Associative iconicity: Sound effects in English speech

Benjamin Anible
University of New Mexico

ABSTRACT. In a series of three studies this paper investigates if SOUND EFFECTS – non-lexical speech sounds – are interpreted in the same way as conventional word forms, using language-specific experience with phonological forms. In Study 1 a sample of 30 sound effects was collected from 17 speakers on YouTube and rated for phonological and semantic properties. Study 2 performed identical analysis with 30 tokens of Japanese ideophones. In both studies phonological content alone was a significant predictor for semantic categorization. Study 3 used survey data measured agreement between participant and researcher semantic judgments of sound effects. An associative form-meaning relationship is proposed that governs the production/perception of English sound effects.

Keywords: iconicity, ideophones, sound symbolism, usage-based, cognitive linguistics.

1. INTRODUCTION. The acoustic realization of a given lexical item's phonetic segments will vary across a sampling of tokens in diverse phonological, morphological or discursive contexts. How listeners are able to reconcile this variation in comprehension is an unresolved question. Usage-based theories of phonological comprehension/production suggest that EXEMPLARS of particular phones are abstracted from the co-occurrence of sounds and meaning over numerous instances (Klatt 1979, Langacker 1987, Ohala & Ohala 1995, Pierrehumbert 2001, Pisoni 1997). Structuralists, on the other hand, suggest that sounds are perceived, stored and accessed separately from their semantic content in a modular system (Chomsky & Halle 1968, McClelland & Elman 1986, Norris 1994, Saussure 1986). While conventionalized lexical items constitute the majority of the sounds speakers produce when engaged in communicative acts, one unique language behavior that can only be explored in a usage-based framework is the phenomena of SOUND EFFECTS. Sound effects vary in their degree of conventionality and iconicity and are problematic to analyze without allowing for gradience and interaction between phonological and semantic content. Because of this, sound effects are usually relegated to the vagaries of performance, instead of the analyzable status of competence in an approach that clearly disambiguates these domains.

Some sound effects are typically called onomatopoeia; *bang, woof,* and *ouch*, for example. In addition to their conventionality, onomatopoeia are somewhat iconic as well – there is little or no variation in their meaning (conventional) and it is more or less clear that they sound like what they mean (iconic). Many languages outside the Indo-European family have a rich inventory of lexical items called IDEOPHONES that fall somewhere between true onomatopoetics and the highly unconventional sound effects that will be examined here. Ideophones have been documented in Japanese (Itô & Mester 1995), Aboriginal Australian languages (Courtenay

1974), Mayan languages (Durbin 1973), and many pidgins and creoles (DeCamp 1974). Other sound effects are less conventional and/or less iconic. They can be used to mimetically represent the sound an object makes, and they can also be used to describe more abstract entities like thoughts, feelings, or processes that do not have an obvious sound that can be imitated.

There is emerging evidence that even speakers in languages without established ideophonic systems access highly schematic sound-meaning mappings in addition to more conventional/arbitrary word forms (Monaghan et al. 2014, Pena et al. 2011). For example, Reilly and Kean (2007), in an English corpus-based analysis, detected patterned phonological characteristics for nouns that varied based on their imageability. Imageability is a measure of how often adults report that a strong mental image is evoked by an item. Etymology, syllable structure, phonological complexity, word length, prosody, phonological and neighborhood density (the number of similar sounding words or lexical competitors that exist in the language) were included in the analysis. The strongest correlations between factors and imageability were from word length and etymology. Specifically longer words were highly correlated with low-imageability nouns and shorter words with high-imageability nouns. Germanic word-origin was correlated with low imageability and Latinate word-origin was correlated with high imageability, indicating a non-arbitrary relationship between word-origin (a characteristic of form) and how easy to visualize the concept was (a characteristic of meaning). Variables with non-significant correlation were total number of phonemes, phonological neighborhood density, and presence of compounding.

In another study, Monaghan, et al. (2014), propose that arbitrariness and systematicity are competing forces in language where highly arbitrary language parts allow for ease of perception since lack of patterning causes low neighborhood densities, but systematicity allows for ease of learning. They hypothesize that a mix of arbitrariness and systematicity is ideal for language learning. The authors use machine and human learning of artificially constructed languages to test this hypothesis and also assess the hypothesis on corpora of English and French. Machine and human learning studies found that arbitrariness in words was advantageous for language learning, but only if contextual information was also provided. For the corpora studies it was found that the onset part of words is more often arbitrary and the coda is systematic - suggesting this is the way the languages studied handle the need for both arbitrary and systematic information in the lexicon.

These findings suggest that phonological-semantic arbitrariness within the lexicon of English, and presumably all spoken languages, is not inviolate. While duality of patterning need not be completely dismissed, theories that rely on discreteness at various levels (such as word processing, reading, and language acquisition) are recommended for reexamination. This paper adds to the mounting evidence for this perspective.

2. BACKGROUND. There are discernible patterns connecting broad semantic properties and phonological forms in many languages. The Chinookan language, Wishram, has a reduction by lowered sonority for diminutive forms of many lexical items as well as expansion and heightened sonority for augmentative forms (Sapir 1911). English speakers in an early psycholinguistic experiment (Sapir 1929) showed an unconscious phonetic symbolism with the semantic properties *smaller* vs. *bigger* being attributed to the size of the “vibrating column of air in the resonance chamber” (high vs. low vowels). Also, Roman Jakobson has noted that in all of the world's languages, plural morphemes are denoted by either an increase or zero derivation from the singular forms, but never a morphemic decrease (1965).

Attempts over the past century to abstract any of these echoic/iconic properties as language universals have met with disappointment, however. Even the seemingly very robust "/a/: big, /i/: small" pattern in English and other Germanic languages must contend with the case of Bahnar (Diffloth 1994), a Mon-Khumer language of Vietnam. In Bahnar, the long and short variants its vowels are used in an extensive expressive system where /i/ is "big" and /a/ is "small."

"enormous"	ii	uu	i	u
"big"	ee	oo	e	o
"small"	εε	ɔɔ	ε	ɔ

Table 1: Bahnar "size" sound-symbolism.

For this reason, the topic of sound-symbolism has long been regarded as somewhat of a pariah by the mainstream of linguistic analysis. Most linguists more or less accept the Saussurean assumption that the nature of a linguistic symbol is a completely arbitrary association of SIGNANS (the material properties – acoustic or visual) with SIGNATUM (the interpretation of those properties). Saussure noted that the ARBITRARY connection between form and meaning was only one of three possible interactions. The other interactions include an ICONIC connection where the form and meaning are physically related and an INDEXIC connection where the form is physically related to a selected portion of the meaning (such as smoke being a signal for fire). Thomas Gamkrelidze (1974) observed that all three of these possibilities; ICONS, INDICES and SYMBOLS, form only vertical interactions. One form is matched with one meaning in one of three ways. What these definitions do not take into account are horizontal interactions:

> Phonetic closeness of familiar terms (mother, father, brother) is a phonetic allusion to the semantic proximity of the corresponding signata (…) In this sense

alone can we speak of the motivation of one series of relations through the other, of the dependence of the relations between the signata (107).

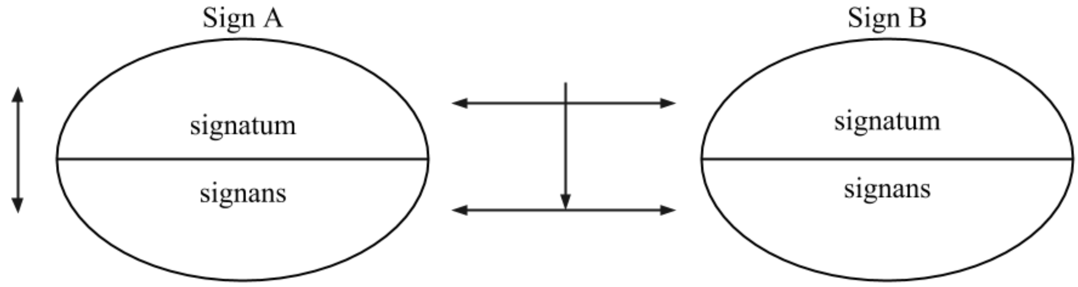

Figure 1: Horizontal relations between signans and signata (Gamkrelidze 1974:107).

Such "horizontal" relationships are both possible in the Sassurian framework, and substantiated in language. EXPRESSIVES – documented in languages other than English such as Japanese, ASL, Bahnar, and many others – are a class of lexical items closely related to sound-symbolism and phonesthemes, but distinct enough to warrant disambiguation. Diffloth (1994:108) calls them "a parallel sub-language grafted on, and parasitic on the conventional one." Many researchers who have documented sound-symbolisms for languages they work in note that they always use only the sounds in the conventional lexicon or a subset of those sounds. Though few signed language researchers necessarily recognize the cross-modality parallelism, the "classifier" systems (Dudis 2007, Suppalla 1986) in signed languages fit nicely into Diffloth's definition. The key is that phonesthemes, ideophones, etc. (or DEPICTIVES as they might be called in sign) create semi-conventional symbolic pairs from existing phonological and semantic schemas of the host language.

Sound effects are arguably neither phonesthemes, expressives, nor depictives though they share a number of similarities with each of these systems. It is clear that sound symbolism in some form exists in many different languages and that high levels of schematization are necessary to capture these phenomena. Given their unconventional nature, only a quantitative analysis can reveal regularities in form.

3. STUDY 1. This study attempted to identify connections between sound and meaning as constituents in an associative network in a sampling of sound effects. A sample of 30 sound effects was collected. Seventeen speakers were sampled in collection. Tokens came from natural student/teacher settings. All tokens were produced by the teachers, only. These teachers were, generally, well known master musicians teaching orchestral conducting, piano, violin, voice, etc. The majority (15/17) of these individuals did not speak English as their native language although the language spoken in all clips was English.

It is likely that a higher number of sound effects are found in the genre of teaching a masterclass in music due to the need to often sing or hum the melody of a piece of music being analyzed. The arbitrary assignment of phonetic segments to a melody is similar in quality to the process underlying sound effects. The difference is that humming a melody is not referential in any way (though the melody itself may be referential in some way, the issue of musical meaning is beyond the scope of this analysis). It seems likely that speakers were primed for making sound effects by their surface similarity to humming and since they were actively discussing the topic of producing sounds.[1]

Sound effects were coded for the phonological quality of their onset. The onset was considered prominent following Wescott's (1971) treatment of labio-velarity and derogation in

English where he notes that onsets and codas have greater degrees of prominence in assigning derogatory meaning when filled by labial or velar phones in informal slang. The particular quality noted for each token was the coarsest granularity level of the sonority hierarchy applicable to onsets – sonorants vs. obstruents. Sonorants include any phone that does not completely close the vocal tract. These are vowels, glides, liquids and nasals. Non-sonorants (obstruents) are stops, affricates and fricatives (Hayes 2009:76). The semantic content was coded at the highest level of schematicity possible following cognitive grammar – either a THING or a PROCESS. A thing is "any product of grouping and reification (Langacker 2008:105). A process is "a complex relationship that develops through conceived time and is scanned sequentially" (Langacker 2008:112).

Phonology	**Context**	**Token**
sonorant onset	"And split of a second [ja:l] and it burst open..."	[ja:l]
obstruent onset	"Not polite gypsy music, very [ʔmm]."	[ʔmm]
Semantic		
thing	(...)and more [ɛʔ], trumpet.	sharp delicateness
process	"Instead of going [wʰɑp], which would tend to..."	raising your hands quickly

TABLE 2. Coding examples.

Results from this study indicate that there was an interaction between the phonetic quality (obstruent/sonorant) of the onset and the quality of the meaning expressed by the sound effect (a thing/a process). See Table 3. A chi-square test of goodness-of-fit was performed to determine whether the type of onset was equally preferred for semantic category. Preference for the onset type was not equally distributed in the population, [$\chi 2$ (1, N = 30) = 8.48, $p < .01$]. Specifically, sound effects coded as things were more likely to have obstruent onsets and sound effects coded as processes were more likely to have sonorant onsets.

	Obstruent Onset	**Sonorant Onset**	*Totals*
Thing	10	6	16
Process	3	11	14
Totals	13	17	30

TABLE 3. Study 1 – Sample of 30 sound effects.

4. STUDY 2. Are the patterns observed in Study 1 universal? If they are, we might expect to find them when looking at similar phenomena from other languages. Study 2 used a random sample of Japanese ideophones taken from a *Dictionary of Iconic Expressions in Japanese* (Kakehi et al. 1996). Fifteen tokens with first syllable sonorant onsets (last 5 entries under "A", "O" and "K") were selected. Classification of the meaning of these ideophones was performed in an identical manner to Study 1. Results showed a strong tendency for sonorant onsets to occur with items that had process type meanings. Items with obstruent onsets had a strong tendency to occur with thing type meanings. See Table 4. Examples 1-3 illustrate sonorant onsets and 4-6 illustrate obstruent onsets. 1 and 2 match the pattern of sonorant/process, but 3 does not. Likewise 4 and 5 match the obstruent/thing pattern, but 6 does not.

	Obstruent	**Sonorant**	*Totals*
Thing	12	3	15
Process	3	12	15
Totals	15	15	30

TABLE 4. Sample of 30 Japanese ideophones

(1) *ata-futa* | “in a flurry” | process

(2) *wasa-wasa* | “trees shaking violently” | process

(3) *on-on* | “sound of loud crying” | thing

(4) *byuun-byuun* | “whistling wind” | thing

(5) *tsuu-tsuu* | “sound of a dial-tone” | thing

(6) *kyoton* | “stare blankly” | process

A chi-square test of goodness-of-fit was performed to determine whether the type of onset was equally preferred for semantic category. Preference for the onset type was not equally distributed in the population, [$\chi 2$ (1, N = 30) = 33.75, $p < .001$]. These results indicate that the associative network of sonorant/process and obstruent/thing may not be unique to highly unconventional English sound effects but may extend to a broader range of related phenomena including more conventionalized items like ideophones.

5. STUDY 3. While Studies 1 and 2 identified a pattern of sound-meaning associations, all results were based solely on a single investigator’s intuitions. Study 3 aimed to test the thing/process classification schema against native speaker judgment using a simple Likert-scale of similarity. Participants were presented with pairs of the sound effect video-clips from Study 1 and asked to judge how similar the sounds were to each other and how similar the meanings were on a scale of 1 to 7. Eight pairs were presented. Four pairs constituted the critical alternation (shown in Table 5) and four additional pairs were crossed with the target conditions. The survey itself can be found by visiting: http://sites.google.com/site/soundeffectsurvey/welcome.

Sound match	**Sound 1**	**Sound 2**	**Meaning 1**	**Meaning 2**	**Token 1**	**Token 2**
matching	obstruent	obstruent	thing	thing	*dohp*	*bong*
matching	sonorant	sonorant	process	process	*wrn*	*wha*
mismatching	obstruent	sonorant	thing	thing	*dwi*	*yip*
mismatching	sonorant	obstruent	process	process	*ooo*	*digadigadig*

TABLE 5. Sound effect similarity survey items

Twenty-four participants completed the survey. The average meaning similarity scores for the pairs with matching onsets was higher [*M*=3.97, *SD*=2.09] than the average meaning similarity scores for pairs with mismatched onsets [*M*=3.0, *SD*=1.91]. The difference between the scores was significant [$t(93)=2.38$, $p < .01$].

In both conditions, each pair had equivalent meanings; things were matched with things and processes with processes. The four control items had mismatched meanings and matching sounds or mismatched meanings and matching sounds. There was a significant trend towards higher meaning similarity ratings for matching sonorant/sonorant or obstruent/obstruent pairs. This suggests that subjects were incorporating phonological information into their semantic judgments. The higher sound similarity ratings for the matched onsets is predictable since the inclusion of a given phone into either sonorant or obstruent categories is based on sound similarity. It also indicates that subjects considered the sonorant or obstruent examples selected to sound similar to each other. Figure 3 shows these values as well as the mean values for ratings of how similar the obstruents and sonorants pairs were rated.

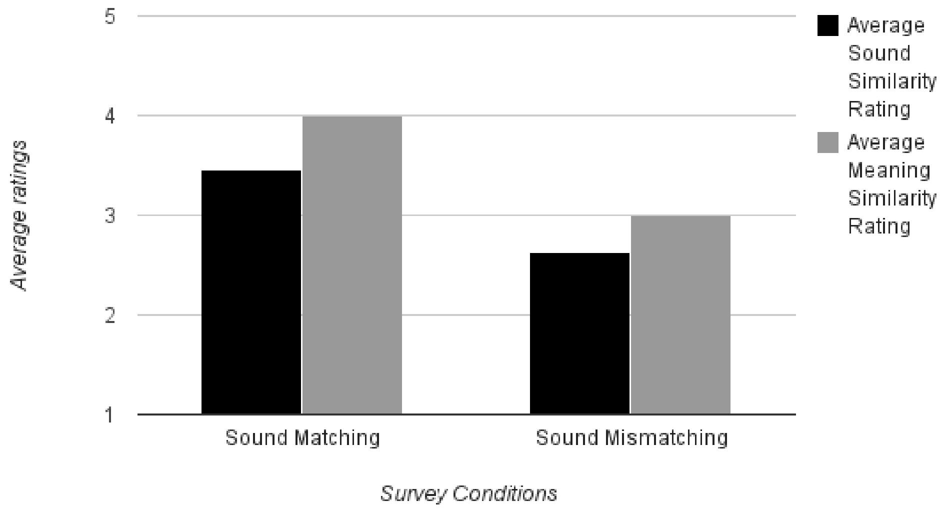

FIGURE 2. Mean agreement for Study 2 experimental conditions.

Participants were also asked to select their level of classical musical training at the end of the survey as either “little/none,” “some,” or “extensive.” Subjects with more experience with the style and manner of music masterclasses might be better able to ascertain the semantic content of the types of sound effects presented in the survey. A one-way ANOVA revealed no significant difference in meaning similarity rating between the three groups for the sound matched condition [$F(2,18) = 1.45$, p = n.s.] nor for the sound mismatched condition [$F(2,18) = .009$, p = n.s.]. Table 6 shows the mean values for the two conditions based on the three groups.

	Sound matched average	***SD***	**Sound mismatched average**	***SD***
Extensive	4.08	2.24	3.17	1.58
Some	3.85	2.13	3.08	1.86
Little/None	4.50	1.80	3.20	1.47

TABLE 6. Average rating differences between musical training groups.

6. DISCUSSION. The studies presented in this paper make a case for an associative sound-meaning network that aids language users in the coherent production and perception of novel sound effects. See Figure 4 for a model of how this network might work. Example tokens are listed for both semantic and phonological interactions. Pairings of semantics and phonology are denoted by their parallel position to each other (such as the pairings of [conducting vigorously] with [ra:h]). Similarity in process-type meanings are shown by a vertical solid line between tokens, as with similarity in thing-type meanings. The activation of the semantic network is represented by the horizontal line between the networks of process and thing meanings. Likewise, the similarity between sonorant onsets is represented by the vertical line (as between [ra:h] and [wah]) and between obstruent onsets. The activation of the phonological network between these two types of onsets is represented by the horizontal lines connecting the separate sonorant and obstruent networks. This latticework is a schematic representation of the high-level exemplar cloud that is activated when a sound effect is perceived and a listener attempts to use their prior experience to determine possible meanings.

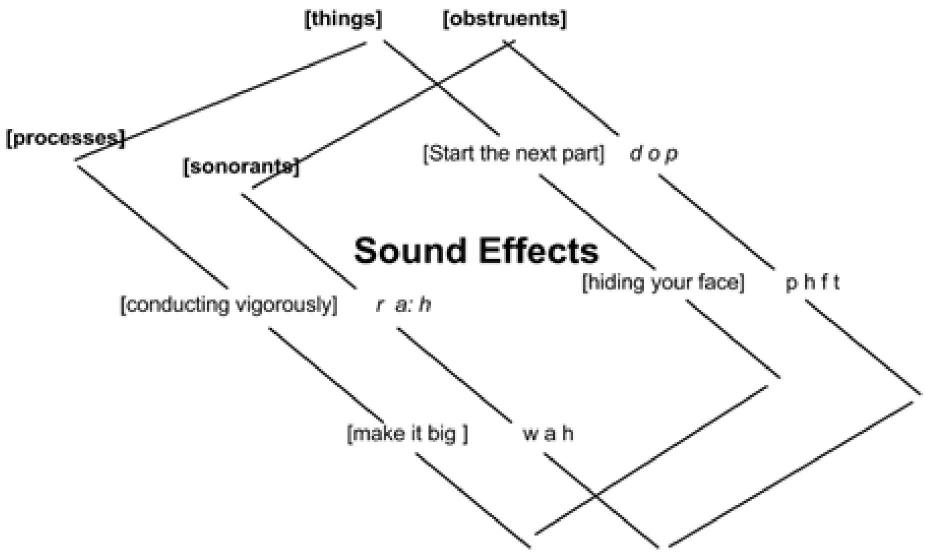

FIGURE 3. Associative network for sound effects.

The manual gestures that accompany sound effects have not been discussed in this analysis, but almost certainly play a role in decoding the meaning of a sound effect. Gestures were, however, coded using rubrics from Cienki's *Image schemas and gesture* (2005). Cienki categorizes gestures by their function. Broadly, there are referential gestures, and discursive/performative gestures. Discursive and performative gestures are viewed as accompanying a verbal utterance by illustrating a discourse level feature (like marking emphasis). Referential gestures are depictive of both concrete/abstract entities/ideas or relations. They are further classified as indexical of OBJECTS, PROPERTIES, BEHAVIORS/ACTIONS and RELATIVE LOCATION. Examples are "a picture frame," "the edge of a ruler," "the rolling of a tire," and "the space behind oneself," respectively.

Analysis of the gestures accompanying the sound effects from Study 1 found a very similar result to the sonorant-process, obstruent-thing pattern. Results show significant correlation between the meaning of the conductor's mimicked vocal gesture and the accompanying manual gesture [$\chi 2(1, N = 28) = 11.81, p<.005$)]: thing meanings tend to be accompanied by object manual gestures, whereas process meanings are accompanied action manual gestures. This demonstrates cross-modal iconicity between the functions of the conductor's vocal and manual gestures. Also interesting is the observation that not all of the multimodal constructions coded were complementary. That is, a small portion of thing meanings were accompanied by action manual gestures, while some process meanings were accompanied by object manual gestures. Further study is warranted.

	Object gesture	**Behavior gesture**	*Totals*
Thing	11	3	14
Process	1	13	14
Totals	14	14	28

TABLE 5. Gesture analysis of items from Study 1

One possible criticism of this analysis is the perspective that the sound effects examined in this study are simply lexicalized – and this is why speakers are able to understand them. Lexicalized items in language are necessarily standardized, repeatable and shared across all language users. Sound effects are schemas that are shared but not standardized or repeatable. Essentially, lexicalized items are semantic entities with a retrievable phonetic realization. Critically, there is a fine grained one-to-one correspondence between form and meaning. While a usage-based perspective does require scalar, rather than categorical distinctions, the phenomena described in this paper are so coarse grained in their semantic/phonological correspondence it would be disingenuous to consider them as lexical items in the traditional sense.

7. CONCLUSION. The correlations between meaning and sound found for English sound effects in Study 1 combined with the correlation in the same direction found for Japanese ideophones in Study 2 suggests a robust network that may span different languages. The results of Study 3 confirm that English listeners show a tendency to use the quality of the onset in determine the meaning of a sound effect encountered for the very first time. The proposal put forward here is that the same cognitive apparatus that allows for the management of acoustic variation for conventionalized lexical items also applies to the production and perception of completely novel sound/meaning pairings. Moreover, this system should not be constrained by modality in any way – signed languages should have similar systems in so far as there are should be schematic handshapes, movements, and locations that are matched with schematic semantics. Sound effects, ideophones expressives, and other “non-standard” forms of sound-meaning pairings are in sore need of study and analysis, given their long estrangement from sanctified linguistic fields of study. Further exploration into the associative networks that speakers and signers alike have access to during comprehension and production will likely challenge many long standing assumptions about the relationship between form and meaning in the lexicon.

REFERENCES

CHOMSKY, NOAM AND MORRIS HALLE. 1968. *The sound pattern of English*. New York: Harper & Row.

CIENKI, ALAN. 2005. Image schemas and gesture. *From Perception to Meaning*, ed by Beate Hampe, 29:421–442. Berlin, New York: Mouton de Gruyter.

COURTENAY, KAREN. 1974. On the nature of the Bambara tone system. *Studies in African Linguistics, Supplement*.13–26.

DECAMP, DAVID. 1974. Neutralization, iteratives, and ideophones: the locus of language in Jamaica. *Pidgins and Creoles: Current Trends and Prospects.*, ed by D. Decamp and I. F. Hancock, 46–50. Washington, D.C.: Georgetown University Press.

DIFFLOTH, GÉRARD. 1994. i: big, a: small. *Sound symbolism*, ed by Leanne Hinton, Johana Nichols, and John J. Ohala. Cambridge: Cambridge University Press.

DUDIS, PAUL. 2007. Types of depiction in ASL. Washington, D.C.: Gallaudet University.

DURBIN, MARSHALL. 1973. Sound symbolism in the Maya language family. *Meaning in Mayan languages: Ethnolinguistic studies*, ed by M.S. Edmonson, 23–49. The Hague: Mouton.

GAMKRELIDZE, THOMAS V. 1974. The Problem of "l"arbitraire du signe'. *Language* 50.102.

HAYES, BRUCE. 2009. *Introductory phonology*. Blackwell textbooks in linguistics 23. Malden, MA ; Oxford: Wiley-Blackwell.

ITÔ, JUNKO AND ARMIN MESTER. 1995. Japanese Phonology. *The Handbook of Phonological Theory*, ed by John Goldsmith, 817–838. Blackwell.

JAKOBSON, ROMAN. 1965. Quest for the Essence of Language. *Diogenes* 13.21–37.

KAKEHI, HISAO, IKUHIRO TAMORI, AND LAWRENCE SCHOURUP. 1996. *Dictionary of Iconic Expressions in Japanese: Vol I: A - J. Vol II: K - Z*. Berlin, New York: De Gruyter

Mouton.

KLATT, DENNIS H. 1979. Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics* 7.279–312.

LANGACKER, RONALD W. 2008. *Cognitive Grammar*. Oxford University Press.

LANGACKER, RONALD W. 1987. *Foundations of cognitive grammar, Vol 1: Theoretical Prerequisites.* Standford, CA: Standford University Press.

MCCLELLAND, JAMES L. AND JEFFREY L ELMAN. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18.1–86.

MONAGHAN, P., R. C. SHILLCOCK, M. H. CHRISTIANSEN, and S. KIRBY. 2014. How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.20130299–20130299.

NORRIS, DENNIS. 1994. Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52.189–234.

OHALA, JOHN J. AND MANJARI OHALA. 1995. Speech perception and lexical representation: the role of vowel nasalization in Hindi and English. *Phonology and Phonetic Evidence*, ed by Bruce Connell and Amalia Arvaniti, 41–60. Cambridge: Cambridge University Press.

PENA, M., J. MEHLER, AND M. NESPOR. 2011. The Role of Audiovisual Processing in Early Conceptual Development. *Psychological Science* 22.1419–1421.

PIERREHUMBERT, JANET B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Typological Studies in Language*, ed by Joan L. Bybee and Paul J. Hopper, 45:137. Amsterdam: John Benjamins Publishing Company.

PISONI, DAVID, B. 1997. Some Thoughts on “Normalization” in Speech Perception. *Talker Variability in Speech Processing*, ed by K. Johnson and J. W. Mullennix, 9–32. San

Diego: Academic Press.

REILLY, JAMIE AND JACOB KEAN. 2007. Formal Distinctiveness of High- and Low-Imageability Nouns: Analyses and Theoretical Implications. *Cognitive Science* 31.157–168.

SAPIR, EDWARD. 1929. A study in phonetic symbolism. *Journal of Experimental Psychology* 12.225–239.

SAPIR, EDWARD. 1911. Diminutive and Augmentative Consonantism in Wishram. *Bureau of American Ethnology* 40.638–645.

SAUSSURE, FERDINAND DE. 1986. *Course in general linguistics*. LaSalle, Ill: Open Court.

SUPPALLA, TED. 1986. The Classifier System in American Sign Language. *Typological Studies in Language*, ed by Colette G. Craig, 7:181. Amsterdam: John Benjamins Publishing Company.

WESCOTT, ROGER W. 1971. Labio-Velarity and Derogation in English: A Study in Phonosemic Correlation. *American Speech* 46.123.

[1] Thanks to Sherman Wilcox for the suggestion look for sound effect data in masterclasses on YouTube.