

Empirically Observed Iconicity Levels of English Phonaesthemes

Nahyun Kwon

This paper aims to reveal to what extent the *diagrammatic iconicity* (i.e., form-meaning correspondences which are created by relating similar sets of forms with similar sets of meanings [Peirce, 1955, p. 104]) of English phonaesthemes (e.g., *gl-* in *glitter*, *glisten*, and *glow*) could manifest *primary iconicity* (i.e., iconicity that involves a sufficient similarity between sign and referent to allow the understanding that the former stands for the latter [Sonesson, 1997]). To serve the aim, the current research conducts a test, using a multiple-choice task in which groups of native English and Korean speakers choose the meanings of phonaesthemes in sets of aurally-presented nonsense English phonaesthetic words. If primary iconicity is carried by a phonaestheme, then both native and non-native listeners should be able to report with some consistency the putative meaning of the nonsense phonaesthetic words. If, on the other hand, a form-meaning correspondence is carried by secondary iconicity (where the existence of the sign-relation, given by convention or by being explicitly pointed out, is the precondition for noticing the similarity between sign and referent [Sonesson, 1997]), then neither language group is expected to deliver high correct guessing rates. The result showed that the purported meanings of *sk-* and *tw-* were correctly guessed by the Korean-speaking participants only, and those of *cl-*, *gl-*, *sw-*, *gr-*, *sn-*, and *sq-* were correctly guessed by the English-speaking participants only. The purported meanings of *sp-* and *tr-* were correctly guessed by neither language group. These findings show that individual phonaesthemes possess varying degrees of (primary) iconicity.

Key words: phonaesthemes, language iconicity, sound-symbolism, behavioural experiments, English

1. Introduction

In many, if not all, natural languages, there are parts of the lexicon whose sound-meaning associations appear to fall short of full arbitrariness. For example, as a form of non-arbitrariness of the sign, English phonaesthemes display *diagrammatic iconicity* (Peirce, 1955, p. 104) by associating the shared form *sw-* with the shared meaning ‘swaying movement’ in a set of phonaesthetic words: *swirl*, *swivel*, *swift*, *swig*, *sweep*, *swallow*, *swarm*, *swim*, *swing*, *swipe*, *switch*, *swoosh*, *swoop*, *swill*, and *swoon* (Parault, 2006, p. 231). On a conceptual level, diagrammatic iconicity displays a less transparent mapping than *imagistic iconicity*, which uses acoustic signals of speech sound only to mimic acoustic phenomena, as in onomatopoeia. The question that arises is, then, to what extent the diagrammatic iconicity of phonaesthemes could reveal *primary iconicity* (i.e., iconicity that involves a sufficient similarity between sign and referent to allow the understanding that the former stands for the latter) allowing speakers of different languages to understand their meaning (Sonesson, 1997). This is a testable hypothesis.

The question of the presence (and the degree) of natural motivation in language has been of marginal interest in modern linguistics, since its foundational assumption has been that the link between linguistic form and meaning is “arbitrary” (Locke, 1975[1690]). However, despite the dominant presumption of the arbitrariness of the linguistic sign (Hockett, 1960, p. 97), the subject of motivated (sound-symbolic) relationships has long continued to be investigated by linguists and anthropologists. The earliest works on this subject include Sapir (1929) and Köhler (1929). Sapir (1929) claims that the high, front vowel /i/ is naturally related to smallness and that the low, back vowel /a/ is related to largeness, based on the finding that 80% of his subjects judged a nonsense word pair *mil* and *mal*, as a small and large table, respectively. In a similar vein, Köhler (1929) claims that there are universally accessible sound-shape correspondences, currently known as the “bouba-kiki effect” (Ramachandran and Hubbard, 2001), given findings that speakers cross-linguistically associate forms like *bouba* with a rounded shape and forms like *kiki* with a spiky shape (see

Ahlner and Zlatev, 2010). Such, possibly universal, sound-size and sound-shape mappings encouraged the idea that there may also be a transparent correspondence between certain sounds and meanings in natural languages, i.e. a high degree of primary iconicity. Accordingly, the number of cross-linguistic investigations into sound symbolism has been on the rise since the early 20th century (Hinton, Nichols, and Ohala, 1994; Nuckolls, 1999; see also the special issue of this journal on sound symbolism [Vol 5, Issue 1]).

In the 21st century, studies of sound symbolism are becoming increasingly influenced by the idea that the methodology must reflect ongoing psychological processes: “More critical evidence for humans’ sensitivity to iconic mapping must come from online studies of language processing that show that iconicity affects lexical processes” (Perniss, Thompson, and Vigliocco, 2010, p. 7). Following this recent trend, several psycholinguistic works on the advantage of iconicity in language learning (e.g., Nygaard, Cook, and Namy, 2009; Kovic, Plunkett, & Westermann, 2010; Monaghan, Mattock, and Walker, 2012) and language development (e.g., Maurer, Pathman, and Mondloch, 2006; Imai et al., 2008; Kantartzis, Imai, and Kita, 2011) have confirmed language users’ sensitivity to the iconicity of linguistic signs.

Proceeding from previous sound-symbolic literature, this study aims to determine the kinds and levels of iconicity of English phonaesthemes on an empirical basis, and to specify possible (universal) perceptions of the diagrammatic iconicity of such phonaesthemes, if there are any. The paper is structured as follows. Section 2 introduces the possible iconic statuses of individual phonaesthemes and reviews previous debates about the iconicity of phonaesthemes. Section 3 provides a general overview of current behavioural experiments. Section 4 examines English-speaking listeners’ delimitation of existing English phonaesthetic word-sets. Sections 5 and 6 report the English- and Korean-speaking listeners’ responses to sets of nonsense English phonaesthetic words, constructed with specific word-initial consonant sequences that were independently allocated to phonaesthetic sets, organised in a multiple-choice task. Section 7 discusses the results and provides a conclusion.

2. Phonaesthemes

2.1. Overview

Many, if not all, natural languages manifest sound-symbolic phenomena through words of sensory imagery, where the speaker’s sensory experiences in some sense resemble the linguistic signs that denote them (Nuckolls, 1999; Perniss, Thompson, and Vigliocco, 2010; Dingemanse, 2012). In the course of research, various terms have been introduced to designate such expressions in different languages. These terms include “expressives” (Diffloth, 1972; Klamer, 2001; Tufvesson, 2011) in South-East Asian languages, “ideophones” (Doke, 1935; Childs, 1988; Dingemanse, 2011) mostly in sub-Saharan African languages and indigenous languages of South America, “mimetics” (Kita, 1997; Hamano, 1998; Akita, 2009) in Japanese, and “phonaesthetic words” that contain *phonaesthemes* (Firth, 1930; Hutchins, 1998; Abelin, 1999; Bergen, 2004; Kwon and Round, 2015) in Indo-European languages.

Among these, phonaesthemes occur inside independent lexical items, unlike the others. For example, Japanese mimetics, which often exhibit regular correspondence between voiced initial consonants and larger or more intense connotation (e.g., *gira-gira* ‘glaring’ vs. *kira-kira* ‘twinkling’) (Akita, 2009, p. 15), can be analysed as having the voicing feature that would characterise them as phonaesthemes. Likewise, Bahnar expressives (Diffloth, 1994) and Korean ideophones (Kwon, 2015), where high vowels are associated with largeness and low vowels are associated with smallness (e.g., Bahnar: *bleel-bleel* vs. *bleel-beel* ‘large vs. small flames appearing intermittently but remaining vivid’; Korean: *pipi* vs. *pepe* ‘a state of bigger vs. smaller things being entwined’), can be analysed as having different vowel

qualities that would characterise them as phonaesthemes. The vital point is that, regardless of which sound-symbolic stems phonaesthemes appear in, they manifest diagrammatic iconicity by constituting a relational similarity (i.e. mapping) between similar sets of forms and similar sets of meanings.

Having seen the diagrammatically iconic nature of phonaesthemes, the following subsection characterises the possible iconic statuses of individual phonaesthemes, attending to Sonesson's (1997) distinction between primary and secondary iconicity.

2.2. Analyzing the iconicity in phonaesthemes

A conceptual analysis of the iconicity of English phonaesthetic words can be built on two of Peirce's (1955) types of icon¹ (see also Haiman, 1980) – *images* and *diagrams* – and Sonesson's (1997) further distinction of iconic signs into primary and secondary iconicity.

In somewhat simplified Peircean terms, an *icon* is a sign that displays a close resemblance to its referent (e.g., a portrait); an *index* is a sign that is, by contiguity, associated with a referent (e.g., smoke as an index of fire). A *symbol*, on the other hand, is a sign that is associated with its object (which may be an abstract idea) by convention (e.g., \$ for dollar). With respect to icons, which are the main concern of this paper, there are imagistic and diagrammatic types of iconicity. Possible examples of imagistic and diagrammatic iconicity are, respectively, onomatopoeic terms (e.g. *meow*, where there is a more or less one-to-one relation between the sound form and some aspect of the referent) and phonaesthetic words, where the similarity is much more schematic. It is worth noting, however, that all iconic signs do not exclude indexical or symbolic (i.e., conventional) relations/grounds, since the three Peircian signs are ideal types, and all actual signs involve combination of the different kinds of grounds (Jakobson, 1971[1965]; Ahlner and Zlatev, 2010). In this regard, Sonesson further categorises iconic signs as primary and secondary. In *primary iconicity*, a sufficient similarity between sign and referent is the precondition for the understanding that the former stands for the latter. In *secondary iconicity*, the existence of the sign-relation, given by convention or by being explicitly pointed out, is the precondition for noticing the similarity between sign and referent. In order to illustrate the difference between primary and secondary iconicity, Ahlner and Zlatev (2010) provide two drawings, shown in Figure 1.

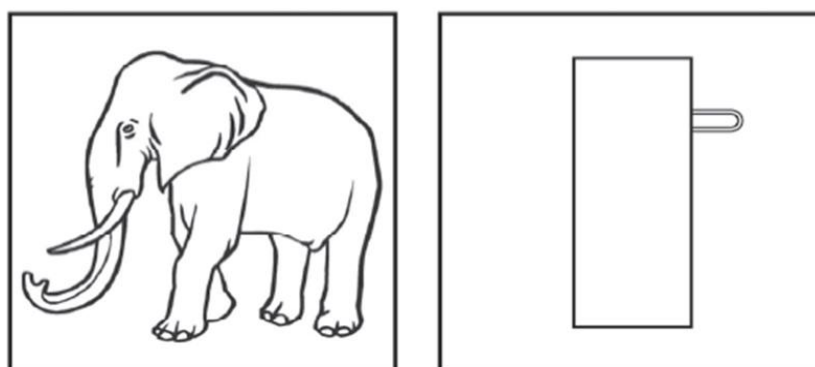


Figure 1. A picture of an elephant as an example of a primary iconic sign, and a doodle for an example of a secondary iconic sign (adapted from Ahlner and Zlatev, 2010, p. 316)

¹ In Peirce's original taxonomy of signs, "metaphors" is listed as a third type of icon, but since the distinction between diagrams and metaphors is controversial (cf. Chen and Su, this volume) and arguably not relevant for sound symbolism, its discussion is omitted in this paper.

For the primary iconic sign of an elephant, what is depicted can be clearly seen from the drawing itself, so the transparent similarity is the ground for establishing the sign function/relation. For the doodle, on the other hand, the interpreter can find some resemblance between the drawing and the depicted object only after he/she has been told what the sign function/relation is (e.g., it could be a depiction of a person playing a trombone in a wardrobe, or a paper clip placed under an envelope).

Discussing the bouba-kiki effect, Ahlner and Zlatev (2010) argue that sound symbolism might involve at least a degree of primary iconicity. With respect to a phonaesthetic phenomenon, for example, the phonaestheme *sn-*, where the nasal sound /n/ seems to evoke its associated semantic domain, the ‘nasal/oral area’, seems to possess a high degree of primary iconicity (Monaghan et al., 2014; Schmidtke, Conrad, and Jacobs, 2014). In line with this, the next sub-section provides a historical overview of conflicting opinions as to which property (“naturalness” vs. conventionality) played a significant role in the establishment of the diagrammatic iconicity of phonaesthetic words in general.

2.3. Debates about the iconicity of phonaesthemes

In the research on phonaesthetic iconicity, two schools of thought have arisen – one group in favour of so-called “naturalness” (Bolinger, 1950; Jakobson and Waugh, 1979; Rhodes, 1994), and the other in favour of conventionality (Firth, 1930; Fordyce, 1988; Markel and Hamp, 1960). For example, Bolinger (1950) places emphasis on the natural form-meaning mapping in the denotations of phonaesthemes upon his observations that phonaesthemes² such as *fl-* (e.g., *flow*, *flush*, *flood*) give the impression of ‘phenomena of movement’ to native English speakers without involving apparent etymology. Jakobson and Waugh (1979) claim that native speakers intuitively feel the phonetic groupings of similar meanings for which “... the patent or latent role played by the ‘intrinsic value’, *videlicet* by the spell of the speech sounds, is undeniable” (p. 198). Rhodes (1994) also stresses the role of natural motivation, particularly in onomatopoeic phonaesthemes, such as *cl-* (e.g., *clank*, *clang*, *click*) and *cr-* (e.g., *crack*, *creak*, *crunch*), by relating the abrupt increase in amplitude of the stop /k/ to their noise-associated meanings.

Other scholars emphasise the conventional nature of phonaesthemes. Firth (1930), for example, considers the phonaestheme *sl-* to be correlated with ‘pejorative’ meanings, as in *slack*, *slouch*, *sludge*, *slime*, *slosh*, due to the regular and habitual sound-meaning pairing occurring in the lexicon. He claims that there is no intrinsic value of *sl-* that could suggest the associated meaning, and that phonaesthemes in general are merely an output of a “phonetic habit” of language users. In a similar vein, Markel and Hamp (1960) suggest that the term phonaestheme should be labelled “psycho-morph” – “a non-morphemic unit of one or more phonemes for which a cultural meaning can be established” (ibid. 55) in which “cultural meaning” indicates a primary role of conventional relationship between form and meaning. As empirical evidence for the conventional basis of phonaesthemes, Fordyce (1988) showed in a psycholinguistic experiment that phonaesthemes are language-specific, as Arabic and Japanese subjects could not guess the meanings of English words containing phonaesthemes, such as *sn-*, *gl-*, *cl-*, and *fl-*, at a significant level. The responses of subjects from the two different language groups also showed a significantly low correlation value, $r = .14$.

Taking into account the arguments made by both the “naturalist” and the “conventionalist”, it would be pertinent to see if the heated debate could be resolved by paying more attention to the differing degrees of primary and secondary iconicity in individual English phonaesthemes (see also Monaghan et al., 2014). Empirical investigations into this conjecture are described in the following sections.

² This corresponds to “sub-morpheme differentials” in Bolinger’s terminology.

3. Overview of the experiments

The membership of a phonaesthetic word family is “fuzzy” (Fordyce, 1988, p. 178). For example, it is not clear whether to include *snob* (Blust, 2003, p. 188), *snooty* or *snoop* (Fordyce, 1988, p. 177) in a *sn-* ‘nasal/oral area’ phonaesthetic word family; they are somewhat related to the notion of nasality but not as apparently as *sneeze*, *snuff*, and *sniffle* (Fordyce, 1988, p. 177). Such fuzziness causes difficulties when one wishes to build a rigorous dataset to investigate the nature of form-meaning pairings in English phonaesthemes.

Fordyce (1988) attempts to draw a clear-cut line between the core and peripheral members in a phonaesthetic word-set using native English speakers’ intuitions. As a result, Fordyce found that native English speakers were able to choose those lexical items that are strongly interconnected with each other in five lists of English words containing the phonological component of the phonaesthemes, *dr-*, *cl-*, *fl-*, *sl-*, and *str-*, and to express the shared meanings of the chosen words. To take an example, more than 90% of his subjects rated that *cling*, *clutch* and *claw* were the lexical items central to the semantic domain of the *cl-* word family and agreed that they shared the meaning related to ‘grasping motion with the hands’ and ‘closeness’. Thus, he suggests that there are lexical items that are central to a phonaesthetic word-set, and that the core members should reveal a tighter form-meaning mapping than the members that are located far from the centre of the semantic domain of a phonaestheme.³

Regardless of this finding, no studies, including Fordyce (1988), have taken into account potential differences in the level of iconicity lying between the core and peripheral phonaesthetic members when investigating the nature of English phonaesthemes. Consequently, this has made their results rather inconclusive. For example, Fordyce’s (1988) study argued that English phonaesthemes are exclusively based on convention. However, in the sense that he examined the basis of phonaesthemes without removing peripheral phonaesthetic members from his stimulus set, it is possible that the role of convention in phonaesthemes was exaggerated by the phonosemantically vague peripheral items. This warrants further investigation, as also pointed out by Fordyce (1988, p. 237) “the slight overall trend toward better than expected correct matching, though not significant, suggests the need for further research into possible sound symbolism among English phonaesthemes”.

Motivated by these considerations, I first delimited the phonaesthetic word-sets through a preliminary experiment, where native English speakers rated how closely individual candidate phonaesthetic words were related to a proposed key word of the given phonaestheme (Section 4). With the selected core exemplars that were judged to possess high semantic interconnection with the relevant key word, I created nonsense phonaesthetic words. Then, I asked native speakers of either English or Korean to guess their shared meanings in sets of aurally presented nonsense English phonaesthetic words in a multiple-choice task (Sections 5-6).

4. Preliminary experiment for a selection of core English phonaesthetic words

The aim of this preliminary experiment was to identify words that belong to the semantic core of phonaesthetic word-sets in English speakers’ mental lexica. In Fordyce’s (1988) experiment on native speakers’ delimitation of English phonaesthetic words, speakers were asked to cluster semantically related words and to give the shared meaning for a set of words which contained the phonological component of a phonaestheme. However, in the preliminary experiment, participants were asked to rate how closely individual candidate phonaesthetic words were related to a key word supplied by the author, for which the

³ This can be related to the notions of prototype-based categories (e.g., Lakoff, 1987) and relative phonaesthematicity of different phonaesthemes (cf. Abelin, 1999).

purported meaning of a phonaestheme was clear. This method aimed to ease the fatigue experienced by the participants, as it involved only a single rating task. In terms of stimulus items, Fordyce randomly selected words that shared the phonological component of a phonaestheme from Webster's Dictionary. In contrast, the current experiment used words that had been assigned to the phonaesthetic category by Hutchins (1998). This was expected to narrow down the semantic domain of stimulus items and thus help to extract words that conformed to the proposed meaning of a phonaestheme in a semantically strict manner.

A prediction that guided the current experiment was that if an English phonaestheme possesses both a semantic core and a periphery, then a list of phonaesthetic words should reveal gradual ratings to its different exemplars.

4.1. Method

4.1.1 Participants

English-speaking students, who were over 18 years of age and enrolled in an introductory linguistics course at the University of Queensland, served as participants. They were mostly speakers of Australian English. The number of participants was 124, including eight native English speakers who had some knowledge of Korean and two native Chinese speakers.⁴ All participants were given course credit for their participation.

4.1.2 Stimuli

For the sake of simplicity, the scope of data was restricted to a prefixal type of phonaestheme in Hutchins' list. Hutchins' (1998) study compiled phonaesthemes from proposals made in previous literature from a period of 70 years. Specifically, 15 phonaesthetic studies (e.g., Bolinger, 1950, 1965; Bloomfield, 1953; Marchand, 1959, 1960; Fordyce, 1988) were cited by Hutchins to generate 46 sets of phonaesthetic words, including 22 prefixal phonaesthemes (e.g., *sn-*, *gr-*, *gl-*), 22 suffixal phonaesthemes (e.g., *-ack*, *-ash*), and 2 circumfixal phonaesthemes (i.e., *sp_t* 'a rush of liquid' in *spat*, *spout*, *spurt*; *str_p* 'a line having breadth in *strap*, *strip*, *stripe*).⁵

In order to make the rating task relatively easy for the participants, the word samples included only monosyllabic words.⁶ This reduced the total number of example words for 22 prefixal phonaesthemes in the list from 465 to 345. The frequency rate for the example words was then controlled by eliminating the two extreme ends of frequency. Highly frequent words, in which the frequency of the lemma was more than 100 times per million words in the British National Corpus (Leech, Rayson, and Wilson, 2001), were eliminated (e.g., *stay*, 174.944; *stop*, 236.833; *still*, 648.485 for *st-* phonaesthetic group). Their exclusions were necessary to reduce the frequency effect on participants' perceptual judgements. Words that occurred less than 50 times in the entire 100-million word cases from the corpus were also removed from the word stimuli, since that indicated a rounding down to zero per million words (e.g., *sty*, 0.261; *stoge*, 0.180 for the *st-* phonaesthetic group); they would be unfamiliar to participants and thus difficult to yield reliable ratings. Based on the frequency measure, an additional 72 words were eliminated and consequently, 273 example words remained for the 22 phonaesthemes.

Overall, I examined 20 candidate phonaesthemes. Of those, 19 were taken from 22 sets

⁴ Due to unforeseen technical difficulties, it was not possible to discard these confounding data. However, given their limited number, their influence on the current result is considered to be minimal.

⁵ There were considerable overlaps of the example words in each phonaesthetic set among the cited studies. For example, for the *gl-* phonaestheme, out of 79 example words drawn from six previous studies, 59 words (75%) showed overlapping appearances across the multiple studies examined.

⁶ Controlling for phone and letter length was not attempted because it would have considerably reduced the number of the stimulus items.

of prefixal phonaesthemes in Hutchins' lists.⁷ The selected candidate phonaesthemes were instantiated by a varied number of example words, within the range of maximum 26 and minimum 6. (The number of example words for each phonaestheme could not be equalised, as Hutchins' original list exhibited a different number of example words for each phonaestheme.) Additionally, one phonaestheme, which had 10 example words remaining after the aforementioned selection process (*wh-* 'noises of air or breath, or forcible movement' in *whip, whoosh, whoop, which*, etc.; Marchand 1969: 415), was included in the final stimulus items, yielding 20 phonaesthetic sets of 276 example words, as shown in Table 1.

Table 1. The selected example words for 20 candidate phonaesthemes (bold highlighting of some phonaesthemes and words is explained in Section 4.3)

<i>Candidate phonaesthemes</i>	<i>Key words</i>	<i>Selected example words</i>
cl-	clang	clank, clash, clap, clack, cling, click , cluck, clamp, clip, clod, clog, clam, clinch, clutch, clasp, clump, clench, claw, clay, cloy
fl-	flow	float, flush, flee, flail, flop, flap, fleet, flit, flag, flex, fling, flare, flash, flip, flick, flat, flaunt, fluster, fleck, flirt, flinch, flake
gl-	gleam	glow, glint, gloss, glare, glaze , glee, glad, glimpse, glance, gloom, gloat, glum, glide, globe, glove
sk-	skim	skid, skip, skate, skimp , scud, scour, skirt, scope, skin, sketch, skew, scat, scoff, scare, skull
st-	stiff	stick, stump, stanch, steep, stall, stuff, stir, stamp, sting, stomp, stash
str-	straight	stripe, strip, stretch, streak, strait, string, strap, stream, stride, strive, strut, strum, strength, strain, stroll, strife, strange, strew, stress
sw-	swing	swish, swoop, swipe, sweep, swirl, swat , swoon, switch, swag, swap, swell, swill, swim
sp-	spit	spat, spew, spurt , spout, spurn
tr-	tread	trudge, trot, tramp , trip, troll, trap
bl-	blow	blast, blurt, blaze, bleat, blob
kr-	crook	creep, crack, crick, cramp, crutch, creak, crouch, cross, cringe, crane, croak, crimp, crag, crow, crash, crawl, crunch, crush, crib, crate, crump, crab, crumb
dr-	drag	drift, droop, drape, drawl, drown, dregs, drug, drain, droll, drench, drool, drip, dry, drum, dram
gr-	growl	grunt, groan, gruff , grim, grouse, grudge, grasp, grope, grab, grip, graft
sl-	slide	slip, slope, slant, slick, slink, sleek, slime, sleet, sludge, slosh, slop, slouch, slough, slash, slow, sling, slack, slam, slay, slit, sloth, slap, slog, slave, slang
sn-	snout	sniff, snort, sneeze, snore, snuff, snarl , sneer, snoop, snub, snob, snack, snap, sneak, snatch, snag, snip
wh-	whack	whip, whoosh, whoop, whisk, whizz, whirl, whine, wheeze, whit
squ-	squeeze	squash, squirt, squirm, squelch , squeal, squeak, squawk, squid, squall, squander
tw-	twist	twirl, twine, tweak , twitch, twang
wr-	writhe	wring, wrap, wrath, wry, wrench
scr-	screech	scream, scrape, scratch, scrawl, scrub, scrounge, screw

⁷ Three of Hutchins' lists of prefixal phonaesthemes (*sm-*, *spl-*, *spr-*) were not chosen because they had less than four example words after the selection process and therefore they were expected to contribute only to a weak degree of diagrammatic iconicity.

4.1.3 Key word

A key word is a representative word for a phonaestheme, in terms of the sense that it signifies. The function of a key word here was to ensure that semantic interconnection between example words for a phonaestheme was objectively measured with one common reference point. In order to qualify as a key word, the example words of a phonaestheme underwent three stages in the selection process.

First, if an example word appeared in Hutchins' semantic gloss of a candidate phonaestheme, it became a key word (Hutchins listed the semantic glosses for the phonaesthemes of interest by combining different researchers' descriptions of them). For instance, *blow* was chosen as a key word for a list of words sharing *bl-* because it appeared in Hutchins' definition of the phonaestheme *bl-* "to blow, swell, or inflate; or to be round, swollen, or globular in shape". The first stage was dependent on Hutchins' gloss of a candidate phonaestheme because it was developed to reflect the commonly cited meaning for the phonaestheme by multiple studies and, therefore, expected to represent the core semantic feature of the phonaestheme.

Second, if the gloss did not contain any words bearing the sound pattern of interest, an example word that had the most occurrences of synonyms sharing the sound pattern of interest in WordNet,⁸ was chosen as a key word. For example, *clang* became a key word for *cl-* because it had the largest number of synonyms sharing word-initial *cl-* (e.g., *clangour*, *clank* and *clash*).

Third, if there were multiple example words having the same number of synonyms bearing the sound patterns of interest, the most frequent word among them became a key word. This third stage was based on the conjecture that high-frequency words would, by virtue of their common occurrence, have a strong effect on the phonaesthetic experience, if indeed there is one. The external evidence linking frequency of use and core semantics is found in Crossley, Salsbury, and McNamara (2010), which demonstrated that the growth of polysemous words (i.e., words with various related senses) and the growth of high-frequency words are positively correlated in second-language learners of English. Under this criterion, *squeeze* (with the frequency of the lemma of 8.189) became a key word for *sq-*.

4.1.4 Procedure and design

In order to minimise participant fatigue, the 20 lists of 276 English example words (256 sample words plus 20 key words) were divided into two different sets and distributed to two sub-groups of the participants: Group A and Group B. Consequently, 62 participants in Group A and 62 participants in Group B read 10 lists, totalling 131 and 125 sample words respectively.⁹ They rated each word on their list by how closely it was related to a key word on a 5-point Likert scale, ranging from strongly related (5) to not related (1). For this rating task, they were specifically told that some words on the list could have various possible meanings (i.e., homonymy/polysemy). For example, *close* can mean shutting or completing (a verb) and nearness (an adjective). For such a case, they were asked to select and rate the most closely related meaning to the given key word, disregarding its part of speech. In the case of encountering a word that they did not know, they were told to select the 'Don't know' option instead of consulting dictionaries (see Appendix for detailed instructions). The experiment was run on a University of Queensland online test system, Blackboard, with no time limit. On this system, the presentation order of the stimulus items for each list was automatically randomised for each participant.

⁸ <http://wordnet.princeton.edu/wordnet/download/current-version/>

⁹ There is a difference in the number of stimulus items per group because each proposed phonaesthetic list instantiated different numbers of example words.

4.2. Results

The participants' ratings for the individual words on each phonaestheme's list were averaged, after any "Don't know" responses, which accounted for 4.1% of the entire responses (649/15872), were excluded (see Table 2).

Table 2. Mean ratings of individual words in each phonaestheme's list (key words in parentheses)¹⁰

cl- (clang)	Mean	fl- (flow)	Mean	gl- (gleam)	Mean	sk- (skim)	Mean	st- (stiff)	Mean
clank	4.06	float	3.59	glow	4.14	skid	2.7	stick	2.4
clash	3.96	flush	2.88	glint	3.93	skip	2.58	stump	1.81
clap	3.12	flee	1.92	gloss	3.57	skate	2.48	stanch	1.69
clack	3.02	flail	1.9	glare	2.94	skimp	2.24	steep	1.66
cling	2.93	flop	1.79	glaze	2.83	scud	1.96	stall	1.55
click	2.74	flap	1.75	glee	2.06	scour	1.78	stuff	1.31
cluck	2.14	fleet	1.71	glad	1.95	skirt	1.68	stir	1.3
clamp	1.88	flit	1.66	glimpse	1.95	scope	1.67	stamp	1.28
clip	1.82	flag	1.57	glance	1.9	skin	1.61	sting	1.25
clod	1.75	flex	1.55	gloom	1.82	sketch	1.58	stomp	1.23
clog	1.69	fling	1.55	gloat	1.5	skew	1.45	stash	1.12
clam	1.68	flare	1.5	glum	1.39	scat	1.4		
clinch	1.64	flash	1.5	glide	1.25	scoff	1.25		
clutch	1.62	flip	1.48	globe	1.21	scare	1.15		
clasp	1.62	flick	1.44	glove	1.08	skull	1.1		
clump	1.55	flat	1.41						
clench	1.51	flaunt	1.4						
claw	1.47	fluster	1.35						
clay	1.34	fleck	1.27						
clay	1.33	flirt	1.25						
		flinch	1.24						
		flake	1.21						
str- (straight)	Mean	sw- (swing)	Mean	sp- (spit)	Mean	tr- (tread)	Mean	bl- (blow)	Mean
stripe	3.05	swish	2.97	spat	4.73	trudge	3.72	blast	3.65
strip	2.53	swoop	2.71	spew	3.61	trot	3.25	blurt	2.08
stretch	2.35	swipe	2.47	spurt	3.42	tramp	2.98	bleat	1.84
streak	2.29	sweep	2.3	spout	2.4	trip	2.48	bleat	1.36
strait	2.14	swirl	2.26	spurn	1.56	troll	1.69	blob	1.31
string	2.1	swat	2.08			trap	1.3		
strap	1.97	swoon	1.71						
stream	1.82	switch	1.6						
stride	1.69	swag	1.56						
strive	1.57	swap	1.51						
strut	1.5	swell	1.48						
strum	1.38	swill	1.44						
strength	1.36	swim	1.43						
strain	1.33								
stroll	1.31								
strife	1.3								
strange	1.24								
strew	1.23								
stress	1.21								

¹⁰ In the *cl-* word family, some of the more highly related words share the same vowel, showing an additional phonological overlap, apart from the proposed word-initial phonaestheme (e.g., *clang*, *clank*, *clash*, *clap*, *clack*).

Table 2. (continued)

cr- (crook)	Mean	dr- (drag)	Mean	gr- (growl)	Mean	sl- (slide)	Mean	sn- (snout)	Mean
creep	2.68	drift	2.72	grunt	4.1	slip	3.95	sniff	3.89
crack	2.26	droop	2.7	groan	3.86	slope	3.63	snort	3.84
crick	2.19	drape	2.28	gruff	3.15	slant	2.84	sneeze	3.36
cramp	2.11	drawl	2.25	grim	2.13	slick	2.58	snore	3.24
crutch	2.03	drown	2.03	grouse	2.05	slink	2.29	snuff	2.85
creak	1.94	dregs	1.86	grudge	2.02	sleek	2.22	snarl	2.58
crouch	1.84	drug	1.77	grasp	1.5	slime	2.13	sneer	2.18
cross	1.82	drain	1.74	grope	1.4	sleet	2.05	snoop	2.16
cringe	1.79	droll	1.69	grab	1.38	sludge	1.94	snub	2.05
crane	1.74	drench	1.53	grip	1.35	slosh	1.93	snob	1.95
croak	1.68	drool	1.47	graft	1.09	slop	1.89	snack	1.48
crimp	1.66	drip	1.37			slouch	1.67	snap	1.31
crag	1.57	dry	1.32			slough	1.46	sneak	1.2
crow	1.53	drum	1.24			slash	1.45	snatch	1.15
crash	1.51	dram	1.21			slow	1.44	snag	1.07
crawl	1.49					sling	1.41	snip	1.07
crunch	1.47					slack	1.38		
crush	1.38					slam	1.37		
crib	1.36					slay	1.37		
crate	1.35					slit	1.37		
crump	1.3					sloth	1.3		
crab	1.15					slap	1.24		
crumb	1.15					slog	1.23		
						slave	1.23		
						slang	1.11		
wh- (whack)	Mean	sq- (squeeze)	Mean	tw- (twist)	Mean	wr- (writhe)	Mean	skr- (screech)	Mean
whip	3.34	squash	3.74	twirl	4.08	wring	2.41	scream	4.57
whoosh	1.93	squirt	2.89	twine	3	wrap	1.85	scrape	2.4
whoop	1.89	squirm	2.52	tweak	2.33	wrath	1.83	scratch	2.13
whisk	1.59	squelch	2.45	twitch	1.51	wry	1.55	scrawl	1.47
whizz	1.57	squeal	1.93	twang	1.38	wrench	1.5	scrub	1.26
whirl	1.44	squeak	1.8					scrounge	1.21
whine	1.42	squawk	1.49					screw	1.2
wheeze	1.26	squid	1.26						
whit	1.09	squall	1.22						

The main prediction was borne out by these results: The graded mean ratings indicate that certain words are closer to the core meaning of a phonaestheme than others. As expected, such gradience hinders one from locating a point where the core category stops and the peripheral category starts. In a boxplot, where the strengths of semantic interconnection between example words and the relevant key word is visually arranged (see Figure 2), the point at which the interquartile range started to appear at the lower half end of the rating scale was clearly seen. Provisionally, this enables us to identify what constitutes the core and peripheral category in the current experiment.

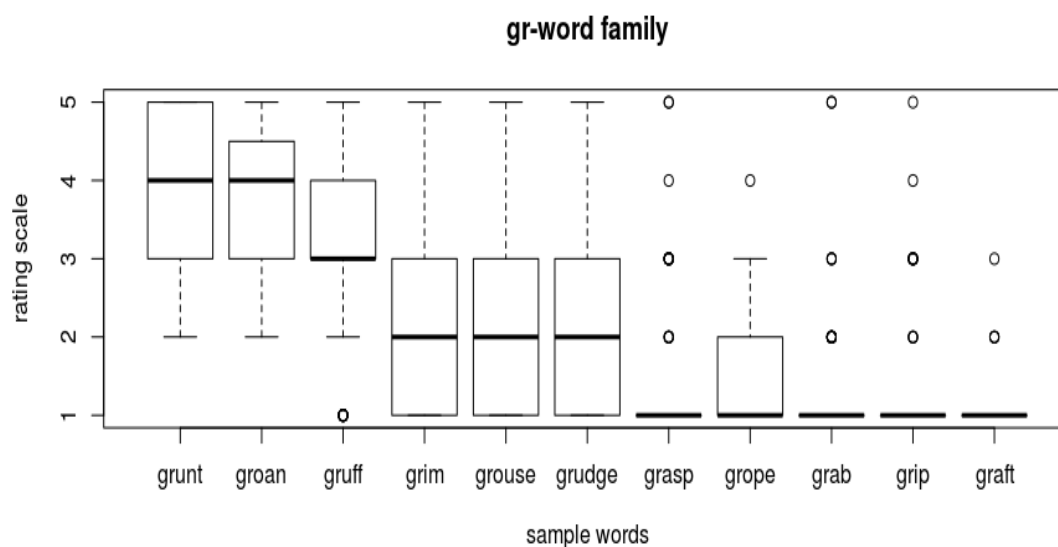


Figure 2. Median ratings for example words for the *gr-* phonaestheme and the spread of responses (key word *growl*)

For the *gr-* phonaestheme shown in Figure 2, the point distinguishing the interquartile ranges of the ratings appearing in the upper versus lower ends of the scale occurred between *gruff* and *grim*. Thus, only the three words (*grunt*, *groan* and *gruff*) whose interquartile ranges occurred in the upper end of the rating scale, and the key word, *growl*, were considered to be the core words central to the purported semantic domain of *gr-*, and any words below the distinguishing point were considered to belong to the peripheral category.

This sorting process made it possible to discard any lists with less than three core example words (excluding key words), since these appeared to have only slightly recurring form-meaning correspondences and which, therefore, were less likely to be a representative example of diagrammatic iconicity. For example, the *fl-* and *sl-* phonaesthetic sets were discarded because they were left with only two example words each (i.e., *float* and *flush* for *fl-*, and *slip* and *slope* for *sl-*) after the sorting process. Consequently, only ten lists of phonaesthemes, with a total of 54 example words, were retained for the following experiment, namely those shown in bold in Table 1.

5. Main experiment: English-speaking listeners' perceptions of English phonaesthemes

The preliminary experiment showed that words which contain a certain phonaestheme are different from each other in terms of their distance from the semantic core of the phonaesthetic category. Those words that are located in a central part of the phonaesthetic category have a strong bond between the phonaestheme and its suggestive meaning, leading to the idea that the phonaestheme occurring in the core members may display a strong degree of form-meaning correspondence. The main experiments examined whether or not the selected example words in the preliminary experiment show varied degrees of primary iconicity from one phonaestheme to another, by assessing the universality of the interpretations for the ten phonaesthemes that the sets of core example words share.

In order to examine whether English and non-English speakers perceive semantic aspects of the ten selected English phonaesthemes occurring in the core members of the phonaesthetic categories, listeners were asked to participate in a multiple-choice task. This

section examines English-speaking participants' perceptions in this regard, while Korean-speaking listeners' perceptions of the phonaesthemes in a corresponding task are described in Section 6.

5.1. Method

5.1.1 Participants

A total of 46 students from an introductory linguistics course at the University of Queensland, who had not participated in the preliminary experiment, served as English-speaking subjects. They reported no extensive knowledge of a second language and no hearing problems. Their participation was reimbursed with course credit.

5.1.2 Stimuli

In order to block the participants' attempts to use their prior knowledge of English, nonsense English phonaesthetic words were created by using 54 core phonaesthetic words drawn from the preliminary experiment described in Section 4 and a multilingual pseudoword generator, Wuggy.¹¹ The program generated multiple candidate pseudowords that matched the letter length, transition frequencies between sub-syllabic segments, and word-initial phonaesthemes of the input words (Keuleers and Brysbaert, 2010). The best nonsense English phonaesthetic word for each input word was then selected, on the criterion that it should have the smallest difference from the input word in terms of the neighbourhood size and its density, as calculated by Wuggy. Since Wuggy does not provide a pronunciation guide for the generated pseudowords, their pronunciations, particularly in regard to vowels, were unclear. In order to overcome this limitation, two native English speakers were asked to agree on the pronunciation of each pseudoword. This pronunciation checking process helped to discard any pseudohomophones and to replace them with the next best candidate pseudowords available. Also, it ensured that all of the pronunciations sounded natural and were in accord with native English speakers' intuitions (they showed 100% agreement as to how each pseudoword should be pronounced). The ten lists of the selected nonsense phonaesthetic words are shown in Table 3.

Crucially, there was a possibility that the nonsense words would lead to activation of real words and this would affect participants' performances in the following multiple-choice task. For instance, *skig* and *sked* could both result in activation of (and access to the meaning of) *skid*, and *skake* could result in activation of *skate*. This may not be equally true for all of the nonsense words. For the assessment of such mediation, I examined OLD20 ("Orthographic Levenshtein Distance 20"), which indicates the neighbourhood density of the pseudoword by means of the average Orthographic Levenshtein Distance between the pseudoword and its 20 closest orthographic neighbours in the lexicon (Yarkoni, Balota, and Yap, 2008), for each nonsense word using the Wuggy program. The OLD20 scores for the ten lists of selected nonsense words are shown in Table 4 below. (A lower OLD20 score indicates a dense neighbourhood [Keuleers and Brysbaert, 2010, p. 632]).

¹¹ <http://crr.ugent.be/programs-data/wuggy>

Table 3. The lists of nonsense phonaesthetic words, given to participants only orally¹²

<i>Proposed semantic gloss/definition, in formulations provided to participants</i>	<i>Phonaestheme</i>	<i>Nonsense words</i>
Two things coming together or separating often producing a noise; the result of such an action	<i>cl-</i>	<i>clant, clane, cland, clat, clabs, climp, clins</i>
Having to do with light or with vision; or something visually salient	<i>gl-</i>	<i>glear, gloy, glind, glose, glore, glard</i>
Pertaining to the surface, edges, or thinness; or superficial, not deep	<i>sk-</i>	<i>sked, skig, skake, skind, skir</i>
To oscillate, undulate, or move rhythmically to and fro	<i>sw-</i>	<i>swibs, swoos, swime, swees, swigh, swar, swire</i>
Bring to a point or send out from a point; or to reject	<i>sp-</i>	<i>spad, spow, spuce, sput</i>
A path, or to walk in a line; or to locomote by foot; or to step forcefully	<i>tr-</i>	<i>trudes, trit, trass, tris, trood</i>
Deep-toned, complaining, or threatening noises	<i>gr-</i>	<i>grent, groon, gruce, grofs</i>
Related to the nose, or breathing; or by metaphorical extension to inquisitiveness	<i>sn-</i>	<i>snice, snoft, sneese¹³, snire, snump, snorl, snous</i>
Something soft, spongy, or compressed; or to constrict, compress, or contract something	<i>sq-</i>	<i>squave, squird, squibe, squelps, squeece</i>
To turn, distort, entangle, or oscillate; or the result of such an action	<i>tw-</i>	<i>twibe, twing, tweal, twire</i>

Table 4. OLD20 values (which indicate the neighbourhood density) for the ten lists of selected nonsense words (existing sample words in parentheses)

clant (clang)	1.6	swibs (swish)	1.75	gruce (gruff)	1.9
clane (clank)	1.45	swoos (swoop)	1.65	grofs (growl)	1.75
cland (clash)	1.55	swime (swipe)	1.75	snice (sniff)	1.7
clat (clap)	1.3	swees (sweep)	1.65	snoft (snort)	1.75
clabs (clack)	1.45	swigh (swirl)	1.8	sneese (sneeze)	2.2
climp (cling)	1.55	swar (swat)	1.15	snire (snore)	1.65
clins (click)	1.5	swire (swing)	1.6	snump (snuff)	1.85
glear (gleam)	1.8	spad (spat)	1.35	snorl (snarl)	1.85
gloy (glow)	1.8	spow (spew)	1.55	snous (snout)	1.7
glind (glint)	1.8	spuce (spurt)	1.7	squave (squash)	1.9
glose (gloss)	1.7	sput (spit)	1.35	squird (squirt)	1.75
glore (glare)	1.65	trudes (trudge)	1.8	squibe (squirm)	1.85
glard (glaze)	1.7	trit (trot)	1.45	squelps (squelch)	2.7
sked (skid)	1.5	trass (tramp)	1.55	squeece (squeeze)	2.6
skig (skip)	1.6	tris (trip)	1.45	twibe (twirl)	1.85
skake (skate)	1.6	trood (tread)	1.8	twing (twing)	1.35
skimp (skimp)	1.7	grent (grunt)	1.7	tweal (tweal)	1.85
skir (skim)	1.4	groon (groan)	1.7	twire (twire)	1.75

N.B. OLD scores of 2.3 and 4.3 were considered relatively low and high, respectively, in Yarkoni, Balota, & Yap (2008).

According to the scores shown in Table 4, there were no considerable differences in the density of real-word neighbourhoods among the nonsense words and, therefore, it was reasonable to expect a low likelihood of differing degrees of neighbourhood effects on participants' performances for the lists of different nonsense words in the multiple-choice task (i.e., all of the nonsense words would be influenced by neighbourhood effect, if there was one, because all of them have relatively low OLD20 scores).

¹² Although many of the nonsense words end in *-s* (e.g., *clabs, clins, swibs, swoos*, etc) and may therefore be heard as plurals, this is not so problematic because participants were explicitly told to infer one general meaning that all the nonsense words in each phonaesthetic list would share, rather than individual meanings for the nonsense words.

¹³ *Sneese* was pronounced as /sni:s/, not /sni:z/ which is homophonous to *sneeze*.

5.1.3 Procedure and design

The ten lists of 54 nonsense phonaesthetic words were recorded by a male native speaker of Australian English. The stimuli were pronounced consistently in rise-fall intonation to control any possible meaning-related prosody or intonation cues.

The participants were asked to hear a list of words that shared a word-initial phonaestheme. They were explicitly told that the nonsense words in each list all shared one general meaning with each other and that their answers do not need to be correct. The experiment was again run on a University of Queensland online test system, Blackboard, with no time limit. Since all of the procedures were conducted online, participants had freedom to choose a time and place that were convenient and comfortable for them to participate in the study. Moreover, they were able to listen to the stimulus sound files multiple times if they wished. There were no minimum requirements for the listening condition, such as a request to wear earphones/headphones.

The multiple-choice task was designed to limit the participants' semantic judgements of the phonaesthetic words, allowing specific phonosemantic information for their perceptual decisions to be observed. Here, they chose the description that they felt was the best fit for each phonaesthetic list, among the ten proposed definitions in Table 3. The semantic glosses for the ten phonaesthemes in Table 3 were based on Hutchins' (1998) study. Some of her definitions contained words that shared the phonological patterns of the phonaesthemes of interest. For example, the phonaesthemes *sn-* and *sq-* (seen in Table 3 above) were originally defined as 'related to the nose, or breathing; or by metaphorical extension to *snobbishness*, inquisitiveness' and 'something soft, spongy, *squishy*, or compressed; or to constrict, compress, contract, or *squeeze* something', respectively. In these cases, the words that contain *sn-* and *sq-* (*snobbishness*, *squishy*, and *squeeze*) were excluded from the semantic glosses in the current multiple-choice task, in order to prevent the participants from having their perceptual decisions influenced by the inclusion in the definition of a word that contained the sound symbol of interest.¹⁴

The presentation order of the phonaesthetic lists was randomised in the Blackboard system. The order of the definitions was, however, fixed across the lists. This contained a potential confound – participants may have assumed that they could not again choose a definition which they had previously chosen, and this may have restricted the subjects' choices toward the end of the lists. However, despite this concern, 96% of the subjects had duplicate answers in their own responses, indicating that the fixed presentation order of the definitions was not likely to have affected the subjects' responses.

5.2. Results

The mean proportions of the participants' choices of the target and non-target definitions for each list of nonsense phonaesthetic words are shown in Table 5. The rows and columns indicate the involved phonaesthemes and their purported meanings, respectively. The shaded cells indicate the proportions in which the participants correctly chose the target definition for the phonaestheme of interest.

¹⁴ The limitation of this procedure is that an exclusion of *snobbish* from the definitions of *sn-* removed a meaning, whereas exclusions of *squishy* and *squeeze* from the definitions of *skw-* did not (since *soft*, *spongy* and *compressed* are still in the definition). However, a possible effect that the deletion of the metaphorically extended meaning 'snobbishness' would have on the outcomes of the experiment is expected to be trivial, since its core meaning 'something related to the nose' is still accessible in the definitions of *sn-*.

Table 5. English-speaking participants' guessing performance in the multiple-choice task, with rows representing the phonosthemes, and the columns the corresponding meanings

	<i>cl-</i>	<i>gl-</i>	<i>sk-</i>	<i>sw-</i>	<i>sp-</i>	<i>tr-</i>	<i>gr-</i>	<i>sn-</i>	<i>sq-</i>	<i>tw-</i>
<i>cl-</i>	33%**	11%	13%	4%	9%	4%	2%	2%	9%	13%
<i>gl-</i>	4%	61%***	7%	0%	7%	2%	4%	7%	9%	0%
<i>sk-</i>	13%	4%	17%	13%	9%	24%	0%	4%	4%	11%
<i>sw-</i>	2%	2%	7%	39%***	4%	2%	2%	13%	20%	9%
<i>sp-</i>	17%	4%	4%	4%	22%	2%	2%	17%	22%	4%
<i>tr-</i>	2%	4%	20%	9%	9%	24%	4%	2%	4%	22%
<i>gr-</i>	9%	0%	0%	4%	9%	9%	48%***	7%	2%	13%
<i>sn-</i>	0%	4%	4%	7%	4%	2%	9%	59%***	7%	4%
<i>sq-</i>	13%	2%	9%	9%	4%	2%	4%	2%	43%***	11%
<i>tw-</i>	20%	7%	13%	4%	13%	7%	0%	4%	7%	26%

* $p < .05$, ** $p < .01$, *** $p < .001$

The percentages of correct and incorrect answers in each list were compared to the chance probability of being correct¹⁵, using a one-tailed binomial test. As a result, it was shown that an above-chance of correct answers appeared for only 6 out of the 10 lists of proposed phonaesthetic words – *cl-*, *gl-*, *sw-*, *gr-*, *sn-*, and *sq-*. No definitions other than the target ones were chosen at an above-chance level for those phonaesthemes. For the remaining four lists of phonaesthemes – *sk-*, *sp-*, *tr-*, and *tw-*, neither incorrect nor correct answers reached statistical significance.

Why did the English-speaking participants recognise purported form-meaning correspondences only for a subset of the phonaesthemes? Is it simply because the six correctly recognised form-meaning correspondences are more conventionalised than the other four? Or is it because they possess a higher degree of primary iconicity? To answer this question, I proceed to the next section, which examines how non-native English speakers perceive the proposed English phonaesthemes.

6. Main experiment: Korean-speaking listeners' perceptions of English phonaesthemes

6.1. Method

6.1.1 Participants

A total of 28 students from Kunsan National University in Korea served as Korean-speaking participants on a voluntary basis. None of them self-rated themselves as fluent English speakers in the language-background questionnaire that they completed prior to the experiment. However, this does not entirely ensure that participants were unable to make inferences from an existing knowledge of English, since all of the Korean-speaking participants were university students in Korea who had been exposed to English as a second language for at least nine years. For future research, participants who were less well educated and thus had little exposure to English would be needed to overcome this limitation.

¹⁵ The chance probability of being correct in the forced choice task is only 1/10 if the ten definitions are all independent of one another, but there are commonalities between some of the definitions that make this less likely. For example, the definition for *tw-* 'to turn, distort, entangle, or oscillate; or the result of such an action' somewhat overlaps with the definition for *sw-* 'to oscillate, undulate, or move rhythmically to and fro'. Given this, for a stricter statistical testing, I raised the chance probability from 0.1 to 0.15.

6.1.2 Stimuli

The audio stimuli for the main experiment were the same as those used in the multiple-choice task described in Section 5. The participants were told that the stimulus words were from a non-existing language. In Korean, syllable-initial consonant clusters are not allowed in either morphophonemic or phonetic levels (Sohn, 1999, p. 159) and this ensures that the stimulus items sound apparently foreign to the participants. The English semantic glosses of the ten phonaesthetic words used in multiple-choice task were translated into Korean by the author, a native Korean speaker.

6.1.3 Procedure and design

The procedure and design are the same as described in Section 5.

6.2. Results

The mean proportions of the Korean participants' choices of the target and non-target definitions for each list of nonsense phonaesthetic words are shown in Table 6. The rows and columns indicate the involved phonaesthemes and their purported meanings, respectively. The shaded cells indicate the proportions in which the participants correctly chose the target definition for the phonaestheme of interest.

Table 6. Korean-speaking participants' guessing performance in the multiple-choice task, with rows representing the phonosthemes, and the columns the corresponding meanings

	<i>cl-</i>	<i>gl-</i>	<i>sk-</i>	<i>sw-</i>	<i>sp-</i>	<i>tr-</i>	<i>gr-</i>	<i>sn-</i>	<i>sq-</i>	<i>tw-</i>	<i>Un-answered</i>
<i>cl-</i>	25%	11%	18%	0%	4%	11%	14%	4%	0%	14%	0%
<i>gl-</i>	0%	25%	11%	7%	0%	7%	7%	11%	18%	11%	3%
<i>sk-</i>	0%	18%	36%**	0%	14%	14%	7%	7%	0%	4%	0%
<i>sw-</i>	4%	4%	7%	21%	21%	0%	4%	14%	25%	0%	0%
<i>sp-</i>	14%	18%	21%	4%	7%	7%	14%	4%	11%	0%	0%
<i>tr-</i>	21%	7%	7%	4%	4%	18%	11%	4%	11%	14%	0%
<i>gr-</i>	0%	11%	0%	14%	18%	11%	7%	14%	18%	7%	0%
<i>sn-</i>	14%	4%	18%	14%	7%	7%	7%	14%	4%	7%	4%
<i>sq-</i>	4%	4%	7%	7%	29%	7%	21%	14%	0%	7%	0%
<i>tw-</i>	29%	7%	4%	4%	0%	7%	4%	4%	11%	32%*	0%

* $p < .05$, ** $p < .01$, *** $p < .001$

The percentages of correct and incorrect answers in each list were compared to the chance probability of being correct ($p_0 = 0.15$) using a one-tailed binomial test. As a result, it was revealed that the correct definitions were chosen at a significant level only for *sk-* and *tw-*. For the remaining phonaesthemes, none of the provided definitions was chosen as an appropriate answer at an above-chance level.

Given the results from the multiple-choice tasks in both language groups, the following section compares in detail the guessing performances of the Korean and English-speaking participants, drawing implications for the roles of (primary) iconicity and conventionality in explaining the findings.

7. Discussion and conclusion

Comparison of the significance levels of the English and Korean-speaking participants' correct-guessing rates for each phonaestheme in the multiple-choice tasks can lead to four possible scenarios, as shown in Table 7: (1) both English- and Korean-speaking listeners' correct guessing rates were above chance; (2) only the English-speaking listeners' correct guessing rate was above chance; (3) only the Korean-speaking listeners' correct guessing rate

was above chance; and (4) both English- and Korean-speaking listeners' correct guessing rates were at (or below) chance level. Each one of these categories could be seen as corresponding to a different level of iconicity: strong primary iconicity (1), conventionalized primary iconicity (2), weak primary iconicity (3), and secondary iconicity (4), as explained in what follows.

Table 7. The phonaesthemes in the hypothesised categories.

	English result	Korean result	Phonaestheme
(1) Strong primary iconicity	High	High	None
(2) Conventionalized primary iconicity	High	Low	<i>cl-, gl-, sw-, gr-, sn-, sq-</i>
(3) Weak primary iconicity	Low	High	<i>sk-, tw-</i>
(4) Secondary iconicity	Low	Low	<i>sp-, tr-</i>

The logic of the scenarios is as follows. Since naturalness is linked to universality, the statistically significant level of correct answers for both subject groups would indicate that a phonaestheme possesses a high degree of primary iconicity (1); no phonaesthemes appeared to be on this level. On the other hand, the case where only the English result is above-chance level indicates that the psychological reality of the phonaestheme is due to conventionalized primary iconicity: the speakers are able to guess the correct meaning on the basis of their knowledge of form-meaning mappings in the English lexicon (2). The interpretation for the case where only the Korean result is above-chance level is complex (3); given that Korean speakers' correct guessing rates of the phonaesthemes are significantly high, it can be interpreted that the phonaesthemes contain primary iconicity to a considerable degree, as with case (1). However, for English speakers, the purported meanings of the phonaesthemes, which are active in existing words, is not being observed in novel words. This may be because, unlike case (1), the phonaestheme is conventionalised as a lexical mapping in English in a way that is contrary to the "natural" iconic mapping. The remaining case (4) is the most arbitrary, under which a phonaestheme possesses only secondary iconicity, which can be perceived only post hoc.

To summarize, as several of the stimulus phonaesthemes (i.e., *cl-, gl-, sw-, gr-, sn-, sq-*) were correctly guessed at a statistically significant level by only the English-speaking listeners, this supports Firth's (1930) idea that phonaesthemes are the results of a "phonetic habit" of language users. However, as the participants were able to establish the mappings to the corresponding motivated meanings for novel expressions, this also testifies to a role for primary iconicity. With regard to the *sk-* and *tw-* phonaesthemes in category (3), although they do not seem to be accessible to native English speakers, they also show a degree of primary iconicity, as they were correctly guessed by the Korean participants. In general, all the phonaesthemes but those in category (4) manifest primary iconicity although they do not appear to hold a generalisable psychological reality, showing the probabilistic attributes of phonaesthemes (Markel and Hamp, 1960; Hutchins, 1998). Altogether, the results of the multiple-choice tasks reveal that English phonaesthemes possess various kinds of iconicity from conventionalized primary iconicity to secondary iconicity, and that the (primary) iconicity level varies from one phonaestheme to another.

Acknowledgements

I wish to thank the editor, Jordan Zlatev, and an anonymous reviewer for their invaluable comments on an earlier version of this paper. Any remaining errors are, of course, mine. I also gratefully acknowledge the financial support of a postdoctoral fellowship for overseas researchers from the Japan Society for the Promotion of Science (JSPS) and a Grant-in-Aid for JSPS Research Fellows (#JP16F16729). Substantial portions of this work were adapted from an unpublished PhD thesis.

References

- Abelin, Å. (1999). *Studies in sound symbolism*. Ph.D. dissertation, Göteborg University.
- Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, 38(1/4), 298-348.
- Akita, K. (2009). *A grammar of sound-symbolic words in Japanese: Theoretical approaches to iconic and lexical properties of mimetics*. Ph. D. dissertation, Kobe University.
- Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, 80(2), 290-311.
- Bloomfield, M. W. (1953). Final root-forming morphemes. *American Speech*, 28, 158-164.
- Blust, R. (2003). The phonestheme η-in Austronesian languages. *Oceanic linguistics*, 42(1), 187-212.
- Bolinger, D. L. (1950). Rime, assonance, and morpheme analysis. *Word*, 6, 117-136.
- Bolinger, D. L. (1965). *Forms of English*. Cambridge, MA: Harvard University Press.
- Childs, G. T. (1988). *The phonology of Kisi ideophones*. Ph.D. dissertation, University of California, Berkeley.
- Crossley, S., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning*, 60(3), 573-605.
- Diffloth, G. (1972). Notes on expressive meaning. *Papers from the Eighth Regional Meeting* (pp. 440-447). The Chicago Linguistic Society.
- Diffloth, G. (1994). I: big, a: small. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 107-113). Cambridge: Cambridge University Press.
- Dingemanse, M. (2011). *The meaning and use of ideophones in Siwu*. Ph.D. dissertation, Radboud University.
- Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics compass*, 6(10), 654-672.
- Doke, C. M. (1935). *Bantu linguistic terminology*. London: Longmans, Green, and Co.
- Firth, J. (1930). *Speech*. London: Ernest Benn.
- Fordyce, J. F. (1988). *Studies in sound symbolism with special reference to English*. Ph.D. dissertation. University of California, Los Angeles.
- Haiman, J. (1980). The iconicity of grammar: isomorphism and motivation. *Language*, 56, 515-540.
- Hamano, S. (1998). *The sound-symbolic system of Japanese*. Stanford, CA: CSLI Publications.
- Hinton, L., Nichols, J., & Ohala, J. J. (1994). *Sound symbolism*. Cambridge: Cambridge University Press.
- Hockett, C. F. (1960). *The origin of speech*. *Scientific American*, 203, 89-96.
- Hutchins, S. S. (1998). *The psychological reality, variability, and compositionality of English phonesthemes*. Ph.D. dissertation, Emory University.
- Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, 109, 54-65.

- Jakobson, R. (1971/1965). *Quest for the essence of language. Selected Writings II*. The Hague: Mouton, 345–359.
- Jakobson, R., & Waugh, L. R. (1979). *The sound shape of language*. Brighton, Sussex: Harvester Press.
- Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science*, 35(3), 575-586.
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627-633.
- Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics*, 35, 379-415.
- Klamer, M. (2001). Expressives and iconicity in the lexicon. *Typological Studies in Language*, 44, 165-182.
- Köhler, W. (1929). *Gestalt Psychology*. New York, NY: Liveright.
- Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, 114(1), 19-28.
- Kwon, N. (2015). *The natural motivation of sound symbolism*. Ph.D. dissertation, University of Queensland.
- Kwon, N., & Round, E. R. (2015). Phonaesthemes in morphological theory. *Morphology*, 25(1), 1-27.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. Harlow: Longman.
- Locke, J. (1690/1975). *An Essay concerning Human Understanding*, In P. H. Nidditch (Ed.). Oxford: Oxford University Press.
- Marchand, H. (1959). Phonetic symbolism in English word formations. *Indogermanische Forschungen*, 65, 146-168.
- Marchand, H. (1960). *The categories and types of present-day English word formation*. University of Alabama Press.
- Marchand, H. (1969). *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. München: Beck.
- Markel, N. N., & Hamp, E. P. (1960). Connotative meanings of certain phoneme sequences. *Studies in linguistics*, 15(1), 47-61.
- Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondences in toddlers and adults. *Developmental Science*, 9, 316-322.
- Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1152-1164.
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language?. *Phil. Trans. R. Soc. B*, 369, 20130299.
- Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, 28, 255–282.
- Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, 112(1), 181-186.
- Parault, S. J. (2006). Sound symbolic word learning in written context. *Contemporary Educational Psychology*, 31(2), 228-252.
- Parault, S. J., & Parkinson, M. (2008). Sound symbolic word learning in the middle grades. *Contemporary Educational Psychology*, 33(4), 647-671.
- Peirce, C.S. (1955). In J. Buchler (Ed.), *Philosophical writings of Peirce* (pp. 99-119). New York: Dover.

- Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology, 1*, 1-15.
- Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia--a window into perception, thought and language. *Journal of consciousness studies, 8*(12), 3-34.
- Rhodes, R. (1994). Aural images. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 276-292). Cambridge: Cambridge University Press.
- Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology, 12*, 225-239.
- Schmidtke, D. S., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in Psychology, 5*, 80.
- Sohn, H. (1999). *The Korean Language*. Cambridge: Cambridge University Press.
- Sonesson, G. (1997). The ecological foundations of iconicity. In R. Irmengard, & C. Gerald (Eds.), Paper presented at *the Semiotics Around the World: Synthesis in Diversity. Proceedings of the Fifth International Congress of the IASS, Berkely* (pp. 739-742). Berlin & New York: Mouton de Gruyter.
- Tufvesson, S. (2011). Analogy-making in the Semai sensory world. *The Senses and Society, 6*(1), 86-95.
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*, 971-979.

Appendix

Participant instructions for the preliminary experiment in Section 4

This survey contains 10 lists of English words. Each list has a key word written above it. You are asked to read each word on the list carefully, and rate one by one how closely each word is related to the key word at the top with the following numbers below. You may refer back to the key word at any time.

Score	Rating
1	Not related at all
2	Less related
3	Somewhat related
4	Related
5	Strongly related

Key words can have various alternative meanings. Rate a word by matching the most closely related meaning of the key word. Some words on the list can also have various possible meanings. For example, 'close' can mean shutting or completing (a verb) and nearness (an

adjective). In this case, regardless of a part of speech, select and rate the most closely related meaning to the key word given. It is important that you use your own judgement and do not consult dictionaries. When you encounter a word that you do not know, please select 'Don't know' option. Participation in this experiment has no time limit so it is recommended to take a short break if you feel stressed by reading lists of many words all at once.

Author address

Nahyun Kwon, Department of Applied Linguistics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Japan, E-mail: n.kwon@uq.edu.au

About the author(s)

Nahyun Kwon is an overseas researcher under Postdoctoral Fellowship of Japan Society for the Promotion of Science at Nagoya University in Japan. Her research interests include the study of motivated form and meaning associations in language and the phonotactics and morphophonology of ideophonic words.