

Pokémonikers: A study of sound symbolism and Pokémon names

Stephanie S Shih, Jordan Ackerman, Noah Hermalin, Sharon Inkelas & Darya Kavitskaya*

Abstract. Sound symbolism flouts the core assumption of the arbitrariness of the sign in human language. The cross-linguistic prevalence of sound symbolism raises key questions about the universality versus language-specificity of sound symbolic correspondences. One challenge to studying cross-linguistic sound symbolic patterns is the difficulty of holding constant real-world referents across cultures. In this study, we address the challenge of cross-linguistic comparison by utilising a rich, cross-linguistic dataset drawn from the Pokémon game franchise. Within this controlled universe, we compare the sound symbolisms of Japanese and English Pokémon names (*pokemonikers*). Our results show a tendency in both languages to encode the same attributes with sound symbolism, but also reveal key differences rooted in language-specific structural and lexical constraints.

Keywords. sound symbolism; iconicity; names; onomastics; phonology; corpus linguistics; cognitive science

1. Introduction. A core assumption about the design of human language is the *arbitrariness of the sign*, which holds that there is no intrinsic relationship between linguistic form and its function or meaning (Saussure 1915). Arbitrariness allows human languages immense expressionistic capability and flexibility.

Languages, however, also exhibit symbolism, where form corresponds to meaning or function: such symbolism flouts the assumption of arbitrariness. Symbolic patterns are quite common across the world's languages (for a survey, see e.g., Dingemanse et al. 2015). Onomatopoeia (e.g., *tick tock*), for example, is a cross-linguistically prevalent type of symbolism. The *bouba/kiki* phenomenon is another example of a cross-linguistically wide-spread symbolism: speakers of many different and unrelated languages have been shown to share the intuition that certain sounds such as *bouba* correspond to large, round-shaped objects while other sounds such as *kiki* tend to correspond to smaller, sharper objects (e.g., D'Onofrio 2014). The presence of sound symbolism in human language has given rise to persistent questions about how language relates to the real world in the human cognitive system. For instance, which linguistic forms correspond to real-world meanings? How cross-linguistic are such symbolic sound-meaning correspondences? And, amongst all the possible real-world referents, why do some beget symbolic correspondences and others do not?

One particular challenge to studying cross-linguistic sound symbolism patterns is that, not only linguistic systems but also perception of real-world references shift across the diversity of human cultures. Thus, finding a stable set of referents where only one side of the sound symbolic equation varies has been a problem. One way that this problem has been addressed is through artificial lab setups. In this paper, we offer one of the first naturalistic observational studies of linguistic iconicity that utilises a controlled universe—the wildly popular *Pokémon* universe—

* Acknowledgements to RA Jem Orgun for his contribution to data in this work. Thank you to Shigeto Kawahara, Alan Yu, Rebecca L Starr, and audiences at SCAMP 2017 and UC Berkeley for discussion. Authors: Stephanie S Shih, University of Southern California (shihs@usc.edu); Jordan Ackerman, University of California, Merced; Noah Hermalin, University of California, Berkeley; Sharon Inkelas, University of California, Berkeley & Darya Kavitskaya, University of California, Berkeley.

where non-linguistic factors are held constant, and only linguistic factors vary between languages. We examine Japanese and English Pokémon names and find shared symbolic patterns as well as differences in symbolism that arise from language-specific structural and lexical constraints. Our results also suggest that the real-world referents that beget sound symbolic correspondences are ones most salient to survival fit in the universe.

2. Data. Data comes from the videogame *Pokémon (Pocket Monsters)*, which was first developed in 1995 and continues to the present. The videogame franchise is popular worldwide, and has spread into other mediums, including trading cards, television shows, and mobile gaming apps. The goal of the game is to collect the complete *Pokédex* by collecting all Pokémon species. Players also train their Pokémon characters to battle. Winning battles increases the power statistics of the characters, helps them grow, and leads to the potential capture of new species for the player’s Pokédex collection.

We call Pokémon names here *pokemonikers*. There are 805 *pokemonikers* in our dataset (a few repeat, because they have differing physical attributes), in both Japanese and English (for a study including Russian and Chinese *pokemonikers*, see e.g., Shih et al., in prep). Information for the Pokémon were taken from the Generation VI (2013–2016) version of the game, from Bulbapedia, a fan-based web encyclopedia of Pokémon data.

Phonological transcription for Japanese names were taken from the orthography. English *pokemonikers* were pronounced by the 3rd author, who is a Pokémon “native speaker” with several years of experience in the Pokémon game ecosystem. Pronunciations were transcribed by an undergraduate RA (Jem Inkelas) and the 4th author.

Each Pokémon has several game-defined physical attributes. Here, we examine the following: weight, height (i.e., length), and total performance statistic, which is a catch-all statistic that includes health points, attack, defense, special attack, special defense, and speed.¹ We also examine the evolutionary stages of the Pokémon. Most Pokémon can undergo evolution to become larger and more powerful, from Stage 1 to Stage 3: for example, *Abra* (Stage 1) evolves to *Kadabra* (Stage 2) evolves to *Alakazam* (Stage 3). There are also baby versions of certain Pokémon, and legendary Pokémon, which do not evolve. Finally, each Pokémon is also assigned a likelihood with which it appears in the game as male or female. In some cases, a Pokémon is always male (e.g., *Braviary*) or always female (e.g., *Blissey*). Other Pokémon occur 50/50 male and female (e.g., *Bunnelby*) or are gender-neutral (e.g., *Staryu*).

2.1. ON THE DEVELOPMENT OF POKÉMON NAMES. According to a game development source (Yin-Poole 2011), Japanese Pokémon names are typically developed first, based largely on word meanings. For example, *Hitokage* is a lizard-shaped creature with a flaming tail: its name blends *hi* ‘fire’ and *tokage* ‘lizard’. While not all of the names are explicitly based on sound symbolic correspondences, some are. For example, *Pikachu* is a combination of the Japanese ideophones *pikapika* ‘sparkle’ and *chuuchuu* ‘squeaking’.

English names are developed largely based on meaning preservation of the Pokémon characteristics. For example, *Hitokage* in English is *Charmander*, a blend of *char* and *salamander*. Some Japanese Pokémon names and English Pokémon names do overlap phonologically ($n=193$, 23.98%): for instance, *Pikachu*, *Kabuto*, and *Darumaka* are the same in both languages. It is important to note, however, that these shared *pokemonikers* are not always Japanese in origin.

¹ Previous work has shown some independence between the various types of performance statistics (see e.g., Kawahara et al. 2018).

Some Japanese names are themselves based on English words: e.g., Japanese *Riifia* ~ English *Leafeon*; Japanese *Annoon* ~ English *Unown*; Japanese *Foretosu* ~ English *Forretress*.

3. Previous work. Kawahara et al. (2018) were the first to note correlations between Pokémon names and the physical attributes of the characters. In their study on Japanese pokemonikers, they examined the number of moras and voiced obstruents in names.² They found that more moras in a pokemoniker corresponded significantly to increases in weight, height, power, and evolutionary stage. Similarly, more voiced obstruents in a pokemoniker also positively correlated with size, power, and evolutionary stage. In a follow-up experimental paradigm, Kawahara and Kumagai (to appear) find that the sound symbolic correlations between moras and voiced obstruents and Pokémon evolutionary stage hold for both Japanese and English speakers when encountering novel Pokémon characters.

4. Current study. In the current study, we examine a greater number of potential phonological correlates to Pokémon attributes, across both English and Japanese. Phonological factors investigated include various measures of name length (i.e., number of syllables, moras, segments, and graphemes), vowel quality (i.e., number of high/low, back/front, and rounded/unrounded vowels), and consonant quality (i.e., number of sonorants, voiced/voiceless obstruents, and various places of articulation).

Correlations between linguistic features and Pokémon attributes were tested using a variety of statistical methodologies, including basic tests (e.g., rank correlations), classification trees, conditional random forest variable importance (Strobl et al. 2008), and regression modeling. This range of methodologies was used in order to ameliorate collinearity affects between related factors. For example, all of our measures of name length are highly collinear, because increasing the number of segments nearly inevitably leads to increases in graphemes, moras, and syllables (see e.g., Grafmiller & Shih 2011 for more on length measures). We report here the phonological features that most robustly associate with Pokémon attributes across tests, which we take to be indicative of the sound iconicities that are most likely to be obvious to learners. It is possible that other cues matter and/or that phonological cues work in concert (see e.g., D’Onofrio 2014), but we reserve exploration of interactions for future work.

4.1. RESULTS. A summary of results is presented in Table 1.

In Japanese, Pokémon weight was found to be positively correlated with mora moras, voiced obstruents, and back vowels ([u, o, a]) in a name. In English, Pokémon weight is positively correlated with more segments, voiced obstruents, and low vowels ([æ, a, aʊ, ai, ɔ]) in a name. In Japanese, Pokémon height is positively correlated with more moras and voiced obstruents in a name, as Kawahara et al. 2018 also found. In addition, we find here that Pokémon height is negatively correlated with more labial consonants in a name. In English, Pokémon height is positively correlated with more alveolar consonants and low vowels, and negatively associated with more high vowels ([i, I, i, u, ʊ]). In Japanese, power is positively correlated with more moras in a name, and, as with height, negatively correlated with more labial consonants. In English, power is also positively correlated with increasing name length (i.e., more segments) and more alveolar consonants. In Japanese, more evolved Pokémon have longer names (more moras), and fewer voiceless obstruents in the name. In English, the only significant phonological correlate with

² Kawahara et al. (2018) also look briefly at the vowel quality of initial vowels and segment length; however, mora count and voiced obstruents were the most significant correlates of Pokémon physical attributes.

evolution was the number of segments: more evolved Pokémon tend to have more segments to their name.

Phonological property	Japanese	English
<i>Name length</i>		
moras	↑ weight, ↑ height ↑ power, ↑ evolution	
segments		↑ weight, ↑ power, ↑ evolution
<i>Vowel quality</i>		
back	↑ weight	(↑ male)
high		↓ height
low		↑ weight, ↑ height
<i>Consonant quality</i>		
labial	↓ height	
alveolar		↑ height, ↑ power
voiced obs.	↑ weight, ↑ height	↑ weight
voiceless obs.	↓ evolution	
sonorant	(↓ male)	

Table 1: Summary of significant correlations between phonological features and Pokémon physical attributes. Arrows indicate direction of correlation: ↑ = positive, ↓ = negative. Correlations in parentheses indicate trending effects that approach significance.

Finally, we found only minor correlations for gender. In Japanese, having more sonorants is negatively correlated with the likelihood of occurring as a male character. In English, having more back vowels is positively correlated with the likelihood of occurring as a male character.

5. Discussion. In our comparison of English and Japanese pokemonikers, we find certain cross-linguistic sound iconicities that are shared between both languages for Pokémon naming. The length of the word is the most common phonological correlate to several of the Pokémon attributes: in general, longer names tend to correlate with increasing size, power, and evolutionary stage for the Pokémon. However, between English and Japanese, we find that the English correlate of word length can be primarily linked to evolutionary stage. For example, by examining only the Pokémon in Stage 2 evolution, the name length and Pokémon weight correlation disappears for English; in Japanese, we find that name length and Pokémon weight is still significantly correlated even within characters of the same stage.

While name length is the most common phonological correlate to Pokémon attributes shared between both languages, there are language-specific structural differences. For example, moras are most strongly correlated with Pokémon attributes in Japanese while segments are the strongest measure of name length in English. This difference, we argue, arises from differences between phonological structure in Japanese and English. In Japanese, the addition of any phonological material will inevitably result in the addition of a mora: there is no way to add phonological material without a resulting increase in prosodic structure. In English, on the other hand, it is possible to add phonological material without increasing prosodic structure size. For example, the Stage 1 Pokémon *Kodakku* evolves into Stage 2 *Gorudakku* in Japanese: 4 moras to 5 moras. In English, the same character *Psyduck* evolves to *Golduck*, with no change to prosodic structure because English allows the addition of a segment to form a syllable coda. This process is not available in Japanese phonology: **Gordakku* would result in an illicit coda consonant.

We also find differences in how Pokémon weight is cued: in Japanese, weight is associated with more back vowels, while in English, it is associated with more low vowels. One possible explanation is that this difference arises from contrast differences in the two languages: there are more featural contrasts, for instance, to distinguish the English vowel inventory from the Japanese vowel inventory. We plan to follow up on contrast-based language-specific differences in sound symbolism in future work.

While both languages have voicing correlations to Pokémon attributes, the voicing symbolisms are more common in Japanese, and correlate with size, power, and stage. In English, only [+voice] seems to matter, and it only cues Pokémon size. This voicing difference between the two languages aligns with previous work on the differences between sound iconicity in English and Japanese speakers (e.g., Iwasaki et al. 2007).

The two languages also feature language-specific sound iconicities for different places of articulation. Japanese features a labial symbolism for smaller, less-evolved Pokémon. This aligns with the Japanese propensity to associate labial sounds with baby items: for example, diaper brands in Japan nearly all feature labial sounds (e.g., *merries*, *mamypoko*, *moony*) (Kumagai & Kawahara 2017). Meanwhile, correlations between alveolar consonants and Pokémon attributes feature in English and not in Japanese. In English, more alveolar consonants cue increases in Pokémon height. We suspect that this arises from the association of [s]-like sounds with serpentine creatures in English, which can be physically longer (height also measures length in Pokémon). While snakes also make a [ʃ] noise in Japanese, the potential strength of such associations may vary cross-linguistically, as we are finding here.

Impressionistically, sound-attribute correspondences in the Japanese pokemonikers tend to have stronger effect sizes and capture more variance than English counterparts. Herein lies another potential difference that arises from language-specific factors: stems in Japanese Pokémon names are based on the already-rich ideophonic lexicon of Japanese (e.g., Hamano 1986), whereas the English lexicon is not as ideophone-rich. For example, *Nyoromo* is a Pokémon based on *nyoronyoro* ideophone for ‘sound of slithering’ and *kodomo* ‘child’; in contrast, the English name is non-ideophonic, *Polliwog* ‘tadpole’.³

Finally, we find that the Pokémon attributes that are more closely associated with sound symbolic patterns seem to be the ones that are more important to achieving game-specific goals. Weight, height, power, and evolutionary stage all have significant phonological correlates. How large and powerful a player’s Pokémon is crucially affects a player’s chances in Pokémon battles and collection for the Pokédex. Meanwhile, gender was not as robustly associated with sound patterns. Gender does not play as large of a role in gameplay: it was not fully introduced until the second generation of the game (though, in later versions, gender and breeding of Pokémon have become more prominent). Conversely, gender often has significant phonological correlations in real life. In English, for example, male and female names demonstrate regular differences in length, stress patterns, and vowel quality (e.g., Slater & Feinman 1985; Cassidy et al. 1999; Wright et al. 2005; Sidhu & Pexman 2015). This difference between sound-attribute correspondences between the Pokéverse and the real world suggests that attributes that are most crucial for evolutionary survival in a universe may have greater propensity to associate to phonological patterns. That is, systematic phonological associations and sound iconicities may occur when it is most useful to distinguish salient attributes.

³ However, it is possible that the origins of *wag* or *wiggle* in *polliwog* may be sound iconic.

6. Conclusion. We have shown here that in both Japanese and English pokemonikers, phonological patterns correlate to features of Pokémon including size, power, and evolutionary stage. By using a universe in which the world referents are held constant, we are able to probe the cross-linguistic similarities and differences between sound symbolic patterns. Our study finds both shared cross-linguistic sound symbolic patterns (e.g., name length with size) and differences that arise from language-specific structures. The results have demonstrated the valuable utility of such corpora that arise from popular media and can have great reach across cultures and languages.

References

- Cassidy, Kimberly Wright; Michael H. Kelly; and Lee'at J. Sharoni. 1999. Inferring gender from name phonology. *Journal of Experimental Psychology: General* 128.362–381. <http://dx.doi.org/10.1037/0096-3445.128.3.362>.
- 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.
- Grafmiller, Jason; and Stephanie S Shih. 2011. New approaches to end weight. Paper presented at the Variation and Typology: New trends in Syntactic Research, Helsinki, Finland.
- Hamano, Shoko Saito. 1986. *The Sound-Symbolic System of Japanese*. University of Florida dissertation.
- Iwasaki, Noriko; David P. Vinson; and Gabriella Vigliocco. 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* 17.53078.
- Kawahara, Shigeto; and Gakuji Kumagai. to appear. Expressing Evolution in Pokémon Names: Experimental Explorations. *Journal of Japanese Linguistics*.
- Kawahara, Shigeto; Atsushi Noto; and Gakuji Kumagai. 2018. Sound symbolic patterns in Pokémon names. *Phonetica* 75.
- Kumagai, Gakuji; and Shigeto Kawahara. 2017. How abstract is sound symbolism: Labiality and diaper names in Japanese. *Proceedings of the 31st Annual Meeting of the Phonetic Society of Japan*. 49–54.
- Saussure, Ferdinand de. 1915. *Course in general linguistics*. New York: McGraw Hill.
- Sidhu, David M.; and Penny M. Pexman. 2015. What's in a Name? Sound Symbolism and Gender in First Names. *PLOS ONE* 10.e0126809. <http://dx.doi.org/10.1371/journal.pone.0126809>.
- Slater, Anne S.; and Saul Feinman. 1985. Gender and the phonology of North American first names. *Sex Roles* 13. 429–440.
- Strobl, Carolin; Anne-Laure Boulesteix; Thomas Kneib; Thomas Augustin; and Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9. 307–317.
- Wright, Sandra K.; Jennifer Hay; and Tessa Bent. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 43. 531–561.
- Yin-Poole, Wesley. 2011. Pokemon gamer interview. *Eurogamer*. <http://www.eurogamer.net/articles/2011-02-22-the-brains-behind-pokemon-black-and-white-interview?page=2>.