# On sound symbolism in baseball player names


Stephanie S Shih
University of Southern California
shihs@usc.edu

Deniz Rudin
University of Southern California
drudin@usc.edu


Draft: 2019 July 14


Recent work has argued that sound symbolism plays a much larger part in language than previously believed, given the assumption of the arbitrariness of the sign.* A slate of recent papers on *Pokémonastics*, for example, has found sound symbolic associations to be rampant in Pokémon names cross-linguistically. In this paper, we explore a real-world dataset that parallels Pokémon, in which human players similarly have physical attributes of weight, height, and power: Major League Baseball. We investigated phonological correlations between baseball player statistics and their given first names, chosen baseball-official first names, and baseball nicknames. We found numerous sound symbolic associations in player-chosen names and nicknames, where conscious design plays a role in choosing a name that may communicate an attribute. These associations were often mediated by language-specific hypocoristic formation processes. We conclude that sound symbolism occurs in real-world naming practices, but only when names are chosen agentively in cognizance of the relevant attributes.


*Keywords*: sound symbolism; iconicity; names; onomastics; phonology; corpus linguistics; cognitive science; English; baseball

## 1 Introduction

The principle of the arbitrariness of the sign holds that human language involves no intrinsic relationships between linguistic form and linguistic meaning or function (e.g., Saussure 1915; Hockett 1959). The existence of sound symbolism, in which the phonological form of a word imitates its referent, suggests that the sign is not so arbitrary after all (e.g., Sapir 1929). These two concepts have long been at odds with each other in linguistic analysis: to what extent are phonological systems intertwined with semantic ones?

Recent work has pushed the idea that sound symbolism plays a much larger part in language than previously believed (e.g., Hinton et al. 1994; Spence 2011; Dingemanse et al. 2015; Lockwood & Dingemanse 2015; Sidhu & Pexman 2018). A slate of recent papers on *Pokémonastics*, for example, has studied the sound symbolic associations that appear rampant in Pokémon names cross-linguistically (e.g., Kawahara et al. 2018; Shih et al. 2018, in prep; Starr et al. 2018; Kawahara & Kumagai, to appear). This work has shown sound symbolic patterns that correlate with the attributes of fictional Pokémon characters in the eponymous video games. For example, Pokémon with longer names tend to be heavier, longer/taller, more powerful, and more evolved than their counterparts with shorter names. This effect holds across many languages, including Japanese, English, Mandarin, Cantonese, Korean, and Russian (Shih et al., in prep).

* To Sharon Inkelas and Vern Paxson: many thanks for taking the first author to her first major league baseball game.

The sound symbolic patterns in Pokémon names (*pokemonikers*) extend beyond name length: they include associations of vowel and consonant qualities, as well as of reduplication and tone. While some of these patterns, like the length effect, are cross-linguistic (though see Shih et al., in prep for specific discussion of this point), other patterns are shown to be language-specific, clearly deriving from the existing phonological sound symbolic associations that already exist in a particular language. Japanese pokemonikers, for instance, demonstrate a negative correlation between the presence of labial consonants and the size and power of Pokémon. Shih et al. (2018) suggest that this stems from the existing association in Japanese between labial sounds and babies (see e.g., Kumagai & Kawahara 2017 on labiality in Japanese diaper brand names). In contrast, the other languages in the Pokémon studies did not show such labiality correlations.

Shih et al. (2018, in prep) assert that the sound symbolic associations in Pokémon names are not incidental, but instead that their purpose is to cue relevant game-specific information about the characters. The Pokémon attributes that were found to have the most robust sound symbolic associations cross-linguistically were those that were most relevant to gameplay, including power and evolutionary stage, and to some extent, character weight and height. On the other hand, Pokémon gender, which was not a part of the original videogames, did not have as many or as strong sound symbolic associations in their corpus. In comparison, human name phonotactics often carry robust cues to gender, which is evolutionarily relevant information in the real world. Female names and male names in English, for example, differ in many aspects, from their name lengths and stress patterns to their propensities of vowels versus consonants (e.g., *Amanda* vs. *Adam*; see e.g., Slater & Feinman 1985; Cassidy et al. 1999; Sidhu & Pexman 2015; for gender-based phonotactics in other languages, see e.g., Starr et al. 2018). Together, Pokémonastics and other recent work on sound symbolism (in names, see Sidhu & Pexman 2015, 2019 for recent overviews; elsewhere in languages, see e.g., Newman 2001; Bergen 2004; Dingemanse 2012; Gasser et al. 2010) suggest that non-arbitrariness is not only a natural but also potentially sizable part of human language, and that it serves linguistic purposes—as arbitrariness does—beyond pure mimicry.[1]

A critical confound in the Pokémonastics work, however, is that all of the Pokémon names are designed for each character. While sound symbolism may or may not be a conscious part of Pokémon name choice, it is undeniable that game designers have in mind particular Pokémon attributes that they want to highlight when choosing suitable names. The Pokémonastics work has shown that the sound symbolism in Pokémon names are scaffolded on existing sound symbolic associations that speakers already have for their respective languages, but it remains unclear to what extent such sound symbolic associations are only prevalent because they are so designed.

In this paper, we explore a real-world dataset that parallels Pokémon, in which human players similarly have physical attributes of weight, height, and power statistics: Major League Baseball.[2] We examine phonological correlations between baseball player statistics and their given first names, chosen baseball-official first names, and baseball nicknames. There is a body of previous research on baseball nicknames (see e.g., Skipper 1981, 1990, 1992; Wilson & Skipper 2013) and orthographic features of baseball player names (see e.g., Newman et al. 2009), but, to the best of our knowledge, this study is the first to focus on sound symbolic patterns in all baseball player names, from their birth names and hypocoristics to their baseball-specific nicknames. Our results show vanishingly few sound symbolic associations between given names and player attributes—

[1] Other purposes not discussed here include the use of sound symbolism in e.g., language acquisition, speech segmentation, and social indexation (e.g., Imai et al. 2008; Kantartzis et al. 2011; Eckert 2017).
[2] Another parallel between Major League Baseball and Pokémon includes rich trading card traditions that accompany both cultures.

as we might expect, since human names, unlike Pokémon ones, are usually assigned at birth before information such as weight, height, and especially baseball aptitude is known. However, we do find numerous sound symbolic associations in player-chosen names and nicknames, where conscious design plays a role in choosing a name that may communicate an attribute (see e.g., Bois 2019a, 2019b for pioneering work on athletes, including baseball players, who go by the hypocoristic *Bob*). Some of these sound symbolic associations are a reflex of existing morphological patterns in English, such as the diminutive construction, but others are not reducible to such 'grammaticalized' instances of sound symbolism. We conclude that sound symbolism occurs in real-world naming practices, but only when names are chosen agentively in cognizance of the relevant attributes.

The paper is organized as follows. Section 2 presents data and methodology. Section 3 presents results of this exploratory study. Sections 4 discusses and concludes, respectively.

## 2 Data and Methodology

Data comes from the Lahman's Baseball Database (Lahman 2018), using `R` package `lahman` (Friendly et al. 2019). Lahman's database includes data for Major League Baseball players from 1871 to the present. For the current study, we used data for hitters beginning at the onset of the live-ball era (1920) and ending in 2017; all players with more than 450 plate appearances total were included.[3]

For baseball player attributes, we were interested in data that would parallel the attributes examined in Pokémon: mainly, player weight, height, and "power".[4] Pokémon total power statistics are an amalgamation of many different characteristics, including speed, attack, and hit points. For a parallel to Pokémon "power", we examined four different statistics that approximate a hitter's strength and effectiveness in baseball. First, *slugging percentage* (a scale from 0–4) is a proxy for hitting power, calculated as the total number of bases divided by at-bats. Second, *batting average* approximates plate discipline and accuracy, measured by the number of hits divided by at-bats. Third, *batting average on balls in play* (BABIP)—batting average calculated on only the balls hit into play—measures a player's hitting skill independent of power and plate discipline. Finally, *on-base percentage plus slugging* (OPS) proxies a player's overall skill at the plate, which encompasses power, accuracy, and plate discipline. For each player, we took the statistics from the year in which they had their highest slugging percentage, to get a picture of their peak performance.

For baseball player names, we had access to three sets of first names for players: their given first names at birth, their chosen baseball-official first names (as registered with the league), and, for some players, their baseball nicknames. For example, Robert Gibson was born *Robert*, but registered as *Bob*, and garnered the nickname *Hoot*. Data for the nicknames were taken from the Wikipedia page listing Major League Baseball nicknames (https://en.wikipedia.org/wiki/List_of_baseball_nicknames). Not all players have nicknames: only 324 nicknames were provided in the Wikipedia list for our data. All first names were phonologically transcribed by a combination of automatic annotation via the Unisyn lexicon (Fitt 2001) and manually by the

[3] Four hundred fifty plate appearances is recommended as the minimum cut-off in the R package documentation due to the strike year of 1994.

[4] Conceivably, one could look at minor league to major league statistics as a measure of evolutionary stage, but we do not currently have such information.

authors. About 294 first names were excluded due to uncertain and unverifiable pronunciation. All told, there are 2557 total players (i.e., hitters) in our dataset.

Henceforth, we use *registered* to refer to all first names officially registered in Major League Baseball ($n$=2557). *Given* refers to first names given at birth ($n$=1064), and *chosen* refers to registered names that differ from those given at birth ($n$=1495). *Nickname* refers to baseball nicknames taken from the Wikipedia list of baseball nicknames ($n$=324), most of which differ from *chosen* or *given* names.

Correlations between phonological features and baseball player attributes were tested using basic correlation tests (Spearman or Kendall, where appropriate) and regression modeling.

## 3 Results

We examined the same phonological features for English as Shih et al. (2018, in prep): the number of each of the following features in a baseball player's (first, nick) name:

| | | |
|---|---|---|
| (1) | Name length: | segments |
| | Vowel quality: | high vowels, low vowels, front vowels, non-front vowels |
| | Consonant quality: | voiced obstruents, voiceless obstruents, sonorants, labial consonants, alveolar consonants, velar consonants |

Results are presented as follows: §§3.1–3.2 provide the results of weight/height and baseball statistics associations with players' registered and given birth names. Section 3.3 discusses results in players with nicknames.

### 3.1 Results: Player weight and height

For players' registered names, we find significant sound symbolic correlations with vowel quality, length, and consonant quality phonological factors. They are discussed below in turn.

#### *3.1.1 Vowel quality and player weight*

The registered names with more high vowels negatively correlate with player weight ($\rho$=–0.089, $p$<0.0001$^{***}$; see Figure 1).[5, 6]

---

[5] * = significant at $p$<0.01; ** = significant at $p$<0.001; *** = significant at $p$<0.0001.

[6] Only one player in our dataset, *Yuniesky Betancourt*, has a name with three high vowels. The high vowel correlation still obtains even with this player excluded ($\rho$=–0.09, $p$<0.0001).

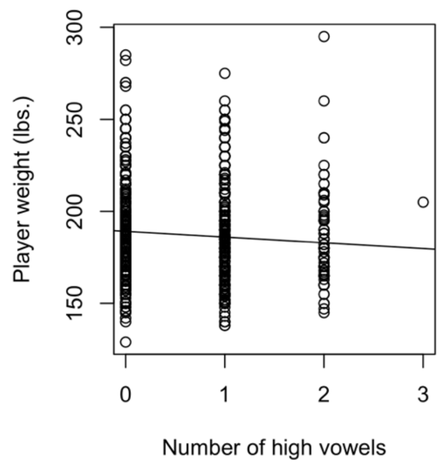


*Figure 1*. Player weight (in pounds) by number of high vowels in registered names.

This effect mirrors the sound symbolic association found in the Pokémon dataset, where having more high vowels in a pokemoniker correlated with smaller Pokémon (Shih et al. 2018).

We find a similar negative correlation for player height with the number of high vowels in a registered name, although the correlation is somewhat weaker than the correlation between player weight and high vowels ($\rho$=–0.086, $p$<0.0001$^{***}$). Given that weight and height can be somewhat correlated in and of themselves, we focus primarily on discussing the weight effects here.

Why would we even expect to find correlations between the phonological patterns of human names and their physical attributes, such as weight? After all, it is unlikely that a name given at birth would reflect aspects of adult form.[7] To investigate this issue, we examined the portion of the data in which a player's chosen name differs from his given name (at birth).

For these players whose names were changed, their chosen names demonstrated sound symbolic correlations. In comparison, their corresponding given names lacked significant sound symbolic associations (though see §3.1.3 for a potential counterexample). For the chosen names in this subset, the high vowel effect is the same as the overall chosen dataset (see Figure 1 for the overall effect): the presence of more high vowels in a name is negatively correlated with weight and height—the weight effect is shown in Figure 2a ($\rho$=–0.135, $p$<0.0001$^{***}$). In contrast, the number of high vowels does not significantly correlate with weight or height in their given birth names (for weight: $\rho$=0.047, $p$=0.07; for height: $\rho$=0.044, $p$=0.09).

For many of the player names that changed from given to chosen, the chosen names with high vowels have an [–i] diminutive suffix: for example, *Bob* to *Bobby* (as in player *Bobby Abreu*, born *Bob*). It turns out that the presence of a diminutive [–i] ending in a name significantly correlates with player weight: those with the ending are lighter than those without ($\chi$=132.79; $p$=0.0001$^{**}$), as shown in Figure 2b. We find also that whether a chosen name has a diminutive suffix is a significantly better predictor of weight than the count of high vowels in the player name ($\Delta AIC$=27). Thus, the results show that the sound symbolic effect for high vowels in player names and player weight is driven primarily by the diminutive suffix in this dataset.

[7] This is not entirely the case, however. Given genetic predispositions, parents are likely to have some idea of a child's future body type at the outset. Additionally, there are potential nurture-based scenarios in which a person 'grows into' their name (see e.g., implicit egotism; Pelham et al. 2002).

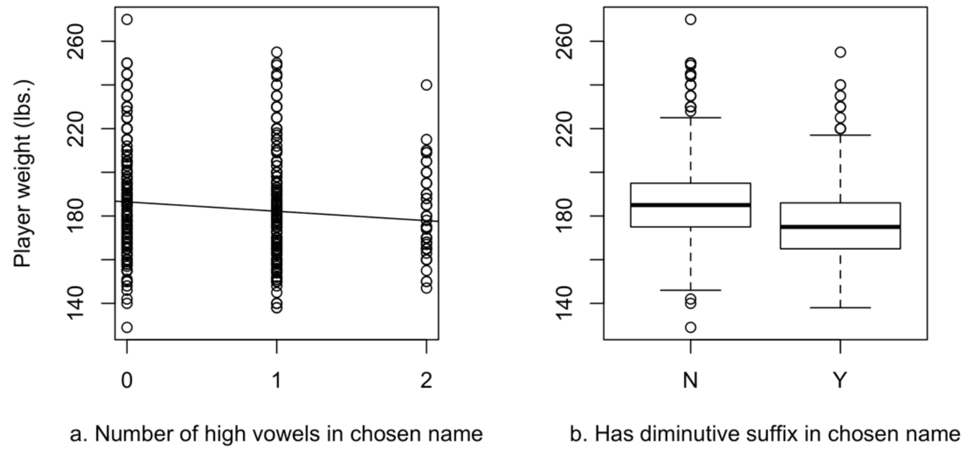


*Figure 2*. Player weight (in pounds) by phonological factor in chosen names for players with different given and chosen names (left to right): (a) number of high vowels; (b) presence of diminutive [–i].

### *3.1.2 Name length and player weight*

In the overall dataset, longer registered names correlate with heavier players ($\rho$=0.094, $p$<0.0001$^{***}$; see Figure 3).[8]

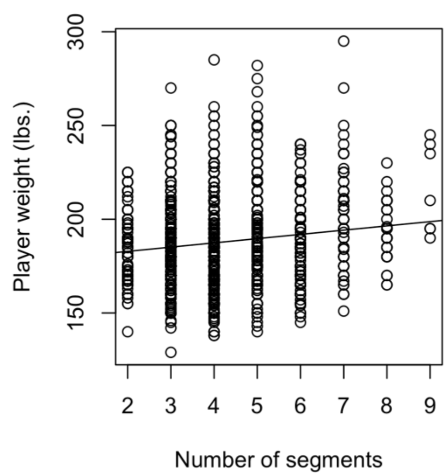


*Figure 3*. Player weight (in pounds) by number of segments in registered names.

As with the high vowel effect in §3.1.1, this length effect parallels sound symbolic behaviour in the pokemonikers, where longer names correlated with heavier Pokémon (e.g., Kawahara et al. 2018).

We also find a weight effect in the subset for different given-to-chosen player names. Unlike the overall registered names, however, there is actually a negative correlation between the length of a name and the corresponding player's weight ($\rho$=–0.089, $p$=0.0005$^{**}$). This is illustrated

[8] When name length is measured as the number of syllables, there is no effect. It is possible that there is more going on here, since hypocoristics in English are often formed via the removal of syllables (rather than segments), but we save this issue of name length specifics for future investigation.

in Figure 4a. Similarly, there's a negative height correlation as well ($\rho$=–0.102, $p$<0.0001***). There is no significant correlation of name length with player weight or height for the players' given birth names ($\rho$=–0.002, $p$=0.94).

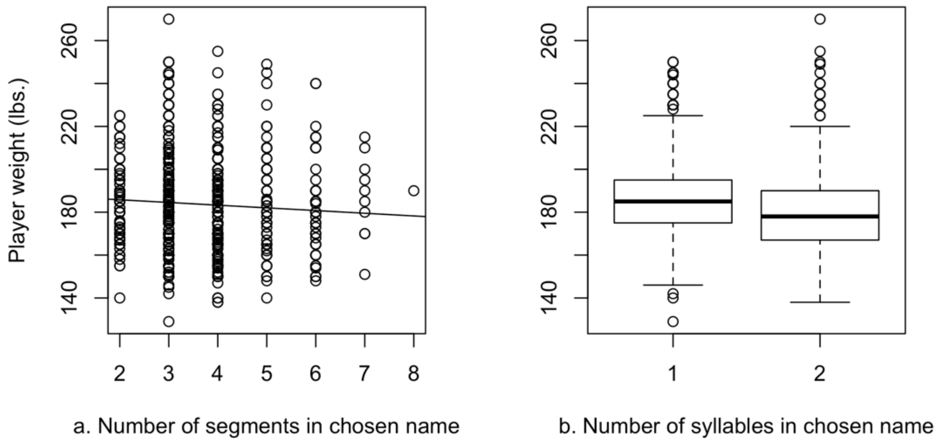


*Figure 4*. Player weight (in pounds) by phonological factor in chosen names for players with different given and chosen names (left to right): (a) number of segments; (b) number of syllables (with 3-syllable names excluded for this graph).

This seemingly opposite effect from the overall dataset—where there was a positive association between name length and weight—is explained by the fact that many of these chosen names differ from their given names in length via shortening (when they aren't otherwise lengthened or kept constant in length by the addition of a diminutive suffix): for example, *Robert* to *Bob* (as in *Bob Skinner*, born *Robert*). As shown in Figure 4b, players whose chosen names have two syllables are on average lighter than those whose chosen names only have one syllable (Figure 4b excludes names with more than 2 syllables, for illustrative purposes). We see here, then, that like the diminutive formation and high vowel effect above, the sound symbolic associations with weight come largely from the existing hypocoristic formation processes in English. That is, sound symbolic effects for weight come into play when a name is chosen by design rather than via "raw" associations between phonological features and meaning.

### *3.1.3 Consonant quality and player weight*

There are two main consonant quality phonological correlations with weight. The first is illustrated in Figure 5.

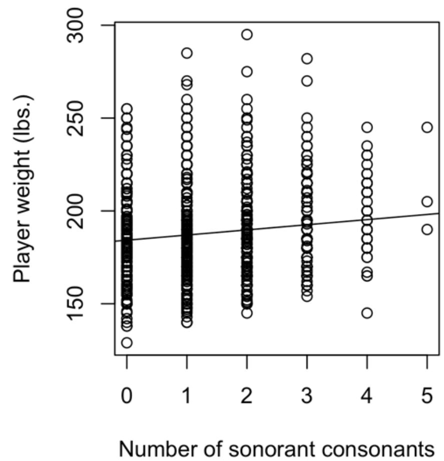


*Figure 5*. Player weight (in pounds) by number of sonorant consonants in registered names.

Registered names with more sonorant consonants significantly correlate with heavier players ($\rho$=0.102, $p$<0.0001$^{***}$). Similarly, English pokemonikers with more sonorant consonants also correlated with heavier Pokémon (see e.g., Shih et al. 2018).

Note that we also find a similar correlation between alveolar consonants and player weight, but it is extremely likely that the alveolar consonant and the sonorant consonant effects are related (and one and the same). The number of sonorant consonants and alveolar consonants in the baseball names are highly correlated: amongst registered names, the correlation between the two features is $\rho$=0.468, $p$<0.0001$^{***}$. Indeed, most of the frequent sonorants are [ɹ, n]. In comparison, there is no correlation between velar consonants and sonorants ($\rho$=–0.03, $p$=0.134), nor is there correlation between labial consonants and sonorants ($\rho$=0.05, $p$=0.011).

Unlike the high vowel or length effects discussed above, the sonorant consonant sound symbolic effect here is interestingly one that does not stem from the use of an existing English hypocoristic formation: that is, there is no regular hypocoristic function in English that adds sonorant consonants. Instead, we find that the sonorant effect stems from the portion of the chosen names that change because a player chooses to use their given middle name or some version of the middle name instead of their first: for example, given first and middle *David Dale* to chosen *Dale* (as in *David Dale Alexander*), or *Leslie Charles* to *Charlie* (as in *Leslie Charles Spikes*). Extremely rarely do we find players' chosen names with added sonorants that differ from both their first and middle given names. An example is *John Collins Ryan*, who registered as *Blondy Ryan*.

The second consonant quality phonological correlation that we find with weight comes from given names. Amongst the given names, names with more velar consonants are positively correlated with heavier players, shown in Figure 6 ($\rho$=0.112, $p$<0.0001$^{***}$).

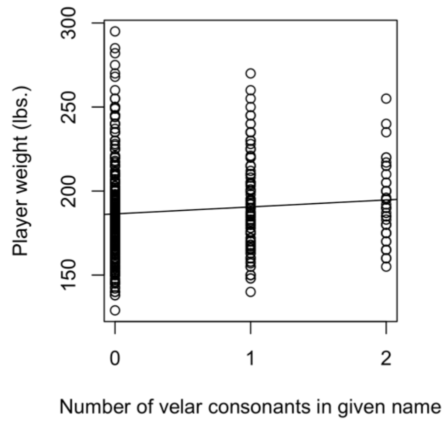


*Figure 6*. Player weight (in pounds) by number of velar consonants in given names.

This velar correlation does not track any of the Pokémon effects found in the previous literature. Nor does it particularly resonate with noted behaviours of velars sound symbolically in English. In the *bouba kiki* phenomenon, D'onofrio (2014) reported greater associations of velar sounds with spiky objects (than, say, labial sounds; see also Sidhu & Pexman 2019), but to the best of our knowledge, no claims have linked velars with specifically weight-based attributes for English speakers.

The velar consonant correlation with weight found here is puzzling, especially as we do not expect the phonology of given names to associate with adult human attributes. We hesitate at this point to put much stock in the presence of velar-weight links in the baseball given name data, especially because most of the names with 2 velar consonants are *Frank* and *Gregory*. The other sound symbolic findings reported herein all involve many more unique name instantiations in the patterns; as such, we can be more confident in their robustness.

## 3.2 Results: Baseball-specific statistics

The results reveal a single baseball-specific correlation between name phonology and baseball statistic. There are no significant phonological correlations for the baseball statistics of slugging percentage, batting average, or on-base percentage plus slugging (OPS). The only baseball statistic that showed a significant phonological correlate was batting average on balls in play (BABIP), which has a positive association with registered name length ($\rho$=0.089; $p$<0.0001$^{***}$), shown in Figure 7.

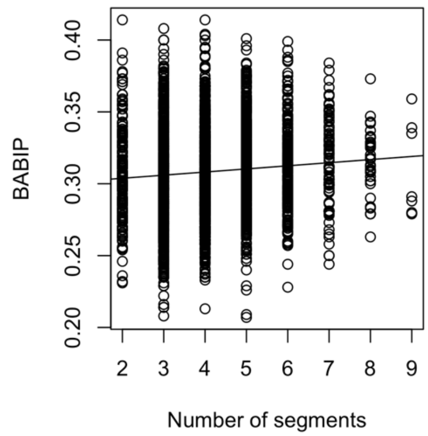


*Figure 7*. Player batting average on balls in play (BABIP) by number of segments in registered names.

It is not completely clear to us why BABIP is the only baseball statistic to show a phonological correlation, out of the 4 statistics investigated here. In fact, it was surprising that there is a phonological correlation to baseball statistic at all: players are, after all, born with given names and likely settle on registered names before they know what their baseball performance will be in their best playing years.

Moreover, the BABIP effect can be attributed largely to the section of players whose chosen registered names *do not* change from their given names. Examining the longest names in these players provides a potential clue to why the BABIP effect may exist. The players with the longest unchanged names include those shown in (2a), versus players with the longest changed given names in (2b).

(2) Longest given names (segment count > 8) and number of occurrences

a. No name change players
*Francisco*
*Alejandro*
*Vladimir*
*Salvador*
*Andrelton*
*Giancarlo*

b. Name change players
*Kristopher, Christopher, Khristopher*
*Alexander*
*Baldomero*
*Salvatore*
*Dagoberto*
*Bernardo*
*Leonardo*
*Cornelius*
*Rigoberto*
*Maximilian*
*Constantino*
*Saturnino*
*Rutherford*
*Sebastian*

We note that the majority of names in (2b) are of European origin, including 18 instances of *Christopher* (lumping together spelling variants) and 10 instances of *Alexander* in the dataset. These are names that have established and common hypocoristics in English (e.g., *Chris*, *Alex*), while none of the longest names in (2a) have established English hypocoristics (with the possible exception of *Sal* for *Salvador*). A possible source for the BABIP effect may be, then, demographic trends in the make-up of major league baseball players resulting in an increase in players without ready-made English hypocoristics, and an increased tolerance for non-shortened, non-English names. Another possible explanation is that players who demonstrate a lot of batting skill do not feel the need to shorten their names. However, we cannot make a conclusive argument for or against either of these potential explanations, and neither of them predicts an effect on BABIP to the exclusion of other batting statistics. It is possible that the BABIP correlation we find here is spurious—the fact that we do not find any other baseball statistic associating with phonological patterning strongly suggests to us that there is no true sound symbolic relationship between baseball player chosen and given names and their hitting aptitudes.

### 3.3 Results: Player nicknames

Players with nicknames represent a much smaller subset of our data ($n$=324). Between player nicknames and their physical and baseball attributes, we find one significant positive association: more voiced obstruents in a nickname correlates with taller players ($\rho$=0.186; $p$=0.0007$^{**}$), as shown in Figure 8. (No corresponding weight correlation was found.)

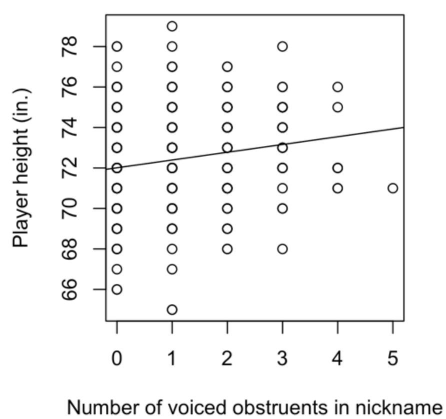


*Figure 8*. Player height (in inches) by number of voiced obstruents in nicknames.

Pokémon names demonstrated similar relationships between voiced obstruents and the Pokémon attributes (see e.g., Kawahara et al. 2018; Shih et al. 2018), where larger and stronger characters had names with more voiced obstruents, particularly in Japanese. English Pokémon names also demonstrated an effect—albeit weaker—of voiced obstruents and size; previous research on English and Japanese sound symbolism has also shown that English speakers share intuitions between voicing and size in sound symbolism (Iwasaki et al. 2007; Kawahara & Kumagai, to appear).

It is curious that we do not find *more* sound symbolic associations amongst the nicknames, since the formation of baseball nicknames (beyond the chosen names) is less constrained by, say, strong faithfulness pressures to given names: e.g., *Crime Dog* from given name *Frederick* and chosen name *Fred* (as in *Fred 'Crime Dog' McGriff*). The dataset of nicknames here is small, and the players who are most likely to garner nicknames are often the best ones, which compresses the range of baseball statistics in the dataset. Furthermore, nicknames not derived from given names are declining in popularity in baseball over the last century, generally (see e.g., Skipper 1981).

Lastly, it is most probable that there are many different attributes that someone would want to communicate using a baseball nickname—many of which we do not examine in the current study. Skipper (1981; see also Skipper 1990) reports, for instance, that in addition to physical attributes (e.g., whether someone is a left-handed player), baseball nicknames in the first half of the 20th century often make reference to familial relationships, ethnicity, personality traits and habits, and geographical location.

## 4 Discussion and Conclusion

Our investigation of sound symbolic patterning in the names of Major League Baseball players found significant phonological correlations with players' physical characteristics only in the chosen names and nicknames, and not in the given first names.[9] The results include a negative correlation between high vowels and player weight, which we argue is due to diminutive hypocoristics being disproportionally chosen by smaller players. We also find a correlation between name length and player weight, which we argue is due to truncation hypocoristics being disproportionally chosen by larger players. For players whose chosen first name differs from, and is not a hypocoristic of, their given first name (e.g., players who choose to go by their middle name instead of their first name), we found a positive correlation between the number of sonorant consonants in the chosen name and player weight. Among nicknames, we found a positive correlation between the number of voiced obstruents and player weight. All of these effects found run parallel to previous findings in the pokémonastics literature. We found no significant correlations between any sound symbolic factor and baseball performance, modulo one correlation with BABIP; we do not have an explanation for this correlation, or for why only BABIP amongst the baseball statistics is involved with it.

Overall, our results support the existence of comparable sound symbolic pressures within baseball player names and Pokémon names, but only when *design* is involved—that is to say, only when the naming choices are made agentively and in cognizance of the relevant attributes. Some of these sound symbolic effects are reducible to 'grammaticalized' sound symbolism, as in the cases of English diminutive and truncation hypocoristics, but others are not, as in the case of non-hypocoristic differences between given and chosen names. It appears as though when people are given a choice of names amongst a circumscribed set of salient alternatives—like the choice between faithfulness to a full first name, application of a hypocoristic process, or going by a middle name—sound symbolic pressures exert some influence on the choice. Our observation of fewer sound symbolic effects in the non-registered nicknames might reflect the fact that the nicknames are not chosen from a circumscribed set of salient alternatives, or might simply reflect the fact that our smaller dataset of nicknames—many of which are assigned by different entities for different reasons at different times in a player's career—allows us less resolution to identify the relevant sound symbolic patterns.

In pokémonastics, the most robust effects are associated with attributes relevant to gameplay, like power and evolution. In baseball player names, we see instead that physical attributes such as weight and height are most correlated with sound symbolic patterns. We hypothesize that this is because baseball players are human beings, who exist for many non-baseball-related purposes. In contrast, Pokémon inhabit a constructed universe that they populate almost exclusively

[9] With the exception of the velar consonant effect in given names (see §3.1.3), which is an unexpected effect given known English sound symbolic patterns.

for gameplay-related purposes. Physical attributes play a role in many areas of a baseball player's life that their batting averages are irrelevant to, and so it is unsurprising that these attributes show the most robust sound symbolic associations. Additionally, baseball players may settle on a preferred first name prior to—and thus while they're still unaware of—their future baseball performance. Given the divergences between the constructed universe inhabited by the Pokémon and the real word inhabited by Major League Baseball players, we take it to be a remarkable finding that baseball players' choices of registered names appear subject to similar pressures as those that are at play in pokémonastics at all.

The findings of this paper are a first look in a naturally-occurring dataset at the associations of human names with physical and performance-based attributes. There are several avenues for future development. Players are more likely to choose new first names than surnames; thus, the immutability of surnames makes it such that we do not expect to find sound symbolic patterns in the surnames of Major League Baseball players. This study only investigated hitting statistics: it would be interesting to explore whether sound symbolic correlations exist with pitching and fielding statistics (cf., Newman et al. 2009 on orthographic correlations in names with pitching performance). The baseball statistics we have examined here are traditional baseball statistics, and are subject to a variety of baseball-specific confounds that keep them from being a true measure of player skill (Thorn & Palmer 1984; Tango et al. 2007). An investigation making use of more sophisticated baseball statistics (e.g., Statcast) could provide more accurate measures on the baseball front. Major League Baseball is not the only baseball league; a comparison with, say, Nippon Professional Baseball, could illuminate informative differences and similarities in sound symbolism cross-linguistically, as pokemonikers do. Likewise, a comparison with professional female athletes could shed light on how gender modulates sound symbolic name associations (see e.g., Wilson & Skipper 2013 on nicknames for professional female baseball players). Our analysis does not take into account fluctuation in league-level statistics (e.g., movement between pitcher-dominated and hitter-dominated eras), providing a potential confound, especially as fluctuation in league-level statistics might be accompanied by fluctuation in the ethnic makeup of Major League Baseball and, as a result, cultural naming practices. Finally, we might wonder whether baseball fans have internalized sound symbolic correspondences. This could be investigated via an experimental study presenting participants with nonce names of fictional baseball players, and measuring expectations about those players' physical attributes and baseball performance. We leave the pursuit of these avenues to future work.

## References


BERGEN, BENJAMIN K. 2004. The Psychological Reality of Phonaesthemes. *Language* 80.290–311.

BOIS, JON. 2019a. *The Bob Emergency: A study of athletes named Bob, Part I.* Video. https://www.youtube.com/watch?v=lvh6NLqKRfs.

BOIS, JON. 2019b. *The Bob Emergency: A study of athletes named Bob, Part II.* Video. https://www.youtube.com/watch?v=dcGamPqUlxl&t=3s.

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. doi:10.1037/0096-3445.128.3.362.

DINGEMANSE, MARK. 2012. Advances in the Cross-Linguistic Study of Ideophones. *Language and Linguistics Compass* 6.654–672.

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.

ECKERT, PENELOPE. 2017. Comment: The most perfect of signs: Iconicity in variation. *Linguistics* 55.1197–1207.
FITT, SUSAN. 2001. Unisyn lexicon. Centre for Speech Technology Research, University of Edinburgh. www.cstr.ed.ac.uk/projects/unisyn.
FRIENDLY, MICHAEL; CHRIS DALZELL; MARTIN MONKMAN; DENNIS MURPHY; VANESSA FOOT; and JUSTEENA ZAKI-AZAT. 2019. *lahman r package*. http://lahman.r-forge.r-project.org/.
GASSER, MICHAEL; NITYA SETHURAMAN; and STEPHEN HOCKEMA. 2010. Iconicity in Expressives: An empirical investigation. *Experimental and Empirical Methods in the Study of Conceptual Structure, Discourse, and Language*, ed. by Sally Rice and John Newman, 163–180. Stanford, CA: CSLI Publications.
HINTON, LEANNE; JOHANNA NICHOLS; and JOHN OHALA. 1994. Sound-symbolic processes. *Sound symbolism*, ed. by Leanne Hinton, Johanna Nichols, and John J. Ohala, 1–14. Cambridge, UK: Cambridge University Press.
HOCKETT, CHARLES. 1959. Animal "languages" and human language. *Human Biology* 31.32–39.
IMAI, MUTSUMI; SOTARO KITA; MIHO NAGUMO; and HIROYUKI OKADA. 2008. Sound symbolism facilitates early verb learning. *Cognition* 109.54–65. doi:10.1016/j.cognition.2008.07.015.
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.
KANTARTZIS, KATERINA; MUTSUMI IMAI; and SOTARO KITA. 2011. Japanese Sound-Symbolism Facilitates Word Learning in English-Speaking Children. *Cognitive Science* 35.575–586. doi:10.1111/j.1551-6709.2010.01169.x.
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* 75.3: 219–244.
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.
LAHMAN, SEAN. 2018. *Lahman's Baseball Database*. http://www.seanlahman.com/baseball-archive/statistics.
LOCKWOOD, GWILYM; and MARK DINGEMANSE. 2015. Iconicity in the lab: a review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology* 6. doi:10.3389/fpsyg.2015.01246. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01246/full.
NEWMAN, LEONARD S.; WILLIAM HERNANDEZ; DARIA A. BAKINA; and ABRAHAM M. RUTCHICK. 2009. Implicit egotism on the baseball diamond: Why Peter Piper prefers to pitch for the Pittsburgh Pirates. *Names: A Journal of Onomastics* 57.175–179. doi:10.1179/175622709X462478.
NEWMAN, PAUL. 2001. Are ideophones really as weird and extra-systematic as linguists make them to be? *Ideophones*, ed. by F. K. Erhard Voeltz and Christa Kilian-Hatz, 251–258. Typological Studies in Language 44. Amsterdam: John Benjamins.
PELHAM, BRETT W.; MATTHEW C. MIRENBERG; and JOHN T. JONES. 2002. Why Susie Sells Seashells by the Seashore: Implicit egotism and major life decisions. *Journal of Personality and Social Psychology* 82.469–487. doi:10.1037//0022-3514.82.4.469.
SAPIR, EDWARD. 1929. A study in phonetic symbolism. *Journal of Experimental Psychology* 12.225–239.
SAUSSURE, FERDINAND DE. 1915. *Course in general linguistics*. New York: McGraw Hill.
SHIH, STEPHANIE S; JORDAN ACKERMAN; NOAH HERMALIN; SHARON INKELAS; HAYEUN JANG; JESSICA JOHNSON; DARYA KAVITSKAYA; SHIGETO KAWAHARA; MIRAN OH; REBECCA L. STARR; andALAN YU. In prep. Cross-linguistic and language-specific sound symbolism: Pokémonastics. Manuscript. University of Southern California.
SHIH, STEPHANIE S; JORDAN ACKERMAN; NOAH HERMALIN; SHARON INKELAS; and DARYA KAVITSKAYA. 2018. Pokémonikers: A study of sound symbolism and Pokémon names. *Proceedings of the Linguistic Society of America*, 3:. doi:http://dx.doi.org/10.3765/plsa.v3i1.4335.
SIDHU, DAVID M.; and PENNY M. PEXMAN. 2015. What's in a Name? Sound Symbolism and Gender in First Names. *PLOS ONE* 10.e0126809. doi:10.1371/journal.pone.0126809.
SIDHU, DAVID M.; and PENNY M. PEXMAN. 2018. Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review* 25.1619–1643. doi:10.3758/s13423-017-1361-1.
SIDHU, DAVID M.; and PENNY M. PEXMAN. 2019. The Sound Symbolism of Names. *Current Directions in Psychological Science*.1–5. doi:10.1177/0963721419850134.
SKIPPER, JAMES K. 1981. An Analysis of Baseball Nicknames. *Baseball Research Journal* 10. http://research.sabr.org/journals/analysis-of-baseball-nicknames.
SKIPPER, JAMES K. 1990. Placenames used as nicknames: A study of Major League Baseball players. *Names: A Journal of Onomastics* 38.1–20.
SKIPPER, JAMES K. 1992. *Baseball Nicknames: A Dictionary of Origins and Meanings*. Jefferson, NC: McFarland & Company.
SLATER, ANNE S.; and SAUL FEINMAN. 1985. Gender and the phonology of North American first names. *Sex Roles* 13.429–440.
SPENCE, CHARLES. 2011. Crossmodal correspondence: A tutorial review. *Attention, Perception, and Psychophysics* 73.971–995.
STARR, REBECCA LURIE; ALAN C.L. YU; and STEPHANIE S SHIH. 2018. Sound symbolic effects in Mandarin and Cantonese personal names and Pokémon names. Paper presented at the 1st conference on Pokémonastics, Keio University, Tokyo.
TANGO, TOM; MITCHEL LICHTMAN; and ANDREW DOLPHIN. 2007. *The Book: Playing the Percentages in Baseball*. Washington, DC: Potomac Books.

THORN, JOHN; and PETE PALMER. 1984. *The Hidden Game of Baseball: A Revolutionary Approach to Baseball and its Statistics*. New York, NY: Doubleday.

WILSON, BRENDA S.; and JAMES K. SKIPPER. 2013. Nicknames and Women Professional Baseball Players. *Names: A Journal of Onomastics* 38.305–322.