any relation to the group of sounds with the same meaning in Japanese. A general tenet of linguistics, therefore, has been that the sounds composing a word are nearly always arbitrary and unrelated to that word's meaning (Fromkin *et al*. 2003). This leaves us with some questions. Is onomatopoeia just a small exception to this principle? Do our strong feelings about what sounds are heavy have any psychological reality for other people, even people who speak different languages?

At the root of this conflict is an issue that has fascinated many scholars: whether sound symbolism is only language specific or whether it also has universal elements. Sound symbolic sub-systems within specific languages are a well supported phenomenon at this point (Childs 1994, Austerlitz 1994, Hamano 1998, Blust 2003). Not only can we find sound-meaning correlations, but most speakers exhibit statistically significant intuitions about these correlations in their language (Parault and Schwanenflugel 2006). Universals of sound symbolism on the other hand seem to be a bit like the Holy Grail: exciting but somewhat speculative. Language-specific sound symbolism can teach us interesting things about how humans actually process language, but if no elements of it are universal, then it is ultimately conventional and arbitrary just like other aspects of language. Universal sound symbolism is exciting because it raises the possibility that we as humans are predisposed to connect certain sounds with certain meanings. This would have greatly facilitated the evolution of language, and could give us important insights on this notoriously difficult puzzle. Yet there seems to be considerable counter evidence to universal sound symbolism and there is no conclusive answer yet as to whether it actually exists.

A fascinating experiment: Brown* et al. *(1955)

One attempt to address the question of the existence of universal sound symbolism was an experiment done in 1955 by Brown *et al*. This work looked for evidence of universal sound symbolism by examining whether people could to some extent assign the correct meaning to foreign words. The authors proposed that if humans are predisposed to represent concept A with sounds of type B, then this should be reflected in natural languages, and furthermore humans should connect sounds of type B with concept A when they encounter them. Clearly people cannot translate unfamiliar foreign words when presented with them, but given a pair of antonyms which describe two ends of a sensory spectrum and the translation of these words in a foreign language, subjects might be able to say that one of the foreign words sounds more like one end of that spectrum than the other. Brown *et al*. presented subjects with twenty-one such pairs in English and their translations in Chinese, Czech, and Hindi. Surprisingly, subjects correctly matched the pairs on average at frequencies significantly above chance for all three languages. This was true whether they read the words and heard them pronounced by a native speaker, or just read the words silently. This controlled for the possibility that the person reading the words aloud might use a certain intonation or tone that could give the subjects clues about a word's meaning. Subjects in this study did not do equally well for all pairs, however. The percentage of subjects that guessed correctly on a given pair ranged from 17% to 97% (the results are presented in Appendix II). In other words, subjects agreed strongly about the way some word pairs should be translated and were correct, agreed strongly about other word pairs and were wrong, and also had mixed judgments about some pairs, but overall they were able to guess the correct translations more often than not.

This study raises a number of interesting questions which have not been convincingly addressed in the intervening years:

1) *Were subjects able to correctly match word pairs based on some sound symbolic quality?* Brown *et al*. determined that differences in word length did not predict subjects' responses, but it is conceivable that a variety of random factors like word length could have been responsible. For example, if words had superficial resemblances to their English translations, this could explain why subjects agreed strongly in some of their responses.
2) *Why did subjects perform much better on some word pairs than others?* If this result was not due to random patterns in the data, it could indicate that certain semantic domains exhibit sound symbolism more widely than others. Identifying these semantic domains could give us clues about the neurological basis of universal sound symbolism.
3) *Would subjects do as well on a wider sample of languages?* The results suggest both that subjects have intuitions about how meaning should be represented by sound, and that their intuitions are reasonably well reflected by patterns in natural languages. If these patterns are in fact widely distributed, then there might be some basis for the hypothesis that sound symbolism played a key role in the origin of human language.

This paper addresses these three questions as a basis for evaluating the existence and possible nature of universal sound symbolism. I show that subjects do use sound qualities to make judgments in an experiment similar to Brown *et al.* by comparing performance on a group of word pairs that contrast a particular sound quality and a group that does not. Further analysis of the data in Brown *et al.* (1955) suggests that subjects in that experiment also predominantly used sound quality over superficial factors. It also suggests that subjects most successfully recognize word pairs which represent basic sensory parameters such as brightness/darkness or sharpness/bluntness. However, despite the success that subjects seem to have using sound quality, attempts to quantify a relationship between sound and meaning in natural language have generally failed to find significant correlations, and they fail in this paper as well. Universal sound symbolism may exist in the form of shared intuitions about sound-meaning associations, although this has yet to be adequately demonstrated. The evidence suggests, however, that these shared intuitions are not consistently reflected in natural languages. I conclude by discussing the limited distribution of sound symbolic patterns in natural language in the context of the competing selective advantages of incorporating sound symbolism and arbitrariness into language.

2 Background

2.1 Types of Sound Symbolism

Before continuing, it is necessary to define sound symbolism more carefully. The term sound symbolism may refer to as many as four related phenomena which are differentiated by Hinton *et al*. (1994). The explanations below are adapted from this source, but include some additional examples and comments.

Corporeal sound symbolism. This describes sounds outside of language which are directly related to meaning by indicating the emotional or physical state of the speaker. Physical state can be indicated by sounds such as coughing or hiccupping. Emotional state can be indicated by pitch level, range and variability, loudness, and tempo (Ostwald 1994). Corporeal sound symbolism has many universal elements and some aspects are even shared with other animals. This kind of sound symbolism is of little interest in this paper apart from the fact that it may share a biological origin with other kinds of symbolism.

Imitative sound symbolism. This category includes all sounds which are imitative of other sounds whether accurate or conventionalized. That is, it includes both someone's direct mimicry of the sound their cat makes, and also the word *meow*. Purely imitative sounds should be universally understood, provided that the source of the sound is widely distributed, but conventional sounds may be different from language to language. However, given their basis in imitation, we might expect that these conventional sounds would be more similar between languages than other vocabulary.

Synesthetic sound symbolism. Synesthesia as a neurological condition is a strong coupling of two types of sensory information (Ramachandran and Hubbard 2001). Synesthetes often see numbers or words as distinctly colored, taste sounds, or see moving colors when they hear music. The term is also often used in relation to cross-sensory metaphors, and it is in this sense, rather than the pathological sense, that it is used here. Synesthetic sound symbolism is the imitation with speech sounds of non-acoustic phenomena. Generally this means that certain vowels, consonants, or suprasegmentals consistently represent certain properties such as size, shape, brightness, texture, speed, etc. The representation is not arbitrary, but is based on some similarity between the sound and the sensory signal it symbolizes. For example, there is commonly thought to be an association of palatal consonants and front and/or high vowels, which have relatively high frequencies (or small wavelengths), with diminutive qualities, small animals, or objects. This pattern has been documented across a large number of languages, and although it is not without exception, has long been a candidate for a universal sound symbolic pattern. Another such cross-cultural example is the association of rounded vowels with rounded objects. These are the types of sound symbolism which would explain why subjects in Brown *et al*. (1955) were able to match sensory foreign word pairs with their English equivalents. While synesthetic sound symbolism in general is less likely to be universal than corporeal or imitative sound symbolism, it is more interesting to investigate. Representation through metaphor is a very powerful tool, so if humans consistently represent certain properties with certain sounds this would have helped early humans develop a common vocabulary.

Conventional sound symbolism. This is the association of certain phonemes, consonant clusters, or syllables with particular meanings. In the literature this phenomenon is often called clustering, and the units of sound are called phonesthemes. So far all languages that have been examined have been shown to have phonesthemes (Bergen 2004), and English is no exception. For example, the words *glitter, glisten, glassy, glow, glimmer, glint, gleam, glance, glare, glower, glimpse* all share the cluster *gl* and a meaning that has something to do with light or sight. *Clash, bash, smash, crash, splash, lash, gnash, mash* all share the syllable *ash* and their meaning involves violent impact. *Flutter, fly, flit, float, flap, flip, fling* all share *fl* and a meaning that has something to do with movement through the air. These phonesthemes do not have strict one-on-one correlations with a meaning however. *Glove* has nothing to do with light, *cash* has no element of impact, and *flavor* has no connotation of movement through the air. Along the same lines, conventional sound symbolism can be cross cultural, but in general is a language-specific phenomenon. At the same time, connections between phonesthemes and meaning do seem to have psychological reality for speakers within a common language. Phonesthemes do have a direct connection to meaning, but they are different from other types of sound symbolism in that, as best we can tell, the brain creates this connection rather than because of any property intrinsic to the sound.

2.2 Language-Specific Sound Symbolism

The focus of this paper is primarily on synesthetic sound symbolism. This type of symbolism is extremely interesting because the connection between meaning and sound is more indirect than purely imitative or corporeal sound symbolism, and yet as we understand more about how the brain functions, we can begin to understand the mechanisms behind it. Also, as mentioned previously, finding that synesthetic sound symbolism has universal elements provides some insight into the evolution of language whereas we already suspect that imitation of sounds could have had a role in this process. A review of conventional, language-specific sound symbolism, however, is also critical to a search for universal sound symbolism because it is conceivable that the two could share common mechanisms. Therefore this section reviews what we know about language-specific symbolism and its psychological basis.

Unlike universal sound symbolism, conventional (language-specific) sound symbolic systems are very well documented. Phonesthemes are well attested not only in English, but also in a diverse group of languages including a number of African languages such as Gbaya, Ijo, Hausa, Kisi, Nembe, Igbo, and Yoruba (Childs 1994), Finnish (Austerlitz 1994), and Japanese (Hamano 1998). Phonesthemes further have a demonstrable psychological reality which is similar to that of morphemes. Bergen

(2004) showed that English phonesthemes, at least word initially, had priming effects very similar to morphemes and distinct from the priming effects seen with unrelated words, words with shared onsets, words with shared meanings, and even words containing pseudo-phonesthemes. Pseudo-phonesthemes were defined as shared onsets in words with a related meaning where only a very small number of words (usually only two) shared that onset and semantic relationship. The fact that phonesthemes more successfully primed subjects than pseudo-phonesthemes allowed Bergen to conclude that the frequency of a phonestheme, or the number of words sharing a related meaning in a group sharing the same onset, is sufficient to give it a morpheme-like status.

Frequency may also play an important role in the origin of phonesthemes. It is unclear exactly how phonesthemes come about, but one plausible hypothesis has been labeled the "snowballing effect" (Blust 2003). Under this model, a group of words sharing the same phonestheme (hereafter called a phonesthematic group) begins with two words with related meanings coming to have a common group of phonemes. This could happen historically by the existence of two words from the same etymological root, by the creation of a new word through blending, or through borrowing. If the resulting pair of words has high token frequency, then the shared phonemes may begin to be psychologically associated with the shared meaning. In addition, it seems that if the shared phoneme or cluster of phonemes is relatively rare in a given language, then it is more likely to become psychologically sound symbolic (Austerlitz 1994). The sound-meaning connection may further start to exert an influence on other words that share the same phonemes and gradually draw meanings into alignment with the group. They may even change the form of words with similar meanings and some similar sound element. Given the importance of frequency demonstrated by Bergen (2004), the larger the group becomes, the stronger these forces should become. In already established groups of phonesthemes, both of these effects can be observed. *Glory* which primarily means "praise, honor, or admiration accorded by common consent" also at some point developed an association with the light of heaven. Bergen references an example of sound change for the group *flag, drag, lag, sag* which share the meaning of "slow, tiring, tedious motion." *Sag* was originally *sacke* until the sixteenth century, at which time it went through an irregular sound change and made this group one word larger. Exactly how a phonesthematic group becomes psychologically established, however, is still a mystery. While high token frequency of an originating two words may be enough to initiate these effects, it could equally be the case that three or four words of common etymology are necessary to seed a phonesthematic group.

The important thing to notice about phonesthematic groups is that they do seem to grow by attracting new members. When we combine this with the observation that phonesthemes seem to be shared between related languages, but not in a wider context (Blust 2003), this seems to agree more with a model where the brain creates arbitrary sound-meaning connections than one in which these connections have some iconic or biological basis. This is not to say that iconic or biologically motivated sound-meaning correlations do not exist. Some phonesthemes may in fact have a biological basis and a more universal distribution. Rather the majority of phonesthemes are not universal and seem to demonstrate that the brain has a bias towards associating sound with meaning.

2.3 Universal Sound Symbolism

The ultimate goal of this paper is to come to some tentative conclusion as to whether universal sound symbolism can exist and if so what form it could take. Before wading into the evidence, it will therefore probably be useful to outline what "universal" might indicate and why it has come to indicate more than one thing. One very straight-forward understanding of universal sound symbolism is that we must be able to identify at least one sound symbolic pattern which is present in every language. This need not be a particular phoneme or cluster correlating with some meaning, but could potentially be a suprasegmental or a contrast between suprasegmentals which is realized with different concrete phonemes in different languages. Ideally the pattern would be applicable uniformly throughout language, that is wherever a word was semantically related to meaning A, it would incorporate the universal sound pattern B. This is not a requirement, however, and if the pattern was restricted to a certain class of words, for example onomatopoeia, it would still qualify as a sound symbolic universal. However the pattern is realized, in order for it to be universal, there may be no exceptions, and there certainly cannot be counterexamples. This perspective on what universal sound symbolism should denote is espoused among others by Gérard Diffloth, who justly criticizes "the incorrect use of the term 'universal' to mean simply 'found in a number of languages'" (Diffloth 1994: 107). As this view is concerned with a universal pattern found in human languages, I will refer to it in the future as linguistic universal sound symbolism.

At this point the reader will be wondering what non-linguistic universal sound symbolism might be an and why it is relevant. Many studies of sound symbolism have focused on the psychological element of it, or in other words how subjects assign meaning to various sounds or sound contrasts. What I will call universal sound symbolic intuition would involve at least one sound symbolic pattern which is intuitive to speakers of every language. There might be variation from person to person with regard to how strongly the pattern is intuitive, and some people tested individually may not seem to have any intuition about the pattern at all. In any reasonably large sample of people, however, there should be a significant tendency to map certain sounds onto certain meanings. This understanding of universal sound symbolism, with an emphasis on the psychological phenomenon instead of an observed pattern among natural languages, is typically held by researchers who hypothesize that sound symbolic intuition has an innate basis (Ohala 1994, Ramachandran and Hubbard 2001).

The degree to which universal sound symbolic intuition, assuming it exists, would have influenced sound symbolism in natural languages is unknown. There are multiple possibilities:

1) It may have had a profound influence on the origin of language and may have constrained language evolution and language acquisition since that time (Ramachandran and Hubbard 2001). Such a scenario would likely produce linguistic universal sound symbolism.
2) On the other hand, sound symbolic intuition could be largely directed at non-linguistic sounds, and could have had no influence on language evolution whatsoever. This would leave only language-specific sound symbolism, which if it had any similarity to universal sound symbolic intuition would be due to chance.
3) A third possibility would be that sound symbolic intuition was initially influential in the creation and establishment of human language, and may continue to have a limited influence when language is deliberately created, but that typically it is directed more towards non-linguistic sounds. In this scenario, the original language or languages would have displayed universally intuitive sound symbolic patterns, but the patterns would have gradually changed and eroded through the process of regular sound change. The patterns are unlikely to still be recognizable today, but we would observe the sound symbolic patterns

incorporated sporadically here and there in cases where people have put thought into how a word should sound. This might be most noticeable in naming systems, diminutive suffixes and other suffixes that denote some characteristic, and possibly slang.

4) A final unsettling but realistic possibility is that universal sound symbolic intuition could have a weak or moderate effect on language evolution, but that patterns in a speaker's native language could also have an effect on people's intuitions (Maurer *et al*. 2006). Sound symbolic intuitions could still be considered universal if they were innate, or if they were learned experientially and were experimentally recoverable in non-linguistic contexts. Patterns in language could change through regular sound changes, however, so we would not find truly universal sound symbolic patterns. Furthermore, in languages in which sound symbolic patterns matching the speakers' innate intuitions were absent or reversed, speakers tested for universal intuitions might not demonstrate them.

In all of the above scenarios universal sound symbolic intuitions exist, but as we can see, the extent to which they are incorporated into natural languages could vary considerably. If we are interested in whether universal intuitions have shaped some of the sound symbolic patterns that we observe in language, however, a minimum requirement is that the patterns be represented above what would be expected by chance.

Having outlined the possible interpretations of universal sound symbolism, it is now possible to review the evidence for and against it. This evidence generally falls into three lines of research which will be discussed at length below. One involves magnitude symbolism and an ethological explanation for it called the Frequency Code

Hypothesis. The second is often called either the Maluma-Takete phenomenon or the Bouba-Kiki phenomenon, but is probably better labeled shape symbolism. The third line of research has addressed the existence of sound symbolic patterns in general without focusing on a specific pattern. Recent research into the neurological basis of synesthesia offers some hypotheses as to the mechanisms behind these types of synesthetic sound symbolism.

2.3.1 Magnitude Symbolism

The first experimental work concerning magnitude symbolism began in the 1920s. Sapir (1929) presented about 500 English speakers and seven Chinese speakers with words that contrasted vowel sounds such as *mil* and *mal* and asked them to judge which one was larger and which smaller. He found that when words with [i] were contrasted against words with [a] at least 80% of his subjects felt that the word containing [a] was larger. Newman (1933) continued this work, discovering that articulatory position of the vowel and the acoustic frequency were both important predictors of the relative size subjects attributed to vowels. In other words, higher frequency or front/high vowels tended to be judged smaller. He was further able to show that subjects correlate brightness and darkness with vowels in the same way that they correlate magnitude.

Given the significance of vowel contrasts for English speakers, Newman also examined the distribution of vowels among size-related words in English. He compiled all of the words in Roget's *Thesaurus* that were listed under Greatness, Smallness, Size, and Littleness and divided them into words evoking largeness and smallness. Separate judges removed words that were ambiguous for size. However, when the word lists were compared, there was no significant distribution of vowels. Brown (1958) repeated this procedure, but accounted for usage frequency and still found no sound-meaning correlation. The implications of these studies seemed to be that magnitude symbolism was a largely psychological phenomenon, at least among English and possibly Chinese speakers. It was further curious because was not based on a subconscious knowledge of vowel distributions in one's native language.

Despite the lack of a correlation in English, scholars continued to show an interest in vowel distribution in natural languages. Jesperson conducted an informal survey in 1933 of a number of languages that highlighted many examples of the association between [i] and diminutive qualities (Nuckolls 1999). A thorough and extensive cross-linguistic survey did not appear until the 1970s. Ultan (1978) surveyed 136 languages for size ablauting, or a change of one phoneme or tone that indicates a difference in size or in a semantically related quality such as distance or number. He found that 27% of the languages he sampled had size ablauting, while 33% had distance ablauting. Almost 90% of the size ablauting systems and 85% of the distance ones conformed to the principle that the diminutive quality was represented by the higher frequency sound. Other semantically and grammatically related concepts such as quantity, number, force, intensity, proximity, and weight among others were also expressed by ablauting but at much lower frequencies.

In a more recent study, Woodworth (1991) studied magnitude symbolism in deictic pronouns using 26 languages chosen to be maximally genetically distant and representative of world's languages. Although not as extensive as Ultan (1978), this sample of languages was more carefully constructed to include diverse language types. Woodworth found that in a little over half of the languages, the proximal (i.e. "this") and distal ("that") pronouns had clear vowel contrasts which could be analyzed with respect to relative frequency. In other cases she was unable to make comparisons because there were more than two deictic pronouns in some languages, there were multiple vowels that differed between the two proximal and distal forms, or the proximal and distal forms had different numbers of vowels. Among those languages that could be analyzed, however, Woodworth found that proximal pronouns used higher frequency vowels than distal pronouns in about 70% of the languages. This pattern was reversed in only about 10% of the languages and the two pronouns had vowels of equivalent frequency in the other cases. She found very similar results for place adverbs (i.e. "here" and "there") and directional affixes (indicating toward or away from the speaker). Ultan and Woodworth both give compelling evidence that at least in some areas of language, magnitude symbolism appears to conform to the pattern described by Sapir and Newman with surprising frequency. The pattern is not universally represented, but nonetheless is represented beyond what would be expected from a chance distribution of vowels.

Berlin (1994) extended the study of size symbolism to ethnozoological nomenclature. He studied bird names in four languages spoken in South and Central America: Huambisa (a Jivaroan language spoken in north central Peru), Wayampí (a Tupian language), Apalái (a Cariban language), and Tzeltal (a Mayan language). For each of these languages he divided the names into those birds that were greater and less than 10 inches long. He found that small birds tended to receive names containing higher frequency sounds while larger birds received names with lower frequency sounds. In Huambisa, for which Berlin had the most information, when he removed birds with clearly onomatopoeic names, 71% of the birds under 10 inches long had names containing [i]. Fish names in Huambisa appeared to conform to the same pattern. Berlin's findings are interesting because ethnozoological nomenclature is an area of language where someone or a group of people often consciously give thought to what a certain animal should be called. The results suggest that when people are creating new names for things around them, they may make use of sound symbolic intuition.

Sound symbolism is very appealing in the corporate world because it suggests the possibility of brand names that will convey certain properties about a product. Consequently it is no surprise that consumer research has fleshed out some of the details of magnitude symbolism, both in terms of what qualities fall under the same sound symbolic pattern as magnitude and in terms of what sound contrasts can evoke those qualities. A study by Klink (2000) investigated whether sounds that indicate something is smaller to English speakers could also indicate that it is lighter (relative to darker), milder, thinner, sharper, softer, faster, colder, more bitter, more feminine, friendlier, weaker, lighter (relative to heavier), and prettier. He hypothesized that between two contrasted sounds, the higher frequency sound would indicate these qualities. Although he did not attempt all possible sound pairings, he found evidence supporting most of his claims. Words containing front vowels were judged to match the above qualities better than words with back vowels. Voiceless stops evoked these qualities better than voiced stops. Fricatives were somewhat, but not always better than stops and the same was true for voiceless and voiced fricatives. In general, vowels consistently gave subjects information about Klink's list of qualities while consonants appeared to be a less reliable source of information, at least for English speakers. It is entirely possible that in other languages, contrasts in consonant frequency are more likely to symbolize magnitude.

The evidence thus far appears to contradict the existence of linguistic universal sound symbolism. Ultan (1978) found much support for size and distance ablauting, but it is important to note that 11% (size) and 16% (distance) of the languages he sampled had ablauting counter to his proposed pattern. This was true of Woodworth's study as well. Diffloth (1994) documents such a counterexample in Bahnar (a Mon-khmer language of Vietnam), where high vowels represent large things and low vowels small ones. Interestingly, he points out that the system in Bahnar could still be considered symbolic as size correlates with the degree to which the tongue fills the mouth rather than with frequency as has been reported in so many other cases. This may add support for the notion that such systems are formed in agreement with cultural or universal notions about sounds and their symbolic value.

2.3.2 The Frequency Code Hypothesis

Magnitude symbolism is best explained by the Frequency Code Hypothesis which has largely been developed by Eugene Morton and John Ohala (Morton 1994, Ohala 1994). This hypothesis looks to other animals and evolutionary biology in order to explain why people would associate large magnitude with low frequency sounds and small magnitude with high frequency sounds. In particular, the hypothesis considers fundamental frequency, which is the frequency of the sound produced by the vocal cords before it is modified in the vocal tract. Morton (1994) best describes the motivation for this pattern in other animals. It probably originated in reptiles and amphibians, where individuals continue to grow throughout their lifetime and larger animals are able to win competitions for mates and other resources. Since smaller animals nearly always lose fights with larger animals and may become seriously injured during such fights, it is to their advantage to avoid them. They can judge the size of potential opponents visually, but the calls of larger, more dominant animals have a lower fundamental frequency than the calls of smaller animals, therefore calling frequency is also an honest indicator of size. Individuals that correctly identify low calls with large individuals are likely to survive to an age when they will themselves be large enough to mate. And indeed, studies have shown that playing the call of a larger toad will cause smaller toads to retreat (Davies and Halliday 1978), demonstrating that sound alone is adequate to indicate size.

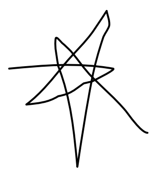
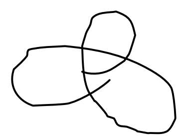
Among birds and mammals, growth stops at sexual maturity and fundamental frequency is often, but not always correlated with body size (Ohala 1994, Morton 1994). However, it still seems to play a role in the communication of size and size-related motivation. For example, animals displaying aggression usually give harsh, low frequency calls, while in friendly, appeasing, or fearful situations they tend to give tone-like, higher frequency calls (Ohala 1994). Large animals will often still be the more aggressive and dominant ones, so that in these more recent animal lineages it remains advantageous to correctly associate frequency with size and dominance.

There is further compelling evidence that acoustic frequency has played an important role in human evolutionary history. This has to do with formant frequencies, which are another element of vocal quality that indicate body size even when fundamental frequency does not. Formants are the result of sound filtered by the vocal tract, allowing only certain frequencies to pass through. They are determined by the length and shape of the vocal tract involved, which is in turn proportional to body size (Fitch 2000). In human males, unlike other primates, the larynx descends at puberty, making the vocal tract longer and the males' formant frequencies lower. This change is accompanied by other developments such as facial hair and broadening of the shoulders which serve to increase the male's body size or the impression of body size. Males that could communicate larger size effectively to others should, as among other animals, have been able to assert dominance and win mates without necessarily having to fight for them. The descent of the larynx was therefore probably selected for because, like other indicators of size, males with these characteristics were more successful and had more children (Ohala 1994). This hypothesis is supported by evidence from birds that have developed tracheal elongation. The trachea in such birds is often looped or coiled within the bird's body, significantly decreasing their formant frequencies. There is some evidence that these elongated tracheas help these birds sound larger, especially at night or in heavy vegetation (Fitch 2000). Thus the descended larynx in human males seems to indicate that frequency is built into human anatomy and evolutionary history.

In order for males with low frequency voices or low formant frequencies to be selected for, other humans must have been able to associate frequency with size and dominance. Ohala (1994) offers this as justification for why the frequency code must be innate in humans, however, it is also possible that it is learned through experience, as long as it is learned by virtually everyone. This is actually very plausible. Frequency is generally an honest indicator of size, even among non-living things. For example, avalanches make a lower sound than small rocks sliding down a hill, and anyone who has blown through two tubes of different lengths knows that the longer one produces a lower sound. All humans are likely to have experiences of associations of this type, whereas the inverse relationship is unlikely to occur often if at all, so it is entirely plausible that everyone learns this kind of association. There is in fact some evidence supporting learning over an innate frequency code in that children have been found not to consistently associate pitch with size until about age 11 (Marks 1987). It seems very likely that associations between frequency and magnitude would be universal, although this has not been rigorously documented, but it remains to be seen whether these associations have an impact on language and intuition in a linguistic context.

2.3.3 Shape Symbolism

Shape symbolism research has centered around matching drawings with sound. The basic design, developed by Kohler in 1929 and later by Fox in 1935, has been to present subjects with two line drawings, such as the two pictured below, along with two words and ask them which word best represents each drawing.

In Kohler (1929) the two words were *baluma* and *takete*. Fox (1935) changed *baluma* to *malumba* to avoid obvious similarity with the English word *balloon*. In a 1947 version of the experiment, Kohler adopted the final forms *maluma* and *takete*. This last example gives the two words a more consistent composition: continuants and rounded vowels contrasted with stops and unrounded vowels. In any case, in all three version of the experiment, speakers of English and German overwhelmingly matched the former word to the rounded shape and the latter word to the pointed one.

In the quest for universal sound symbolism, this experimental design was attempted with speakers of various foreign languages. Among these studies, Davis (1961) repeated the experiment with school children near Lake Tanganyika who spoke Kitongwe (a Bantu language) and Swahili. *Maluma* was once again changed, this time to *uloomu* because of similarity to a word in Kitongwe. Davis found that the children did not match the rounded picture with *uloomu* quite as often as the English school children that he tested, but that they still had a significant tendency to do so. Another very informal test with Songe speakers in Papua New Guinea, however, did not find that subjects matched pictures and words with any level of significance (Rogers and Ross 1975). Roughly half of the subjects matched *maluma* with the pointed shape and half with the curved shape.

According to Westbury (2005), all of these studies on shape symbolism have shared two main flaws. These were the use of just two words to generalize about symbolic sounds, and the use of an experimental design transparent to subjects. Westbury bypassed these problems by using a lexical decision task similar to that used in Bergen (2004). While it is not clear whether the effect demonstrated by Kohler and others is dependent on consonants or vowels or both, Westbury chose to focus on a theoretical association between stops and pointy shapes and continuants and rounded shapes. Thirty subjects were asked to determine whether a string of letters was a word or not. The string was centered within a frame that was either spiky or curvy around the outside, and the experiment measured whether the shape of the frame interfered with the decision task. Interestingly, there was an interference effect for non-words but not for real words such that subjects took significantly longer to identify strings of continuants and vowels in spiky frames than in curvy frames and the reverse was true for strings of stops and vowels. Westbury's work not only confirmed that people generally find stops representative of spikiness and continuants representative of curviness, but also introduced the interesting idea that words are affected differently than other sounds by sound symbolic intuition.

One important question about sound symbolism concerns whether it is based on innate neural connections or connections that are learned early in life. While size-frequency connections seem to develop in late childhood, this is not necessarily the case for other types of associations. A very recent paper examined this issue using shape symbolism to test the judgments of 2.5 year olds (Maurer *et al.* 2006). The authors found that toddlers, like the college students in their study, matched rounded shapes with names that had rounded vowels significantly above chance. Interestingly, toddlers did not match rounded shapes with rounded vowels quite as often as adults, suggesting that they may be learning about shape-sound connections and polishing their associations of this type. Unfortunately toddlers already have significant linguistic experience, so this study does not show that shape symbolic intuition is innate, but it does indicate that if it is not innate, it must be learned very early in life. This differs from previous findings that children do not develop this intuition until about age 9 (Davis 1961), and suggests that when age-appropriate methods are used, very small children may be able to reliably make such discriminations. This also calls into question the study suggesting that children do not consistently associate size and frequency until age 11 (Marks 1987), and suggests that alternative methods might find that younger children also relate size and frequency.

2.3.4 General Studies

Brown *et al*. (1955) was actually not the first experiment in which English speaking subjects were asked to match a pair of foreign, sensory opposites to their English equivalents. The original idea that subjects should do well at such a task if sound symbolism were represented in natural languages occurred much earlier and was repeatedly tested in a series of largely unpublished studies. These included one by Tsuru in 1934 using Japanese word pairs, followed by Müller (1935) with Swahili and Bantu word pairs, Allport (1935) with Hungarian pairs, and Rich (1953) with Japanese and Polish pairs (Brown *et al*. 1955). All of these studies found that subjects performed at rates over 50%, although not all of them tested for statistical significance. Brown *et al*. (1955) was an attempt to expand the languages that had been tested in this manner and to correct methodological issues which could have accounted for the above-chance results. The most important of these corrections was to have someone who was unfamiliar with the experiment translate the list of English word pairs into the tested foreign language. This precaution was designed to minimize the chance that the translator would choose translations on the basis of their similarity to English. If the foreign words reliably resembled their English translations in just a few cases, this could potentially be enough to allow the test subjects to perform at levels above chance. The authors also tested for the effects of presentation by giving one group the test in written form and having a native speaker read the words, while giving a second group only the written format.

Brown *et al*. (1955) was therefore a much more rigorous incarnation of the word-pair matching experiment than previous versions, but since the overall success rate continued to be only slightly above chance, around 55-60% depending on the language, questions of methodology continued to arise. Maltzman *et al*. (1956) introduced a new aspect whereby subjects were asked to match an antonymic pair of Croatian words with their translations in Japanese. They compared this to the normal condition where subjects matched English word pairs to their equivalents in Japanese and Croatian, and they found that subjects only performed at above-chance levels when the English word pairs were involved. Brackbill and Little (1957) further changed the procedure by using a list of 50 high frequency words rather than pairs of antonyms. These were not strictly sensory words, as in other experiments, but also function words like *when, first, this, etc.* They translated their list into Chinese, Japanese, and Hebrew, and then paired half of the words in each language with their translation in another language, and half with a word that meant something different. Subjects were presented with two paired words and asked whether they meant the same thing or not. They were told that half of the words were paired with their correct translation. Not surprisingly, subjects failed to guess correctly at a level above chance.

Brown and Nuttal (1959), in reviewing the three methodologies conclude that successful guessing must be based on the knowledge of how various phonetic spectra relate to sensory spectra.[1] For example, assuming the frequency code hypothesis to describe reality, if we were considering the pair *big-small*, we would know that words meaning *small* should have higher frequency sounds than words meaning *big*. If we were presented with the word pair *kit-ket*, we would know to pick the word containing *i*. However, if the contrast was between *ket* and *kat*, then we would know to pick the word containing *e*. If we were only given *ket* and asked whether it meant *big*, we would have no way of knowing for sure whether the sounds in the word were of relatively higher or lower frequency than its antonym. Similarly, if we were given the pair *kit-sot* and *set-kat*, but we were not told that these were both translations of *small-big*, then we would not know whether to pair the words such that the first letters matched or such that the vowel contrast was the same. The authors suggest that because English speaking subjects do somewhat better than chance when presented with the Brown *et al.* (1955) method, but do not when presented with the Maltzman (1956) and Brackbill and Little (1957) methods, cross-cultural sound symbolic patterns must operate by the association of a phonetic spectrum, or spectra, with a sensory spectrum. This also explains how languages with different phonemic inventories could all hold to the same sound symbolic pattern—as long as they have some of the sounds along the relevant phonetic spectrum, they could follow the same sound symbolic pattern as another language using different phonemes on that spectrum.

[1] Brown and Nuttal (1959) discuss phonetic spectra, but the sensory spectra could just as easily be represented by the presence or absence of a single feature. A spectrum is a good descriptor of magnitude symbolism and the frequency code, but in shape symbolism the distinction is between stops and continuants or between rounded and unrounded vowels, and there are no gradations between these features.

2.3.5 The Neural Basis of Sound Symbolism

Although the synesthetic in synesthetic sound symbolism refers to cross-modal metaphors rather than clinical synesthesia, studies of the latter have actually begun to provide some insight into potential mechanisms behind sound symbolism. Connections have been made between clinical synesthesia and sound symbolism at least since the 1950s, however, Brown (1958) argued that the two should not be associated because individual synesthetes of the same variety vary widely in their associations. For example, if two synesthetes see colors when they hear chords, one might find that a C major chord is green while the other finds it to be gold. Therefore it seems unlikely that sound symbolism is a very common, mild form of synesthesia because it shows a relatively high level of consistency, at least among English speakers.

There is growing evidence, however, that sensory modalities are not the discrete, separate modalities that they were once thought to be. In a 2001 review, Shimojo and Shams document a number of cases in which sensory regions in the brain have been shown to be relatively plastic. For example, in early onset deaf individuals the regions of the brain normally devoted to hearing are often used for vision in addition to the normal visual areas. Furthermore, the primary and secondary visual cortical areas are activated by reading Braille in blind but not sighted subjects. Sensory modalities have also been shown to interfere with and alter other sensory modalities. This is commonly experienced with a visual cue altering the perception of location or quality of a sound, as when a ventriloquist makes it appear that someone else is speaking. However, sounds can also alter visual input. For example, if multiple beeps accompany a single, brief flash of light, subjects will perceive the flash as multiple flashes. Sounds are also used to disambiguate visual stimuli. If in a cartoon, two balls travel towards each other, fuse, and then separate, subjects will see the two balls bounce off each other if there is an accompanying sound of appropriate sharpness and duration. If there is no sound they will see the balls travel past each other. All of this is just to say that the senses are relatively interconnected in normal people so that hypotheses about sound symbolism that involve cross-modal neural connections may not be that far fetched. The cross-modal neural connections underlying sound symbolism would not have to be the same as those that we think are unnaturally strengthened in synesthetes, but synesthesia has nonetheless inspired hypotheses about the neural basis of sound symbolism because it is a well established example of cross-modal connection.

The major hypothesis regarding the neural basis of sound symbolism was presented as part of a larger paper on synesthesia by Ramachandran and Hubbard (2001). They propose that there are a number of non-arbitrary connections in the brain between the maps of speech lip and tongue movements and the representations of certain phonemes or visual qualities in auditory and visual areas. They present a number of observations to support the existence of such sensory-to-motor connections. The most familiar observation is probably the translation of audio input into movement, or dance. Another observation is that there are rare forms of synesthesia in which particular sounds evoke the automatic adoption of specific postures. If we suppose that synesthesia is an unnaturally strong form of normal neural interconnectivity, then this would support the existence of sensory-to-motor connections. Finally, Ramachandran and Hubbard cite the existence of mirror neurons, which fire in humans and monkeys both when they are watching someone perform a task with their hand or mouth and when they themselves are performing it. Mirror neurons have also been shown to fire when someone is listening to sounds that are easily identified with the task, including speech sounds (Rizzolatti and Craighero 2005). Given the existence of neurons that fire in audio, visual, and motor contexts, the existence of innate, non-arbitrary connections between audio and visual input and the motor map for the production of corresponding speech sounds seems possible. As discussed earlier, however, there is no reason that these connections would have to be innate. It is equally possible that they could form as a result of experiential learning during childhood, as long as the sound-quality pairing was found in the natural world.

2.3.5 Conclusions from Sound Symbolism Research

The literature suggests that linguistic universal sound symbolism, that is a sound symbolic pattern expressed in natural languages without exception, probably does not exist. Magnitude symbolism, described by the frequency code hypothesis, has very compelling reasons to be universal, but we have yet to find an area of language in which there are not exceptions and counterexamples to it. Vowel frequency, the strongest indicator of size to English speakers, does not seem distributed such that words affiliated with smallness have higher vowel frequencies. Magnitude is widely expressed in size ablauting systems, and deictic pronouns, but there are exceptions and counterexamples. The frequency of vowels in bird names in some languages corresponds to the birds' size, but this is not true for all bird names. In every way that we find magnitude symbolism incorporated into language, it always has exceptions, and this suggests that if we find it incorporated in other aspects of language, there will be exceptions there as well. Shape symbolism in natural language has not really been studied, but if speakers of Songe really do not have intuitions matching those of English, German, and Kitongwe speakers, then that suggests that their language will not follow the same pattern either. We cannot rule out linguistic universal sound symbolism, but if all hypothetically universal sound symbolic intuitions are incorporated into language by a similar mechanism, then the evidence suggests that that mechanism will always lead to some exceptions.

Universal sound symbolic intuitions seem likely, but are far from proven. The frequency code hypothesis is well supported by evidence from other animals and human anatomy. It seems likely that even if intuitions about size and frequency weren't found in other animals and weren't innate, that humans would learn about them from their environment. The evidence that magnitude symbolism is expressed widely across languages also supports the notion that humans have matching widespread intuitions. It would be hard explain the prevalence of magnitude symbolism in language if it were not motivated by some psychological factor. Currently there have been studies on speakers of relatively few languages, however, and a much wider cross-linguistic sample would be needed to corroborate the frequency code hypothesis. Shape symbolism is similarly in need of a wider cross-linguistic sample. The evidence that speakers of Songe do not share intuitions about shape symbolism with English speakers suggests that this pattern may not be universal, however it does not completely rule out the possibility. Tsur (2006) argues that normally when we process language, we do not attend to sound quality, but only to the abstract sequence of phonemes. He claims there is a "poetic" mode of listening which we can use when we want to attend to sound quality as well as the sequence of abstract phonemes. To make use of sound symbolic intuition, it makes sense that one would use the "poetic" mode of listening and consider actual sound quality. In many studies of sound symbolism, subjects are asked to consider sound quality when they make judgments about meaning. In the Rogers and Ross (1975) study on Songe speakers, subjects were only asked to say which drawing they thought was a *maluma* and which a *takete*, not which name sounded like it matched one of the drawings. Subjects may not have attended to sound quality at all, and given the results they seem to have responded at random. A more careful experiment with speakers of this language seems warranted, and along with a wide survey of speakers of different languages could give us a better idea for whether intuition about shape symbolism is universal.

One issue that remains unclear is the extent to which sound symbolic intuitions have been incorporated into natural language. If they have acted as a general constraint on language evolution, then we should expect to see sound symbolic patterns throughout language. The patterns should not only show up in the somewhat specialized contexts like naming or ablauting systems where at least magnitude symbolism has been largely found so far. They should be apparent in words throughout the lexicon, including the adjectives that label the sensory qualities being symbolized (i.e. *big* and *small* in the case of magnitude symbolism). At the moment this does not seem to be the case. Two attempts uncover a constraint on size-related words have found no significant distribution of high frequency vowels. But before concluding that sound symbolic intuition does not act on sensory labels, it will be worthwhile to reconsider the evidence from Brown *et al*. (1955).

3 Sound Symbolism and Guessing Foreign Word Pair Meaning

Brown *et al.* (1955) is a fascinating study because it suggests that speakers' inner sound-meaning associations are reflected by patterns in natural languages. It is not difficult to understand why English speakers would all have similar sound-meaning associations such that they would match foreign word pairs with their translations consistently. Their experiences with language are relatively similar, and they should have the same knowledge of English phonesthemes or other patterns in the distribution of sounds in English. It is even understandable how in some cases, people that speak different languages might have similar judgments about what sounds best imitate natural sounds, visual and tactile qualities, or size. Although this has not been satisfactorily demonstrated, there is compelling logic behind the Frequency Code hypothesis and hypotheses about neural interconnectivity. But the idea that these sound-meaning correlations are reflected in natural languages to such an extent that people can correctly translating a pair of antonyms, even when limited to certain sensory antonyms, is incredible. If there are clear constraints on the sounds making up sensory adjectives, why don't these words clearly resemble each other from language to language? If there are not constraints, how can test subjects assign them meanings correctly? If subjects in the Brown *et al*. study were predominantly using honest sound symbolic cues to match pairs with their translations then this has implications for how some parts of language are created. Therefore this next section aims to experimentally determine whether sound symbolic cues are generally responsible for correct translation of antonymic foreign word pairs.

3.1 Methods

It is most likely that subjects use a combination of different cues when they match foreign word pairs with their meanings. These could include

a. characteristics of the written word such as length or shape
b. resemblance to an English word
c. sound symbolic patterns specific to English
d. cross-cultural or universal sound symbolic patterns

It is unlikely that subjects use only sound symbolic patterns because in any case that is not perfectly controlled, they probably use a combination of all of the above cues. Instead of trying to completely rule (a) and (b), it is better to ask whether sound symbolic patterns are one of the cues that subjects use to correctly match foreign words, and possibly if they are necessary cues for subjects to do so. This can be done by comparing the success on a group of foreign words that display a known sound symbolic pattern to the success on a group that does not display this pattern. If success with the "symbolic" group of words is much higher than with the "non-symbolic" group, this would indicate that sound symbolic patterns contribute to successful matching. Provided that other cues equally promote correct matching in the two groups, which they should if they are random, if subjects guess correctly at a frequency significantly above chance for the symbolic but not for the non-symbolic group, this would further suggest that sound symbolic patterns are responsible for their success.

The best described sound symbolic patterns are those associated with magnitude symbolism. According to the Frequency Code hypothesis, smallness and related qualities are associated with any sounds that are relatively high in frequency (Ohala 1994). In order to simplify the process of assigning words to symbolic and non-symbolic categories, however, I chose to limit this criterion to vowel contrasts for this experiment. This is reasonable given that most studies have focused on vowel but not consonant contrasts (Nuckolls 1999).

The symbolic and non-symbolic word lists were created using translations for *big*, *huge*, *small*, and *tiny* in sixteen different languages. Languages were chosen based on the selection of dictionaries available in the Swarthmore College library. Languages were only used if they were non-Indo-European to rule out the possibility of cognates with English and to minimize the chances that experimental subjects might have studied them. In addition, only dictionaries using roman characters and with descriptions of pronunciation were used. All one and two syllable definitions for the four words were copied down in the order in which they were listed in the dictionary.

Three or more syllable words were not included to minimize unnecessary complication. With large numbers of syllables it becomes unclear what sounds will most effect a subject's judgment. Word pairs were created using the first translation for *big* and the first translation of *small* or *tiny* with the same number of syllables. In this way, words were roughly balanced for length, although one word might still take up more space on the page than its pair. This method also prevented me from forming pairs on the basis of whether I thought they sounded symbolic or not. Instead pairs were based on the order of dictionary listings and were separated into the symbolic or non-symbolic group by a consistent metric. Ultan (1978) found that small- big contrasts were not only denoted by the high, front vowel [i] and the low, back vowel [a], but more generally by a front vs. back vowel contrast or occasionally by a high vs. low vowel contrast. This can be illustrated by the English *clink/clank*, where *clank* is the heavier sound, but the vowel is a low, front vowel, not a low, back vowel. In order to capture both such contrasts then, words were given a score equal to their number of front vowels and high vowels. High, front vowels were counted once for each feature. If the word meaning "small" had a higher score than the word meaning "big", a pair was counted as symbolic. If the scores were the same, or "big" had a larger score, then the pair was considered non-symbolic. So for example, in Malay *big* = *besar* = 1 point because it has a front vowel and a low, back vowel. S*mall* = *kechil* = 3 points because it has a front vowel and a high, front vowel, so this word pair would be considered symbolic. On the other hand, in Arabic *big* = *kabi:r* = 2 points because there is a high, front vowel. Long vowels were considered the same as short vowels for this analysis and were presented to subjects as short vowels. The Arabic *small* = *saghi:r* = 2 points again because of the high, front vowel, so these two words are equivalent and the pair would be considered non-symbolic. A total of 20 word pairs were compiled and the first eight pairs in each of the symbolic and non-symbolic groups were used in the final two lists. Of the eight pairs in each list, seven were two syllable words and one pair contained one syllable words.

In addition to the sixteen foreign word pairs, twenty-five nonsense word pairs were created and included in the survey. These words were added in the hopes of diluting any obvious patterns and also to pursue some interesting tangents. Five of these were designed to test how similarity to an English word would affect how subjects assigned meaning. Two synonyms of *small* and three synonyms of *big* were each modified slightly. The synonyms of *small* were contrasted against a nonsense word with higher frequency phonemes while the synonyms of *big* were contrasted against a nonsense word of lower frequency. If subjects chose consistently with the Frequency Code, they would choose a different word than if they were affected by the similarity to an English word. An additional ten nonsense words were designed to test the relative strengths of consonants and vowels as signals of size. English speakers judge voiceless stops as smaller than voiced stops, and fricatives as smaller than stops (Klink 2000). Therefore for five words, a word with a voiceless stop and a low or back vowel was paired with a word containing a voiced stop and the vowel [i]. This was called the "voiceless vs. voiced stop" group. In the other five words, a word with a fricative and a low or back vowel was paired with a word containing a stop and the vowel [i]. This was labeled the "fricative vs. stop" group. Finally, five words were designed as a positive control to be symbolic with both consonant and vowel contrasts following the frequency code. The remaining five words were designed as a negative control to be non-symbolic using vowels and consonants of similar frequencies.

The orthography of all words presented to subjects was adjusted so that native English speakers would read words with at least approximately correct pronunciation. For example, the high, back vowel [u] was represented as *oo* in positions where it would not be pronounced that way. The Basque orthography *tx* was converted to its English equivalent *ch*. The Korean vowel *eu* (/ʌ/) was represented as *u*, and the rounded front, high vowel [ü] was written as *i*. Although in most words [a] would tend to be pronounced as [æ] by English speakers, there were no cases in which an [æ] instead of an [a] would greatly change the frequency contrasts in the pair of antonyms, so no changes were made to this orthography.

The final survey included 41 word pairs randomized for order and for whether "big" or "small" was presented first. The sequence order was determined from random numbers generated at http://www.random.org, a web site that generates random numbers using atmospheric data. The order of the words in each pair was determined by tossing a coin. The survey was sent by email as a Word document to 60 college students between the ages of 18 and 22 who were all native speakers of American English. Subjects were asked to mark the word in each pair that they thought sounded like it meant "small". Subjects were asked to report any languages that they knew, had studied, or were familiar with. The format of the survey can be viewed in Appendix I.

Performance on the foreign word pairs was measured in two ways. A binomial test was used to evaluate whether subjects matched pairs correctly at frequencies greater than chance for all foreign word pairs and for the symbolic and non-symbolic groups separately. To directly compare performance on symbolic and non-symbolic word pairs, the number of words subjects correctly matched for each group was compared using a Wilcoxon signed-rank test for non-normal data. This test takes into account individual variation, in this case by analyzing each individual's performance on the symbolic word pairs relative to their performance on the non-symbolic word pairs. Thus if most subjects scored better on the symbolic word pairs than they did on the non-symbolic word pairs, this would lead to a significant result even if the mean scores on the two groups of word pairs were very similar. The five groups of invented words were not analyzed with any statistical test because of the small number of words tested in each one. Instead, the average number of "correct" guesses was calculated for each group and then compared with the other groups. As in Brown *et al*. (1955), the percentage of subjects that correctly matched each pair was also analyzed using a binomial test. Given the large number of statistical tests conducted, I chose to use a Bonferroni correction such that P-values less than 0.001 were considered significant.

3.2 Results

In total, 56 students completed the survey. Two subjects were familiar with Hungarian, and one was studying Arabic, therefore their answers on these questions were discarded from the analysis. Overall, subjects matched foreign word pairs correctly with their meaning more often than chance (Table 1). This was also true for the group of symbolic foreign word pairs, but not for non-symbolic pairs.

Table 1. The average percentage of correctly matched pairs for all foreign words and for symbolic and non-symbolic word pairs separately (n = 56 subjects).

Group	**% Correct**	**Significance of Departure from Chance**
Foreign word pairs	59.9%	$p < 0.001$
Symbolic word pairs	70.1%	$p < 0.001$
Non-symbolic word pairs	49.8%	$p = 0.04$*

*Not significant after the Bonferroni correction

The percentage of subjects that matched a given symbolic foreign word pair correctly ranged from 41.1% to 92.9% (Table 2). This range for non-symbolic foreign word pairs was 30.4% to 69.6%. In general subjects matched significantly more symbolic foreign word pairs correctly (median = 6) than non-symbolic word pairs (median = 4; $t = 5.81$, $df = 55$, $p < 0.0001$).

Table 2. The percentage of subjects (n= 56) that correctly chose the word meaning "small." "Big" is presented first in each word pair.

Word Pair	**Language**	**% correct**
"Symbolic"		
nagy- kicsi	Hungarian	75.5% *
handi- chiki	Basque	92.9% *
chempo – chooncoon	Tibetan	41.1%
besar- kechil	Malay	64.3%
mwuk- ngin	Trukese	71.4%*
nui- iki	Hawai'ian	76.8%*
babba- tsigil	Hausa	62.5%
nene- nini	Kikuyu	78.6%*
"Non-symbolic"		
kabir- saghir	Arabic	58.2%
iri- kicik	Turkish	30.4%*
kuda- jagun	Korean	39.3%
kulu- nono	Bemba	69.6%*
agi- nta	Igbo	39.3%
khulu- ncane	Zulu	62.5%
ge- ah	Cayuga	35.7%
waru- uta	Tarahumara	64.3%

*P< 0.001

Among the invented words, the percentage of subjects "correctly" matching a given word pair with its meanings ranged from 30.4% to 87.5% (Table 3). The average percentage of subjects choosing the higher frequency word among symbolic control pairs was 80.6%. The average for word pairs in the fricative vs. stop group was 76.4%, 63.4% for the voiceless vs. voiced stop group, and 53.6% for the English-like word pairs. The average percentage of subjects choosing the "correct" word of the non-symbolic word pairs was 45.0%.

Table 3. The percentage of subjects (n= 56) that chose the word containing the higher frequency vowel(s) for 25 invented words. For non-symbolic control words, one word was designated "correct" for scoring purposes. The word containing the higher frequency vowel, or the "correct" word is presented second.

Word Pair	**% choosing higher frequency vowel**
Positive Control: symbolic	
balaz- kilsu	73.2% *
godan- ipich	83.9% *
gujre- shekri	76.4% *
orveb- filka	87.5% *

vadong- inkush	82.1% *
Negative Control: non-symbolic	
felpo- athek	35.7%
zench- chedj	46.4%
watla- kwona	53.6%
brelko- egralp	30.4% *
ob- da	58.9%
English-like	
smar- shil	62.5%
ahob- hugi	58.9%
dabo- ibig	53.6%
dondag- gigank	30.4%*
minag- ziji	62.5%
Voiceless vs. Voiced Stops	
pako- bidu	55.4%
fok- thig	67.9%
atin- udin	60.0%
opor- dimir	85.7% *
tras- grish	48.2%
Fricative vs. Stop	
yaze- yeggi	73.2% *
cho- twi	75.0% *
shozat- kizta	66.1%
haju- habi	81.8% *
aflet- pliki	85.7% *

*$P < 0.001$

3.3 Discussion

Subjects in this experiment were able to correctly choose which word meant "small" at frequencies greater than chance when a word pair was symbolic. Furthermore, when the pair was non-symbolic, subjects failed to do better than chance. This indicates that speakers used vowel frequency to make at least some choices. There are also indications that subjects may have used other types of sound symbolic cues to correctly translate pairs. Although exactly half of the word pairs had symbolic vowel contrasts, the overall pairs were matched correctly 60% of the time. Also for 11 out the 16 pairs more than 50% of the subjects made correct translations. Some pairs that were labeled "non-symbolic" may still have had symbolic consonant contrasts which helped subjects translate them correctly. For example, *kabir-saghir* contrasts a fricative with a stop in the initial syllable which would agree with the frequency code. Although the percentage of subjects correctly translating this pair was not significant, it did approach 60%. The prominence of contrasts in accented syllables may also have outweighed the overall vowel contrasts in certain pairs, further contributing to high scores on some non-symbolic pairs. These factors along with some chance ones may have contributed to the overall success rate being above 50%.

In addition the results suggest that other criteria which subjects may use on such tests are not consistent enough among a large group of words to be sufficient for success. In the case of individual word pairs, there may have been factors unrelated to sound symbolism which lead subjects to guess correctly. For instance, in the pair *waru-uta* both words are very similar aside from the fact that *uta* is shorter. Word length in this case may explain the 64% of subjects who guessed correctly on this pair. This is also true for *iri-kicik*, where 70% of subjects guessed incorrectly that the shorter word *iri* meant small. However, in the context of the larger group of words, word length does not play a factor because these two word pairs cancel each other out. In fact, there was no factor or combination of factors which allowed subjects to consistently guess correctly and do better than chance across the non-symbolic word pairs.

The results from responses to the invented words indicate that conflicting signals reduce a group of test subjects' ability to correctly match pairs to differing degrees. When a fricative-stop contrast ran counter to the relative vowel frequencies in a pair of words, subjects seemed to ignore this information and focus on the vowel contrasts. A contrast between voiced and voiceless stops, however, seemed to have a stronger effect, consistent with Klink (2000). English-like words had the strongest effect in countering a signal from symbolic vowels. Sharing one syllable with an English word appeared to be sufficient to keep more than 60% of the subjects from choosing the word with higher frequency vowels to mean "small." Sharing two syllables, as in the *dondag-gigank* pair appeared to be enough to have a significant percentage of subjects choose against the frequency code. In the one case where I accidentally created a word containing an English diminutive prefix, *dim-*, that also reinforced the information from the vowel contrasts, the result was a very high score (*opor-dimir* 86%). These results confirm the fact that a likeness to an English word could certainly explain why many subjects correctly translate a certain word pair, and suggest that highly significant results should be carefully checked for evidence of this caveat.

4 Analysis of Brown Data

4.1 Methods

Part a

Although on average subjects in the foreign word pair test in the previous section seem to have used sound symbolic cues to match pairs, this does not necessarily explain the results of Brown *et al*. (1955). The Brown *et al.* results were therefore analyzed to rule out several trivial cues that subjects might have used and strengthen the conclusion that the results were largely due to concordance with subject's intuitions about sound-meaning correlations. Given that resemblance to English can have a strong effect on judgment, foreign words were examined for resemblance to English and whether they shared their first letter or an internal consonant cluster with a member of the English word pair or an obvious synonym. In addition, the effect of length was measured in two ways. In the first, the data were examine for instances in which subjects matched the longer foreign word to the longer English word and a significant percentage of subjects answered correctly. In the second, word length was examined for word pairs that related to magnitude in any way. While the use of longer words to represent larger magnitude is a form of symbolism, it is different from the types of sound symbolism that have been examined so far. Therefore it was counted among the random factors that might lead a significant number of subjects to translate a word in the same way, even though it may not be a random factor.

Part b

The word pairs presented in Brown *et al.* were also analyzed to determine whether pairs for which subjects did especially well on multiple languages had any common element. This information would indicate the semantic domains in which we would be most likely to observe universal sound symbolism. "Especially well" was arbitrarily defined as 80% of subjects correctly guessing a given pair. In addition, a pair of antonyms was only selected if subjects did especially well for at least two of the three languages tested. An average score (% subjects that guessed correctly) across the three languages tested was calculated for each word pair that qualified.

4.2 Results

Part a

Although there were some resemblances to English words among the dataset, these explained less than half of the correct choices. In a total of 15 cases, one of the foreign words presented shared a first letter with its English translation. Brown *et al.* (1955) administered the test under two conditions (written and oral presentation vs. just written presentation), so in total there were 30 instances out of 126 total possible instances in which subjects were presented with a foreign word that shared a first letter with its correct translation (see Appendix II). Of these 30 instances, a significant percentage of subjects gave a correct translation 13 times. In other words, subjects matched a foreign word with an English translation because they shared a first letter in no more than 43% of cases in which such a strategy was possible. Of six instances in which an internal part of the foreign word was similar to its English translation, a significant percentage of subjects guessed correctly in three of those six cases. Word length by either measure did not seem to play a large role. In total there were 22 instances in which subjects were presented with a foreign word visually longer than its antonym whose correct translation in English was also visibly longer than its antonym. Subjects only matched the longer foreign word with the longer English word in 7 of the 22 instances. Eight of the 21 word pairs used in the experiment were clearly a type of magnitude. For 8 of the 24 translations of these pairs, the longer word corresponded to the larger magnitude. However, subjects guessed correctly significantly more than chance in only four of the cases where length and meaning were correlated. Overall, of all 44 instances where a significant percentage of subjects guessed word meaning correctly, 18, or 41%, could potentially be explained by word length or chance resemblance to English.

Part b

For 5 out of 21 word pairs 80% or more of the subjects correctly translated the word pair into at least two languages. These word pairs and the average % of subjects that correctly matched them to English across Chinese, Czech, and Hindi were: *hard-soft* (80%), *blunt-sharp* (77%), *bright-dark* (73%), *fast-slow* (70%), and *long-short* (69%). In addition, *warm-cool* (69%) showed significant results for all three languages, but fewer than 80% of subjects guessed correctly for the three languages.

4.3 Discussion

Part a

The results suggest that the subjects in Brown *et al*. (1955) successfully assigned translations to at least some foreign word pairs on the basis of sound symbolism. It is striking that roughly 40% of their correct responses could be explained by seemingly random factors, and it is possible that this is an underestimate. There is always the possibility of a similarity to English which I did not recognize, but which was clear to those in the study. It is unlikely, however, that all of the correct responses could be explained by similarity to some English word not apparent to me, especially when the similarity must have been apparent to a very high percentage of subjects in the study. In addition, if universal sound symbolic patterns exist, then we would expect there to be similarities to English among the foreign word pairs. Some of what I am conservatively assuming to be random similarity might in fact be symbolic. It seems safe to agree with the conclusion in Brown *et al*. (1955) that at least in some cases, subjects must be using sound symbolic patterns in the foreign word pairs to assign correct translations.

Part b

Interestingly, the five most successful word pairs from Brown *et al*. were all very basic sensory terms. This is in contrast to other pairs in the experiment such as *beautiful-ugly*, *happy-sad*, or *sweet-sour*. *Bright-dark*, *fast-slow*, and *long-short* could all be considered types of magnitude, so it is not surprising that subjects have strong intuitions about them. *Hard-soft* and *blunt-sharp* are also basic sensory terms that might be represented by synesthetic symbolism, both of a tactile-auditory sort. *Blunt-sharp* is particularly reminiscent of the maluma-takete phenomenon as these adjectives would describe the pointy and curvy pictures if they were instead three-dimensional. We might expect that tactile-auditory symbolism would be widespread since the tongue movements that produce speech sounds would be associated with tactile sensations. Given that this notion is supported by the high success rates subjects had on the *blunt-sharp* and *hard-soft* pairs cross-linguistically, it may be worthwhile in the future to examine the distribution of tactile-auditory symbolism in a larger set of natural languages.

5 Antonym Pair Conformity to the Frequency Code

The results so far are consistent with the idea that the Brown *et al.* results were likely due to universal sound symbolic patterns incorporated in natural language. Subjects did well on the sound symbolic word pairs in section 3, and they were not able to do well without vowel frequency cues indicating the word that meant small, suggesting that they have also needed sound symbolic cues to do well in previous experiments. Whether such cues would be sufficient for subjects to do well on a wider sample of languages is another matter. This section examines the incorporation of the frequency code into natural languages since this pattern is the best studied and is the best defined.

5.1 Methods

The only way to definitively determine how widely the frequency code is expressed in natural languages is to sample a very large number of them. Even if one were just randomly choosing one word that meant "small" and one that meant "big" from each language, this would be a daunting task. In reality, one would need to sample all of the words meaning "small" and "big" from each language, score them, account for word length, and somehow give each language a score based on this information. This would only give us information about adjectives like *big, huge,* *enormous*, *small, little, tiny* however, and we would not have any idea about whether other areas of the language have been affected by the putative sound symbolic constraint. Clearly other ways to estimate the frequency code's pervasiveness in natural languages are necessary.

The word list generated for the foreign word pair experiment above can be used to give a very rough estimation of how significantly the frequency code is represented among words meaning "big" and "small." Nothing can be concluded from the list itself because we have no sense of how often in a randomly chosen pair of words one of the words will have more front and/or high vowels than the other. Furthermore, the sample of languages is small, and only one word pair is taken from each language. Instead, I compared the list with two other lists of word pairs, compiled using the exact same method and the same dictionaries, with the addition that three and four syllable words were also used if one or two syllable translated pairs could not be created. For both of these comparison lists, I chose simple words that I expected to find in all dictionaries. One list contrasted translations of *river* against translations of *sun* while the other contrasted *eat* against *walk*. There is no reason to expect that in either case one of the words will consistently have a higher score than the other, as measured in section 3.1, so this should provide a baseline from which to see if "big" and "small" deviate from chance. A χ-squared analysis was used to determine whether the occurrence of symbolic word pairs in the big-small list was significantly different than in the eat-walk or river-sun lists. Because the word corresponding to "small" in the *eat-walk* and *river-sun* lists was arbitrary, I determined symbolic and non-symbolic word pairs first with one word corresponding to "small", then the other, and used the condition in which the most word pairs were symbolic for the analysis. This was the most conservative approach since it decreased the difference in the number of symbolic word pairs between the three lists.

5.2 Results

Word pairs were translated into a total of 22 languages, but it was not possible to analyze the translated pairs in every one of these languages. It was unfortunately not always possible to create a translated pair where each word had the same number of syllables, and it was unclear how pairs with different numbers of syllables should be compared. It was especially difficult to create translated pairs with equal numbers of syllables for the nouns as these entries generally only had one translation. Therefore the total number of translated pairs that could be analyzed in the big-small list was 20, in the eat-walk list 20, and in the river-sun list it was 15. The big-small list of word pairs contained a larger proportion of symbolic pairs than either of the other two lists, but this difference was not statistically significant ($\chi^2 = 2.79$, df = 2, $p > 0.05$; Fig. 1).[2]

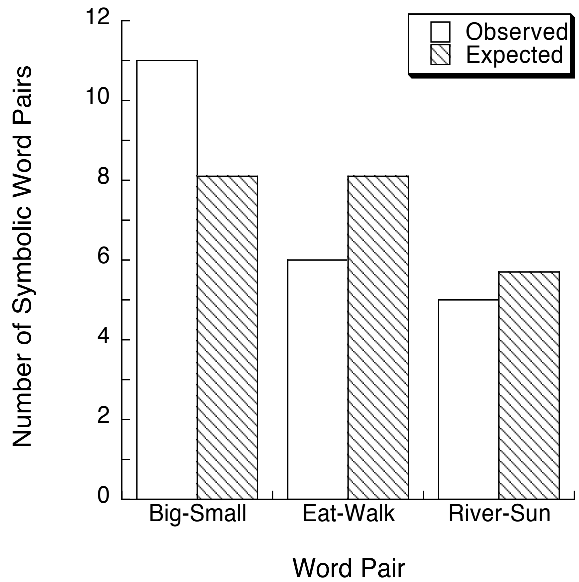


Figure 1. The number of big-small word pairs following the sound symbolic pattern described in section 4.1 was not significantly greater than chance. The expected number of symbolic pairs is based on the total number of languages each word pair was translated into. In total, *big-small* was translated into 20 languages, eat-walk into 20, and river-sun into 15 languages.

[2] In this χ-squared analysis, the null hypothesis is that the distribution of symbolic pairs across the three lists is due to chance alone. We can calculate an expected number of symbolic pairs for each list, and then if the observed number of symbolic pairs diverges enough from the expected number, we can reject the null hypothesis. The expected number of symbolic pairs is based on the total number of languages each list was translated into and on the average proportion of symbolic pairs to total word pairs. So if in the *big-small* list, about half of the pairs were symbolic and in the other two lists about a third were symbolic, then the expected number of symbolic pairs for each list will be between a half and a third of the total

5.2.3 Discussion

The above result suggests that sound symbolic intuition does not place a constraint on the creation of sensory adjectives, or that if it does, the constraint is very subtle. The proportion of symbolic pairs in each list was greater in the *big-small* list than in the other two—about half of the *big-small* pairs were symbolic while only about a third of the pairs were symbolic in the other two lists. If the proportion of symbolic pairs in each of the *big-small*, *eat-walk*, and *river-sun* lists remained constant at one half and one third as the number of languages sampled increased, the difference between the lists would become significant after analyzing about 50 languages. Even if half of the *big-small* pairs being symbolic was significantly greater than what we would expect by chance, though, this would indicate a very subtle constraint on language evolution.

One artifact of the methods that could have weakened the results was the way in which symbolic pairs were defined. To begin with, only vowels were considered, so that if there was a contrast between higher and lower frequency consonants but no vowel contrast, the pair would have been considered non-symbolic. Somehow taking consonants into account as well might have increased the number of symbolic pairs in the *big-small* list, although it likely would have increased the numbers in the other two lists as well.

Another factor that should be considered is the effect the sampling method used to create pairs for the analysis could have on the results. It may be that in many cases the big-small pairs analyzed were not really a pair in the minds of speakers of each given language. For example, "big" meaning "famous" may have been listed as a translation in one of the dictionaries and paired with a word meaning "small." No one

---

number of translated pairs in that list. The expected number will be the same for *big-small* and *eat-walk* because these lists were translated into the same number of languages, but the expected number for *river-sun* will be smaller because this list could only be translated into 15 languages.

would normally pair these two words together as opposites, and so there might not be a sound symbolic constraint on them. Similarly, "big" meaning "expansive" could have been paired with "small" meaning "short in stature." These are still adjectives that convey magnitude and might normally follow a sound symbolic pattern when contrasted with their natural opposites "narrow" and "tall in stature", but do not always when they are paired with each other. In fact, if *big-small* word pairs had been symbolic significantly more than expected, this would have indicated a strong constraint on size word creation since it would mean that almost all "big" related words used a certain subset of vowels and all "small" related words used another subset of vowels. However, this was not the case, so while we might find a constraint with a larger sample size, at best it would be a weak one.

6 Sound symbolism, Arbitrariness, and the Evolution of Language

As the above results suggest that sound symbolism is at best inconsistently incorporated into natural languages, it is worth considering why we would expect it to be incorporated at all. As mentioned earlier, the idea of universal sound symbolism has been particularly interesting in the context of the origin of language. In order for communication between two individuals to work, both need to share the same understanding of what the signals they are using mean. In moving from a situation where the is no communication to one where there is communication, participants must first develop a common set of basic signals. Computer simulations suggest that even if humans created arbitrary signals to refer to things around them, they could have come to a common understanding about what those signals meant (Steels and Kaplan 2002) and subsequently developed a language (Nowak and Komarova 2001). The development of language in this type of simulation is dependent on a chance event whereby the signs representing some feature of language come to have a common element and humans are able to construct mental rules based on the shared common element. If humans shared common assumptions about how certain sounds should be used to indicate distance from the speaker or to name certain properties or objects this would have speeded up the process of coming to a consensus and in some circumstances might have increased the incidence of chance resemblances. Once language was established, sound symbolic words would also be processed faster and would be more easily comprehensible (Bergen 2004, Hinton *et al*. 1994). Therefore groups of people that utilized sound symbolism would probably have developed language earlier and would have used it more effectively once developed, leading them to become more successful than other groups. The selective advantage of incorporating sound symbolism into language, especially during the initial process of development, makes it seem likely that we would observe sound symbolism in language.

On the other hand, there are reasons why arbitrariness in language should be selected for which would tend to counter the selective advantage of sound symbolism. One is that not all concepts can be metaphorically related to a set of sounds. We can imitate simple qualities like magnitude and texture with sound, and perhaps we could even derive names for things based on their salient qualities, but how should *sixty-four*, *interesting,* or *justice* be represented? At a certain point, sounds must be arbitrarily assigned to meanings in order for people to express their ideas. So the ability to create and learn words composed from arbitrary sounds would be selected for, and in fact must have been selected for in order for us to have developed a language that can express complex concepts.

A second selective advantage of arbitrariness, somewhat paradoxically, also arises from the need for the signs in communication to be shared between participants.

In order for any type of communication to function, all individuals involved must agree on the meaning of whatever signals are involved. If there is disagreement, then communication fails. Selection will therefore favor individuals that learn the communication system of their parents precisely over those who invent appropriate-sounding words for concepts that they want to express (Pinker and Bloom 1990). While this would tend to preserve sound symbolism if early language was highly sound symbolic, children must also be able to learn arbitrary forms. Children that learn whatever language their parents speak and pay no attention to whether words "sound right" will be better able to communicate than children that have trouble learning words that do not "sound right" to them. For this reason, it is possible that the separation of language acquisition from intuitions about sound symbolism and how words should sound would be adaptive.

It is further conceivable that the necessity for children to learn language without potential interference from notions about what "sounds right" could occur by separation of the language faculty as a whole from sound symbolic intuitions. This is supported somewhat by evidence that we have different modes for listening to speech and non-speech sounds. When we are listening to speech sounds, we generally deduce and attend to the sequence of abstract phonemes. Most of the acoustic information that allowed us to deduce the phonemes is either shut out or occasionally it can be subconsciously registered (Tsur 2006). Thus sound symbolic cues (like frequency) in the acoustic signal would seem to receive little or no attention during normal language processing. More evidence supporting a separation between the language faculty and sound symbolic intuition comes from evidence from a lexical decision task described previously. Stimuli (words and non-words) were presented in spiky or curvy frames. When the sounds composing the stimuli symbolically matched the frame, decision times were shortened if the stimulus was a non-word, but not if it was a real word (Westbury 2005). The interference from a frame that did not match a stimulus' symbolic quality would seem similar in origin to the interference that a child might experience learning a word when the word's meaning disagreed with the word's symbolic quality. Since words were processed much faster than non-words, there is some evidence that words are processed before sound symbolic intuition can have an effect, and this would be advantageous for children learning language.

As with most types of behavior, language is under selective pressures that favor opposite things. Sound symbolism would be favored because it tends to make language easier to process and more understandable. On the other hand, arbitrariness will be favored because it is necessary for full expression, and learning language as if it were arbitrary will tend to be the most successful language learning strategy. The competing selective pressures favoring sound symbolism and arbitrariness, however, may not currently be of equal strengths. It is worth noting that the incorporation of sound symbolism into language would be most adaptive at the time language arose. Once language and the common vocabulary were well established, there would not seem to be as strong a selective advantage for the incorporation of sound symbolism, and the selective pressure toward arbitrariness may have become dominant. This may explain why there currently seems to be a separation between the language faculty and sound symbolic intuition, and also why sound symbolism only seems to be inconsistently incorporated into natural languages.

7 Conclusion

The body of research investigating universal sound symbolism has so far provided equivocal evidence for its existence. Speakers of a handful of unrelated languages share some sound-meaning associations, but there have not been enough studies with non-English speakers to rule out that these associations are language-specific and that agreement so far has been due to chance similarities in the languages studied. English, Chinese, and Japanese speakers seem able to guess the translations of pairs of antonyms in a foreign language more often than chance. From the results of this paper it seems that at least for magnitude-related words, they are able to do so predominantly by using the frequency code, which is a sound symbolic pattern. However, the same pairs of adjectives which they were able to judge correctly seem to be symbolic only slightly and not significantly more often than a standard pair of nouns or verbs. Other studies examining the distribution of high frequency vowels among magnitude words have also failed to find a significant correlation with diminutive words. There does not seem to be a strong constraint to follow sound symbolic patterns on the majority of the lexicon. On the other hand, there is evidence for a constraint on some parts of it. Ultan's findings suggest that ablauting is highly constrained when it indicates size or distance. Woodworth's findings also suggest a constraint on deictic pronouns, place adverbs, and directional affixes. In this case it is simpler to postulate that the constraint is universal and explain the exceptions than to explain the independent creation of so many forms following the same pattern. This is the strongest evidence for a universal sound symbolism to date, and the apparently widespread existence of exceptions and counterexamples may be explained by the selective advantages of arbitrariness in language.

References

Austerlitz, R. 1994. Finnish and Gilyak sound symbolism – the interplay between system and history. See Hinton *et al.* 1994, pp. 249- 256.

Bergen, B.K. 2004. The psychological reality of phonaesthemes. *Language* 80. 290-311.

Brackbill, Y., Little, K.B. 1957. Factors determining the guessing of meanings of foreign words. *Journal of Abnormal and Social Psychology* 54. 312-318.

Brown, R.W., Black, A. H., Horowitz, A.E. 1955, Phonetic symbolism in natural languages, *Journal of Abnormal and Social Psychology* 50. 388-393.

Brown, R.W. 1958. "Phonetic symbolism and Metaphor." In: *Words and Things*. Glencoe, IL: The Free Press, pp. 110- 154.

Brown, R.W., Nuttall, R. 1959, Methods in phonetic symbolism experiments, *Journal of Abnormal and Social Psychology* 59. 441-445.

Childs, G.T. 1994. African ideophones. See Hinton *et al.* 1994, pp. 181-194.

Davies, N.B., Halliday, T.R. 1978. Deep croaks and fighting assessment in toads *Bufo bufo*. *Nature* 274. 683-685.

Davis, R. 1961. The fitness of names to drawings: a cross-cultural study in Tanganyika. *British Journal of Psychology* 52. 259-268.

Fitch, W.T. 2000. The evolution of speech: a comparative review. *TRENDS in Cognitive Sciences* 4. 258-267.

Fromkin, V., Rodman, R., Hyams, N. 2003. "What is Language?" *Language*. Boston, MA: Heinle, pp. 3-12.

Hamano, S. 1998. *The sound-symbolic system of Japanese*. Cambridge, UK: Cambridge University Press.

Hinton, L., Nichols, J., Ohala, J. eds. 1994. *Sound symbolism*. Cambridge, UK: Cambridge University Press.

Klink, R.R. 2000. Creating brand names with meaning: the use of sound symbolism. *Marketing Letters* 11. 5-20.

Maltzman, I., Morrisett, L. Brooks L.O. 1956. An investigation of phonetic symbolism. *Journal of Abnormal and Social Psychology* 53. 249-251.

Marks, L.E., Hammeal, R.J., Bornstein, M.H., Smith, L.B. 1987. Perceiving similarity and comprehending metaphor. *Monographs of the Society for Research in Child Development* 52. 1-100.

Maurer, D., Pathman, T., Mondloch, C.J. 2006. The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science* 9. 316-322.

Morton, E.S. 1994. Sound symbolism and its role in non-human vertebrate communication. See Hinton *et al.* 1994, pp. 348- 362.

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

Nowak, M.A., Komarova, N.L. 2001. Towards and evolutionary theory of language. *TRENDS in Cognitive Sciencs* 5. 288- 295.

Nuckolls, J.B. 1999. The case for sound symbolism. *Annual Review of Anthropology* 28. 225-252.

Ohala, J. 1994. The frequency code underlies the sond-symbolic use of voice pitch. See Hinton *et al.* 1994, pp. 325- 347.

Ostwald, P.F. 1994. Some observations on the function of sound in clinical work. See Hinton *et al.* 1994, pp. 309- 324.

Parault, S.J., Schwanenflugel, P.J. 2006. Sound-symbolism: a piece of the puzzle of word learning. *Journal of Psycholinguistic Research* 35. 329-351.

Pinker, S., Bloom, P. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13. 707-784.

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. 2005. The mirror-neuron system. *Annual Review of Neuroscience* 27. 169-192.

Rogers, S.K., Ross, A.S. 1975. A cross cultural test of the Maluma-Takete phenomenon. *Perception* 4. 105.

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

Shimojo, S., Shams, L. 2001. Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology* 11. 505-509.

Steels, L., Kaplan, F. 2002. AIBO's first words: the social learning of language and meaning. *Evolution of Communication* 4. 3-32.

Tsur, R. 2006. Size-sound symbolism revisited. *Journal of Pragmatics* 38. 905-924.

Ultan, R. 1978. Size-sound symbolism. In J.H. Greenberg (Ed.), *Universals of Human Language vol. 2*. (pp. 525- 568). Stanford, CA: Stanford University Press.

Westbury, C. 2005. Implicit sound symbolism in lexical access: evidence from an interference task. *Brain and Language* 93. 10-19.

Woodworth, N.L. 1991. Sound symbolism in proximal and distal forms. *Linguistics* 29. 273-299.

Foreign Language Dictionaries

Arabic
Awde, N., Smith, K. 2004. *Arabic: Practical Dictionary*. New York: Hippocrene Books.

Basque
Aulestia, G., White, L. 1990. *English-Basque Dictionary*. Reno, NV: University of Nevada Press.

Bemba
Hoch, E. 1960. *Bemba Pocket Dictionary*. Tabora, Tanganyika: TMP Printing Department.

Cayuga
Froman, F., Keye, A., Keye, L., Dyck, C. 2002. *English-Cayuga, Cayuga-English Dictionary*. Toronto: University of Toronto Press.

Creek
Martin, J.B., Mauldin, M.M. 2000. *A Dictionary of Creek/Muskogee*. Lincoln, NE: University of Nebraska Press.

Hausa
Newman, R.M. 1990. *An English-Hausa Dictionary*. New Haven, CT: Yale University Press.

Hawaiian
Pukui, M.K., Elbert, S.H. 1986. *Hawaiian Dictionary*. Honolulu, HI: University of Hawaii Press.

Hungarian
Magay, T., László, K. 1996. *NTC's Hungarian and English Dictionary*. Chicago: NTC Publishing Group.

Igbo
Echeruo, M.J.C. 1998. *Igbo-English Dictionary*. New Haven, CT: Yale University Press.

Japanese
Adachi, T. 1999. *Kodansha's Furigana Japanese Dictionary*. Tokyo: Kodansha International.

Kikuyu
Barlow, A.R. 1975. *English-Kikuyu Dictionary*. T.G. Benson (Ed.). Oxford: Oxford University Press.

Khmer
Keesee, A.P.K. 1996. *An English-Spoken Khmer Dictionary*. London: Kegan Paul International.

Korean
Lee, J., Lee K. 2005. *Korean: Dictionary and Phrasebook*. New York: Hippocrene Books.

Malay
Winstedt, R. 1965. *A Practical Modern English-Malay Dictionary*. Kuala Lumpur, Singapore: Marican and Sons.

Mongolian
Hangin, J.G. 1970. *A Concise English-Mongolian Dictionary*. Bloomington, IN: Indiana University Publications.

Oneida
Michelson, K., Doxtator, M. 2002. *Oneida-English, English-Oneida Dictionary*. Toronto: University of Toronto Press.

Tarahumara
Thord-Gray, I. 1955. *Tarahumara-English, English-Tarahumara Dictionary*. Coral Gables, FL: University of Miami Press.

Tibetan
Goldstein, M.C. 1984. *English-Tibetan Dictionary of Modern Tibetan*. Berkeley, CA: University of California Press.

Trukese
Goodenough, W.H., Sugita, H. 1980. *Trukese-English Dictionary*. Philadelphia, PA: American Philosophical Society.

Turkish
Iz, F., Hony, H.C. 1992. *The Oxford Turkish Dictionary*. Oxford: Oxford University Press.

Zapotec
Britton, A.S. 2003. *Zapotec-English, English-Zapotec*. New York: Hippocrene Books.

Zulu
Doke, C.M., Malcolm, D.M., Sikakana, J.M.A. 1958. *English-Zulu Dictionary*. Johannesburg, South Africa: Witwatersrand University Press.

Appendix I: Survey

Directions: This experiment tests your intuitions about the meanings of foreign words based on how they sound. Each of the word pairs below is a translation of the English pair “big- small” or “small-big” into a different language. For each pair, please read the two words aloud to yourself, so you can hear how they sound, then bold/underline/circle the one you think means “small.” Do not think about a pair for more than 5-10 seconds- go with your first intuition. If at any point during the test you feel that you have looked at too many pairs and no longer have intuitions, please do something else for a while and finish the test later.

Please fill in any languages that you know, have studied, or are familiar with:____

i. (Example) grande **pequeño**

1)	udin	atin
2)	handi	chiki
3)	ncane	khulu
4)	saghir	kabir
5)	tras	grish
6)	jagun	kuda
7)	haju	habi
8)	iki	nui
9)	zench	chedj
10)	smar	shil
11)	pako	bidu
12)	nta	agi
13)	nini	nene
14)	inkush	vadong
15)	brelko	egralp
16)	ah	ge
17)	kulu	nono
18)	yaze	yeggi
19)	besar	kechil
20)	watla	kwona
21)	opor	dimir
22)	ibig	dabo
23)	waru	uta
24)	athek	felpo
25)	chooncoon	chempo
26)	aflet	pliki
27)	hugi	ahob
28)	kicsi	nagy
29)	kicik	iri
30)	minag	ziji
31)	mwuk	ngin
32)	kilsu	balaz
33)	godan	ipich
34)	da	ob
35)	cho	twi
36)	gujre	shekri
37)	fok	thig
38)	filka	orveb
39)	dondag	gigank
40)	babba	tsigil
41)	kizta	shozat

Appendix II: Brown Data

Percentage of correct translations for each pair in three languages for experimental conditions A (words written and spoken, n= 86) and B (words written only, n= 16)

		% Correct			% Correct			% Correct	
English	Chinese	A	B	Czech	A	B	Hindi	A	B
1. beautiful	mei	88*	70	krása	57	31	khubsurat	64*	50
ugly	ch'ou			osklivost			badsurat		
2. blunt	tun	78*	70	tupy	81*	83*	gothil	68*	83*
sharp	k'uai			spichaty			tez		
3. bright	liang	67*	90*	svetly	64*	77	chamakdar	51	90*
dark	an			tmavy			dhundhala		
4. coarse	ts'u	65*	70	hruby	21*	44	mota	48	31
fine	his			drobny			achha		
5. down	hsia	10*	31	dolu	56	50	niche	75*	83*
up	shang			nahoru			upar		
6. drunk	tsui	66*	50	opily	21*	70	nashe men	80*	77
sober	hsing			strizlivy			sanjida		
7. dry	kan	72*	70	suchy	44	50	sukha	42	44
wet	shih			mokry			bhiga		
8. fast	k'uai	83*	83*	richly	87*	83*	tez	27*	57
slow	man			pomaly			sust		
9. fat	fei	31*	57	tlusty	69*	77	mota	66*	57
thin	shou			tenky			patala		
10. gold	chin	57	57	zlato	19*	57	sona	42	64
iron	t'ieh			zelezo			loha		
11. bad	huai	34*	64	zly	62	57	kharab	64*	31
good	hao			hodny			achha		
12. happy	huan	38	50	radostny	57	64	khush	17*	38
sad	pei			smutny			ranjida		
13. hard	kang	97*	83*	tvrdy	76*	96*	sakht	61	64
soft	jou			mekky			narm		
14. light	ch'ing	93*	90*	lehky	66*	77	halka	36*	57
heavy	chung			tezky			wazani		
15. long	ch'ang	55	44	dlouhy	80*	70	lamba	90*	70
short	tuan			kratky			chhota		
16. many	to	73*	57	mnoho	55	25	bahut	88*	90*
one	yi			jeden			ek		
17. strong	ch'iang	37	64	silny	28*	64	mazbut	34*	31
weak	jo			slaby			kamzor		
18. sweet	t'ien	58	51	sladky	24*	25	mitha	88*	70
sour	suan			kysely			khatta		
19. thunder	lei	23*	31	hrom	92*	96*	garaj	62	77
lightning	tien			blesk			chamak		
20. warm	nuan	73*	50	teply	69*	77	garam	66*	77
cool	liang			chladny			thanda		
21. wide	kuan	37	90*	siroky	43	57	chaura	76*	51
narrow	chai			uzky			tang		

* $p \leq 0.01$