Pokémon Meets Psychology and Linguistics: Experimental and Theoretical Exploration of the *Bouba-Kiki* Effect*

Shigeto Kawahara
Keio University

ABSTRACT. The *bouba-kiki* effect is a sound symbolic connection that is well-studied in psychology, whereby certain nonce names (e.g. *bouba* or *maluma*) tend to be associated with round shapes, whereas other nonce names (e.g. *kiki* or *takete*) tend to be associated with angular shapes. Building on previous studies on this effect, the current study addresses the question of which phonological features affect the perception of roundness, and how these effects interact with each other. The current experiment is situated within the Pokémonastics research paradigm, in which researchers explore the nature of sound symbolism using Pokémon characters. An experiment using nonce names with native speakers of English shows that labiality, stop voicing and vowel backness all increase the probability of roundness judgments. Bayesian modeling of the results shows that these phonological factors influence the sound symbolic pattern in a cumulative fashion, and that the cumulativity can be considered to be plausibly linear, with no strong evidence for the interaction terms to be playing a very substantial role in the model. To model the experimental results from the perspective of generative phonology, a MaxEnt analysis using Optimality Theoretic constraints is developed.

Keywords: sound symbolism, the *bouba-kiki* effect, English, Pokémonastics, cumulativity, linearity, MaxEnt

1. Introduction

1.1 The synopsis and the goals of the paper

The *bouba-kiki* effect is a sound symbolic connection that is widely studied in psychology and cognitive science, in which certain nonce names (e.g. *bouba*) tend to be associated with round figures, whereas other names (e.g. *kiki*) tend to be associated with angular figures (Ramachandran and Hubbard 2001). The current study extends upon the Pokémonastics paradigm, in which researchers use Pokémon names to explore the nature of sound symbolism (Shih et al. 2019), and examines the *bouba-kiki* effect from the perspective of experimental and theoretical linguistics. Building on D'onofrio (2014), the specific questions that are addressed in the current study are (1) which phonological features affect the perception of roundness, and (2) how these effects interact with each other. The current study shows that labiality, stop voicing and vowel backness all affect the perception of roundness for the native speakers of English, and they do so cumulatively. To model these experimental results, I develop a MaxEnt analysis of the pattern using Optimality Theoretic constraints, as a case study of generative phonological analyses of sound symbolism (e.g. Kawahara 2020a).

A general starting point of this project, following some previous proposals (Alderete and Kochetov 2017; Kawahara 2020a,b), is that studying sound symbolism from the perspective of theoretical phonology can be informative, both for formal phonologists and researchers on sound symbolism. Against this general theoretical backdrop, this paper has three goals. First, the current paper shows that a set of phonological features that formal phonologists routinely deploy is useful in deciphering the nature of the sound symbolic connection between sounds and shapes. Second, the current study addresses the question of cumulativity, the question of whether multiple factors can additively influence the mapping between one representation to another. In the context of formal phonology, this issue bears on the question of whether linguistic systems should be characterized as numeric optimization (as in Harmonic Grammar: Legendre et al. 1990) or lexicographic optimization (as in Optimality Theory: Prince and Smolensky 2004). The former framework predicts that linguistic effects are fundamentally cumulative, whereas the latter framework posits that such cumulative patterns are in principle impossible.[1] Third, this paper develops a generative phonological analysis of sound symbolism (Kawahara 2020a), which deploys MaxEnt Harmonic Grammar (e.g. Hayes 2020;

Hayes and Wilson 2008) coupled with Optimality Theoretic constraints (Prince and Smolensky 2004). This analysis is shown to be successful in modeling the probabilistic patterns of the experimental results. The fact that the same mechanism—i.e. MaxEnt Harmonic Grammar—can model probabilistic phonological patterns and sound symbolic patterns implies that there may be a non-trivial parallel between the two phenomena.

Despite the fact that sound symbolism is currently attracting intensive attention in a wide range of research disciplines (e.g. Dingemanse et al. 2015), the issue of cumulativity is still heavily under-addressed (Kawahara 2020b). A descriptive contribution of this study is thus to address this gap in the literature. See Kawahara (2020a) and Kawahara and Breiss (to appear) for recent reviews of cumulative nature of sound symbolism, as well as reports of experiments which show that several phonological factors affect sound symbolic judgments cumulatively.

1.2 Background on the *bouba-kiki* effect

The observation that certain sounds tend to be associated with particular shapes go back at least to Köhler (1929), who found that nonce words like *takete* tend to be associated with an angular figure and nonce words like *maluma* tend to be associated with a round figure. These sound-shape correspondences have received renewed interest in psychology, due to the influential work by Ramachandran and Hubbard (2001), who used *bouba* and *kiki* as their stimuli. D'onofrio (2014) reexamined this effect to explore which linguistic features contribute to the perception of roundness, and showed that for English speakers, consonant place of articulation, consonant voicing, and vowel backness all had a tangible influence. The current study is a rather direct replication of D'onofrio's study within Pokémonastics, a research paradigm within which we explore the general nature of sound symbolism using Pokémon names (Kawahara et al. 2018; Shih et al. 2019). Due to space limitation, I am unable to review the research advantages of this paradigm, for which I would like to refer the readers to Shih et al. (2019) and Kawahara and Breiss (to appear).

Using Pokémon characters addresses the issue of whether the *bouba-kiki* effect holds with realistic objects rather than solely with very abstract shapes (D'onofrio 2014). The use of Pokémon rather than other imageable objects may be rather arbitrary, except that the current study was intended to extend the scope of the general Pokémonastics paradigm.

2. Methods

2.1 The stimuli

Table 1: Stimulus structure of the current experiment

PoA	Voicing	Backness	Stimuli	
labial	voiceless	front	PEEP	PEP
labial	voiceless	back	POOP	POP
labial	voiced	front	BEEB	BEB
labial	voiced	back	BOOB	BOB
alveolar	voiceless	front	TEET	TET
alveolar	voiceless	back	TOOT	TOT
alveolar	voiced	front	DEED	DED
alveolar	voiced	back	DOOD	DOD
velar	voiceless	front	KEEK	KEK
velar	voiceless	back	KOOK	KOK
velar	voiced	front	GEEG	GEG
velar	voiced	back	GOOG	GOG

Table 1 shows the stimulus structure of the current experiment, which closely follows that of D'onofrio (2014). Three phonological dimensions were varied: place of articulation (PoA:

labial vs. alveolar vs. velar), stop voicing (voiced vs. voiceless) and vowel backness (front vs. back). These three factors were fully crossed in order to examine the cumulative nature of the sound-shape connections at issue. Two vowels were tested for each front vs. back vowel type ([i], shown orthographically with EE and [ɛ] shown with E vs. [u] shown with OO and [ɔ] shown with O). Two items were created for each condition by adding a prefixal "uh-" (representing schwa) and a suffixal "-uh". This overall design resulted in 48 test items: 3 PoA * 2 voicing * 2 backness * 2 vowels * 2 items.

2.2 The procedure

The experiment was distributed online using SurveyMonkey. The consent form was shown on the first page. The participants were instructed to imagine a situation in which they were helping the Pokémon company to come up with new Pokémon names. They were presented with three sample pairs of non-existing angular and round Pokémon characters; one sample pair is reproduced here in Figure 1. During the main session, they were presented with one nonce name per trial, and were asked to decide whether the name is better suited for an angular Pokémon or a round Pokémon. Although the stimuli were presented orthographically, the participants were encouraged to read the names in their head before responding. They were told that there were no right or wrong answers. They were told that all the stimuli were non-existing names, and if some of them sounded similar to existing words, they were asked to try not to take their meanings into account. The stimulus "stems" (those shown in Table 1) were shown in capital letters, whereas affixal "uh" portions were shown with small letters (e.g. PEEPuh and uhPEEP). The participants were asked to read the "uh" portions like the second vowel in *sofa*. The order of the stimuli was uniquely randomized per each participant.

Figure 1: A sample pair of an angular and a round Pokémon character presented to the participants. Three such pairs were shown in the experiment. These pictures were drawn by a digital artist, *toto-mame*, and were used in the experiment with permission.

2.3 The participant

The responses were collected using SurveyMonkey's Buy Function. Only those who met the following five conditions were allowed to participate: they (1) were native speakers of English, (2) knew what Pokémon is, (3) never studied sound symbolism before, (4) had not heard about the *bouba-kiki* effect or *takete-maluma* effect, and (5) had never participated in an experiment on sound symbolism before. The data from 156 participants were obtained.

2.4 Statistical analyses

The results were analyzed using Bayesian mixed effects logistic regression using `brms` package (Bürkner 2017). The binary dependent variable was whether the response was round or not. The model had three fixed effects: PoA, voicing, and vowel backness. The model included random intercepts for participants and items, and all the interaction terms of the fixed effects as well as by-participant slopes of all the fixed effects and the interaction terms. The random slopes and intercepts were uncorrelated. Bayesian modeling is deployed in the current analysis because we can fit a complex model with multiple interaction terms without convergence issues (see e.g. Kruschke 2014 for an accessible introduction to Bayesian modeling). Moreover, Bayesian modeling allows us to directly assess how substantial the interaction terms are. To the extent that we find cumulative interactions between multiple phonological factors, its linearity can be examined by exploring how robust the interaction terms are in the model (Breiss and Albright 2020; Kawahara and Breiss to appear). If we used a frequentist analysis, the best we could conclude about linearity is that we cannot reject the null hypothesis that multiple effects interact linearly.

Bayesian modeling yields a posterior distribution of possible values for each estimated parameter. As a useful heuristic, we can interpret Credible Intervals (CIs, also known as HDIs, High Density Intervals) for the bounds of this distribution, which represent a range of values that each estimate can take. If the 95% CIs do not include 0, we interpret it to mean that its effect meaningfully affects the responses. If the 95% CIs include 0, on the other hand, we can examine their posterior distributions more closely to explore how substantial the effect is.

For the current analysis, the default, weakly informative priors were used. Four chains were run with 6,000 iterations, the first 2,000 of which were discarded as warmup. All the R-hat values were 1, indicating that the four chains mixed successfully. The R Markdown files as well as the raw data files are available at the following osf repository (https://osf.io/4k7pj/), in accordance with the open science initiative in linguistics (e.g. Winter 2020).

3. Results

Figure 2 illustrates the results, in which each dot represents round responses for each item averaged over all the participants. We observe that labials tend to be more likely to be associated with round Pokémon more often than alveolars or velars (the left panels vs. the middle/right panels). Voiced stops tend to be more likely to be associated with round Pokémons than voiceless stops (the bottom panels vs. the top panels). Back vowels induced more round responses than front vowels (the comparison within each panel).

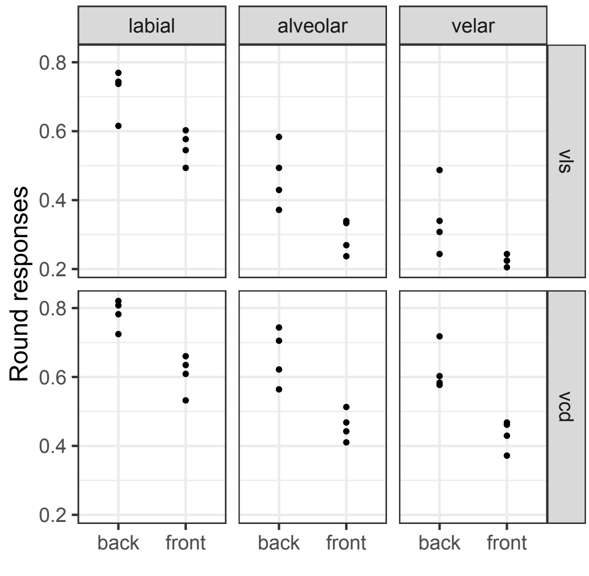

Figure 2: The descriptive results. Each dot represents round responses for each item averaged over all the participants.

Figure 3 shows the 80% CIs (thick regions) and 95% CIs (thin lines) for each parameter estimate. The exact 95% CI values can be inspected from the accompanying R Markdown file. The top coefficient is the intercept, representing the baseline condition (alveolar, voiced, back). Since this result is based on logistic regression, the positive intercept indicates that this condition had a round response that is higher than 50%. The rest shows the estimates for the slopes of the fixed effects as well as their interactions. Positive slopes of the fixed effects represent increase in round responses, whereas negative slopes, decrease in round responses (see Winter 2020: Ch.12 for how to interpret the results of logistic regression models).

The second and third coefficients show the difference between the baseline PoA (=alveolars) and labials/velars, respectively. The 95% CIs for the labial vs. alveolar comparison shows that labials were more likely to be associated with the round Pokémon characters. The next row of CIs show, on the other hand, that alveolars and velars do not differ from each other very much. The fourth row of CIs shows that voiceless stops, when compared to voiced stops, decrease the round responses. The fifth row of CIs shows that front vowels are less likely to be associated with round Pokémons than back vowels. To summarize, then, the results suggest that labiality, stop voicing, and vowel backness are all associated with the perception of roundness. The results show, moreover, that each of these phonological features affected the responses *cumulatively,* in that each factor increased the probabilities of round responses.

The rest of the CIs represent the posterior estimates of the interaction terms, which generally distribute around 0. These results show that the current cumulative pattern may be considered to be plausibly linear (in the log-odds domain); i.e. the combinatorial effects of multiple factors may be predicted based on the effects of these individual factors (see Breiss and Albright 2020 and Kawahara and Breiss to appear for related discussion on (non-)linearity; see also the R Markdown file for more careful discussion on this point, which includes ROPE analyses: Kruschke 2014). This result means that, in the context of constraint-based linguistic analyses, we may not need to posit locally-conjoined constraints, which are equivalent to interaction terms in statistical models (Shih 2017).

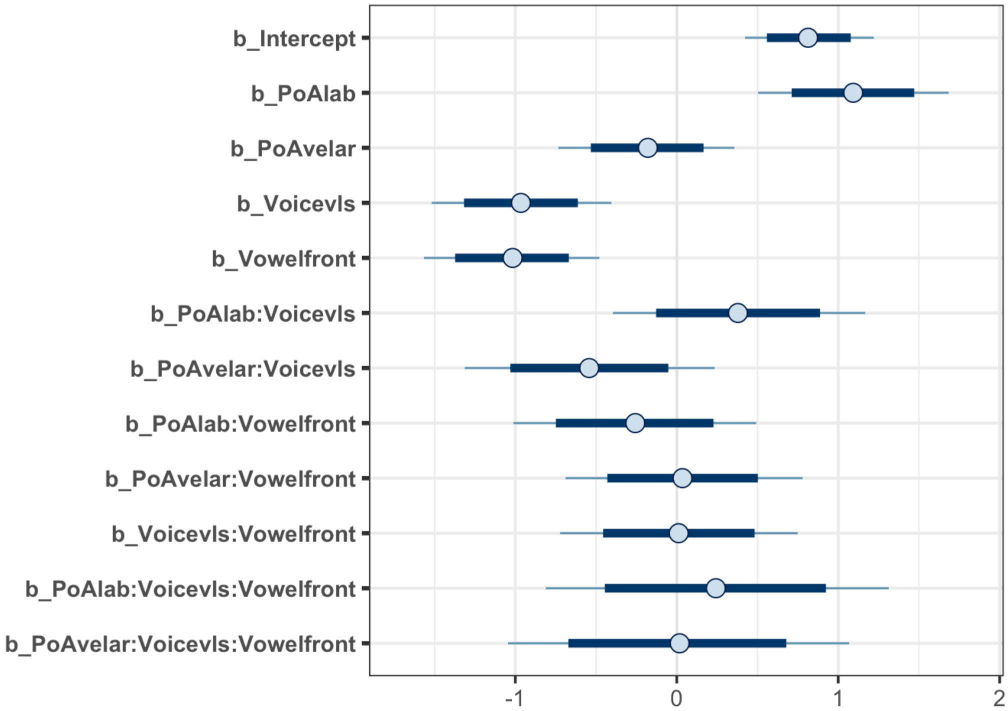

Figure 3: The posterior parameter estimates for all the fixed factors and their interaction terms. The thick regions represent the 80% Credible Intervals, whereas the thin bars represent the 95% CIs. See the R Markdown file for complete details.

The current results mostly accord well with D'onofrio's (2014) results, except that we did not find a difference between velars and alveolars, unlike D'onofrio, who found that velars were judged to be more roundish than alveolars. I do not have a clear explanation of where this difference comes from, and a targeted follow-up experiment is necessary to settle this question. The current results also make sense from the general perspective of iconicity. Labials and back vowels, which generally involve a lip rounding gesture in English, can be iconically mapped onto round figures. The fact that voiced stops were more likely to be associated with round figures than voiceless stops may arise from their acoustic differences, in which the bursts of voiceless stops are characterized by aperiodic energy, which literally look "spikey" on waveforms (Kawahara and Shinohara 2012). Voiced stops, with their periodic energy, can be mapped onto round shapes, or, be neutral in their symbolic values.

4. A MaxEnt analysis

Kawahara (2020a) proposes that we can conceive of sound symbolic mappings from sounds to meanings just like phonological input-output mappings. Kawahara (2020a) furthermore shows that MaxEnt Harmonic Grammar (e.g. Hayes and Wilson 2008), coupled with the phonological constraints that are familiar from the Optimality Theoretic research tradition (Prince and Smolensky 2004) are useful in modeling sound symbolic patterns. Following this spirit, this section develops a MaxEnt analysis of the current experimental results.[2] To do so, I posit the four constraints stated in (1).

(1)The constraints used in the MaxEnt analysis
*VcdAng: Assign a violation mark for each voiced stop in an angular object name.
*LabAng: Assign a violation mark for each labial consonant in an angular object name.
*BackAng: Assign a violation mark for each back vowel in an angular object name.
*Rnd: Assign a violation mark for each round object name.

(2) The MaxEnt Tableaux

			0.70	1.00	0.74	0.99					
			*VcdAng	*LabAng	*BackAng	*Round					
sound	shape	observed	C1	C2	C3	C4		harmony	e^-H	Z	predicted
PEEP	round	55.4				1		0.99	0.37	0.74	50.3
PEEP	angular	44.6		1				1.00	0.37	0.74	49.7
BEEB	round	60.9				1		0.99	0.37	0.56	67.0
BEEB	angular	39.1	1	1				1.70	0.18	0.56	33.0
TEET	round	29.5				1		0.99	0.37	1.37	27.2
TEET	angular	70.5						0.00	1.00	1.37	72.8
DEED	round	45.8				1		0.99	0.37	0.87	42.9
DEED	angular	54.2	1					0.70	0.50	0.87	57.1
KEEK	round	22.4				1		0.99	0.37	1.37	27.2
KEEK	angular	77.6						0.00	1.00	1.37	72.8
GEEG	round	43.3				1		0.99	0.37	0.87	42.9
GEEG	angular	56.7	1					0.70	0.50	0.87	57.1
POOP	round	71.6				1		0.99	0.37	0.55	68.0
POOP	angular	28.4		1	1			1.74	0.18	0.55	32.0
BOOB	round	78.4				1		0.99	0.37	0.46	81.0
BOOB	angular	21.6	1	1	1			2.44	0.09	0.46	19.0
TOOT	round	47.0				1		0.99	0.37	0.85	43.9
TOOT	angular	53.0			1			0.74	0.48	0.85	56.1
DOOD	round	65.9				1		0.99	0.37	0.61	61.2
DOOD	angular	34.1	1		1			1.44	0.24	0.61	38.8
KOOK	round	34.5				1		0.99	0.37	0.85	43.9
KOOK	angular	65.5			1			0.74	0.48	0.85	56.1
GOOG	round	62.0				1		0.99	0.37	0.61	61.2
GOOG	angular	38.0	1		1			1.44	0.24	0.61	38.8

The first three constraints in (1) formally express the sound symbolic connections identified in the experiment. The fourth constraint is a baseline constraint determining the general preference for angular shapes. Due to space limitation, I do not go through the underlying mathematical details—or its theoretical architecture/strengths—of MaxEnt Harmonic Grammar, for which readers are referred to e.g. Hayes (2020), Hayes and Wilson (2008) and Kawahara (2020a) (see also Anttila and Magri 2018). The MaxEnt tableaux of the current experimental results are shown in (2). The observed values are taken from the grand average of each phonological condition. The optimal weights of the four constraints in (1) were calculated using the Solver function of Excel by minimizing the log-likelihood of the model with respect to the current data. The weights found for these constraints are shown at the top row in (2). The Excel analysis file is available on the osf repository mentioned above.

The correlation between the observed values obtained in the experiment and those values that are predicted by the MaxEnt analysis is shown in Figure 4, which shows a good match between the two values. Whether each constraint is justified to be included in this MaxEnt analysis is accessed by log-likelihood ratio tests, which showed that all four constraints significantly increased the fit of the model ($\chi^2(1)$=33.3, 37.6, 59.7, 75.1, all $p < .001$).

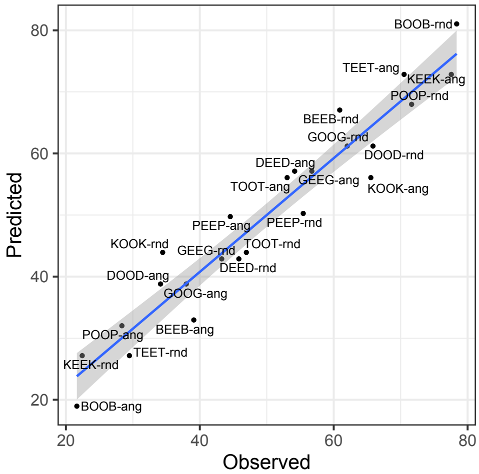

Figure 4: The correlation between the observed values in the experiment and the values predicted by the MaxEnt analysis.

The success of this analysis shows that MaxEnt with OT constraints is a useful tool to model sound-shape correspondences. To the extent that MaxEnt provides a useful tool for phonological analyses (e.g. Hayes 2020; Hayes and Wilson 2008), it shows that there may be some meaningful parallel between sound symbolic patterns and phonological patterns, two patterns which have hitherto been considered as two completely separate phenomena (Alderete and Kochetov 2017; Kawahara 2020a, b).

5. Conclusion

The current experiment has shown that English speakers symbolically associate labiality, stop voicing and vowel backness with round Pokémon characters, largely replicating D'onofrio (2014). The results show that these phonological factors cumulatively influence the sound symbolic patterns, and that the cumulativity can plausibly be linear, with no strong evidence for interaction terms to be playing a substantial role. A MaxEnt analysis with OT constraints shows that the sound-shape mappings can be modelled just like phonological input-output mappings. Overall, I hope to have shown that linguistic analyses, in particular the use of phonological features and MaxEnt, are useful in exploring the nature of sound symbolism.

****

Notes

[*] This project is supported by JSPS grants #17K13448 and #18H03579. I thank Canaan Breiss, Donna Erickson, Gakuji Kumagai and two anonymous reviewers for their comments. Remaining errors are mine.

[1] We can also understand this issue from a more general perspective of decision-making: a regression-based model predicts cumulative patterns, whereas a fast-and-frugal decision-making model predicts a non-cumulative pattern (Gigarenzer and Gaissmaier 2011). In the former framework, it is predicted that all relevant factors are taken into consideration when determining the sound symbolic value of a particular word, whereas the latter framework predicts that only the most salient segment plays the decisive role (Kawahara and Breiss to appear).

[2] Since MaxEnt is mathematically equivalent to logistic regression, there is some conceptual overlap between the Bayesian logistic regression analysis presented in section 3 and the MaxEnt analysis presented here. The former is a statistical analysis to examine the experimental results; the latter is a generative phonological analysis.

References

Anttila, Arto and Giorgio Magri (2018) Does MaxEnt overgenerate? Implicational universals in Maximum Entropy Grammar. *Proceedings of the Annual Meeting of Phonology 2017*.

Alderete, John and Alexei Kochetov (2017) Integrating sound symbolism with core grammar: The case of expressive palatalization. *Language* 93: 731–766.

Breiss, Canaan and Adam Albright (2020) Cumulative markedness effects and (non-)linearity in phonotactics. Ms. UCLA and MIT.

Bürkner, Paul-Christian (2017) brms: An R package for Bayesian multilevel models using Stan. R package.

Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morten H. Christiansen and Padraic Monaghan (2015) Arbitrariness, iconicity and systematicity in language. *Trends in Cognitive Sciences* 19: 603–615.

D'Onofrio, Annette (2014) Phonetic detail and dimensionality in sound-shape correspondences: Refining the bouba-kiki paradigm. *Language and Speech* 57: 367–393.

Gigerenzer, Gerd and Wolfgang Gaissmaier (2011) Heuristic decision making. *Annual Review of Psychology* 62: 451–482.

Hayes, Bruce (2020) Deriving the wug-shaped curve: A criterion for assessing formal theories of linguistic variation. Ms. UCLA.

Hayes, Bruce and Colin Wilson (2008) A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39: 379–440.

Kawahara, Shigeto (2020a) A wug-shaped curve in sound symbolism: the case of Japanese Pokémon names. *Phonology* 37.

Kawahara, Shigeto (2020b) Sound symbolism and theoretical phonology. *Language and Linguistic Compass* 14(8). e12372.

Kawahara, Shigeto and Canaan Breiss (to appear) Exploring the nature of cumulativity in sound symbolism: Experimental studies of Pokémonastics with English speakers. *Journal of Laboratory Phonology*.

Kawahara, Shigeto, Atsushi Noto and Gakuji Kumagai (2018) Sound symbolic patterns in Pokémon names. *Phonetica* 75: 219–244.

Kawahara, Shigeto and Shinohara, Kazuko (2012) A tripartite trans-modal relationship between sounds, shapes and emotions: A case of abrupt modulation. *The proceedings of the 34th annual meeting of Cognitive Science Society*: 569–574.

Kruschke, John K. (2014) *Doing Bayesian data analysis*. Amsterdam: Elsevier.

Legendre, Géraldine, Yoshiro Miyata and Paul Smolensky (1990) Harmonic grammar – a formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*: 388–395.

Köhler, Wolfgang (1929) *Gestalt psychology*. New York: Liveright.

Prince, Alan and Paul Smolensky (2004) *Optimality theory: Constraint interaction in Generative Grammar*. Malden and Oxford: Blackwell.

Ramachandran, Vilayanur S. and Edward M. Hubbard (2001) Synesthesia–a window into perception, thought, and language. *Journal of Consciousness Studies* 8: 3–34.

Shih, Stephanie S (2017) Constraint conjunction in weighted probabilistic grammar. *Phonology* 34.243–267.

Shih, Stephanie S, Jordan Ackerman, Noah Hermalin, Sharon Inkelas, Hayeun Jang, Jessica John- son, Darya Kavitskaya, Shigeto Kawahara, Miran Oh, Rebecca L Starr and Alan Yu (2019) Cross-linguistic and language-specific sound symbolism: Pokémonastics. Ms. University of Southern California, University of California, Merced, University of California, Berkeley, Keio University, National University of Singapore and University of Chicago.

Winter, Bodo (2020) *Statistics for linguists: An introduction using R*. New York: Routledge.