MaxEnt, its quantitative signature and sound symbolism (Or, can you draw wug-shaped curves with Pokémonastics?)

Abstract

Whether linguistic patterns show cumulative effects or not is one of the most actively debated issues in contemporary phonological studies. This paper attempts to shed new light on this debate from a novel perspective by studying sound symbolism, systematic associations between sounds and meanings. The current experiment shows that when Japanese speakers judge Pokémons' evolution status based on nonce names, the judgments are affected both by their mora counts and the presence of a voiced obstruent. The effects of mora counts instantiate counting cumulativity, while the interaction between these two factors instantiates ganging-up cumulativity. These patterns together result in what Hayes (2020) refers to as "wug-shaped curves," a quantitative signature predicted by MaxEnt grammar. The paper shows that the experimental results can indeed be successfully modeled using MaxEnt grammar with Optimality Theoretic constraints. I also examine Stochastic Optimality Theory in light of the current experimental results and show that this theory faces an interesting set of challenges. Finally, in this paper I make a novel methodological proposal for general phonological inquiry and sound symbolism research. The current study was inspired by Hayes (2020), a proposal made within formal phonology. The experiment ended up revealing important, hitherto understudied aspects of sound symbolism, and in turn, it revealed how cumulativity manifests itself in linguistic patterns. The current exploration thus shows that formal phonology and research on sound symbolism can mutually inform one another.

1 Introduction

1.1 The direct source of inspiration: Hayes (2020)

This study arose as a result of a confluence of various research interests, but the direct source of inspiration was Hayes (2020), who has asked whether we can draw "wug-shaped curves" when certain aspects of linguistic patterns are plotted in a certain way.

Traditional linguistics theories in the twentieth century primarily focused on categorical generalizations, which assumed that grammar makes only dichotomous distinctions between grammatical forms and ungrammatical forms (Chomsky 1957; Sprouse 2007). The crucial distinction that was the target of phonological analyses had been a dichotomous distinction between impossible forms (e.g. *bnick*) and possible/existing forms (e.g. *brick* or *blick*) (Chomsky & Halle 1968; Halle 1978). Probabilistic or stochastic generalizations were hardly the focus of phonological analyses, although in practice exceptions to phonological generalizations were usually acknowledged and handled by some means (e.g. Kisseberth 1970).

A growing body of recent studies have shown, however, that linguistic knowledge is deeply stochastic in nature, which seems evident in syntax (e.g. Bresnan & Hay 2008; Kellar 2006; Schütze 1996) as well as in phonology (e.g. Boersma & Hayes 2001; Coetzee & Pater 2011; Cohn 2006; Daland et al. 2011; Hayes & Londe 2006; Pierrehumbert 2001; Zuraw 2000). It is not uncommon that the same word can be produced differently in different social or discourse contexts, which has been the focus of the extensive sociolinguistics studies (e.g. Guy 2011). Some phonological processes can apply with different probabilities in different contexts, and these probabilities can be predicted based on the interaction of various (morpho-)phonological factors (e.g. *t/d*-deletion in English: Guy 1991). Some phonotactic sequences are neither completely grammatical nor ungrammatical, but can instead be intermediate; indeed, controlled phonotactic judgment experiments typically reveal a continuous, gradient pattern (e.g. Daland et al. 2011).

Therefore, the field of phonology has recently witnessed a dramatic rise of interests in formal grammatical models which can account for such stochastic phonological generalizations. Among those, most widely employed frameworks include (1) Stochastic Optimality Theory (Boersma 1998; Boersma & Hayes 2001; Zuraw 2000), (2) Noisy Harmonic Grammar (Boersma & Pater 2016; Coetzee & Kawahara 2013; Hayes 2017), and (3) Maximum Entropy Harmonic Grammar (henceforth MaxEnt) (Goldwater & Johnson 2003; Zuraw & Hayes 2017). Teasing apart these stochastic models of phonology is one of the most actively debated issues in contemporary phonological studies (Anttila & Magri 2018; Anttila et al. 2019; Breiss 2019; Breiss & Albright 2020; Hayes 2017, 2020; Jäger & Rosenbach 2006; Jäger 2007; O'Hara 2017; Pizzo 2015; Smith & Pater 2017; Zuraw & Hayes 2017 among many others).

To bear on this debate, Hayes (2020) took an abstract, top-down approach and asked the following question: if we take the MaxEnt grammar framework seriously, what predictions does it make in terms of its quantitative signature (i.e. the probabilistic pattern that it typically generates)? To be more specific, suppose that there is a scalar constraint, S, that is gradiently violable—i.e., its violations can be assessed on a numerical scale—and a binary constraint, B.[1] Further suppose that these constraints are in direct conflict with each other; i.e. the satisfaction of S entails the violation

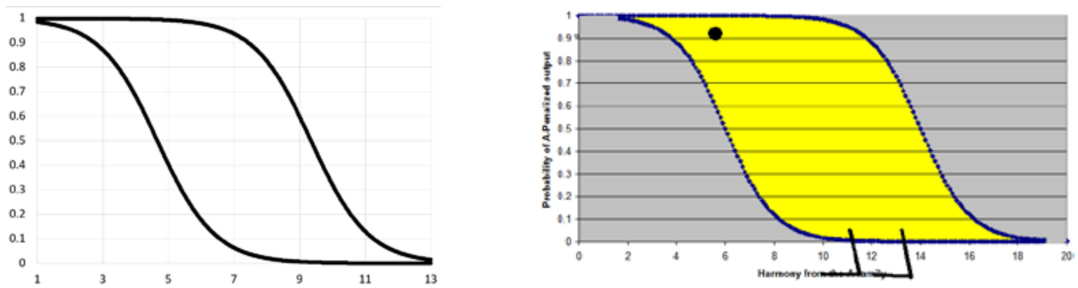
[1] Hayes (2020) uses different names (VARIABLE and ONOFF) for these two constraints.

B, and vice versa. When we simulate the probabilities of the candidate that obeys B and violates S as a function of the number of violations of S, we get a sigmoid (s-shaped) curve, as shown in Figure 1 (see McPherson & Hayes 2016 and Zuraw & Hayes 2017 for the equation which derives this curve). In reality, the constraint violation profile of S is discrete (ranging from 1 to 7 in Figure 1), but for the sake of illustration, Figure 1 continuously plots for all values, not just the integers. This curve is characterized by the fact that the y-axis values do not change very much when the x-axis values are small (from 1 to 3) or large (from 5 to 7), but there is a radical change in the middle range (from 3 to 5).

Figure 1: A sigmoid curve predicted by the MaxEnt grammar with a scalar constraint S and a binary constraint B, which are directly opposed to each other. Adapted from Hayes (2020: 5). The axis labels were edited by the author.

Hayes (2020) further considers a case in which two sets of inputs are relevant—each set consists of inputs with the constraint violation profiles that are identical to those in Figure 1, but the two sets differ in terms of whether they violate an additional "perturber" constraint (P) or not. This scenario creates two sigmoid curves, as in Figure 2(a). Hayes (2020) calls these "wug-shaped curves," because, as illustrated in Figure 2(b), they resemble the beloved animal well known in the phonology tradition, since the classic work by Berko (1958).

Hayes (2020) shows that the wug-shaped curves are natural outcomes of the MaxEnt grammar, and are also predicted under some versions of Noisy Harmonic Grammar, but not under Stochastic Optimality Theory. Therefore, this top-down approach to examine quantitative signatures of linguistic generalizations offers one strategy to distinguish between different stochastic grammatical models. If we are to find wug-shaped curves in linguistic patterns, then it provides support for MaxEnt grammar or Noisy Harmonic Grammar over Stochastic Optimality Theory. Hayes (2020), building upon McPherson & Hayes (2016) and Zuraw & Hayes (2017), argues that such wug-shaped curves are commonly observed in phonology, and elsewhere in other domains of linguistic patterns, such as speech perception (classic categorical perception: Liberman et al. 1957 *et*



(a) (b)

Figure 2: Wug-shaped curves with two sigmoid functions. Adapted from Hayes (2020: 7).

seq.) and diachronic changes in syntax (Kroch 1989; Zimmermann 2017).

Inspired by Hayes (2020) and the body of research cited by him, this paper asks whether we can draw wug-shaped curves in the patterns of sound symbolism, systematic/iconic associations between sounds and meanings (Hinton et al. 2006). If the answer to this question turns out to be positive, then it supports the idea that MaxEnt is suited to model the knowledge that lies behind sound symbolism (Kawahara et al. 2019). Moreover, to the extent that MaxEnt is suited as a model of phonological knowledge (e.g. Hayes & Wilson 2008; McPherson & Hayes 2016; Zuraw & Hayes 2017 among many others), it implies that the same mechanism may lie behind phonological patterns and sound symbolic patterns; i.e. that there is a non-trivial parallel between phonological patterns and sound symbolic patterns.

1.2 Other issues addressed in the current study

1.2.1 Cumulativity

Stepping back from Hayes (2020) a bit, one general theoretical issue that lies behind the wug-shaped curves is that of cumulativity. Cumulativity is one of the most actively debated issues in current linguistic theorization, because addressing this issue potentially helps us tease apart Optimality Theory (henceforth OT) (Prince & Smolensky 1993/2004) with ranked constraints from other constraint-based theories with numerically weighted constraints, such as Harmonic Grammar (Breiss 2019; Breiss & Albright 2020; Farris-Trimble 2008; Hayes et al. 2012; Jäger & Rosenbach 2006; Jäger 2007; McPherson 2016; McPherson & Hayes 2016; Pater 2009; Zuraw & Hayes 2017).

It is convenient to distinguish two types of cumulativity, counting cumulativity and ganging-up cumulativity (Jäger & Rosenbach 2006; Jäger 2007), as they present different types of challenges to OT. Counting cumulativity is a case in which multiple violations of the same constraint add up. Suppose, in the context of OT, that Constraint A dominates Constraint B, then OT predicts that a single violation of Constraint A takes precedence over any number of violations of Constraint B—this is a consequence of the strict domination of constraint rankings, one central tenet of OT (Prince & Smolensky 1993/2004). In reality, however, it is not uncommon that a language tolerates one violation of a particular constraint but not two violations. For instance, the native phonology of Japanese tolerates one voiced obstruent within a morpheme, but not two voiced obstruents (a.k.a. Lyman's Law: Ito & Mester 2003). Such observations are commonly accounted for in OT by positing OCP constraints (Leben 1973; Ito & Mester 1986; Myers 1997) or self-conjoined constraints, which are violated if and only if there are two instances of the same structure (Alderete 1997; Ito & Mester 2003). Grammatical frameworks related to OT, which use numerical weights instead of rankings, can account for counting cumulativity without positing an additional mechanism.[2]

A ganging-up cumulativity would be a case in which Constraint A dominates both Constraints B and C, but simultaneous violations of Constraints B and C would "gang-up" to take precedence over the violation of Constraint A. Such a scenario is not predicted under OT, again because of the strict domination of constraint rankings. To analyze a ganging-up cumulativity effect, OT generally requires local conjunction of Constraints B and C (Crowhurst 2011; Smolensky 1995, 1997). For example, the loanword phonology of Japanese tolerates voiced obstruent geminates in isolation, as well as two voiced obstruent singletons. However, voiced obstruent geminates undergo devoicing when they co-occur with another voiced obstruent. Nishimura (2006) thus proposes to locally conjoin *VOICEDOBSGEM and OCP(voice) within the domain of stem. Frameworks with numerically weighted constraints can demonstrably account for this ganging-up cumulativity pattern without stipulating complex locally conjoined constraints (Pater 2009) (see also Potts et al. 2010).

In short, whether phonological patterns show counting/ganging-up cumulativity is an important issue that is actively debated in contemporary phonological theorization, because it bears on the issue of whether the grammatical model should be based on rankings or weights. More broadly speaking, the question is whether the optimization algorithm deployed in the linguistic system is based on lexicographic ordering or numeric ordering (Tesar 2007).

The current paper attempts to shed new light on this debate by examining a pattern that has hitherto been barely analyzed from this perspective; namely, sound symbolism, or systematic/iconic relationships between sounds and meanings (Hinton et al. 2006). The primary question that is addressed in this study is thus whether sound symbolism shows cumulative effects, both in terms of

[2] One widely-shared idea in OT (and much of the pre-OT literature) is that there can be a constraint that penalizes two instances of a particular structure, but there are no constraints that penalize three instances (e.g. Ito & Mester 2003). It is not clear how this restriction can be imposed upon OT grammar, once we add a tool like self-conjunction in the analytical arsenal. Weight-based theories predict no essential differences between one instance vs. two instances and two instances vs. three instances, as will be shown in further detail in §4. It used to be believed that phonological systems do not count beyond two (e.g. McCarthy & Prince 1986), although this thesis was recently challenged by Paster (2019). See Kawahara et al. (2020), McPherson & Hayes (2016), Paster (2019) as well as the experimental results below, for cases which apparently count beyond two.

counting cumulativity and ganging-up cumulativity.

This is an empirical question that is important to address for its own sake, because there are only a few studies which directly addressed the (non-)cumulative nature of sound symbolism, and therefore this is one aspect of sound symbolism that is only poorly understood. There are some impressionistic reports regarding counting cumulativity in the literature—more segments of the same kind evoke stronger sound symbolic images (Hamano 2013; Martin 1962; McCarthy 1983). Thompson & Estes (2011) addressed whether sound symbolism is categorical or gradient by way of experimentation, and found some evidence for cumulativity in their results. A recent experimental study by Kawahara & Kumagai (to appear) found evidence for counting cumulativity, when they studied various sound symbolic values of voiced obstruents in Japanese. D'Onofrio (2014) examined the well-known *bouba-kiki* effect (Ramachandran & Hubbard 2001), in which certain classes of sounds are associated with round figures, whereas other classes of sounds are associated with angular figures. She found that vowel backness, consonant voicing and consonant labiality all contribute to the perception of roundness, instantiating a case of ganging-up cumulativity (Kawahara 2020). No studies, to the best of my knowledge, have addressed the question of whether counting cumulativity and ganging-up cumulativity can co-exist in the same sound symbolic system (though see Kawahara et al. 2020, which is discussed in some detail in §2).

In a sense, this question—whether the same pattern can show counting cumulativity and ganging-up cumulativity at the same time—is the one that is addressed by Hayes (2020): each of the sigmoid curves can arise when counting cumulativity is in action,[3] and the separation of the two curves is a sign of ganging-up cumulativity. Apart from sound symbolism, Experiment 4 of Breiss (2019) shows that we observe both counting cumulativity and ganging-up cumulativity in phonotactic learning patterns in an artificial language learning experiment. Case studies reported in McPherson & Hayes (2016) and Zuraw & Hayes (2017) can also be understood as simultaneously involving counting cumulativity and ganging-up cumulativity. Apart from these studies, there are not many case studies that have directly addressed this question, especially in the domain of sound symbolism. One aim of this paper is to address this gap in the literature.

The issue of cumulativity in sound symbolism is interesting to address from a theoretical perspective as well. To the extent that cumulativity is a general property of phonological patterns (Breiss 2019; Hayes 2020; McPherson & Hayes 2016; Zuraw & Hayes 2017), and to the extent that sound symbolic effects show similar cumulative properties, then a non-trivial parallel between phonological patterns and sound symbolic patterns can be instantiated (Kawahara 2020). This parallel would lend some credibility to the hypothesis that sound symbolism is a part of "core" linguistic knowledge, as recently argued by several researchers (Alderete & Kochetov 2017; Jang 2019; Kawahara to appear; Kumagai 2019; Shih 2019). This is a rather radical conclusion, given

[3]A sigmoid curve entails counting cumulativity, but not vice versa. Counting cumulativity, for example, can manifest itself as a linear function rather than an s-curve function.

the fact that sound symbolism has long been considered as residing outside the purview of theoretical linguistics.

1.2.2 Pokémonastics

In addition to addressing the issue of cumulativity in sound symbolism, the current study can also be considered as a case study of the Pokémonastics research paradigm, within which researchers explore the nature of sound symbolism using Pokémon names (Kawahara et al. 2018; Shih et al. 2019). I refer the readers to Shih et al. (2019) for the discussion of several advantages of this research paradigm, and provide minimal background information necessary for what follows. Pokémon is a game series first released in Nintendo Inc in 1996 and became very popular world wide since then. In this game series, players collect and train fictional creatures called Pokémon, which is a truncation of [**poke**tto **mon**sutaa] 'pocket monster.' One feature that becomes crucial in what follows is that some Pokémon characters undergo evolution, and when they do so, they generally become larger, heavier and stronger. When they evolve, moreover, they are called by a different name; for instance, *Pikachu* becomes *Raichu*.

Kawahara et al. (2018) show that when we systematically examine their names from the perspectives of sound symbolism, post-evolution characters have longer names than pre-evolution characters. They call this "the longer the stronger principle," and attribute this observation to one of the previously known sound symbolic principles, "the iconicity of quantity," (Haiman 1980, 1984), in which larger quantity is expressed by longer phonological material. They also show that post-evolution Pokémon characters are more likely to have names with voiced obstruents than pre-evolution characters. This observation is likely to be related to the observation that in Japanese, voiced obstruents often sound symbolically denote large quantity and/or strength (Hamano 1998; Kawahara 2017). The experiment below examines these two sound symbolic effects in further detail by way of experimentation.

The current paper turns out to be a convenient place to highlight a few research advantages that were implicit in the previous Pokémonastics studies. First, as reported in detail below, it was possible to collect responses from more than 800 participants over a single night. Being able to collect this many participants itself is an non-trivial research advantage. The fact that this many people participated in the experiment without any compensation shows that Pokémonastics experiments can be a fun and interesting experiment, and therefore it may provide us with an opportunity to popularize linguistics among the general public.

1.3 The goals of this paper: An interim summary

As discussed above, this study arose as a result of a confluence of various research interests. The take-home messages of the current study are summarized in (1)-(4).

(1) Wug-shaped curve

a. We can draw wug-shaped curves with Pokémonastics.

(2) Cumulativity

a. Sound symbolism shows counting cumulativity.
b. Sound symbolism shows ganging-up cumulativity.
c. These two types of cumulativity can co-exist within a single system.

(3) Theoretical implications

a. MaxEnt Harmonic Grammar with Optimality Theoretic constraints can naturally account for these two types of cumulativity.
b. Stochastic OT can also model the observed patterns, but it requires additional tweaks.
c. Sound symbolic mappings can be modeled using the same mechanism as phonological generalizations.
d. Phonological theories and research on sound symbolism can inform one another.

(4) Pokémonastics

a. Pokémonastics has a methodological advantage in that we can collect data from a large group of participants over a short period of time.
b. Pokémonastics may appeal to a general audience, potentially contributing to popularizing linguistics.

1.4 The roadmap

To achieve the goals summarized in (1)-(4), the rest of the discussion is developed as follows. §2 reports the methods of an experiment which was designed to address the question of whether we can draw wug-shaped curves in a sound symbolic pattern within the framework of Pokémonastics research. Along the way, we witness one research advantage of this paradigm, i.e. we can obtain a large number of participants over a short period of time. We will observe in §3 that the results clearly instantiate wug-shaped curves, showing both counting cumulativity and ganging-up cumulativity. §4 develops a MaxEnt grammar analysis of the experimental results, by making use of three Optimality Theoretic constraints which each corresponds to S, B and the perturber constraint P, schematically illustrated in §1.1. §5 discusses how Stochastic OT may account for the current results, showing that in order to model the wug-shaped curves, Stochastic OT requires several tweaks. The final section offers general concluding remarks.

2 Methods

One precursor of the current experiment is Kawahara et al. (2020), who report a judgment experiment on the strengths of Pokémon move names. They manipulated mora length from 2 moras to 7 moras, and show that the longer the move names, the stronger they were judged to be. They also manipulated the presence/absence of a voiced obstruent placed at word-initial position, and found that move names with voiced obstruents were judged to be stronger. Their results are reproduced in Figure 3, which instantiate both counting cumulativity (the effect of mora counts) and ganging-up cumulativity (the additive effects of the two factors). However, their experiment is not suited to address the question of whether we can draw wug-shaped curves, nor were their results amenable to a MaxEnt analysis, because the judged values were continuous—what we need instead is the probability distributions of categorical outcomes.

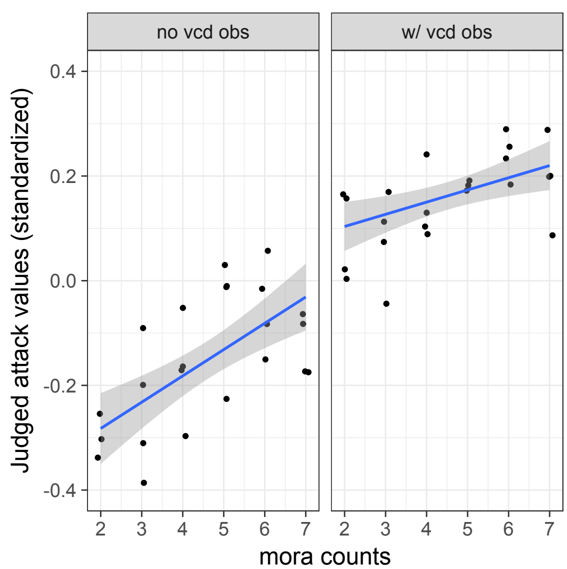


Figure 3: The effects of mora counts and word-initial voiced obstruents on judged attack values in nonce Pokémon move names. The y-axis shows standardized judged strengths, which are continuous. Adapted from Kawahara et al. (2020), their Figure 4.

The current study builds upon Kawahara et al. (2020), but in order to obtain a binary categorical response, the experiment asked the participants to judge whether each stimulus name is better suited for a pre-evolution character or post-evolution character. To obtain more reliable estimates of each condition, more items were included for each condition. As we will see below, moreover, the current experiment collected responses from many more participants, highlighting one research advantage of Pokémonastics.

2.1 Stimuli

Table 1 lists the stimuli, in which dots represent mora boundaries. The experiment manipulated two variables: mora counts and the presence of a voiced obstruent placed at word-initial position. The mora count was varied in order to examine the counting cumulativity, and relatedly, to examine whether it would result in a sigmoid curve. Mora counts varied from 2 to 6, corresponding to minimum and maximum lengths for Pokémon names.[4] This experiment manipulated mora counts rather than segment counts or syllables counts, because mora counts is what was used in the previous studies (Kawahara et al. 2018, 2020; Shih et al. 2019), and moreover, mora is demonstrably the most psycholinguistically salient prosodic counting unit in Japanese (Otake et al. 1993). The perturbing factor (see §1.1) was the presence/absence of a voiced obstruent placed at the name-initial position. As shown in Table 1, 6 items were included in each cell. All the names were created using a nonce name generator, which randomly combines Japanese moras to create new names.[5] This random generator was used to preclude the potential bias by the experimenter to select the stimuli that were likely to support their hypothesis prior to the experiment (Westbury 2005). No voiced obstruents appeared word-internally. No geminates, long vowels, or coda nasals appeared anywhere in the stimuli; i.e. all syllables were open syllables.

[4]This minimum length is likely to be related to the prosodic minimum word effect that is active elsewhere in Japanese phonology (Ito 1990).

[5]`http://sei-street.sakura.ne.jp/page/doujin/site/doc/tool_genKanaName.html`

Table 1: The list of stimuli. Dots represent mora boundaries.

	No voiced obstruent	With voiced obstruent
2 moras	[su.tsu]	[ze.ke]
	[ju.se]	[za.me]
	[no.çi]	[ɡu.ka]
	[jo.ni]	[ɡi.ke]
	[ho.mu]	[ba.ɾu]
	[ni.mi]	[ɡo.ɸu]
3 moras	[ku.ɕi.me]	[bu.ɾo.se]
	[jo.ɾu.so]	[ɡo.se.he]
	[se.sa.ɾi]	[bo.ma.sa]
	[ɾe.to.na]	[bi.nu.ki]
	[mu.su.ha]	[ɡu.ne.ju]
	[ɾi.to.no]	[da.su.ɾo]
4 moras	[ku.ki.me.se]	[be.ni.ɾo.ɾu]
	[so.ha.ko.ni]	[bi.to.ɾe.ni]
	[ɾi.se.mi.ɾa]	[za.ni.te.ja]
	[ɾa.ɕi.no.ɾo]	[ɡa.çi.ke.ɾo]
	[ko.te.nu.ne]	[da.ka.i.mi]
	[a.mo.çi.ni]	[do.i.wa.nu]
5 moras	[ha.ku.te.çi.no]	[bi.so.ɸu.sa.ta]
	[ɾo.ta.ɾa.na.to]	[da.ɾa.su.to.ki]
	[so.ka.ne.ni.ɾe]	[de.mu.sa.te.he]
	[ɾu.ɾi.ha.me.ke]	[zu.to.tsu.ɾi.su]
	[me.ju.na.u.ɾi]	[ɡi.a.so.ta.e]
	[sa.na.çi.ta.ni]	[de.nu.ɾa.so.me]
6 moras	[ju.ɾo.ka.mu.mo.ja]	[ɡu.se.ɸu.çi.ɾa.mo]
	[te.su.ɸu.ɾa.ku.su]	[ɡo.na.ɸu.to.ko.so]
	[mu.ku.ho.ɾo.ho.te]	[do.ja.to.sa.mi.ta]
	[ɾa.ha.ɾi.tɕi.ɾu.tsu]	[da.na.ɾi.no.mi.ki]
	[ne.nu.he.mo.sa.nu]	[ɡu.ko.tsu.ni.u.mi]
	[ɾu.no.nu.ɾo.te.tɕi]	[zo.te.he.so.ju.ɾa]

2.2 Procedure

The experiment was distributed as an online experiment using SurveyMonkey. Within each trial, the participants were given one nonce name and asked to judge whether that name is better for a pre-evolution character or a post-evolution character, i.e., the task was to make a binary decision. The stimuli were presented in the Japanese *katakana* orthography, which is used to represent real Pokémon names. They were asked to base their decision on their intuition, without thinking too much about “right” or “wrong” answers. The order of the stimuli was randomized for each participant.

2.3 Participants

The experiment was advertised on a Pokémon fan website.[6] A total of 857 participants completed the experiment over a single night. Since some previous Pokémonastics experiments were advertised on the same website (e.g. Author et al., 2020), 124 of them reported that they either had participated in another Pokémonastics experiment or studied sound symbolism before. Three participants were non-native speakers of Japanese. After excluding the data from these speakers, the data from the remaining 730 participants entered into the subsequent analysis.

2.4 Analysis

For statistical analysis, a logistic linear mixed effects model was fit with the response (pre-evolution vs. post-evolution) as the dependent variable (Jaeger 2008). The fixed independent variables include the mora count and the presence of a voiced obstruent as well as its interaction. The mora count was centered, because it was a continuous variable. Random factors were participants and items. The model with maximum random structure with both slopes and intercepts (Barr et al. 2013) did not converge; hence a simpler model with only random intercepts was interpreted.

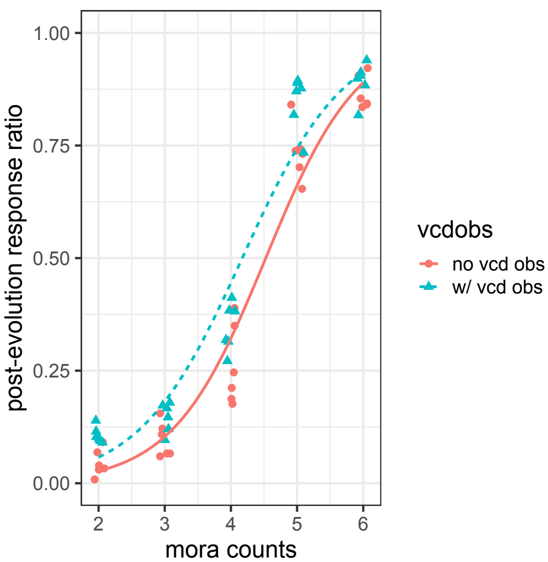
3 Results

Figure 4 shows the results. Figure 4(a) plots “post-evolution response ratios” for each item, averaged over all the participants. The items for the condition with voiced obstruents are shown with green triangles and the items for the condition without a voiced obstruents are shown with red circles. A logistic curve is superimposed for each voicing condition (green dotted line = with a voiced obstruent; red solid line = without a voiced obstruent).

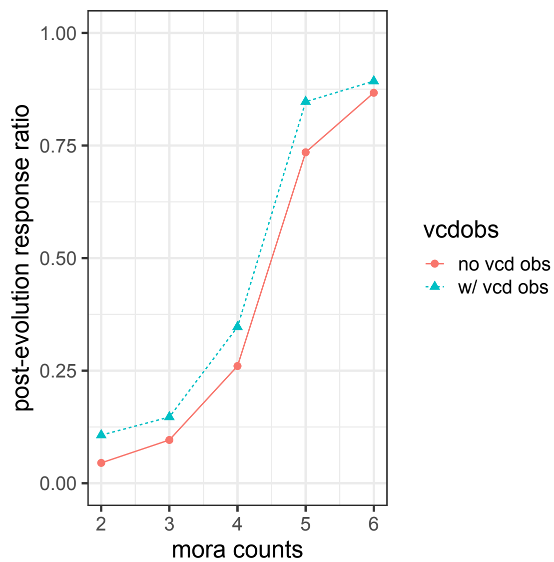
These results look like wug-shaped curves, schematically illustrated in Figure 2, consisting of two s-shaped curves slightly separated from each other.[7] The relationships between the x-axis and y-axis appear to be closer to s-shaped curves than to a linear function, in that the slope is evidently steepest in the middle range. This observation is also clear in Figure 4(b), which illustrates the overall pattern by presenting grand averages for each condition—this analysis does not presuppose that sigmoid curves would fit the data points well. The slopes are rather steep between the 3-mora

[6] `http://pokemon-matome.net`

[7] One question that can be explored in future experimentation is whether we can draw a “stripey wug” consisting of three sigmoid curves, if we include conditions that contain two voiced obstruents.



(a)



(b)

Figure 4: (a) The by-participant averages for each item. The items with a voiced obstruent are shown with green triangles and those without a voiced obstruent are shown with red circles. To avoid overlap, the points were horizontally jittered by 0.1. Logistic curves are superimposed—the green dotted line represents the condition with a voiced obstruent, whereas the red solid line represents the condition without a voiced obstruent. (b) The line-plots with grand averages for each condition.

condition and the 5-mora condition. On the other hand, the slopes are not very steep between the 2-mora condition and the 3-mora condition and between the 5-mora condition and 6-mora condition. As Hayes (2020: 3) puts it, "certainty is evidentially expensive"—it requires very strong evidence to be certain that a particular name is for a pre-evolution character or for a post-evolution character.

The model summary of the linear mixed effects model appears in Table 2. It shows that the two main factors are statistically significant: the longer the names, the more likely that they are judged to be a better name for post-evolution characters, and names with a voiced obstruents are judged to be better suited for post-evolution characters. The significant interaction effect between these two factors shows that the strengths of the effects of voiced obstruents differ depending on how long the names are; for example, the difference between the two voicing condition is not as clear for 6-mora long names in Figure 4.

Table 2: The summary of the logistic linear mixed effects model.

	β	$s.e.$	z	p
intercept	-0.78	0.06	-12.74	$< .001$
mora count	1.48	0.018	78.98	$< .001$
vcd obs	0.55	0.027	19.89	$< .001$
mora count $\times$ vcd obs	-0.13	0.025	-5.41	$< .001$

The effect of mora count instantiates a clear case of counting cumulativity (Jäger & Rosenbach 2006), in which each mora count additively contributes to the post-evolution judgment. This effect of mora count is evident both when there is a voiced obstruent name-initially and when there is not. The effect of a voiced obstruent in name-initial position manifests itself as a shift between the two s-curves. The effects of mora counts and a voiced obstruent together instantiate a ganging-up cumulativity (Jäger & Rosenbach 2006)—both factors contribute to the judgement of evolvedness. Overall, the current results show that counting cumulativity and ganging-up cumulativity can co-exist within a single sound symbolic system. This conclusion is compatible with the results of an artificial language learning experiment on phonotactic learning reported by Breiss (2019), as well as some phonological observations made by Hayes (2020), McPherson & Hayes (2016), and Zuraw & Hayes (2017). See also Breiss (2019) and Kawahara (2020) for summaries of cumulative effects in phonological alternations as well as wellformedness judgment patterns of surface phontactics.

While the results in Figure 4 seem to instantiate a clear case of "wug-shaped" curves, one may wonder if the results could have been otherwise. The answer is positive, as there were multiple alternative patterns that could have arisen from the current experimental design. For example, the effect of mora counts could have been cumulative but linear, instead of being sigmoid. Indeed, the effect of mora counts in the existing Pokémon names actually looks to be more linear than sigmoid (see the Appendix).

Or, the results could have been non-cumulative. This point is related to another important aspect of sound symbolism; namely, its stochastic nature (Dingemanse 2018; Kawahara et al. 2019). It was not the case, for example, that there was a "length threshold" such that any names shorter than that threshold were judged to be pre-evolution names. Neither was it the case that the presence of a voiced obstruent made the names post-evolution names 100% of the time. Instead, both mora counts and voiced obstruents gradiently increased the probabilities of each name being judged as a post-evolution name.[8] At a more general level, Gigerenzer & Gaissmaier (2011) discuss several cases in which when people make decisions, they take "a fast and frugal" decision heuristics—they take into account the most important information only and disregard other information (just as OT with strict domination would do). If people had applied such a heuristics decision making approach in the current experiment, the results would have been neither stochastic nor cumulative. See §5 for further discussion.

Finally, to conclude this section, the stochastic nature of sound symbolism may remind formal phonologists of a growing body of evidence that many if not all phonological generalizations have to be stated in a stochastic or probabilistic way; e.g., some structures tend to be preferred over others, and some alternations occur with different probabilities in different environments (see §1.1). The current results thus reveal an interesting parallel between phonological patterns and sound symbolic patterns.

4 An analysis with Maximum Entropy Model

The experimental results reported in §3 appear to instantiate wug-shaped curves, a quantitative signature of MaxEnt grammar model; the results thus appear to lend support for this grammatical model from the perspective of sound symbolism. To provide more concrete support for the MaxEnt grammar model, this section develops an analysis of the experimental results using MaxEnt grammar with Optimality Theoretic constraints.[9] One fundamental idea behind this analysis is that

[8] A very small subset of the participants (17 out of 730, ca. 2%) showed categorical patterns in which they assigned names of all mora lengths to either pre-evolution characters or post-evolution characters 100% of the time, and hence showed no intermediate responses. No participants judged names with voiced obstruents to be post-evolution character names 100% of the time.

[9] Wug-shaped curves are also predicted under some versions of Noisy Harmonic Grammar, depending upon precisely how noise is added to the calculation of harmony scores (Hayes 2017, 2020). Since wug-shaped curves do not themselves distinguish MaxEnt grammar from Noisy Harmonic Grammar (both are versions of Harmonic Grammar), I focus on the former framework. The distinction that the current study bears on is ranking (lexicographic) vs. weighting (numeric) rather than MaxEnt Harmonic Grammar vs. Noisy Harmonic Grammar.

Another quantitative framework that can model stochastic generalizations in phonology is the inverted-exponential model proposed by Guy (1991), which derives different probabilities by positing that an optional phonological rule can apply different numbers of times in different morphological conditions. I set this analysis aside in the paper for sound symbolic connections—mapping from sounds and meanings—can be understood as involving essentially the same mechanism as the phonological input-output mappings (Kawahara et al. 2019; Kawahara 2020). The model makes use of the sort of constraints that are familiar from the OT tradition (Prince & Smolensky 1993/2004). To underscore the parallel between phonological analyses and sound symbolic analyses, I adapt a particular formalism to formulate constraints in the OT tradition (McCarthy 2003).

4.1 A brief review of MaxEnt grammar

This section briefly reviews how MaxEnt grammar works in the context of linguistic analyses. (This section repeats section X of Author 2020; readers who are familiar with MaxEnt grammar can safely skip this section.) MaxEnt grammar is similar to OT (Prince & Smolensky 1993/2004) in that a set of candidates is evaluated against a set of constraints. Unlike OT, however, constraints are weighted rather than ranked. Consider a toy example in (5). The set of candidates that are evaluated are listed in the leftmost column. The top row lists the set of constraints that are relevant, and each constraint is assigned a particular weight. The tableau shows the violation profiles of each constraint—which candidate violates which constraints how many times.

(5) A toy example tableau of MaxEnt grammar

	Constraint A Weight = 3	Constraint B Weight = 2	Constraint C Weight = 1	H-score	eHarmony	*Z*	predicted percentages
Candidate 1	1			1*3=3	e^{-3}= 0.0498	0.0565	88
Candidate 2		2	1	2*2+1*1=5	e^{-5}= 0.0067	0.0565	12

Based on the constraint violation profiles, for each candidate x, its Harmony score (H-score(x)) is calculated using the formula in (1):

$$\text{H-score}(x) = \sum_{i}^{N} w_i C_i(x) \ \ (N \text{ is the number of the constraints}) \tag{1}$$

where w_i is the weight of the i-th constraint, and $C_i(x)$ is the number of times candidate x violates the i-th constraint. For example, Candidate 2 in the tableau (5) violates Constraint B twice and Constraint C once; its H-score is therefore $2 * 2 + 1 * 1 = 5$.

The H-scores are negatively exponentiated (eHarmony, e^{-H} or $\frac{1}{e^H}$: Wilson 2014), which is proportional to the probability of each candidate. Intuitively, the more constraint violations a

three reasons: (i) it is not clear how a rule-based approach can be used to model sound symbolic connections, (ii) the current probabilistic patterns have nothing to do with morphological differences, and (iii) this exponential model does not derive sigmoid curves (McPherson & Hayes 2016).

candidate incurs, the higher the H-score, and hence the lower the eHarmony (e^{-H}). Therefore, more violations of constraints lead to lower probability of that candidate. The eHarmony values are relativized against the sum of the eHarmony values of all the candidates, which is referred to as Z:

$$Z = \sum_{j}^{M} (e^{-H})_j \text{ (M is the number of the candidates)} \tag{2}$$

In the example in (5), Z is $0.0498 + 0.0067 = 0.0565$. The predicted probability of each candidate $x_j, p(x_j)$, is $\frac{eHarmony(x_j)}{Z}$.

4.2 The MaxEnt analysis of the current results

Like other analyses in OT and other related frameworks, the current MaxEnt analysis of sound symbolism consists of inputs, outputs and constraints that evaluate the mapping between these two levels of representations. The inputs are phonological forms, and the outputs are their sound symbolic meanings, here either pre-evolution character names or post-evolution character names. The set of constraints deployed in the current analysis is shown in (6).[10] The constraints deployed are essentially similar to the markedness constraints in OT in that they evaluate the wellformedness of output structures. The format of the constraints follows that of McCarthy (2003).

(6) Constraints deployed in the current analysis

a. *LONGPRE: Assign a violation mark for each mora in a pre-evolution character name.

b. *VCDPRE: Assign a violation mark for each voiced obstruent in a pre-evolution character name.

c. *POST: Assign a violation mark for each post-evolution name.

The first constraint prevents long names from being used for pre-evolution characters. This constraint is a formal expression of "the longer the stronger" principle or "the iconicity of quantity" principle (Haiman 1980, 1984). The constraint is a single gradient/scalar constraint (Hsu & Jesney 2017; McPherson & Hayes 2016) in that it is a reflection of one principle whose violations can be assessed on a numerical scale.[11] This constraint corresponds to the scalar constraint S that was

[10] If one is concerned that notions like "pre-evolution" and "post-evolution" are too language/culture-specific to be mentioned in the OT-style constraints, they can each be replaced with "small entity" and "large entity," since Pokémon characters generally become larger after evolution. Size, together with shapes, is one semantic dimension that is most clearly signaled by sound symbolism across many languages (Sidhu & Pexman 2018).

[11] The use of a scalar constraint is not new, even in the OT research tradition. The HNUC constraint proposed by Prince & Smolensky (1993/2004) can be understood as this type of constraint, although Prince & Smolensky did not use actual numbers. See, for example, de Lacy (2006) and Gouskova (2004) who offer extensive discussion on how various phonological scales should be formally captured in the OT analyses, although they deploy a family of

used to schematically illustrate the wug-shaped curves in §1.1. The second constraint is a formal expression of the preference that names with voiced obstruents are used for post-evolution character names, and this corresponds to the perturber constraint P that was used in §1.1. The last constraint is a *STRUC constraint (Prince & Smolensky 1993/2004) that penalizes post-evolution character names in general, which corresponds to the binary constraint B discussed in §1.1. Daland (2015) recommends that we include a *STRUC constraint in a MaxEnt grammar analysis. For the current analysis we need this constraint, because there has to be some constraint that favors pre-evolution character names. All of these constraints are statistically motivated by a log-likelihood ratio test discussed below.

Hayes (2020) recommends that we conceive of constraints as evidence to make a decision about which candidate to choose. The constraints posited in (6) do precisely this: the first two constraints offer evidence to choose post-evolution names when the candidates are long (*LONGPRE) or when they contain a voiced obstruent (*VCDPRE). The last constraint helps us to choose a pre-evolution name in general. The weights associated with each constraint reflect the strengths, or cogency, of each evidence.

The MaxEnt tableaux for all types of inputs are shown in (7). The leftmost column shows each phonological form, and the second column shows how each phonological form is mapped onto two meanings: pre-evolution character names vs. post-evolution character names. The constraint violation profiles are shown in the third-fifth columns. The observed percentages of each condition, shown in the rightmost column, were taken from the grand averages obtained in the experiment. Based on the constraint profiles and the observed percentages of each output form, the optimal weights of these constraints were calculated using MaxEnt Grammar tool (Hayes et al. 2009). The weights that were obtained by this analysis are shown at the top row of the tableaux. These weights, together with the constraint profiles, allow us to calculate harmony scores, e-harmony scores, and predicted percentages, using the procedure reviewed in §4.1.

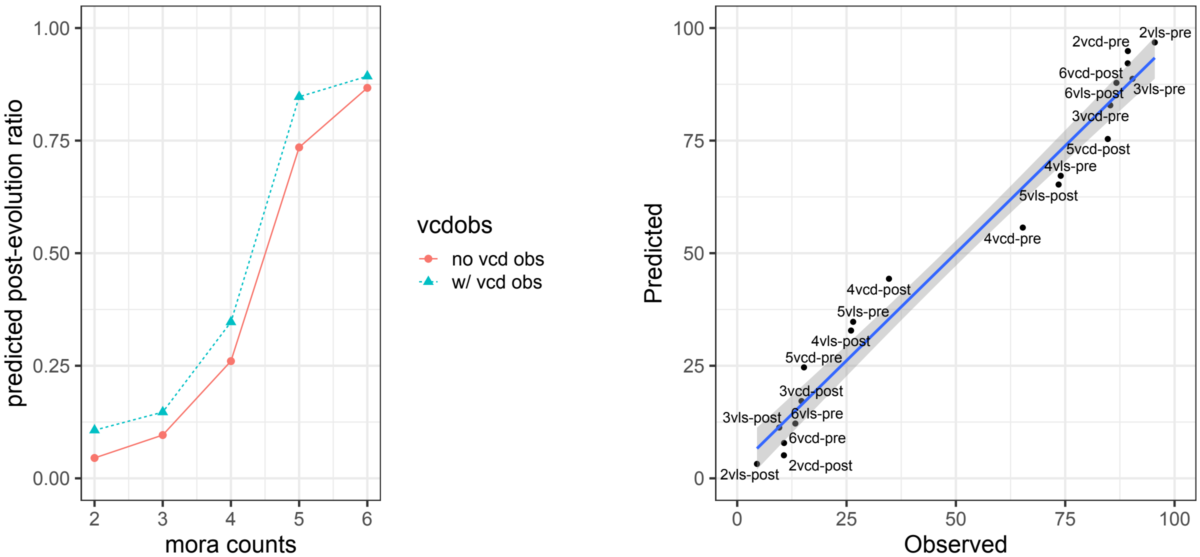
constraints instead of a single scalar constraint (see §5). See McCarthy (2003) for a review of gradient constraints in Optimality Theory and criticisms against them. See also McPherson & Hayes (2016: 149) for other examples of scalar constraints that have been used in linguistic theories.

(7) The MaxEnt Tableaux

		w = 1.35	w = 0.49	w = 1.98				
Input	Output	*LongPre	*VcdPre	*Post	H-score	eHarmony	Predicted	Observed
2 moras, vls	Pre	2			2.69	0.068	96.80	95.48
	Post			1	6.10	0.002	3.20	4.52
3 moras, vls	Pre	3			4.04	0.018	88.72	90.39
	Post			1	6.10	0.002	11.28	9.61
4 moras, vls	Pre	4			5.38	0.005	67.18	73.97
	Post			1	6.10	0.002	32.82	26.03
5 moras, vls	Pre	5			6.73	0.001	34.76	26.51
	Post			1	6.10	0.002	65.24	73.49
6 moras, vls	Pre	6			8.08	0.0003	12.18	13.29
	Post			1	6.10	0.002	87.82	86.71
2 moras, vcd	Pre	2	1		3.18	0.042	94.89	89.32
	Post			1	6.10	0.002	5.11	10.68
3 moras, vcd	Pre	3	1		4.53	0.011	82.84	85.30
	Post			1	6.10	0.002	17.16	14.70
4 moras, vcd	Pre	4	1		5.87	0.003	55.68	65.30
	Post			1	6.10	0.002	44.32	34.70
5 moras, vcd	Pre	5	1		7.22	0.001	24.64	15.27
	Post			1	6.10	0.002	75.36	84.73
6 moras, vcd	Pre	6	1		8.57	0.0002	7.84	10.71
	Post			1	6.10	0.002	92.16	89.29

We observe that the observed and the predicted values are very close to each other. To visualize the success of this MaxEnt analysis with OT-style constraints, Figure 5(a) shows the probability curves that are predicted by the MaxEnt analysis (cf. Figure 4(a)). Figure 5(b) plots the correlation between the observed and the predicted values obtained from the MaxEnt analysis, which shows a good fit between the two measures.

One general advantage of MaxEnt analyses is that it allows us to assess the necessity of each constraint using a well-established statistical method (Breiss & Hayes to appear; Hayes et al. 2012; Hayes & Jo 2020) (see also Shih 2017 for a related idea). We can do so by comparing two grammatical models—for the current analysis, we compare the full model with the three constraints and a smaller model with two of the three constraints. By removing one of the three constraints, we then have three simpler two-constraint models. We can compare the full model with each of the three simpler models by way of log-likelihood ratio tests, which would tell us whether the full model fits the data better than the simpler models to a statistically significant degree.



(a) (b)

Figure 5: (a) The probability curves predicted by the results of the MaxEnt analysis. (b) The correlation between the observed and the predicted percentages obtained from the MaxEnt analysis.

The results of these log likelihood ratio tests are shown in Table 3, which demonstrates that all three constraints are statistically motivated to explain the data—in other words, each of the constraints plays a role in the explanation of the data, in addition to what is explained by the other two constraints (see the Appendix of Breiss & Hayes to appear for further discussion of this analysis).

Table 3: The results of the log-likelihood ratio tests. The log-likelihood of the best fitting model with the three constraints was -432.3.

	Δlikelihood	$\chi^2(1)$	p
*LongPre	249.88	499.77	$<.001$
*VcdPre	4.10	8.20	$<.01$
*Post	260.81	521.63	$<.001$

Finally, a more complex model was tested with a fourth constraint, which is equivalent to the locally conjoined version of *LongPre and *VcdPre (cf. Shih 2017), but it did not improve the model fit at all. The MaxEnt grammar tool actually assigned 0 weight to the conjoined constraint. This is a welcome result, since the interaction of the effects of voiced obstruents and those of mora counts followed directly from the architecture of the MaxEnt model itself, obviating the need to posit a specific constraint to capture the interaction between the two factors (see also

Zuraw & Hayes 2017).

5 Analyses with Stochastic Optimality Theory

While Hayes (2020) as well as Zuraw & Hayes (2017) have shown that patterns with wug-shaped curves cannot be modeled well with Stochastic OT (Boersma 1998; Boersma & Hayes 2001), this section reports several attempts to fit a Stochastic OT model to the current data. In Stochastic OT, each constraint is assigned a particular ranking value, which is perturbed by a Gaussian noise at each time of evaluation. Each evaluation is computed just as in Classic OT with strict domination, predicting a single winner per each evaluation trial. The probability distributions of variable outputs are calculated over multiple evaluation cycles.

To analyze the current experimental results using Stochastic OT, first, the same data structure that was used for the MaxEnt analysis in (7) was fed to OTSoftware (Hayes et al. 2014) with the Gradual Learning Algorithm as the learning algorithm. The initial ranking values of all constraints were set to be 100 (the default value). The initial plasticity and the final plasticity were set to be 0.01 and 0.001, respectively. There were 1,000,000 learning trials, and the grammar was tested for 1,000,000 cycles in order to get the predicted probability distribution.

This learning simulation resulted in the following ranking values: *LONGPRE = 99.6, *VCDPRE = 98.2, *POST = 100.4. All the constraints were active in at least one of the evaluation trials. The problem with this analysis using Stochastic OT is that it was not able to model the effects of mora counts at all; indeed, Stochastic OT does not handle counting cumulativity effects well in general (Hayes 2020; Jäger 2007). For all the conditions without voiced obstruents, regardless of the mora counts, post-evolution candidates were predicted to win 39.5% of the time and pre-evolution candidates were predicted to win 60.5% of the time. For all the conditions with voiced obstruents, post-evolution characters were predicted to win 46.6% of the time, whereas the pre-evolution characters were predicted to win 53.4% of the time. Stochastic OT was thus able to model the effect of voiced obstruents (39.5% vs. 46.6%), which seems to reflect the actual observed post-evolution response values averaged across all the mora length conditions (40.1% vs. 46.8%). However, it was unable to learn the effects of mora counts.

The failure to model the counting cumulativity effects of mora counts stems from the fact that Stochastic OT is no different from Classic OT (Prince & Smolensky 1993/2004) at each time of evaluation. OT does not distinguish between, for example, one violation mark vs. two violation marks (* vs. **) and one violation mark vs. four violation marks (* vs. ****). Therefore, if *POST dominates *LONGPRE at a particular time of evaluation, then the pre-evolution candidate is predicted to win at that particular time of evaluation, no matter how many violations of *LONGPRE the pre-evolution candidate incurs. Similarly, if *LONGPRE dominates *POST, the post-evolution candidate wins no matter how long the pre-evolution candidate is. The number of violations simply does not add up in Classic OT or Stochastic OT, because of strict domination. For these reasons, it was not able to account for the counting cumulativity effects of mora counts.

This problem can be (partially) remedied by splitting up *LongPre into a set of separate constraints which each penalizes a pre-evolution name with a particular mora length; i.e. *LongPre3Mora, *LongPre4Mora, *LongPre5Mora, and *LongPre6Mora (see footnote 21 of McPherson & Hayes 2016 as well as Boersma 1998, de Lacy 2006, and Gouskova 2004).[12] A new learning simulation was run with the same parameter settings. With the expanded set of constraints, it learned the following values: *LongPre3Mora = 97.2, *LongPre4Mora = 99.7, *LongPre5Mora = 103.7, *LongPre6Mora = 103.2, *VcdPre = 98.7, *Post =101.6. When we plot the predicted values obtained based on these ranking values, the Stochastic OT analysis created two separate sigmoid curves for the two voicing conditions, as shown in Figure 6. However, these sigmoids formed the "open jaw" pattern, in which we observe the convergence of the two curves at one end and divergence between the two curves at the other end, and the difference between the two curves grow monotonically toward the left end (compare this pattern with the result of the MaxEnt analysis, Figure 5(a)).

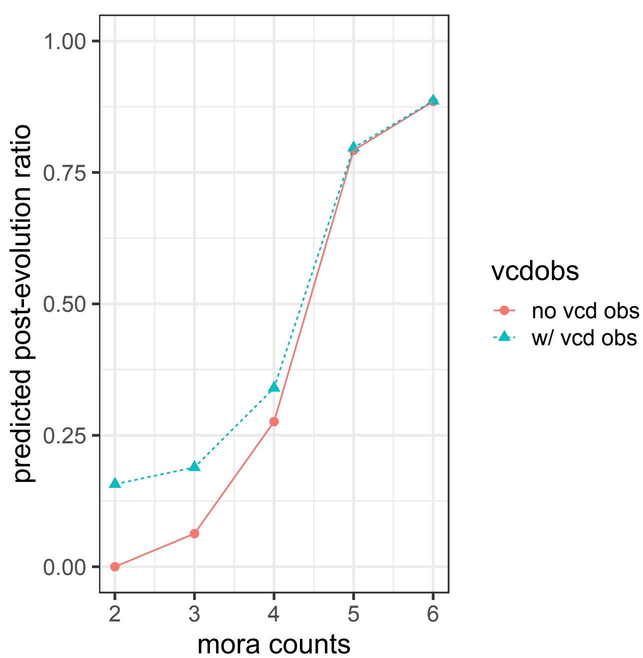


Figure 6: The probability patterns predicted by the GLA, when *LongPre is split into a family of different constraints.

The problem comes from the fact that the ranking value of the perturber constraint—

[12] *LongPre2Mora is vacuous and hence unnecessary.

*VcdPre—is too far away from the ranking values of *LongPre5Mora, *LongPre6Mora, and *Post, essentially resulting in "near strict domination." As a result, *VcdPre does not have a visible influence on 5-mora long names and 6-mora long names. This problem is a general one (Hayes 2020): the perturber constraint can have one ranking value, and hence has a hard time exerting its influence across the whole x-axis range, when it is placed near the lower end of the constraint value continuum.

This aspect of Stochastic OT was identified by Zuraw & Hayes (2017) in their quantitive analysis of French liaison. Indeed the general constraint profiles for the current analysis are similar to their analysis of French. The set of *LongPreX constraints and *VcdPre are synergistic in that they both favor post-evolution names, and the other constraint, *Post, favors pre-evolution names. Zuraw & Hayes (2017: 530) offer an intuitive explanation of how this type of constraint violation profile results in a pattern like the one in Figure 6. Citing an unpublished work by Giorgio Magri, they characterize this pattern as "[two curves] will be uniformly converging in one direction and diverging in the other...where [the] differences...grow monotonically toward the right of the plot" (p. 530). The pattern in Figure 6 looks precisely like what Zuraw & Hayes describe, with a very minor difference that the divergence is largest on the left (rather than on the right) of the plot in Figure 6.

Bruce Hayes pointed out (p.c.) that Stochastic OT may perform better if the perturber P constraint (*VcdPre) is reformulated in such a way that it agrees with the binary constraint B (*Post). Following this suggestion, I reformulated *VcdPre as a constraint that penalizes a post-evolution name unless it starts with a voiced obstruent, as in (8).

(8) The new perturber constraint

a. StartWithVcdObsPost: Assign a violation mark for each post-evolution name which does not start with a voiced obstruent.

This new constraint reminds us of a positional markedness constraint, which for example, requires a low-sonority segment in onset positions (Smith 2002). The ranking values that the GLA learned with this new perturber constraint are: *LongPre3Mora = 64.6, *LongPre4Mora = 65.9, *LongPre5Mora = 69.9, *LongPre6Mora = 70.4, StartWithVcdObsPost = 66.6, *Post = 67.5, which yields two sigmoid curves shown in Figure 7.

The two curves look better separated in Figure 7 than in Figure 6, because the ranking value of the perturber constraint, StartWithVcdObsPost, is placed in the middle range of the whole constraint ranking continuum in this second analysis. We can see that the difference between the two sigmoid curves is largest for 4-mora long names, and the difference becomes smaller as the name gets shorter or longer. If we had a larger range of x-axis values, the separation of the two curves should eventually disappear at both ends, predicting a "cucumber curve."

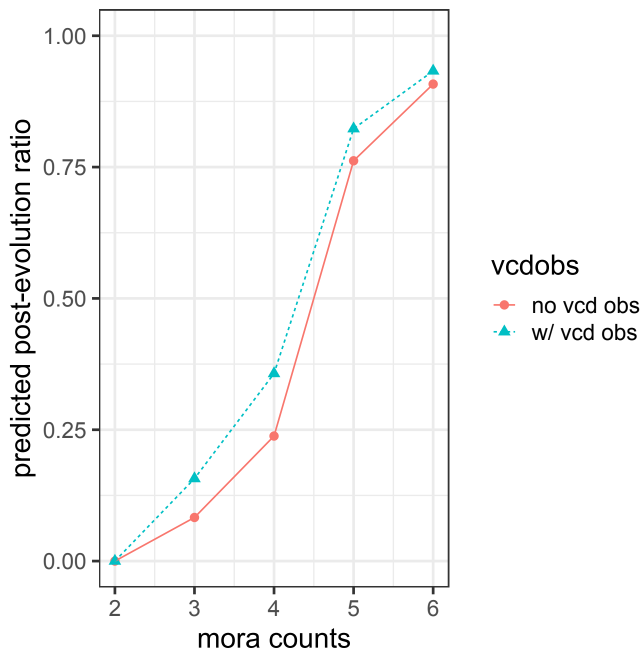


Figure 7: The probability patterns predicted by the GLA with the new perturber constraint in (8).

In short, Stochastic OT requires that we split the scalar constraint (*LONGPRE) into a set of multiple constraints (Boersma 1998; McPherson & Hayes 2016) to account for the counting cumulativity effect. In addition, it also requires a particular relationship between the perturber constraint P and the binary constraint B. Furthermore, the problem identified by Hayes (2020) is a general one: the perturber constraint can have only one ranking value, so its influence is localized. When it is placed in the middle of the ranking value continuum, as in Figure 7, we may observe a global separation of the two curves, as long as the x-axis range is sufficiently limited. If the x-axis has a wider range, however, it is predicted that the perturber cannot influence the whole x-axis range.

Finally, to conclude this section on Stochastic OT, the log-likelihood—a measure of deviation between the observed data and the model predictions—of the two Stochastic OT models are calculated, which were -459.6 and -546.8.[13] These values are lower than that of the MaxEnt model (-432.3) (log-likelihood values that are closer to 0 are better). Moreover, the Stochastic OT models and the MaxEnt model differ in terms of the number of free parameters (i.e. the number of constraints): 6 vs. 3. Therefore, the AIC (Akaike Information Criterion) is calculated for each model, yielding 931.2, 1105.5 (the two Stochastic OT models) and 870.7 (the MaxEnt model)—a model

[13]The use of these statistical measures—log-likelihood and AIC—to compare grammatical models is inspired by Shih (2017) and Zuraw & Hayes (2017). When I calculated log-likelihood, since we cannot take log of 0, it was replaced with $\frac{1}{10^6}$.

with lower AIC makes a better prediction about the data.[14]

6 Concluding remarks

6.1 Summary

We started with a question that was recently raised by Hayes (2020). In order to compare various stochastic linguistic models, thinking abstractly about what quantitative predictions these theories make is useful. Taking MaxEnt as an example, Hayes (2020) shows that we should be able to draw wug-shaped curves under certain circumstances. The current experiment addressed this prediction in the domain of sound symbolism, and has shown that we can indeed draw wug-shaped curves when certain variables are systematically manipulated for the judgment of evolvedness in Pokémon names. To the extent that wug-shaped curves are typical quantitative signatures of MaxEnt, it shows that MaxEnt is a grammatical framework that is suited to model sound symbolic patterns in natural languages (Kawahara et al. 2019; Kawahara 2020). To put the results in a more theory-neutral fashion, Japanese speakers take into account different sources of quantitative information (mora counts and a voiced obstruent) in a cumulative way, more specifically, in the way that is naturally predicted by MaxEnt.

Viewed from a slightly different, albeit related perspective, the current experiment addressed the general issue of cumulativity in sound symbolism. The effects of mora counts instantiated a clear case of counting cumulativity, in that each mora count contributes to the judgment of evolvedness in an additive way. The overall patterns also instantiate a ganging-up cumulativity in that the effects of voiced obstruents and those of mora counts additively contributed to the judgement of evolvedness. Such cumulative patterns are natural consequences of the MaxEnt grammar.

6.2 Phonological patterns and sound symbolic patterns

To the extent that MaxEnt is a useful tool to model phonological patterns including both input-output mappings and surface phonotactics judgment patterns, as many previous studies have already shown (e.g. Hayes & Wilson 2008; McPherson & Hayes 2016; Zuraw & Hayes 2017), the overall results point to an intriguing parallel between phonological patterns and sound symbolic patterns. Traditionally, sound symbolism barely received serious attention from formal phonologists (Alderete & Kochetov 2017; Kawahara to appear). However, the current results suggest no inherent differences between sound-meaning mappings and phonological input-output mappings

[14] There is one caveat: *LONGPRE in the MaxEnt model can take a wider range of values than the set of *LONGPREXMORA constraints in the Stochastic OT model, because the former is a scalar constraint and the latter is a binary constraint.

(as well as wellformed judgments of surface phonotactic patterns). Phonological patterns and sound symbolic patterns share two important properties—stochasticity and cumulativity—both of which naturally follow from the MaxEnt grammar. This conclusion in turn implies that sound symbolism may not be as irrelevant to formal phonological theory as has been assumed in the past, echoing the claim recently made by several researchers (Alderete & Kochetov 2017; Jang 2019; Kawahara to appear; Kumagai 2019; Shih 2019).[15]

If this hypothesis is on the right track, one question that arises is how closely these two systems are related to one another. I am unable to offer a full fledged answer to this general question, but can partially address it by asking a more concrete question: whether sound symbolic constraints of the sort that are used in the current paper can coerce phonological changes. Alderete & Kochetov (2017) argue that such patterns do exist. Patterns of expressive palatalization, often found in baby-talk registers, exhibit properties that are different from “regular” phonological palatalization processes; e.g. the former can target all the coronal segments in a word without a clear trigger like a high front vowel (e.g. /o**s**akana-**s**aN/ → [o**ɕ**akana-**ɕ**aN] ‘fish-y’ in Japanese). They thus argue that expressive palatalization patterns are caused by sound symbolic requirements, instead of constraints that are purely phonological, and propose a family of EXPRESS(X) constraints, which demands that a particular meaning is expressed by a particular sound. Expressive palatalization may thus instantiate a case in which sound-symbolic constraints coerce phonological changes. See Jang (2019) and Kumagai (2019) for other possible examples of this sort.

6.3 Addressing domain-specificity of phonological knowledge

If MaxEnt is suited to model phonological patterns and sound symbolic patterns, it raises an important question regarding the domain-specificity of phonological (and sound symbolic) knowledge. MaxEnt is not a theory of phonology *per se*; indeed, it is equivalent to a general statistical device known as a log-linear model or multinomial logistic regression (Breiss & Hayes to appear; Jurafsky & Martin 2019; Shih 2017). It was proposed as a model of human cognition in general (Smolensky 1986), and is widely used in natural language processing (Berger et al. 1996). Hayes (2020: 2) proposes to characterize MaxEnt as a “mathematically close embodiment of common sense.” In this view, MaxEnt is general a tool to model decision-making procedures (though cf. Gigerenzer & Gaissmaier 2011 for other decision making strategies). If this is the case, then, modeling phonology—at least the mapping between two levels of representation—may not require a domain-specific device (i.e. Universal Grammar) (see Archangeli & Pulleyblank 2015; Carr 2016; Mielke 2008 for related proposals to reduce/eliminate the role of UG in phonological theory).

[15]To this, we can also add studies of metrics conducted by phonologists. To the extent that metrics can be a topic of phonological inquiry, which in fact they have been, I do not see any fundamental reason to exclude sound symbolism from phonological inquiry either. See e.g. Hayes et al. (2012) for a MaxEnt analysis of metrics.

There still remains a good possibility that we need a domain-specific device, for instance, in order to derive a universal set of distinctive features or, relatedly, a set of constraints in *CON*. I cannot be conclusive here, but there are proposals that both features (Boersma 1998; Mayer & Daland 2019; Mielke 2008) and constraint sets (Hayes 1999; Hayes & Wilson 2008) may be deducible from the learning data and phonetic considerations, although one can still argue that such induction mechanisms have to be domain-specific. The parallel between phonological patterns and sound symbolic patterns identified in this work may shed new light on this question of domain-specificity, because sound symbolism has been argued to be a specific instance of a domain-general synesthetic cross-modal perception (Bankieris & Simner 2015; Cuskley & Kirby 2013; Ramachandran & Hubbard 2001; Spence 2011).[16] In short, to the extent that MaxEnt provides a useful tool to model linguistic patterns—be they phonological or sound symbolic—we should keep thinking about to what extent a domain-specific device is necessary to account for phonological and sound symbolic patterns.

Indeed, the general MaxEnt framework may offer a useful tool with which to address the question of how domain-general factors and domain-specific factors interact to shape our linguistic knowledge. Recall that in the MaxEnt framework, we can statistically access the necessity of each factor by way of log-likelihood ratio tests. Once we can identify domain-specific factors and domain general factors in action, we can tease apart the contributions of these two types of factors—and their interactions—in a quantitative fashion. Put slightly differently, MaxEnt may offer a quantitive means to examine how competence and performance function together to shape linguistic patterns. Domain-specificity of phonological knowledge should not be taken for granted, and explicit discussion on this point can only help further our understanding of the nature of phonological knowledge.

6.4 Some remarks on Pokémonastics

In addition to offering these theoretical insights, the current study highlights two methodological advantages of the Pokémonastics research paradigm, which were implicitly mentioned in the previous Pokémonastics studies. One is that we were able to obtain a very large data set over a short period of time. It is rare to be able to analyze the data from 730 participants for a controlled linguistic experiment—thanks to this large number of N, we can be confident that our estimates are

[16]For views that sound symbolism may operate at the level of distinctive features, see Hamano (1998), Jakobson (1978), Kumagai & Kawahara (2020) and Nobile (2015). The set of constraints that regulate sound-symbolic mappings is unlikely to be universal, as certain sound symbolic patterns are language-specific (Akita 2015; Bremner et al. 2013; Imai & Kita 2014; Iwasaki et al. 2007; Saji et al. 2019); perhaps the clearest example is the *gl-* sequence in English which is used in many words that are related to the notion of light (Bergen 2004). There are patterns of sound symbolism that are demonstrably universal, however, and such patterns are likely to be grounded in articulatory and/or acoustic properties of sounds. Hence the constraints that regulate such universal, phonetically grounded connections can potentially be induced by a mechanism proposed by Hayes (1999).

rather reliable.

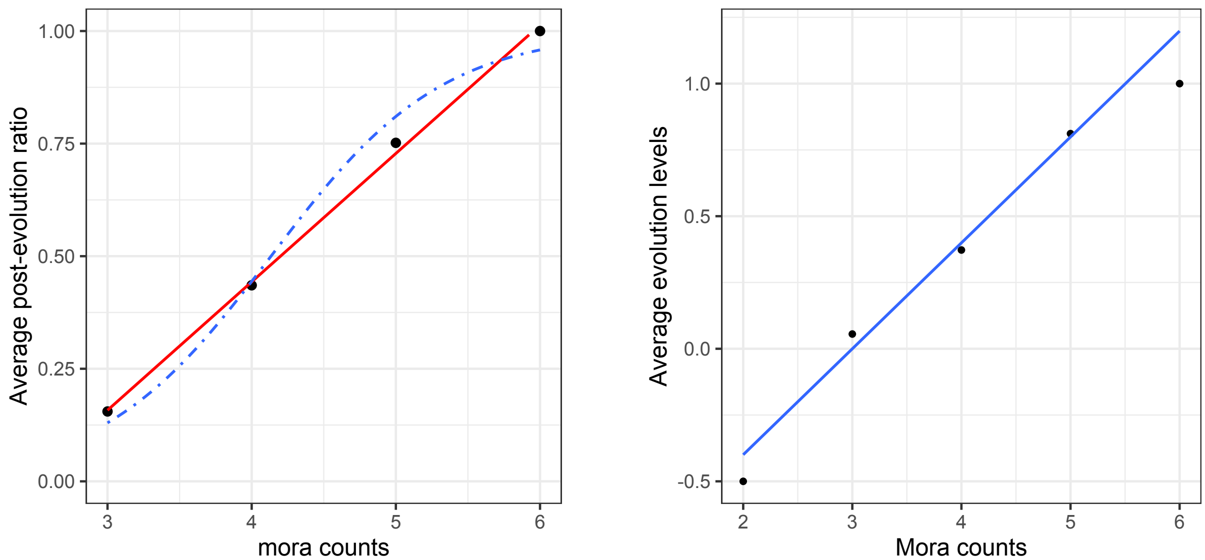
Furthermore, all the participants took part in the experiment without any compensation. They were willing to help out the experiment probably because the experiment was about something that they love; i.e. Pokémon. At the end of the experiment, many participants reported that they are very eager to hear what the experiment was about and how the results would turn out. Therefore, this Pokémonastics research may have potential to contribute to popularizing linguistics. Being able to show the general public that we can learn much about our phonological knowledge through Pokémon can be existing news.

6.5 The final conclusion

Finally, I would like to close this paper by putting forward the following methodological thesis: phonological theory can inform research on sound symbolism. While sound symbolism is currently studied very actively, most of such research is being conducted by psychologists, cognitive scientists and cognitive linguists, and few formal phonologists pay serious attention to sound symbolism. However, the current research, inspired by Hayes (2020), has revealed important aspects of sound symbolism—their cumulative nature and how they can be modeled by MaxEnt. Hayes (2020) offers an abstract "top-down" approach, which takes one theory seriously and considers its consequences. If it weren't for this approach, I would not have conducted the current experiment. More generally speaking, then, phonological theory can inform research on sound symbolism in important ways. In turn, I hope to have shown that sound symbolism may offer a new testing ground to examine the cumulative nature of linguistic patterns, and in this sense, studies of sound symbolism can inform phonological theories as well. All in all, I hope to have shown with the current case study that phonological theories and research on sound symbolism both can and should mutually inform one another.

Appendix: Patterns in the existing names

One may wonder how the existing patterns of Pokémon names behave with respect to the issues discussed in the main text. To address this question, I have used the dataset compiled by Kawahara et al. (2018), which includes all the data up to those characters included in the 6th generation, for which there are about 700 characters. Some Pokémon characters do not undergo evolution at all, and those were removed from the analysis. Some other Pokémons were "baby" Pokémons, which were introduced as a pre-evolution version of an already existing character at a later series. While there are not so many of them (N=16), they were also excluded. Pokémon can undergo evolution twice, and hence as long as they are evolved once, they were counted as post-evolution. There was only one name that is 6-moras long, so this data point has to be interpreted



(a) (b)

Figure 8: (a) The relationship between the mora length and the averaged probabilities of post-evolution in the existing names, in which evolution is coded as a binary variable. (b) The correlation between the number of moras and the average evolution levels, in which evolution is coded as a four-way variable (see text).

with caution. The total N was 585 in this analysis.

In order to examine whether we observe a sigmoid curve, Figure 8(a) plots the relationship between the mora length and the averaged probabilities of post-evolution. Both a linear function (solid, red) and a logistic curve (blue, dotted) are superimposed. There does not seem to be a good reason to believe that the s-curve fits the data better than the linear function. The analysis reported by Kawahara et al. (2018), which makes use of a four-way distinction in terms of evolution—baby Pokémon, no-evolution, evolved once, and evolved twice (coded as -1, 0, 1, 2, respectively)—likewise shows a similar linear trend, as shown here as Figure 8(b) (edited by the author based on their Figure 7).

Based on the inspection of Figure 8, we may tentatively conclude that s-curves (and hence wug-shaped curves) emerge in experimental settings, despite the absence of such patterns in the existing names.

Acknowledgements

The paper would not have materialized without the help of many people, who commented on various incarnations of this project, asked insightful questions, and/or offered analytical help and suggestions. They include XXX [ANONYMIZED FOR THE SAKE OF ANONYMOUS REVIEW BUT YOUR NAME CAN BE ADDED HERE]. I would like to thank Kero for putting the current online experiment on his blog, which helped me to gather the data in a very efficient manner. This project is supported by the grant YYY. All remaining errors are mine.

References

Akita, Kimi. 2015. Sound symbolism. In Jan-Ola Östman & Jef Verschueren (eds.), *Handbook of pragmatics, installment 2015*, Amsterdam and Philadelphia: John Benjamins.

Alderete, John. 1997. Dissimilation as local conjunction. In Kiyomi Kusumoto (ed.), *Proceedings of the North East Linguistics Society 27*, 17–31. Amherst: GLSA.

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

Anttila, Arto, Scott Borgeson & Giorgio Magri. 2019. Equiprobable mappings in weighted constraint grammars. *SIGMORPHON 2019* .

Anttila, Arto & Giorgio Magri. 2018. Does maxent overgenerate? Implicational universals in Maximum Entropy Grammar. *Proceeings of the Annual Meeting of Phonology 2017* .

Archangeli, Diana & Douglas Pulleyblank. 2015. Phonology without universal grammar. *Frontiers in Psychology* .

Bankieris, Kaitlyn & Julia Simner. 2015. What is the link between synaesthesia and sound symbolism? *Cognition* 136. 186–195.

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278.

Bergen, Benjamin K. 2004. The psychological reality of phonaesthemes. *Language* 80. 290–311.

Berger, Adam L, Stephen A. Della Pietra & Vincent J. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics* 22. 39–71.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14. 150–177.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interaction between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32. 45–86.

Boersma, Paul & Joe Pater. 2016. Convergence properties of a Gradual Learning Algorithm for Harmonic Grammar. In John J. McCarthy & Joe Pater (eds.), *Harmonic Grammar and Harmonic Serialism*, 389–434. London: Equinox.

Breiss, Canaan. 2019. Cumulativity by default in phonotactic learning: Experimental and computational evidence. Ms. University of California, Los Angeles.

Breiss, Canaan & Adam Albright. 2020. When is a gang effect more than the sum of its parts. Talk presented at LSA 2020.

Breiss, Canaan & Bruce Hayes. to appear. Phonological markedness effects in sentential formation. *Language* .

Bremner, Andrew J., Serge Caparos, Jules Davidoff, Jan de Fockert, Karina J. Linnell & Charles Spence. 2013. "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition* 126. 165–172.

Bresnan, Joan & Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua* 118(2). 245–259.

Carr, Phillip. 2016. Universal grammar and syntax/phonology parallelism. *Lingua* 116(5). 634–656.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam & Moris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.

Coetzee, Andries W. 2012. Variation: Where laboratory and theoretical phonology meet. In Abigail Cohn, Cécile Fougeron & Marie K. Huffman (eds.), *Oxford handbook of laboratory phonology*, 62–75. Oxford: Oxford University Press.

Coetzee, Andries W. & Shigeto Kawahara. 2013. Frequency biases in phonological variation. *Natural Language and Linguistic Theory* 30(1). 47–89.

Coetzee, Andries W. & Joe Pater. 2011. The place of variation in phonological theory. In John A. Goldsmith, Jason Riggle & Alan Yu (eds.), *The handbook of phonological theory, 2nd edition*, 401–431. Oxford: Blackwell-Wiley.

Cohn, Abigail. 2006. Is there gradient phonology? In Gisbert Fanselow, Caroline Fery, Matthias Schlesewsky & Ralf Vogel (eds.), *Gradience in grammar: Generative perspectives*, 25–44. Oxford: Oxford University Press.

Crowhurst, Megan. 2011. Constraint conjunction. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 1461–1490. Oxford: Blackwell-Wiley.

Cuskley, Christine & Simon Kirby. 2013. Synesthesia, cross-modality, and language evolution. In Julia Simner & Edward Hubbard (eds.), *Oxford handbook of synesthesia*, Oxford: Oxford University Press.

Daland, Robert. 2015. Long words in maximum entropy phonotactic gammars. *Phonology* 32(3). 353–383.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis & Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* 28(2). 197–234.

de Lacy, Paul. 2006. *Markedness: Reduction and preservation in phonology*. Cambridge: Cambridge University Press.

Dingemanse, Mark. 2018. Redrawing the margins of language: Lessons from research on ideophones. *Glossa* 3(1). 4, doi:org/10.5334/gjgl.444.

D'Onofrio, Annette. 2014. Phonetic detail and dimensionality in sound-shape correspondences: Refining the *bouba-kiki* paradigm. *Language and Speech* 57(3). 367–393.

Farris-Trimble, Ashley. 2008. *Cumulative faithfulness effects in phonology*: Indiana University Doctoral dissertation.

Gigerenzer, Gerd & Wolfgang Gaissmaier. 2011. Heuristic decision making. *Annual Review of Psychology* 62. 451–482.

Goldwater, Sharon & Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Workshop on Variation within Optimality Theory* 111–120.

Gouskova, Maria. 2004. Relational markedness in OT: The case of syllable contact. *Phonology*

21(2). 201–250.

Guy, Gregory. 1991. Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change* 3. 1–22.

Guy, Gregory. 2011. Variability. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 2190–2213. Oxford: Blackwell-Wiley.

Haiman, John. 1980. The iconicity of grammar: Isomorphism and motivation. *Language* 56(3). 515–540.

Haiman, John. 1984. *Natural syntax: Iconicity and erosion*. Cambridge: Cambrdige University Press.

Halle, Morris. 1978. Knowledge unlearned and untaught: What speakers know about the sounds of their language. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic theory and psychological reality*, 294–303. Cambridge: MIT Press.

Hamano, Shoko. 1998. *The sound-symbolic system of Japanese*. Stanford: CSLI Publications.

Hamano, Shoko. 2013. Hoogen-ni okeru giongo-gitaigo-no taikeiteki-kenkyuu-no igi. In Kazuko Shinohara & Ryoko Uno (eds.), *Chikazuku oto-to imi: Onomatope kenkyuu-no shatei*, Tokyo: Hitsuzi Syobo.

Hayes, Bruce. 1999. Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. In Michael Darnell, Edith Moravscik, Michael Noonan, Frederick Newmeyer & Kathleen Wheatly (eds.), *Functionalism and formalism in linguistics, vol. 1: General papers*, 243–285. Amsterdam: John Benjamins.

Hayes, Bruce. 2017. Varieties of noisy harmonic grammar. *Proocedings of AMP* .

Hayes, Bruce. 2020. Assessing grammatical architectures through their quantitative signatures. Talk presenetd at BLS.

Hayes, Bruce & Jinyoung Jo. 2020. Balinese stem phonotactics and the subregularity hypothesis. ms. University of California, Los Angeles.

Hayes, Bruce & Zsuzsa Londe. 2006. Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology* 23. 59–104.

Hayes, Bruce, Bruce Tesar & Kie Zuraw. 2014. Otsoft 2.5. software package, `http://www.linguistics.ucla.edu/people/hayes/otsoft/`.

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

Hayes, Bruce, Colin Wilson & Ben George. 2009. Maxent grammar tool. Software. `http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/`.

Hayes, Bruce, Colin Wilson & Anne Shisko. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88(4). 691–731.

Hinton, Leane, Johanna Nichols & John Ohala. 2006. *Sound symbolism, 2nd edition*. Cambridge: Cambridge University Press.

Hsu, Brian & Karen Jesney. 2017. Loanword adaptation in Québec French: Evidence for weighted scalar constraints. *Proceedings of the 34th West Coast Conference on Formal Linguistics* 249–258.

Imai, Mutsumi & Sotaro Kita. 2014. The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651).

Ito, Junko. 1990. Prosodic minimality in Japanese. In Michael Ziolkowski, Manual Noske & Karen Deaton (eds.), *Proceedings of Chicago Linguistic Society: Parasession on the syllable in* *phonetics and phonology*, 213–239. Chicago: Chicago Linguistic Society.

Ito, Junko & Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17. 49–73.

Ito, Junko & Armin Mester. 2003. *Japanese morphophonemics*. Cambridge: MIT Press.

Iwasaki, Noriko, David P. Vinson & Gabriella Vigiliocco. 2007. What do English speakers know about *gera-gera* and *yota-yota*? A cross-linguistic investigation of mimetic words for laughing and walking. *Japanese Language Education Around the Globe* 17. 53–78.

Jaeger, Florian T. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59. 434–446.

Jäger, Gerhad. 2007. Maximum Entropy Models and Stochastic Optimality Theory. In Joan W. Bresnan (ed.), *Architectures, rules, and preferences: Variations on themes*, 467–479. CSLI.

Jäger, Gerhad & Anette Rosenbach. 2006. The winner takes it all—almost: Culmulativity in grammatical variation. *Linguistics* 44(5). 937–971.

Jakobson, Roman. 1978. *Six lectures on sound and meaning*. Cambridge: MIT Press.

Jang, Hayeun. 2019. How cute do I sound?: The iconic function of segmental alternation in Korean baby-talk register, *aegyo*. Ms. University of Southern California.

Jurafsky, Daniel & James H. Martin. 2019. *Speech and language processing (3rd edition, draft).* `https://web.stanford.edu/~jurafsky/slp3/`.

Kawahara, Shigeto. 2017. *Introducing Phonetics through Sound Symbolism*. Tokyo: Hitsuzi Syobo.

Kawahara, Shigeto. 2020. Cumulative effects in sound symbolism. Ms. Keio University.

Kawahara, Shigeto. to appear. Sound symbolism and theoretical phonology. *Language and Linguistic Compass* .

Kawahara, Shigeto, Hironori Katsuda & Gakuji Kumagai. 2019. Accounting for the stochastic nature of sound symbolism using Maximum Entropy model. *Open Linguistics* 5. 109–120.

Kawahara, Shigeto & Gakuji Kumagai. to appear. What voiced obstruents symbolically represent in Japanese: Evidence from the Pokémon universe. *Journal of Japanese Linguistics* .

Kawahara, Shigeto, Atsushi Noto & Gakuji Kumagai. 2018. Sound symbolic patterns in Pokémon names. *Phonetica* 75(3). 219–244.

Kawahara, Shigeto, Michinori Suzuki & Gakuji Kumagai. 2020. Sound symbolic patterns in Pokémon move names in Japanese. *ICU Working Papers in Linguistics 10. Festschrift for Prof. Junko Hibiya in the occasion of her retirement from ICU* 17–30.

Kellar, Frank. 2006. Linear Optimality Theory as a model of gradience in grammar. In Gisbert Fanselow, Caroline Féry, Ralf Vogel & Matthias Schlesewsky (eds.), *Gradience in grammar: Generative perspectives*, 270–287. Oxford: Oxford University Press.

Kisseberth, Charles. 1970. The treatment of exceptions. *Papers in Linguistics* 2. 44–58.

Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(2). 199–244.

Kumagai, Gakuji. 2019. A sound-symbolic alternation to express cuteness and the orthographic Lyman's Law in Japanese. *Journal of Japanese Linguistics* 35(1). 39–74.

Kumagai, Gakuji & Shigeto Kawahara. 2020. How abstract is sound symbolism? Labiality and diaper names in Japanese [in Japanese]. *Journal of the Linguistic Society of Japan* 157.

Leben, Will. 1973. *Suprasegmental phonology*: MIT Doctoral dissertation.

Liberman, A. M., K. S. Harris, H. S. Hoffman & B. C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*

54(5). 358–368.

Martin, Samuel. 1962. Phonetic symbolism in Korean. In N. Poppe (ed.), *American studies in Uralic and Altaic linguistics*, Indiana University Press.

Mayer, Connor & Robert Daland. 2019. A method for projecting features from observed sets of phonological classes. *Linguistic Inquiry* Online first.

McCarthy, John J. 1983. Phonological features and morphological structure. In J. Richardson, M. Marks & A. Chukerman (eds.), *Proceedings from the parasesion on the interplay of phonology, morphology and syntax*, 135–161. Chicago: CLS.

McCarthy, John J. 2003. OT constraints are categorical. *Phonology* 20(1). 75–138.

McCarthy, John J. & Alan Prince. 1986. Prosodic morphology. Ms., University of Massachusetts and Rutgers University.

McPherson, Laura. 2016. Cumulativity and ganging in the tonology of Awa suffixes. *Language: Phonological Analysis* 92(1).

McPherson, Laura & Bruce Hayes. 2016. Relating application frequency to morphological structure: The case of Tommo So vowel harmony. *Phonology* 33. 125–167.

Mielke, Jeff. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Myers, Scott. 1997. OCP effects in Optimality Theory. *Natural Language and Linguistic Theory* 15(4). 847–892.

Nishimura, Kohei. 2006. Lyman's Law in loanwords. *On'in Kenkyu [Phonological Studies]* 9. 83–90.

Nobile, Luca. 2015. Phonemes as images: An experimental inquiry into shape-sound symbolism applied to the distinctive features of French. In Masako Hiraga, William Herlofsky, Kazuko Shinohara & Kimi Akita (eds.), *Iconicity: East meets west*, 71–91. John Benjamins.

O'Hara, Charlie. 2017. How abstract is more abstract? Learning abstract underlying representations. *Phonology* 34(2). 324–345.

Otake, Takashi, Giyoo Hatano, Anne Cutler & Jacques Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32. 258–278.

Paster, Mary. 2019. Phonology counts. *Radical* 1. 1–61.

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33. 999–1035.

Pierrehumbert, Janet B. 2001. Stochastic phonology. *GLOT* 5. 1–13.

Pizzo, Presley. 2015. *Investigating properties of phonotactic knowledge through web-based experimentation*: University of Massachusetts, Amherst Doctoral dissertation.

Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt & Michael Becker. 2010. Harmonic grammar with linear programming: From linear systems to linguistic typology. *Phonology* 27(1). 1–41.

Prince, Alan & Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden and Oxford: Blackwell.

Ramachandran, Vilayanur S. & Edward M. Hubbard. 2001. Synesthesia–a window into perception, thought, and language. *Journal of Consciousness Studies* 8(12). 3–34.

Saji, Noburo, Kimi Akita, Katerina Kantartzis, Sotaro Kita & Mutsumi Imai. 2019. Cross-linguistically shared and language-specific sound symbolism in novel words elicited by locomotion videos in Japanese and English. *PloS ONE* 14(7). e0218707, doi.org/10.1371/journal.pone.0218707.

Schütze, Carlson. 1996. *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Shih, Stephanie S. 2017. Constraint conjunction in weighted probabilistic grammar. *Phonology* 243–268.

Shih, Stephanie S. 2019. The contribution of sound symbolic evidence to lexically-conditioned phonology. Talk presented at NELS 50. `https://stephsus.github.io/Shih_NELS2019.pdf`.

Shih, Stephanie S., Jordan Ackerman, Noah Hermalin, Sharon Inkelas, Hayeun Jang, Jessica Johnson, Darya Kavitskaya, Shigeto Kawahara, Miran Oh, Rebecca L Starr & 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.

Sidhu, David & Penny M. Pexman. 2018. Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review* 25(5). 1619–1643.

Smith, Brian W. & Joe Pater. 2017. French schwa and gradient cumulativity. Ms. University of California, Berkeley and University of Massachusetts, Amherst (ROA-1702).

Smith, Jennifer. 2002. *Phonological augmentation in prominent positions*: University of Massachusetts, Amherst Doctoral dissertation.

Smolensky, Paul. 1986. Information processing in dynamical systems: Foundations of harmony theory. In D. Rumelhart, J. McClelland & PDPR Group (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1: Foundations, 194–281. Cambridge, MA: Bradford Boooks/MIT Press.

Smolensky, Paul. 1995. On the internal structure of the constraint component CON of UG. Talk presented at the University of California, Los Angeles (ROA-86).

Smolensky, Paul. 1997. Constraint interaction in generative grammar II: Local conjunction, or random rules in universal grammar. Handout of talk given at Hopkins Optimality Theory Workshop/Maryland Mayfest, Baltimore.

Spence, Charles. 2011. Crossmodal correspondences: A tutorial review. *Attention, Perception & Psychophysics* 73(4). 971–995.

Sprouse, Jon. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1. 123–134.

Tesar, Bruce. 2007. A comparison of lexicographic and linear numeric optimization using violation difference ratios. Ms. Rutgers University.

Thompson, Patrick D. & Zachary Estes. 2011. Sound symbolic naming of novel objects is a graded function. *Quarterly Journal of Experimental Psychology* 64(12). 2392–2404.

Westbury, Chris. 2005. Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language* 93. 10–19.

Wilson, Colin. 2014. Maximum entropy models. Tutorial presentation, MIT.

Zimmermann, Richard. 2017. *Formal and quantitative approaches to the study of syntactic change: Three case studies from the history of English*: University of Geneva Doctoral dissertation.

Zuraw, Kie. 2000. *Patterned exceptions in phonology*: University of California, Los Angeles Doctoral dissertation.

Zuraw, Kie & Bruce Hayes. 2017. Intersecting constraint families: An argument for Harmonic Grammar. *Language* 93. 497–548.